JMIR Mental Health

Journal Impact Factor (JIF) (2023): 4.8
Volume 9 (2022), Issue 3 ISSN 2368-7959 Editor in Chief: John Torous, MD, MBI

Contents

Original Papers

Reviews

Original Paper

Telehealth Services for Substance Use Disorders During the COVID-19 Pandemic: Longitudinal Assessment of Intensive Outpatient Programming and Data Collection Practices

Kate Gliske[1], BA, MA, PhD; Justine W Welsh[2], BS, MD; Jacqueline E Braughton[1], BS, MA, PhD; Lance A Waller[3], BS, PhD; Quyen M Ngo[1], BA, MS, PhD

[1]Butler Center for Research, Hazelden Betty Ford Foundation, Center City, MN, United States

[2]Department of Psychiatry and Behavioral Services, Emory University School of Medicine, Emory University, Atlanta, GA, United States

[3]Department of Biostatistics and Bioinformatics, Emory University Rollins School of Public Health, Emory University, Atlanta, GA, United States

Corresponding Author:
Kate Gliske, BA, MA, PhD
Butler Center for Research
Hazelden Betty Ford Foundation
15251 Pleasant Valley Road
BC4
Center City, MN, 55012
United States
Phone: 1 6512134347
Email: KGLISKE@hazeldenbettyford.org

Abstract

Background: The onset of the COVID-19 pandemic necessitated the rapid transition of many types of substance use disorder (SUD) treatments to telehealth formats, despite limited information about what makes treatment effective in this novel format.

Objective: This study aims to examine the feasibility and effectiveness of virtual intensive outpatient programming (IOP) treatment for SUD in the context of a global pandemic, while considering the unique challenges posed to data collection during an unprecedented public health crisis.

Methods: The study is based on a longitudinal study with a baseline sample of 3642 patients who enrolled in intensive outpatient addiction treatment (in-person, hybrid, or virtual care) from January 2020 to March 2021 at a large substance use treatment center in the United States. The analytical sample consisted of patients who completed the 3-month postdischarge outcome survey as part of routine outcome monitoring (n=1060, 29.1% response rate).

Results: No significant differences were detected by delivery format in continuous abstinence (χ^2_2=0.4, *P*=.81), overall quality of life ($F_{2,826}$=2.06, *P*=.13), financial well-being ($F_{2,767}$=2.30, *P*=.10), psychological well-being ($F_{2,918}$=0.72, *P*=.49), and confidence in one's ability to stay sober ($F_{2,941}$=0.21, *P*=.81). Individuals in hybrid programming were more likely to report a higher level of general health than those in virtual IOP ($F_{2,917}$=4.19, *P*=.01).

Conclusions: Virtual outpatient care for the treatment of SUD is a feasible alternative to in-person-only programming, leading to similar self-reported outcomes at 3 months postdischarge. Given the many obstacles presented throughout data collection during a pandemic, further research is needed to better understand under what conditions telehealth is an acceptable alternative to in-person care.

(JMIR Ment Health 2022;9(3):e36263) doi:10.2196/36263

KEYWORDS

telehealth; substance use disorder; COVID-19; substance use treatment; feasibility study; routine outcome monitoring data; mental health; addiction; digital health; telemedicine; outpatient program; virtual health; addiction treatment; virtual care; patient outcomes

Introduction

Substance use disorder (SUD) is a chronic relapsing disease associated with numerous psychosocial harms and health sequalae. Addiction was a leading global cause of disability and death prior to the COVID-19 pandemic [1], which has since disproportionately impacted individuals suffering from SUD. Recent studies indicate that individuals with SUD may be more susceptible to severe disease and have higher rates of mortality and postvaccination breakthrough infections [2-4]. Isolation, uncertainty, and financial instability have also compounded substance use and the challenges of early recovery [5-7]. These vulnerabilities have reinforced the critical need for ongoing and safe access to treatment throughout the course of the pandemic through virtual services.

Prior to the COVID-19 pandemic, virtual services showed promise but were slow to develop. Early applications have shown promise as a means of preventing premature dropout from SUD treatment [8]. The pandemic rapidly accelerated the implementation of telehealth services for mental health and substance use treatment [9], and both providers and participants have viewed these types of services favorably [10,11]. Unfortunately, little is known about the actual efficacy or effectiveness of individual SUD treatment in telehealth treatment settings [12-14], and even less about traditional group treatment formats [15].

The onset of the pandemic became a catalyst for addiction treatment programs to quickly pivot to provision of services through telehealth formats despite limited data to guide their delivery. Change was facilitated by paradigm shifts in federal, state, and local policies and in organizational and provider practices [16,17]. Although these policies allowed for the continuity of care through available technology, stakeholders within the addiction field are now facing decisions on which elements of policies and programs to sustain, adapt, or discontinue. Continuation of these policies is dependent upon rigorous assessment of clinical data to define the new standard of SUD treatment through virtual platforms. Unfortunately, the pandemic had a devastating impact on research, with ongoing disruptions to recruitment and study progress, as well as a dramatic reduction in survey participation and response rates across many fields of study [18-20].

There is still a significant need for research related to the application and assessment of telehealth for SUD. Unfortunately, best practices for patient outcome collection for SUD treatment in mixed settings have yet to be established. In this paper, we describe how the COVID-19 pandemic presented a novel opportunity to bridge the gap and assess the effectiveness of a virtual intensive outpatient programming (IOP) for substance use treatment through the examination of short-term postprogram outcomes of adults who received IOP services through different delivery formats at the largest SUD treatment provider in the United States.

Methods

Study Design and Population

The Hazelden Betty Ford Foundation (HBFF) is the largest national provider of addiction services in the United States. The HBFF utilizes evidence-based practices through a multidisciplinary and integrated approach to addiction treatment across varying levels of care. In 2019, the HBFF piloted a single virtual intensive outpatient group with planned expansion of virtual services in 2020, as informed by routine outcome monitoring (ROM) data. The HBFF has an established infrastructure and process for collection of ROM data that has been used and refined since 1974. ROM data provide an understanding of real-world conditions, offering applied generalizability to community-based treatment settings where the majority of care is provided [21,22]. These data can be designed as a feedback loop, intended to quickly translate findings into treatment implementation [23,24]. In a rapidly evolving global pandemic, this type of real-world feedback is invaluable to informing the refinement of virtual treatment, despite potentially lower response rates than a formal randomized controlled trial [22].

This study presents 3-month findings (n=1060, 29.1%) from a 12-month longitudinal assessment of patients, 18 years and older, who were discharged from IOP between January 2020 and March 2021 (N=3642). Patients were separated by 3 distinct treatment delivery settings in response to pandemic changes: (1) in-person care only (n=957, 26.3%); (2) hybrid, in-person, and virtual care (n=541, 14.9%); and (3) virtual care only (n=2144, 58.9%).

Ethics Approval

The study was reviewed and approved by Emory University's Institutional Review Board (STUDY00001822) and was determined to have met the human research exemption since all data were collected within the context of the HBFF's standard ROM practices.

Data Collection Procedures

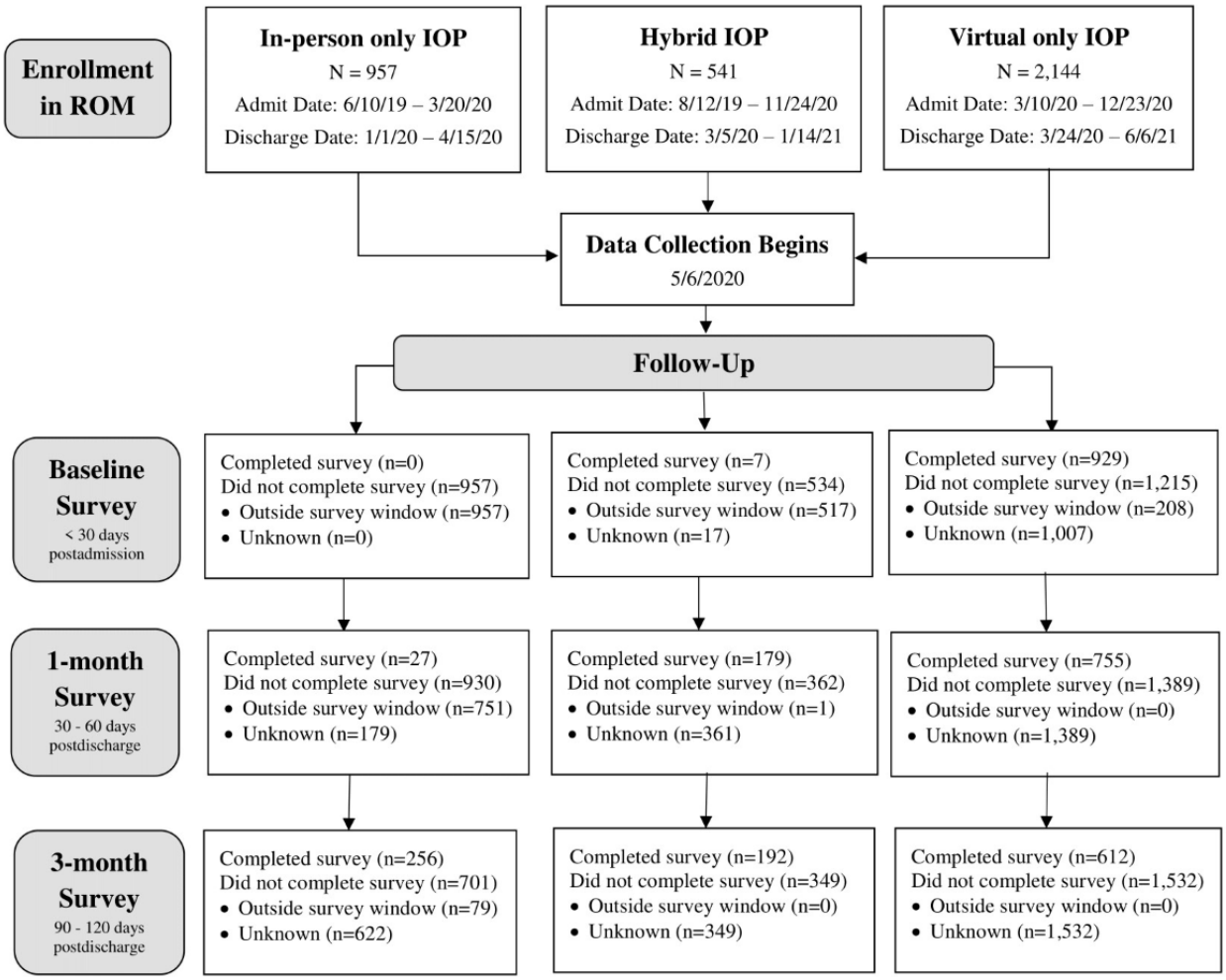
Trained research data collection specialists (RDCS) utilized a systematic and manualized process for data collection. Web-based surveys were automatically assigned to IOP patients at 1, 3, 6, 9, and 12 months postdischarge, with survey completion windows open for 30 days. Survey links were emailed to patients, with reminder prompts every 3 days for up to 2 weeks. Patients were contacted by an RDCS every 4-7 days to complete the survey over the phone or to encourage patients to complete it online if not initially completed. Patients were still prompted to provide responses if admitted to a different level of care.

Impact of the Pandemic on Data Collection

In response to a notable decline in response rates at the beginning of the pandemic, RDCS spent over 1200 hours attempting to contact patients to complete the 3-month survey throughout the course of the virtual IOP study. The timing of data collection was also impacted. Virtual IOP began in March 2020, while data collection began in May 2020, largely due to

reallocation of resources to facilitate the transition of direct care to virtual services. To retrospectively capture an in-person IOP comparison group, all patients discharged from an in-person IOP on or after January 1, 2020, were opted into receiving IOP outcome surveys. However, due to the 30-day survey windows, the majority of in-person and hybrid patients were excluded from completing the baseline survey at admission and the 1-month postdischarge survey, impacting response rates for those time points (Figure 1).

Figure 1. Data completeness from baseline to 3-month follow-up. IOP: intensive outpatient program; ROM: routine outcome monitoring.



Measures

Demographic information was collected from patient electronic medical records. The majority of the full sample of patients were White (n=3296, 91.3%) and male (n=2258, 62.0%), with a mean age of 39.1 (SD 13.5) years.

Outcome Measures

A variety of self-reported outcome measures were used to assess for health and well-being at 3-month follow-up. Continuous abstinence from drugs and alcohol (*abstinent since discharge* vs *relapsed*) during the follow-up period was assessed using a question from a modified Form 90 Alcohol Questionnaire (Form 90-AQ) [25], adapted to ask about the use of any substances and to include the specific time period for clarity: “Have you used any drugs or alcohol since your last survey on (last survey date)?” Compliance to prescribed anticraving medications (eg, buprenorphine, naltrexone, or acamprosate) was assessed with a single-item, binary question: “Have you taken your anticraving medication, as prescribed?” Peer support group engagement was measured by 1 item from the Alcoholics Anonymous (AA) Involvement Scale [26] and adapted to include a reference to other peer group support beyond AA: “About how often have you been attending 12-step/peer support/mutual aid group meetings since you were discharged?”. Participants rated their frequency of attendance on a 6-point scale: *daily*, *4 or more times per week*, *1-3 times per week*, *2-4 times per month*, *once a month or less*, and *never*.

Quality of life was measured using the 4-item self-reported Centers for Disease Control Healthy Days Survey [27,28]. An additional question assessing overall quality of life was also added: “How would you rate your overall quality of life?” Patients were asked to rate their overall quality of life and quality of general health using a 5-point Likert rating, from 1 (*poor*) to 5 (*excellent*), and indicated the number of days out of the previous 30 that they experienced poor mental or physical health. Higher numbers of unhealthy days indicated a lower

quality of life. Psychological well-being was assessed by the summed composite of the 8-item Flourishing Scale [29]. For each item, patients rated their level of agreement on a 7-point Likert scale, from 1 (*strongly disagree*) to 7 (*strongly agree*). The scale yielded high internal consistency (α=.94). Financial well-being was measured using the 5-item Consumer Financial Protection Bureau (CFPB) Financial Well-Being Scale, with higher scores illustrating greater perceived financial well-being [30]. The CFPB Financial Well-Being Scale showed good internal consistency (α=.85). Patients' confidence in their ability to stay sober was measured using an adapted form of the Brief Situational Confidence Questionnaire (BSCQ) to create a sobriety self-efficacy scale [31]. The 7-point Likert response categories were reworded to maintain consistency across the different scales, and the original BSCQ question 5, "I could probably go back to social drinking or other moderate drug use if I wanted to," was removed as initial interitem correlations and α values indicated that this question did not adequately add to the measure of sobriety self-efficacy. After removal of question 5, the adapted scale of sobriety self-efficacy showed high internal consistency (α=.89).

Statistical Analyses

Analyses were performed using IBM SPSS Statistics version 28 [32]. Data were examined using chi-square tests of independence and one-way ANOVAs to ascertain the relationship between IOP delivery setting and patient outcomes. Direct comparisons between the different settings of IOP were not indicated because the virtual IOP study was not prospectively designed and was instead reactively implemented as a result of the pandemic. This pandemic reality resulted in differences in the timing of care, akin to a cohort effect within a single year, where those in-person care reached the 3-month survey earlier in the pandemic (eg, May-July 2020), while those in virtual-only care completed the 3-month survey over a much longer period (eg, June 2020-June 2021).

Results

Sample Characteristics

Sample characteristics are reported by IOP modality in Table 1. Differences between IOP settings emerged in biological sex, age, and length-of-stay distributions. In comparison to in-person and virtual groups, the hybrid group members were more likely to be male (133/192 [69.3%] in the hybrid group vs 161/256 [62.9%] in the in-person group and 363/612 [59.3%] in the virtual group). The virtual group had a greater number of individuals aged from 45 to 64 years (293/612 [47.9%] in the virtual group vs 92/256 [35.9%] in the in-person group and 67/192 [34.9%] in the hybrid group). Individuals in the hybrid group had significantly longer lengths of stay (mean 74.67 [SD 41.78] days) than those who participated in in-person IOP (mean 53.88 [SD 34.79] days) or virtual IOP (mean 54.67 [SD 33.31] days). No significant differences were detected between formats by race, ethnicity, marital status, employment status, education level, whether a patient was discharged against staff/medical advice, use of insurance for services, or type or number of active SUD diagnoses.

Demographic and clinical characteristics of noncompleters of the 3-month survey were compared with those who completed the survey (Table 2). There were no significant differences in completer status in regard to biological sex, race, identification as Latinx, or the highest level of education attained. However, a few differences emerged in age, marital/relational status, and employment type. Completers of the 3-month survey were slightly older (mean age 42.26 [SD 12.93] years) than noncompleters (mean age 38.26 [SD 13.17] years). In addition, completers were more likely to be employed full-time (672/1060 [63.4%] vs 1400/2503 [55.9%]) and to be married/cohabiting (550/1060 [51.9%] vs 985/2503 [39.8%]).

A greater number of differences arose in regard to clinical characteristics. Those who completed the 3-month survey were less likely to have multiple active SUDs (336/1060 [31.7%] vs 1014/2503 [40.5%]) and to get discharged against staff/medical advice (107/1060 [10.1%] vs 531/2503 [21.2%]). Completers also showed a longer length of stay in IOP care (mean 58.10 [SD 36.16] days) than noncompleters (mean 49.57 [SD 38.73] days). There was no difference between those who stepped down into IOP from a higher level of programming within the HBFF (eg, residential vs day treatment).

Table 1. Baseline characteristics of 3-month outcomes survey respondents (N=1060).[a]

Characteristics	In-person only (N=256)	Hybrid (N=192)	Virtual only (N=612)	Overall (N=1060)
Biological sex, n (%); χ^2_2=6.3, *P*=.04				
Male	161 (62.9)	133 (69.3)	363 (59.3)	657 (62.0)
Nonbinary	N/A[b]	N/A	N/A	N/A
Missing	N/A	N/A	N/A	0
Age (years), n (%);χ^2_6=19.5, *P*=.003				
18-25	34 (13.3)	33 (17.2)	64 (10.5)	131 (12.4)
26-44	120 (46.9)	86 (44.8)	241 (39.4)	447 (42.2)
45-64	92 (35.9)	67 (34.9)	293 (47.9)	452 (42.6)
65+	10 (3.9)	6 (3.1)	14 (2.3)	30 (2.8)
Missing	N/A	N/A	N/A	0
Race[c], n (%); χ^2_2=11.3, *P*=.88				
American Indian or Alaska Native	2 (0.8)	1 (0.5)	6 (1.0)	9 (0.8)
Asian or Asian American	3 (1.2)	2 (1.0)	3 (0.5)	8 (0.8)
Black or African American	3 (1.2)	4 (2.1)	12 (2.0)	19 (1.8)
Native Hawaiian or other Pacific Islander	1 (0.4)	0	1 (0.2)	2 (0.2)
White	233 (91.0)	180 (93.8)	560 (91.5)	973 (91.8)
Biracial or multiracial (2+)	6 (2.3)	3 (1.6)	9 (1.5)	18 (1.7)
Other	7 (2.7)	2 (1.0)	13 (2.2)	22 (2.1)
Missing	1 (0.4)	0	8 (1.3)	9 (0.9)
Ethnicity, n (%); χ^2_2=2.3, *P*=.31				
Hispanic or Latinx or Spanish origin	16 (6.3)	6 (3.1)	29 (4.7)	51 (4.8)
Not Hispanic or Latinx or Spanish origin	232 (90.6)	180 (93.8)	551 (90.0)	963 (90.8)
Missing	8 (3.1)	6 (3.1)	32 (5.2)	46 (4.3)
Marital status[c], n (%); χ^2_4=6.3, *P*=.18				
Single/never married	87 (34.0)	78 (40.6)	204 (33.3)	369 (34.8)
Cohabiting	4 (1.6)	3 (1.6)	11 (1.8)	18 (1.7)
Married/life partner	125 (48.8)	95 (49.5)	312 (50.9)	532 (50.1)
Married but separated	14 (5.5)	6 (3.1)	19 (3.1)	39 (3.7)
Divorced	21 (8.2)	9 (4.7)	57 (9.3)	87 (8.2)
Widowed	3 (1.2)	1 (0.5)	1 (0.2)	5 (0.5)
Missing	2 (0.8)	0	8 (1.3)	10 (0.9)
Employment status[c], n (%); χ^2_2=1.9, *P*=.76				
Full-time employment/Self-employed	164 (64.1)	119 (62.0)	389 (63.6)	672 (63.4)
Part-time employment	12 (4.7)	4 (2.1)	21 (3.4)	38 (3.6)
Home and family manager	6 (2.3)	3 (1.6)	9 (1.5)	18 (1.7)
Student (full- or part-time) or retired	18 (7.0)	21 (11.0)	36 (5.9)	74 (7.0)
Unemployment, actively seeking a job	16 (6.3)	6 (3.1)	25 (4.1)	47 (4.4)
Unemployment, not seeking a job	36 (14.1)	36 (18.8)	113 (18.5)	185 (17.5)
Missing	4 (1.6)	3 (1.5)	19 (2.9)	26 (2.5)

Characteristics	In-person only (N=256)	Hybrid (N=192)	Virtual only (N=612)	Overall (N=1060)
Education level[c], n (%); χ^2_4=8.4, *P*=.08				
Some high school or less, no diploma	1 (0.4)	3 (1.5)	6 (1.0)	10 (1.0)
High school diploma or equivalent (General Educational Development [GED])	17 (6.6)	23 (12.0)	66 (10.8)	106 (10.0)
Some college, no degree	31 (12.1)	34 (17.7)	73 (11.9)	138 (13.0)
Associate degree/vocational-technical studies	15 (5.9)	14 (7.3)	36 (5.9)	65 (6.1)
College graduate/bachelor's degree	85 (33.2)	56 (29.2)	166 (27.1)	307 (29.0)
Graduate/professional degree	35 (13.7)	18 (9.4)	61 (10.0)	114 (10.8)
Missing	72 (28.1)	44 (22.9)	204 (33.4)	320 (30.2)
Length of IOP[d] stay, mean (SD); Welch $F_{2,411.943}$=19.67, *P*<.001				
Average length of stay (in days)	53.88 (34.79)	74.67 (41.78)	54.67 (33.31)	58.10 (38.156)
Missing	N/A	N/A	N/A	0
Discharged against staff advice, n (%); χ^2_2=0.8, *P*=.67				
Yes	24 (9.4)	17 (8.9)	66 (10.8)	107 (10.1)
No	232 (90.6)	175 (91.1)	546 (89.2)	953 (89.9)
Missing	N/A	N/A	N/A	0
Used insurance for services, n (%): χ^2_2=2.0, *P*=.36				
Yes	247 (96.5)	185 (96.4)	599 (97.9)	1,031 (97.3)
Self-pay	9 (3.5)	7 (3.6)	13 (2.1)	29 (2.7)
Missing	N/A	N/A	N/A	0
Active SUD[e] diagnosis, n (%)				
Alcohol use disorder; χ^2_2=0.2, *P*=.93	227 (88.7)	172 (89.6)	542 (88.6)	941 (88.8)
Cannabis use disorder; χ^2_2=0.3, *P*=.88	51 (19.9)	42 (21.9)	128 (20.9)	221 (20.8)
Cocaine use disorder; χ^2_2=5.0, *P*=.08	22 (8.6)	15 (7.8)	30 (4.9)	67 (6.3)
Hallucinogen use disorder; χ^2_2=1.5, *P*=.48	2 (0.8)	3 (1.6)	4 (0.7)	9 (0.8)
Inhalant use disorder; χ^2_2=4.5, *P*=.10	0	1 (0.5)	0	1 (0.1)
Opioid use disorder; χ^2_2=0.7, *P*=.71	19 (7.4)	17 (8.9)	56 (9.2)	92 (8.7)
Sedative use disorder; χ^2_2=0.3, *P*=.85	24 (9.4)	15 (7.8)	53 (8.7)	92 (8.7)
Other stimulant use disorder; χ^2_2=2.4, *P*=.30	31 (12.1)	15 (7.8)	59 (9.6)	105 (9.9)
Other psychoactive use disorder; χ^2_2=1.0, *P*=.61	4 (1.6)	4 (2.1)	7 (1.1)	15 (1.4)
Missing	N/A	N/A	N/A	0
Number of co-occurring SUD diagnoses[c], n (%); χ^2_6=8.4, *P*=.21				
1	168 (65.6)	134 (69.8)	422 (69.0)	724 (68.3)
2	61 (23.8)	31 (16.1)	135 (22.1)	227 (21.4)
3	21 (8.2)	20 (10.4)	37 (6.0)	78 (7.4)
4	6 (2.3)	7 (3.6)	18 (2.9)	31 (2.9)

Characteristics	In-person only (N=256)	Hybrid (N=192)	Virtual only (N=612)	Overall (N=1060)
Missing	N/A	N/A	N/A	0

[a]Clinical variables associated with a patient's treatment measured included the patient's length of IOP stay (in days), whether the patient was discharged against staff advice (yes/no), whether the patient used insurance or self-pay to finance their treatment (yes/no), the patient's active SUD diagnoses (eg, alcohol, opioids), and the number of co-occurring SUD diagnoses.

[b]Variables where categories were collapsed into 2-4 levels in order to test for group differences due to small cell sizes.

[c]N/A: not applicable.

[d]IOP: intensive outpatient programming.

[e]SUD: substance use disorder.

Table 2. Sample characteristics of completers of the 3-month follow-up survey and noncompleters.[a]

Variables	Completers (N=1060)		Noncompleters (N=2503)		Statistics		
	n (%)	Mean (SD)	n (%)	Mean (SD)	*F* test (*df*)	χ^2 (*df*)	*P* value
Demographic characteristics							
Biological sex: male	657 (62.0)	N/A[b]	1554 (62.1)	N/A	N/A	0.01 (1)	.93
Age (years)	1060 (100.0)	42.26 (12.93)	2503 (100.0)	38.26 (13.17)	69.35 (1,3561)	N/A	<.001
Race: White	973 (91.8)	N/A	2249 (89.8)	N/A	N/A	3.2 (1)	.07
Ethnicity: Latinx	51 (4.8)	N/A	135 (5.4)	N/A	N/A	0.6 (1)	.45
Employment						16.1 (2)	<.001
Full-time	672 (63.4)	N/A	1400 (55.9)	N/A	N/A	N/A	N/A
Unemployed	232 (21.9)	N/A	677 (27.0)	N/A	N/A	N/A	N/A
Education						1.8 (2)	.40
General Educational Development (GED) or less	116 (10.9)	N/A	306 (12.2)	N/A	N/A	N/A	N/A
Some college or bachelor's degree	510 (48.1)	N/A	1180 (47.1)	N/A	N/A	N/A	N/A
Marital status						50.6 (2)	<.001
Married/life partner	550 (51.9)	N/A	985 (39.4)	N/A	N/A	N/A	N/A
Single	369 (34.8)	N/A	1160 (46.3)	N/A	N/A	N/A	N/A
Clinical characteristics							
Single active SUD[c]	724 (68.3)	N/A	1489 (59.5)	N/A	N/A	24.6 (1)	<.001
Discharged against staff advice	107 (10.1)	N/A	531 (21.2)	N/A	N/A	62.6 (1)	<.001
Length of stay	1060 (100.0)	58.10 (36.16)	2503 (100.0)	48.57 (38.73)	37.60 (1,3561)	N/A	<.001
Step down into IOP[d]	559 (52.7)	N/A	1259 (50.3)	N/A	N/A	1.8 (1)	.18

[a]Mean (SD) reported for continuous variables and proportions (%) of samples reported for categorical variables. Pairwise differences calculated with chi-square tests and ANOVAs, as appropriate.

[b]N/A: not applicable.

[c]SUD: substance use disorder.

[d]IOP: intensive outpatient programming.

Multivariate Comparisons

A few differences emerged between IOP settings across multiple domains of functioning (Table 3). There was no significant difference by setting in self-reported continuous abstinence, with over two-thirds of the sample (680/960, 70.8%) reporting no drug or alcohol use since discharge. Approximately one-third (332/1060, 31.3%) overall reported still being prescribed an anticraving medication. Of those, no difference in medication compliance emerged between in-person, hybrid, or virtual IOP respondents. Individuals across all settings reported attending peer support meetings, on average, of 1 or 2 times per week. Further, there were no differences across settings in the *overall* perceived quality of life or in the total number of poor physical and mental health days at 3-month follow-up. Finally, there were no differences detected between IOP setting and the individuals' confidence in their ability to stay sober, financial well-being, or psychological well-being.

The only significant difference by IOP setting that emerged was in the overall quality of one's *general health*, where those in the hybrid group (mean 4.08 [SD 0.75]) were more likely than those in the virtual group (mean 3.89 [SD 0.83]) to report a higher level of general health.

Table 3. Differences by IOP[a] setting at 3-month follow-up.[b]

Variable	In-person (N=256)		Hybrid (N=192)		Virtual (N=612)		Statistics		
	n/N (%)	Mean (SD)	n/N (%)	Mean (SD)	n/N (%)	Mean (SD)	*F* test (*df*)	χ^2 (*df*)	*P* value
Continuous abstinence	162/231 (70.1)	N/A[c]	131/180 (72.8)	N/A	387/549 (70.5)	N/A	N/A	0.4 (2)	.81
Craving medication compliance	58/75 (77.3)	N/A	46/55 (83.6)	N/A	154/202 (76.2)	N/A	N/A	1.4 (2)	.50
Peer support meeting attendance	235/256 (91.8)	2.54 (1.51)	180/192 (93.8)	2.56 (1.51)	567/612 (92.6)	2.60 (1.52)	0.35 (2,979)	N/A	.86
Overall quality of life	231/256 (90.2)	3.99 (0.83)	176/192 (91.7)	4.09 (0.74)	422/612 (69.0)	3.95 (0.81)	2.06 (2,826)	N/A	.13
Overall quality of general health	225/256 (87.9)	4.00 (0.84)	175/192 (91.1)	4.08 (0.75)	520/612 (85.0)	3.89 (0.83)	4.19 (2,917)	N/A	.01
Total number of poor physical and mental health days	224/256 (87.5)	3.70 (6.66)	176/192 (91.7)	3.59 (6.50)	531/612 (86.8)	3.83 (6.43)	0.10 (2,928)	N/A	.90
Self-efficacy for staying sober	220/256 (85.9)	5.93 (1.16)	173/192 (90.1)	6.00 (1.06)	551/612 (90.0)	5.95 (1.27)	0.21 (2,941)	N/A	.81
Psychological well-being	219/256 (85.5)	45.09 (9.40)	170/192 (88.5)	45.83 (7.51)	532/612 (86.9)	44.91 (8.86)	0.72 (2,918)	N/A	.49
Financial well-being	220/256 (85.9)	48.40 (8.52)	170/192 (88.5)	46.73 (7.96)	380/612 (62.1)	47.20 (8.16)	2.30 (2,767)	N/A	.10

[a]IOP: intensive outpatient programming.

[b]Mean (SD) reported for continuous variables and proportions (%) of samples reported for categorical variables. Pairwise differences calculated with chi-square tests and ANOVAs, as appropriate.

[c]N/A: not applicable.

Discussion

Principal Findings

This study is the first of its kind to assess telehealth for SUD in the IOP setting in a large cohort of patients (N=1000+). No meaningful differences in outcome measures were identified between delivery settings at 3-month follow-up, with individuals reporting similar levels of continuous abstinence, quality of life, and social/emotional well-being. Our findings in regard to continuous abstinence were consistent with previous studies following patients at 3-6 months postdischarge from IOP (eg, 65/103 [63.1%]) [33]. These results are promising and suggest a potential continuing role for virtual IOP as an effective component in addiction treatment settings. Advocacy is needed to maintain these services as a standard offering within the SUD treatment continuum.

Historically, peer-based connections and the therapeutic milieu have been integral parts of addiction treatment. Concern has been expressed by the addiction treatment community regarding the shift to virtual services and its impact on group engagement and patient-centered outcomes [34]. These preliminary results demonstrate the feasibility of offering services virtually. Further research is necessary to obtain feedback on patient experience and measures of group cohesion, such as secure emotional expression, as they apply to virtual addiction treatment [35].

Our findings aid in establishing a platform for future evaluation of data collection processes that inform the effective development of standardized protocols for routine outcomes data practices, including frequency of contact, method of outreach, and training of staff. Standardized protocols must consider the context for accurate interpretation of collected data. For example, differences in response rates emerged in this study based on the timing of data collection in relation to the global pandemic and due to unanticipated staff burden and should be interpreted in this context. At 3-month follow-up, response rates were lowest for those in in-person IOP (256/957, 26.8%). These rates are likely attributable to these being completed by patients between May and August 2020, timing that coincided with major city- and statewide lockdowns and great uncertainty about the unknowns presented by the pandemic. Furthermore, because of the large opt-in of the in-person cohort at the beginning of data collection (May 2020), many participants at 3-month follow-up were given access to the survey well into the 30-day response window, reducing the likelihood that they would have adequate time to complete it prior to survey close. In contrast, response rates for those in virtual IOP (612/1532, 28.5%) were most impacted by a higher-than-projected admission rate of individuals into IOP throughout the study period, resulting in

a larger sample size than originally anticipated. As a result, there were proportionally lower staffing levels than would be typically allocated for the final sample size, which may explain the lower response rates for the virtual group.

Strengths and Limitations

A key strength of this study is the breadth of data collected from such a large number of patients receiving SUD treatment during a period of extensive change. Although these results are mainly descriptive, these analyses are necessary to carefully evaluate the impacts of a global shift in treatment approach. A number of limitations should be considered when interpreting results. Similar to existing research used with ROM data [36-39], response rates at 3-month follow-up were low. Nonresponse bias is a risk inherent to survey analysis. However, even in studies with high response rates, research has shown that nonresponse rates are not always directly predictive of nonresponse bias [37,40-42]. Research supports that highly resource-intensive recruitment deployed to capture late responders does not necessarily alter the outcomes found at lower, less resource-intense response rates, and this type of recruitment can be both cost- and time-prohibitive [40,42]. Despite COVID-induced high nonresponse rates, we expect these data to accurately reflect the effectiveness of IOP services during the COVID-19 pandemic. Substance use at follow-up was based on retrospective self-report of use. Utilizing additional methods for verification, such as urine drug analyses, would strengthen the validity of these reports in future studies. As discussed earlier, due to the sudden onset of the pandemic and subsequent data collection for this unique cohort, baseline and 1-month follow-up data were unable to be collected from a substantial portion of the full sample. The missing data influenced our ability to directly compare the effectiveness of IOP across delivery type, given the inherent confounding effect of the timing of patient care in relation to the unfolding public health crisis. We recommend future prospective studies be designed to compare in-person and virtual treatment directly, with inclusion of a formal evaluation of the ideal conditions for patient success (ie, dosage, treatment duration, frequency).

Finally, although our sample was representative of HBFF program participants, it differs from the general population in a few specific ways that are important to acknowledge when considering the generalizability of the findings. First, patients in the sample were primarily White and non-Latinx. As a result, there may have been too few non-White/Latinx participants to detect differences. However, this sample directly compares to findings from the 2020 National Survey on Drug Use and Health (NSDUH) [43] in terms of full- and part-time employment (United States: 67.3% vs sample: 67.0%) and bachelor's degree attainment (United States: 27.7% vs sample: 29%) among adults 18+ years old with a SUD in the past year. This demonstrates that addiction affects individuals across varying educational and employment statuses. This is in contrast to stereotypes of individuals struggling with addiction, which may characterize this as a disease of the uneducated and unemployed. Even though a higher percentage of patients employed full-time completed the 3-month survey (672/1060 [63.4%] vs 1400/2503 [55.9%]), post hoc analyses revealed no difference in outcomes based on employment status; therefore, we believe this had minimal impact on our findings. Future research should endeavor to improve the representation of racial and ethnic minorities in order to improve generalizability across a wider cross section of demographic variables.

Conclusions

Results from this study suggest that virtual outpatient care for the treatment of SUDs is a feasible alternative to in-person care, leading to similar rates of self-reported continuous abstinence, health, and well-being in patients at 3-month follow-up. This study should serve as a baseline for the assessment and refinement of the role of virtual services in the field of addiction treatment in order to better understand under what circumstances telehealth can function as an effective alternative to the established in-person standard of care.

Acknowledgments

The authors wish to thank Siara Sitar, who assisted in formatting the manuscript, and Clare Schmidt, for her contributions to the Methods section.

Conflicts of Interest

JWW received consulting fees from Applied Clinical Intelligence LLC (ACI Clinical). QMN, KG, and JEB are employees of the Hazelden Betty Ford Foundation (HBFF). LAW has no disclosures to report.

References

1. Degenhardt L, Whiteford HA, Ferrari AJ, Baxter AJ, Charlson FJ, Hall WD, et al. Global burden of disease attributable to illicit drug use and dependence: findings from the Global Burden of Disease Study 2010. Lancet 2013 Nov;382(9904):1564-1574. [doi: 10.1016/s0140-6736(13)61530-5]
2. Wang L, Wang Q, Davis PB, Volkow ND, Xu R. Increased risk for COVID-19 breakthrough infection in fully vaccinated patients with substance use disorders in the United States between December 2020 and August 2021. World Psychiatry 2022 Feb 05;21(1):124-132 [FREE Full text] [doi: 10.1002/wps.20921] [Medline: 34612005]
3. Wang QQ, Kaelber DC, Xu R, Volkow ND. COVID-19 risk and outcomes in patients with substance use disorders: analyses from electronic health records in the United States. Mol Psychiatry 2021 Jan;26(1):30-39 [FREE Full text] [doi: 10.1038/s41380-020-00880-7] [Medline: 32929211]

4. Dorjee K, Kim H, Bonomo E, Dolma R. Prevalence and predictors of death and severe disease in patients hospitalized due to COVID-19: a comprehensive systematic review and meta-analysis of 77 studies and 38,000 patients. PLoS One 2020 Dec 7;15(12):e0243191 [FREE Full text] [doi: 10.1371/journal.pone.0243191] [Medline: 33284825]
5. Hurley EA, Piña K, Cegielski V, Noel-MacDonnell JR, Miller MK. Recovering from substance use disorders during the early months of the COVID-19 pandemic: a mixed-methods longitudinal study of women in Kansas City. J Subst Abuse Treat 2021 Oct;129:108378 [FREE Full text] [doi: 10.1016/j.jsat.2021.108378] [Medline: 34080549]
6. Marsden J, Darke S, Hall W, Hickman M, Holmes J, Humphreys K, et al. Mitigating and learning from the impact of COVID-19 infection on addictive disorders. Addiction 2020 Jun 28;115(6):1007-1010. [doi: 10.1111/add.15080] [Medline: 32250482]
7. Sherman AC, Williams ML, Amick BC, Hudson TJ, Messias EL. Mental health outcomes associated with the COVID-19 pandemic: prevalence and risk factors in a southern US state. Psychiatry Res 2020 Nov;293:113476 [FREE Full text] [doi: 10.1016/j.psychres.2020.113476] [Medline: 33198047]
8. Tarp K, Bojesen AB, Mejldal A, Nielsen AS. Effectiveness of optional videoconferencing-based treatment of alcohol use disorders: randomized controlled trial. JMIR Ment Health 2017 Sep 29;4(3):e38 [FREE Full text] [doi: 10.2196/mental.6713] [Medline: 28963093]
9. Ellis LA, Meulenbroeks I, Churruca K, Pomare C, Hatem S, Harrison R, et al. The application of e-mental health in response to COVID-19: scoping review and bibliometric analysis. JMIR Ment Health 2021 Dec 06;8(12):e32948 [FREE Full text] [doi: 10.2196/32948] [Medline: 34666306]
10. Molfenter T, Roget N, Chaple M, Behlman S, Cody O, Hartzler B, et al. Use of telehealth in substance use disorder services during and after COVID-19: online survey study. JMIR Ment Health 2021 Feb 08;8(2):e25835 [FREE Full text] [doi: 10.2196/25835] [Medline: 33481760]
11. Sugarman DE, Busch AB, McHugh RK, Bogunovic OJ, Trinh CD, Weiss RD, et al. Patients' perceptions of telehealth services for outpatient treatment of substance use disorders during the COVID-19 pandemic. Am J Addict 2021 Sep 17;30(5):445-452 [FREE Full text] [doi: 10.1111/ajad.13207] [Medline: 34405475]
12. Huskamp HA, Busch AB, Souza J, Uscher-Pines L, Rose S, Wilcock A, et al. How is telemedicine being used in opioid and other substance use disorder treatment? Health Aff (Millwood) 2018 Dec;37(12):1940-1947 [FREE Full text] [doi: 10.1377/hlthaff.2018.05134] [Medline: 30633671]
13. Bergman BG, Kelly JF. Online digital recovery support services: an overview of the science and their potential to help individuals with substance use disorder during COVID-19 and beyond. J Subst Abuse Treat 2021 Jan;120:108152 [FREE Full text] [doi: 10.1016/j.jsat.2020.108152] [Medline: 33129636]
14. Ashford RD, Bergman BG, Kelly JF, Curtis B. Systematic review: digital recovery support services used to support substance use disorder recovery. Human Behav Emerg Tech 2019 Apr 30;2(1):18-32. [doi: 10.1002/hbe2.148]
15. King VL, Stoller KB, Kidorf M, Kindbom K, Hursh S, Brady T, et al. Assessing the effectiveness of an internet-based videoconferencing platform for delivering intensified substance abuse counseling. J Subst Abuse Treat 2009 Apr;36(3):331-338. [doi: 10.1016/j.jsat.2008.06.011] [Medline: 18775625]
16. Drug Enforcement Administration. DEA Qualifying Practitioners and DEA Qualifying Other Practitioners Announcement. URL: https://tinyurl.com/yckm2ea5 [accessed 2022-03-08]
17. Substance Abuse and Mental Health Services Administration. Opioid Treatment Program (OTP) Guidance. URL: https://www.samhsa.gov/sites/default/files/otp-guidance-20200316.pdf [accessed 2021-12-20]
18. Tolaney S, Lydon C, Li T, Dai J, Standring A, Legor KA, et al. The impact of COVID-19 on clinical trial execution at the Dana-Farber Cancer Institute. J Natl Cancer Inst 2021 Nov 02;113(11):1453-1459 [FREE Full text] [doi: 10.1093/jnci/djaa144] [Medline: 32959883]
19. Horrell LN, Hayes S, Herbert LB, MacTurk K, Lawhon L, Valle CG, et al. Telemedicine use and health-related concerns of patients with chronic conditions during COVID-19: survey of members of online health communities. J Med Internet Res 2021 Feb 18;23(2):e23795 [FREE Full text] [doi: 10.2196/23795] [Medline: 33539307]
20. Krueger T, Gogolewski K, Bodych M. Assessing the risk of COVID-19 epidemic resurgence in relation to the Delta variant and to vaccination passes. medRxiv 2021:1-34. [doi: 10.1101/2021.05.07.21256847]
21. Kelly PJ, Deane FP, Davis EL, Hudson S, Robinson LD, Keane CA, et al. Routine outcome measurement in specialist non-government alcohol and other drug treatment services: establishing effectiveness indicators for the NADAbase. Drug Alcohol Rev 2021 May 05;40(4):540-552. [doi: 10.1111/dar.13191] [Medline: 33155315]
22. van der Lem R, de Wever WW, van der Wee NJ, van Veen T, Cuijpers P, Zitman FG. The generalizability of psychotherapy efficacy trials in major depressive disorder: an analysis of the influence of patient selection in efficacy trials on symptom outcome in daily practice. BMC Psychiatry 2012 Nov 8;12(1):192. [doi: 10.1186/1471-244x-12-192]
23. Boswell JF, Kraus DR, Miller SD, Lambert MJ. Implementing routine outcome monitoring in clinical practice: benefits, challenges, and solutions. Psychother Res 2015 Jul 26;25(1):6-19. [doi: 10.1080/10503307.2013.817696] [Medline: 23885809]
24. Lambert MJ, Whipple JL, Kleinstäuber M. Collecting and delivering progress feedback: a meta-analysis of routine outcome monitoring. Psychotherapy (Chic) 2018 Dec;55(4):520-537. [doi: 10.1037/pst0000167] [Medline: 30335463]

25. Miller W. Form 90: A Structured Assessment Interview for Drinking and Related Behaviors. 1996. URL: https://pubs.niaaa.nih.gov/publications/projectmatch/match05.pdf [accessed 2022-03-08]
26. Tonigan JS, Connors GJ, Miller WR. Alcoholics Anonymous Involvement (AAI) scale: reliability and norms. Psychol Addict Behav 1996 Jun;10(2):75-80. [doi: 10.1037/0893-164x.10.2.75]
27. Centers for Disease Control and Prevention. Health-Related Quality of Life (HRQOL). URL: https://www.cdc.gov/hrqol/methods.htm [accessed 2022-03-08]
28. Moriarty DG, Zack MM, Kobau R. The Centers for Disease Control and Prevention's Healthy Days Measures: population tracking of perceived physical and mental health over time. Health Qual Life Outcomes 2003 Sep 02;1(1):37 [FREE Full text] [doi: 10.1186/1477-7525-1-37] [Medline: 14498988]
29. Diener E, Wirtz D, Tov W, Kim-Prieto C, Choi D, Oishi S, et al. New well-being measures: short scales to assess flourishing and positive and negative feelings. Soc Indic Res 2009 May 28;97(2):143-156. [doi: 10.1007/s11205-009-9493-y]
30. Consumer Finance Protection Bureau. Measuring Financial Well-Being. A Guide to Using the CFPB Financial Well-Being Scale. URL: https://www.consumerfinance.gov/data-research/research-reports/financial-well-being-scale/ [accessed 2021-12-14]
31. Delaney DJ, Bernstein MH, Harlow LL, Farrow M, Martin RA, Stein LAR. The Brief Situational Confidence Questionnaire for alcohol: a psychometric assessment with incarcerated youth. Psychol Assess 2020 Mar;32(3):254-264. [doi: 10.1037/pas0000780] [Medline: 31697110]
32. IBM Corp. IBM SPSS Statistics 24. URL: https://www.ibm.com/support/pages/release-notes-ibm-spss-statistics-240#relnotes__description [accessed 2021-11-30]
33. Bottlender M, Soyka M. Efficacy of an intensive outpatient rehabilitation program in alcoholism: predictors of outcome 6 months after treatment. Eur Addict Res 2005 Jul 9;11(3):132-137. [doi: 10.1159/000085548] [Medline: 15990430]
34. Mark TL, Treiman K, Padwa H, Henretty K, Tzeng J, Gilbert M. Addiction treatment and telehealth: review of efficacy and provider insights during the COVID-19 pandemic. Psychiatr Serv 2021 Oct 13:appips202100088. [doi: 10.1176/appi.ps.202100088] [Medline: 34644125]
35. Joyce AS, MacNair-Semands R, Tasca GA, Ogrodniczuk JS. Factor structure and validity of the Therapeutic Factors Inventory–Short Form. Group Dyn Theory Res Pract 2011 Sep;15(3):201-219. [doi: 10.1037/a0024677]
36. de Beurs E, den Hollander-Gijsman ME, van Rood YR, van der Wee NJA, Giltay EJ, van Noorden MS, et al. Routine outcome monitoring in the Netherlands: practical experiences with a web-based strategy for the assessment of treatment outcome in clinical practice. Clin Psychol Psychother 2011 Jan 12;18(1):1-12. [doi: 10.1002/cpp.696] [Medline: 20238371]
37. Hendra R, Hill A. Rethinking response rates: new evidence of little relationship between survey response rates and nonresponse bias. Eval Rev 2019 Oct;43(5):307-330. [doi: 10.1177/0193841X18807719] [Medline: 30580577]
38. Hoenders RHJ, Bos EH, Bartels-Velthuis AA, Vollbehr NK, van der Ploeg K, de Jonge P, et al. Pitfalls in the assessment, analysis, and interpretation of routine outcome monitoring (ROM) Data: results from an outpatient clinic for integrative mental health. Adm Policy Ment Health 2014 Sep 25;41(5):647-659. [doi: 10.1007/s10488-013-0511-7] [Medline: 23884455]
39. Stiles WB, Barkham M, Twigg E, Mellor-Clark J, Cooper M. Effectiveness of cognitive-behavioural, person-centred and psychodynamic therapies as practised in UK National Health Service settings. Psychol Med 2006 Feb 14;36(4):555-566. [doi: 10.1017/s0033291706007136]
40. Amico KR. A poor metric for study rigor in hosted intervention designs. Am J Public Health 2009 Sep;99(9):1567-1575. [doi: 10.2105/ajph.2008.134767]
41. Groves RM, Peytcheva E. The impact of nonresponse rates on nonresponse bias: a meta-analysis. Public Opin Q 2008 May 02;72(2):167-189. [doi: 10.1093/poq/nfn011]
42. Davern M, McAlpine D, Beebe TJ, Ziegenfuss J, Rockwood T, Call KT. Are lower response rates hazardous to your health survey? An analysis of three state telephone health surveys. Health Serv Res 2010 Oct;45(5 Pt 1):1324-1344 [FREE Full text] [doi: 10.1111/j.1475-6773.2010.01128.x] [Medline: 20579127]
43. Substance Abuse and Mental Health Services Administration. 2020 NSDUH Detailed Tables. URL: https://www.samhsa.gov/data/report/2020-nsduh-detailed-tables [accessed 2022-03-07]

Abbreviations

AA: Alcoholics Anonymous
CFPB: Consumer Financial Protection Bureau
HBFF: Hazelden Betty Ford Foundation
IOP: intensive outpatient programming
RDCS: research data collection specialists
ROM: routine outcome monitoring
SUD: substance use disorder

Edited by J Torous; submitted 07.01.22; peer-reviewed by D Sugarman, P Yellowlees; comments to author 24.01.22; revised version received 04.02.22; accepted 04.02.22; published 14.03.22.

Please cite as:
Gliske K, Welsh JW, Braughton JE, Waller LA, Ngo QM
Telehealth Services for Substance Use Disorders During the COVID-19 Pandemic: Longitudinal Assessment of Intensive Outpatient Programming and Data Collection Practices
JMIR Ment Health 2022;9(3):e36263
URL: https://mental.jmir.org/2022/3/e36263
doi:10.2196/36263
PMID:35285807

Original Paper

Longitudinal Relationships Between Depressive Symptom Severity and Phone-Measured Mobility: Dynamic Structural Equation Modeling Study

Yuezhou Zhang[1], MSc; Amos A Folarin[1,2,3,4,5], PhD; Shaoxiong Sun[1], PhD; Nicholas Cummins[1], PhD; Srinivasan Vairavan[6], PhD; Rebecca Bendayan[1,3], PhD; Yatharth Ranjan[1], MSc; Zulqarnain Rashid[1], PhD; Pauline Conde[1], BSc; Callum Stewart[1], MSc; Petroula Laiou[1], PhD; Heet Sankesara[1], BSc; Faith Matcham[7], PhD; Katie M White[7], BSc; Carolin Oetzmann[7], MSc; Alina Ivan[7], MSc; Femke Lamers[8], PhD; Sara Siddi[9,10,11], PhD; Elisabet Vilella[10,12], PhD; Sara Simblett[13], PhD; Aki Rintala[14,15], PhD; Stuart Bruce[16], PGDip; David C Mohr[17], PhD; Inez Myin-Germeys[14], PhD; Til Wykes[13,18], PhD; Josep Maria Haro[9,10,11], MD; Brenda WJH Penninx[8], PhD; Vaibhav A Narayan[6], PhD; Peter Annas[19], PhD; Matthew Hotopf[3,7,18], PhD; Richard JB Dobson[1,2,3,4,5], PhD; RADAR-CNS consortium[20]

[1]Department of Biostatistics & Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[2]Institute of Health Informatics, University College London, London, United Kingdom

[3]NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London, London, United Kingdom

[4]Health Data Research UK London, University College London, London, United Kingdom

[5]NIHR Biomedical Research Centre at University College London Hospitals NHS Foundation Trust, London, United Kingdom

[6]Janssen Research and Development LLC, Titusville, NJ, United States

[7]Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[8]Department of Psychiatry, Amsterdam Public Health Research Institute and Amsterdam Neuroscience, Amsterdam University Medical Centre, Vrije Universiteit and GGZ inGeest, Amsterdam, Netherlands

[9]Teaching Research and Innovation Unit, Parc Sanitari Sant Joan de Déu, Fundació Sant Joan de Déu, Barcelona, Spain

[10]Centro de Investigación Biomédica en Red de Salud Mental, Madrid, Spain

[11]Faculty of Medicine and Health Sciences, Universitat de Barcelona, Barcelona, Spain

[12]Hospital Universitari Institut Pere Mata, Institute of Health Research Pere Virgili, Universitat Rovira i Virgili, Reus, Spain

[13]Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom

[14]Center for Contextual Psychiatry, Department of Neurosciences, Katholieke Universiteit Leuven, Leuven, Belgium

[15]Faculty of Social Services and Health Care, LAB University of Applied Sciences, Lahti, Finland

[16]RADAR-CNS Patient Advisory Board, King's College London, London, United Kingdom

[17]Center for Behavioral Intervention Technologies, Department of Preventive Medicine, Northwestern University, Evanston, IL, United States

[18]South London and Maudsley NHS Foundation Trust, London, United Kingdom

[19]H. Lundbeck A/S, Copenhagen, Denmark

[20]See Acknowledgments

Corresponding Author:
Richard JB Dobson, PhD
Department of Biostatistics & Health Informatics
Institute of Psychiatry, Psychology and Neuroscience
King's College London
De Crespigny Park, Denmark Hill
London, SE5 8AF
United Kingdom
Phone: 44 20 7848 0473
Email: richard.j.dobson@kcl.ac.uk

Abstract

Background: The mobility of an individual measured by phone-collected location data has been found to be associated with depression; however, the longitudinal relationships (the temporal direction of relationships) between depressive symptom severity and phone-measured mobility have yet to be fully explored.

Objective: We aimed to explore the relationships and the direction of the relationships between depressive symptom severity and phone-measured mobility over time.

Methods: Data used in this paper came from a major EU program, called the Remote Assessment of Disease and Relapse–Major Depressive Disorder, which was conducted in 3 European countries. Depressive symptom severity was measured with the 8-item Patient Health Questionnaire (PHQ-8) through mobile phones every 2 weeks. Participants' location data were recorded by GPS and network sensors in mobile phones every 10 minutes, and 11 mobility features were extracted from location data for the 2 weeks prior to the PHQ-8 assessment. Dynamic structural equation modeling was used to explore the longitudinal relationships between depressive symptom severity and phone-measured mobility.

Results: This study included 2341 PHQ-8 records and corresponding phone-collected location data from 290 participants (age: median 50.0 IQR 34.0, 59.0) years; of whom 215 (74.1%) were female, and 149 (51.4%) were employed. Significant negative correlations were found between depressive symptom severity and phone-measured mobility, and these correlations were more significant at the within-individual level than the between-individual level. For the direction of relationships over time, Homestay (time at home) (ϕ=0.09, P=.01), Location Entropy (time distribution on different locations) (ϕ=−0.04, P=.02), and Residential Location Count (reflecting traveling) (ϕ=0.05, P=.02) were significantly correlated with the subsequent changes in the PHQ-8 score, while changes in the PHQ-8 score significantly affected (ϕ=−0.07, P<.001) the subsequent periodicity of mobility.

Conclusions: Several phone-derived mobility features have the potential to predict future depression, which may provide support for future clinical applications, relapse prevention, and remote mental health monitoring practices in real-world settings.

(JMIR Ment Health 2022;9(3):e34898) doi:10.2196/34898

KEYWORDS

depression; mobile health; location data; mobility; dynamic structural equation modeling; mHealth; mental health; medical informatics; modeling

Introduction

Depression is a prevalent and serious mental health disorder that is a leading cause of disability worldwide [1]. It can cause physical health and psychological function problems, resulting in loss of productivity and a high social burden [2-5]. Currently, diagnosis of depression relies on skilled clinicians and self-report questionnaires, which have limitations that include subjective bias and dynamic information loss [6]. Consequently, many people with depression do not receive timely and effective treatment [7], and more efficient methods for detecting and monitoring depression are needed. Recently, the use of mobile phones with embedded sensors for depression detection and monitoring, to provide new ways for supporting both depressed people and clinicians, has been investigated [8].

We focused on exploring how phone-collected location data could link individuals' mobility and depression. Past survey-based studies found that mobility is significantly and negatively associated with depression [9-11]. Several longitudinal survey–based studies reported a bidirectional relationship between depression and mobility over time, that is, decreased mobility worsened subsequent depressive symptoms and vice versa [10,11]. If the changes in mobility that occur before changes in depression can be captured by mobile phone technologies, early intervention can take place, which could prevent depression relapse or deterioration. Therefore, it is valuable to investigate relationships between depressive symptom severity and phone location data over time.

In recent years, there have been several studies [12-22] exploring the associations between depressive symptom severity and mobility features extracted from phone-collected location data that have shown that mobility measured by phones is negatively associated with the severity of depressive symptoms which is consistent with past survey-based studies; however, not many have explored the direction of the relationships between depression and mobility over time. Meyerhoff et al [22] recently found that phone-derived mobility features were correlated with subsequent changes in depression, but not vice versa. However, the autoregressive nature of depressive states and mobility levels [23-25] and the influence of individual differences may affect the results. In addition, the limitations of many previous phone-based studies [12-14,18-21] included relatively small and homogeneous (eg, university students) populations and the lack of comparison of between-individual and within-individual differences. To address these limitations, we aimed to explore the relationships and the direction of relationships over time between phone-derived mobility features and depressive symptom severity on a large multicenter data set.

Methods

Study Design

We used a large longitudinal data set of an EU research program called Remote Assessment of Disease and Relapse–Major Depressive Disorder, which explored the utility of remote measurement technologies in long-term (up to 2 years) depression monitoring [26]. We first used existing mobility features and then designed several new mobility features, which were extracted from this data set. Then, we assessed the relationships and direction of the relationships between depressive symptom severity and mobility features over time

using dynamic structural equation models [27]. Furthermore, we investigated the effects of individual differences (such as demographics) on the models at the between-individual level.

Study Participants and Settings

All participants in the study had at least one diagnosis of depression in the most recent 2 years and were recruited from 3 countries (Netherlands, Spain, and the United Kingdom); additional details descriptions are reported in [28]. Participants' passive data (eg, location, steps, and sleep) and active data (eg, questionnaires) were respectively collected via passive remote measurement technologies and active remote measurement technologies apps provided by an open-source platform (RADAR-base) [29]. A patient advisory board comprising service users co-developed the study and were involved in the choice of measures, the timing, and issues of engagement and in developing the analysis plan.

Ethics

Ethical approval was obtained from the Camberwell St. Giles Research Ethics Committee (17/LO/1154) in London, from the *Fundacio Sant Joan de Deu* Clinical Research Ethics Committee (CI: PIC-128-17) in London, and from the *Medische Ethische Toetsingscommissie VUms* (2018.012–NL63557.029.17) in the Netherlands.

Phone Location and Depression Questionnaire Data

We focused on phone location data and data from the 8-item Patient Health Questionnaire (PHQ-8) [30]. The passive remote measurement technologies app measured participants' location coordinates (longitude and latitude) using 2 providers (GPS and network sensors) periodically every 10 minutes. To protect participants' private information, raw locations were obfuscated by adding a unique and random reference location which was assigned to each participant at the start of the study [31]. The participant's self-reported depressive symptom severity was measured via the PHQ-8, with a score between 0 and 24 [30], which was assessed through the active remote measurement technologies app every 2 weeks (thus, the 2 weeks preceding each PHQ-8 record was the PHQ-8 interval).

Data Inclusion Criteria

Several factors may affect our analysis, such as the COVID-19 pandemic, location data accuracy, and missing data. Notably, the COVID-19 pandemic and related lockdown policies greatly impacted European people's mobility behaviors [32]. Therefore, according to suggestions in previous studies [6,14,16,19,33] and our experiences, we selected a subset of the data set [26] using the 3 criteria: (1) data from before February 2020 (prior to COVID-19 interventions in Europe) [6,33] were included, (2) location records with an error larger than 165 meters were removed [14,16], and (3) the amount of missing location data in a given PHQ-8 interval was limited to 50% [14,16,19].

Data Preprocessing

We calculated the distances between consecutive location records and the instantaneous speeds at all location records. The distance between 2 consecutive location records was computed by using the Haversine formula [34]. The instantaneous speed was approximated by dividing the distance by the time between 2 consecutive location records. We regarded one location record as a stationary point if its instantaneous speed was less than 1 km/h; otherwise, we considered it a moving point [14,19].

The second procedure was location clustering. Since the density-based spatial clustering of applications with noise method [35] can treat low-density location points as outliers, avoiding overestimating the number of locations clusters [14], we used this method for location clustering, using hyperparameters and the method for handling unequal sampling intervals from [14].

Feature Extraction

We extracted 11 mobility features (Table 1) from location data in each PHQ-8 interval (14 days), of which 4 features (3 frequency-domain features to reflect periodic characteristics of mobility and 1 feature to represent the number of temporary residential locations during the past 14 days) are new.

Table 1. A list of mobility features used in this study and their short descriptions.

Feature	Description
Location Variance	Variance of longitude and latitude coordinates
Moving Time	Percentage of time spent in moving
Moving Distance	Distance between all location points weighted by available time
Number of Clusters	The number of location clusters found using density-based spatial clustering of applications with noise
Location Entropy	Entropy of time distribution over different locations
Normalized Entropy	Location Entropy normalized by the number of clusters
Homestay	Percentage of time spent at home
Residential Location Count	The number of temporary residential locations
Long-term Rhythm	Percentage of frequency bins within the long-term period (>1 day) of spectrum for longitude and latitude coordinates
Circadian Rhythm	Percentage of frequency bins within the circadian period (24 hours) of spectrum for longitude and latitude coordinates
Short-term Rhythm	Percentage of frequency bins within the short-term period (<1 day) of spectrum for longitude and latitude coordinates

Time-Domain Features

Location Variance

The Location Variance represented the variability of each participant's locations [19] and was calculated as log(*Var*(*Lon*)+*Var*(*Lat*)), where log is the logarithm, and *Var*(*Lon*) and *Var*(*Lat*) represent the variances of the longitude and latitude coordinates, respectively, in one PHQ-8 interval.

Moving Time

The Moving Time represented the percentage of time that a participant spent in moving in one PHQ-8 interval [19]. The feature was computed by dividing the sum duration for all moving points by the sum of available time in one PHQ-8 interval.

Moving Distance

The Moving Distance was adjusted by dividing the total distance by the available time (in hours) in one PHQ-8 interval. In previous studies [18,19], the total distance obtained by accumulating distances between all location records; however, this total distance was affected by the missing data rate.

Number of Clusters

The number of the unique location clusters that a participant visited in one PHQ-8 interval was calculated using density-based spatial clustering of applications with noise [14].

Location Entropy

Location Entropy represented the distribution of time spent by a participant at different location clusters in one PHQ-8 interval [19] and was calculated as

where p_i is the percentage of time spent at location cluster *i*, thus the greater the average time, the higher the Location Entropy and vice versa [19].

Normalized Entropy

Because the number of location clusters varies across participants and the number of clusters is positively correlated with Location Entropy [14,16,19], we also used Normalized Entropy which was given by *Normalized Entropy* = *Location Entropy* / log (*Number of Clusters*)

Homestay

In previous studies [13,14,16,18,19,21], each participant was assigned only one home location, which was the most visited location cluster between 12 AM to 6 AM; however, in our study, due to the long follow-up time and community-based population, participants may have more than one residential location in one PHQ-8 interval (for example, for reasons, such as traveling, business trips, or moving to a new house). Therefore, we adjusted the method of determining the residential locations. We first selected all location clusters visited at night (12 AM to 6 AM) in one PHQ-8 interval. Then, if multiple clusters were visited in the same night, the location cluster with the most location records was selected as the home location. This step partially excluded the impact of activities at night. The Homestay was the time spent at all stationary location points belonging to all home locations as the percentage of the available time in one PHQ-8 interval.

Residential Location Count

This new feature represented the number of residential locations. Since temporary home locations could reflect traveling [36], we used the number of residential locations in one PHQ-8 interval to reflect traveling.

Frequency-Domain Features

People's life rhythms (such as circadian rhythm, sleep rhythm, and social rhythm) are related to depression [37]. We proposed 3 frequency-domain features to reflect the periodicity of participants' mobility. To compute frequency-domain features, we used linear interpolation and the fast Fourier transformation to get the spectrums of longitude and latitude data, respectively (Figure 1). The frequency axis of the spectrum was scaled in cycles per day to reflect the number of periodic patterns that occurred daily. To explore the periodic rhythms of different period lengths, we used the same frequency-domain division as in our previous publication [6], that is, frequency bands of low frequency (0 to 0.75 cycles per day), middle frequency (0.75 to 1.25 cycles per day), and high frequency (>1.25 cycles per day). The power in the middle frequency was used to represent the strength of the circadian rhythm (around 1 cycle/day) of the participant's mobility. Likewise, the power in low frequency and high frequency represent the long-term (>1 day) periodic rhythm and short-term (<1 day) rhythm, respectively. We extracted 3 features to reflect the percentages of these 3 periodic rhythms (long-term, circadian, and short-term rhythms) in individuals' mobility. We summed the power in the same frequency band of longitude and latitude, then divided it by the sum of the total spectral power of longitude and latitude. The formulas of these 3 features are

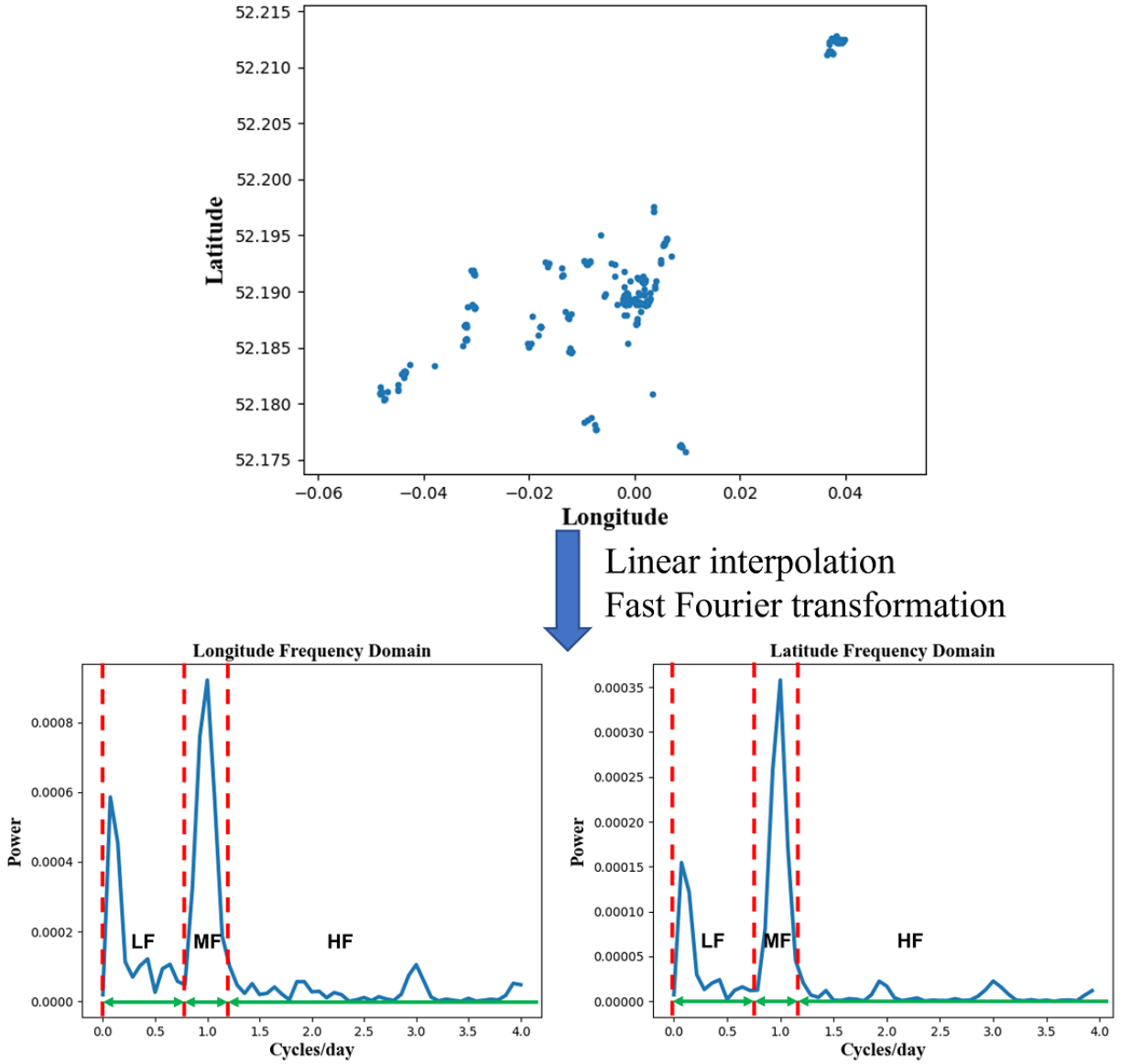
$$\textit{Long-term Rhythm} = (PSD_{lon}(LF) + PSD_{lat}(LF)) / (PSD_{lon}(Total) + PSD_{lon}(Total))$$

$$\textit{Circadian Rhythm} = (PSD_{lon}(LF) + PSD_{lat}(LF)) / (PSD_{lon}(Total) + PSD_{lon}(Total))$$

$$\textit{Short-term Rhythm} = (PSD_{lon}(LF) + PSD_{lat}(LF)) / (PSD_{lon}(Total) + PSD_{lon}(Total))$$

where PSD_{lon} and PSD_{lat} represent the power spectral density of longitude and latitude, respectively, and *LF*, *MF*, *HF*, and *Total* are the low frequency, middle frequency, high frequency, and total spectral power, respectively. If the individuals' mobility is regular, the Long-term Rhythm or Circadian Rhythm will be high, otherwise, Short-term Rhythm will be high.

Figure 1. Transformation of location data from the time domain to the frequency domain. LF: low frequency (0-0.75 cycles/day); MF: middle frequency (0.75-1.25 cycles/day); HF: high frequency (>1.25 cycles/day).

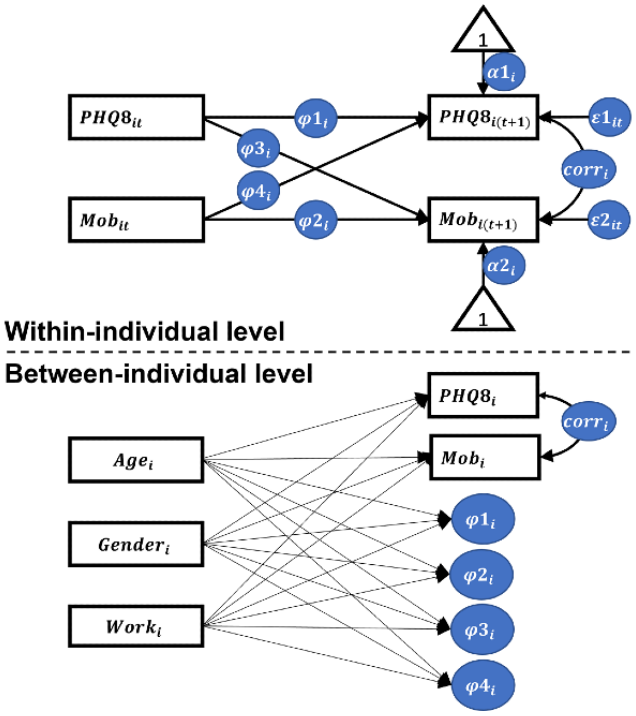


Data Analyses

We used dynamic structural equation modeling to explore the relationships and the direction of relationships between mobility features and PHQ-8 scores over time. Dynamic structural equation modeling is a broad integrated framework that blends multilevel, time-series, and structural equation modeling [27,38,39] and which has shown to be particularly useful for intensive longitudinal data [38,39]. Specifically, the 2-level vector autoregressive model can estimate the lagged effects and cross-lagged effects between 2 outcome variables while considering the variability at both within-individual and between-individual levels [27,39]. The lagged effect is the impact of one variable on itself over time, which was used to represent the autoregressive nature of depressive states and mobility levels [23-25]. The cross-lagged effect is the impact of one variable on the other variable over time, which was used to explore the direction of relationships between mobility features and PHQ-8 score. In this study, we only considered the Lag-1 model (Figure 2), that is, the lagged effects and cross-lagged effects between a time point *t* and the immediately subsequent (2 weeks later) time point (*t* + 1).

We built a vector autoregressive model with each mobility feature and PHQ-8 score as outcome variables and used age, gender, and work status as covariates [40-42] at the between-individual level for adjusting individual differences. The correlations between the PHQ-8 score and the mobility feature (Figure 2) at both within-individual and between-individual levels were also estimated by the vector autoregressive model. We established a total of 11 vector autoregressive models for all mobility features. All *P* values of coefficients in vector autoregressive models and correlations were adjusted using the Benjamini-Hochberg method [43] for multiple comparisons. Findings were considered significant at adjusted *P* value <.05. Vector autoregressive models were implemented in Mplus (version 8) [44] and multiple comparison corrections were performed in R software (version 3.6.3).

Figure 2. Path diagram of the vector autoregressive model. $PHQ8_{it}$ and Mob_{it} represent the score of 8-item Patient Health Questionnaire and a mobility feature, respectively, of participant *i* at time point *t*. Age, gender, and work status were considered covariates at the between-individual level.

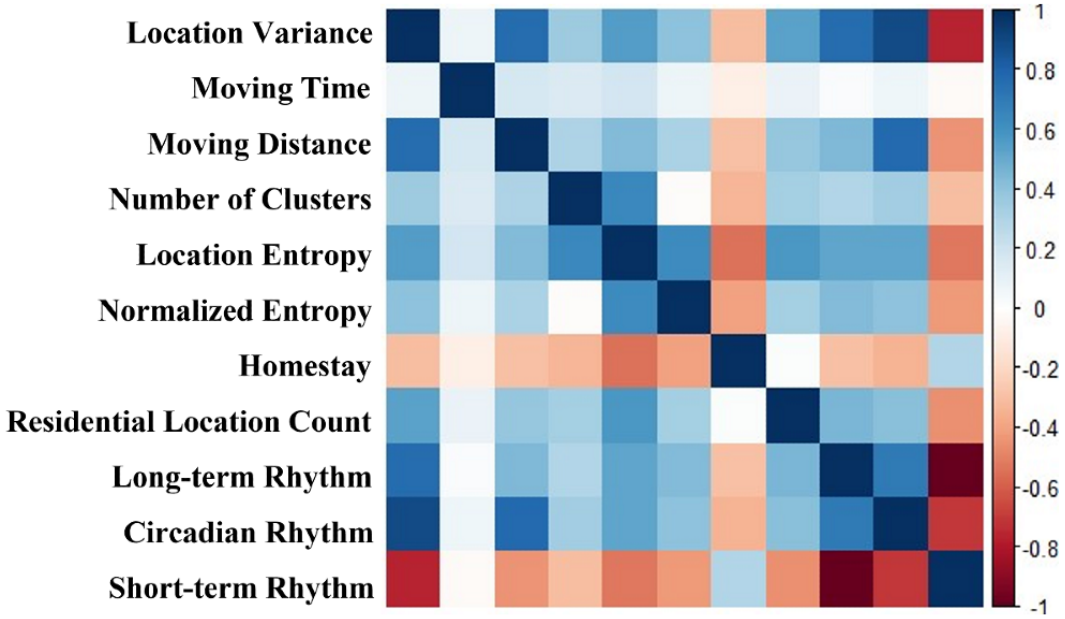


Results

Data Summary

The 2341 PHQ-8 intervals of 290 participants collected between November 2017 and February 2020 were included in our analysis. The sample had a median age of 50.0 (IQR 34.0, 59.0) years, with 215 (74.14%) female participants and 149 (51.38%) employed participants, with a median of 10 (IQR 5, 15) PHQ-8 scores and a median of 8.0 (IQR 3.0, 14.0) PHQ-8 intervals for each participant. The pairwise Spearman correlations between all 11 mobility features are presented in Figure 3.

Figure 3. A heatmap of pairwise Spearman correlations between all 11 mobility features extracted in this paper.



Vector Autoregressive Models

Correlation

Except for Moving Time (*P*=.11), all mobility features were significantly correlated with the PHQ-8 score at the within-individual level (Table 2); Homestay (ρ=0.11, *P*<.001) and Short-term Rhythm (ρ=0.07, *P*=.004) were positively correlated, while other mobility features were negatively correlated. Between individuals, Location Variance (ρ=−0.22, *P*=.04) and Moving Distance (ρ=−0.26, *P*=.04) were significantly and negatively correlated with PHQ-8 scores.

Table 2. Mobility features' correlations with PHQ-8 scores at within- and between-individual levels.

Mobility feature	Within-individual level		Between-individual level	
	ρ	Adjusted *P* value	ρ	Adjusted *P* value
Location Variance	−0.10	<.001	−0.22	.04
Moving Time	0.03	.11	−0.09	.28
Moving Distance	−0.08	.002	−0.26	.04
Number of Clusters	−0.09	.001	−0.02	.44
Location Entropy	−0.15	<.001	−0.09	.22
Normalized Entropy	−0.05	.02	−0.14	.11
Homestay	0.11	<.001	0.10	.20
Residential Location Count	−0.09	.001	−0.09	.27
Long-term Rhythm	−0.07	.004	−0.17	.09
Circadian Rhythm	−0.12	<.001	−0.16	.11
Short-term Rhythm	0.07	.004	0.16	.09

Lagged and Cross-lagged Effects

There were significant and positive lagged effects exist in both PHQ-8 scores (ϕ1=0.45-0.51, *P*<.001) and mobility features (ϕ2=0.11-0.53, *P*<.001) (Table 3). For cross-lagged effects, PHQ-8 scores were significantly and negatively correlated with the subsequent Circadian Rhythm of mobility (ϕ3=−0.07, *P*<.001), while Location Entropy (ϕ4=−0.04, *P*=.02), Homestay (ϕ4=0.09, *P*=.01), and Residential Location Count (ϕ4=0.05, *P*=.02) were significantly correlated with subsequent PHQ-8 scores.

Table 3. Lagged and cross-lagged effects between mobility features and PHQ-8 scores estimated by vector autoregressive models.

Mobility feature	Lagged effects				Cross-lagged effects			
	ϕ1	Adjusted *P* value	ϕ2	Adjusted *P* value	ϕ3	Adjusted *P* value	ϕ4	Adjusted *P* value
Location Variance	0.49	<.001	0.2	<.001	−0.03	.22	0.02	.23
Moving Time	0.47	<.001	0.53	<.001	0.02	.22	0.02	.31
Moving Distance	0.48	<.001	0.38	<.001	0.03	.21	0.03	.21
Number of Clusters	0.49	<.001	0.3	<.001	0.005	.50	−0.01	.32
Location Entropy	0.47	<.001	0.22	<.001	−0.01	.33	−0.04	.02
Normalized Entropy	0.46	<.001	0.14	<.001	−0.004	.44	0.003	.45
Homestay	0.45	<.001	0.34	<.001	−0.01	.30	0.09	.01
Residential Location Count	0.51	<.001	0.11	<.001	−0.01	.34	0.05	.02
Long-term Rhythm	0.49	<.001	0.21	.001	−0.05	.06	0.001	.45
Circadian Rhythm	0.48	<.001	0.11	<.001	−0.07	<.001	0.03	.12
Short-term Rhythm	0.48	<.001	0.11	<.001	0.05	.06	−0.03	.34

The Influence of Individual Differences

Older and employed participants had significantly lower intercepts of the PHQ-8 score than younger and unemployed participants (Table 4). For mobility features, age was significantly and negatively correlated with Number of Clusters (γ=−0.12, *P*=.01), Location Entropy (γ=−0.18, *P*<.001), and Residential Location Count (γ=−0.16, *P*<.001), while work status was significantly correlated with most mobility features (except for Moving Time [*P*=.42] and Residential Location Count [*P*=.09]). For lagged effects, older participants had significantly lower lagged effects on Moving Distance (γ=−0.16, *P*=.02) and Homestay (γ=−0.14, *P*=.03) than younger participants. Female participants had significantly lower lagged effects on Location Entropy (γ=−0.15, *P*=.02) and Residential Location Count (γ=−0.24, *P*=.01) than male participants. Compared with unemployed participants, employed participants have significantly lower lagged effects on the PHQ-8 score (γ=−0.14, *P*=.03) and significantly higher lagged effects on Normalized Entropy (γ=0.25, *P*=.01). For cross-lagged effects, age was significantly and negatively correlated with the ϕ3 coefficient of Circadian Rhythm (γ=−0.49, *P*=.004) in the corresponding vector autoregressive model.

Table 4. Significant effects of individual difference at the between level of the vector autoregressive models. Only significant effects of at least one covariate are reported.

Characteristic	Age		Female		Employed	
	γ	Adjusted *P* value	γ	Adjusted *P* value	γ	Adjusted *P* value
Effects on the intercept of						
Patient Health Questionnaire–8	−0.21	<.001	0.07	.09	−0.10	.01
Location Variance	−0.08	.06	0.03	.29	0.12	.01
Moving Distance	0.01	.47	−0.01	.40	0.07	.01
Number of Clusters	−0.12	.01	0.02	.36	0.09	.03
Location Entropy	−0.18	<.001	0.01	.40	0.20	<.001
Normalized Entropy	−0.09	.09	−0.01	.45	0.26	<.001
Homestay	0.01	.32	0.03	.16	−0.15	<.001
Residential Location Count	−0.16	<.001	0.04	.17	0.06	.09
Long-term Rhythm	−0.07	.07	0.02	.34	0.14	.01
Circadian Rhythm	−0.07	.08	0.06	.10	0.13	<.001
Short-term Rhythm	0.10	.06	−0.06	.13	−0.16	<.001
Effects on the lagged effect of						
Patient Health Questionnaire–8	0.01	.47	−0.07	.13	−0.14	.03
Moving Distance	−0.16	.02	−0.04	.31	−0.08	.06
Location Entropy	−0.01	.46	−0.15	.02	0.02	.38
Normalized Entropy	0.09	.19	−0.19	.05	0.25	.01
Homestay	−0.14	.03	−0.09	.13	0.05	.27
Residential Location Count	0.01	.48	−0.24	.01	−0.04	.36
Effects on the cross-lagged effect of						
Circadian Rhythm (ϕ3)[a]	−0.49	.004	0.01	.48	0.164	.25

[a]ϕ3 represents the effect of the Patient Health Questionnaire–8 on the subsequent mobility feature.

Discussion

Principal Findings

This study provides a comprehensive understanding of the relationships and the direction of the relationships between depressive symptom severity and phone-measured mobility over time by using dynamic structural equation modeling on a large longitudinal data set and considering correlations at both individual and population levels, lagged effects (the autoregressive nature over time), cross-lagged effects (direction of the relationships over time), and the influences of individual differences (demographic characteristics).

Most mobility features extracted in this paper were significantly correlated with the PHQ-8 score at the within-individual level (Table 2), which indicated that, for a participant, the higher the severity of depressive symptoms, the lower mobility. This is consistent with both past survey-based [9] and phone-based studies [18,19]. These findings reaffirmed that the link between depressive symptom severity and mobility can be captured by mobile phones. However, many of the mobility features' correlations with PHQ-8 score were not significant at the between-individual level, possibly due to the significant effects of individual differences (age and work status) on both PHQ-8 score and mobility features (Table 4). Notably, features of Location Variance (ρ=−0.22, *P*=.04) and Moving Distance (ρ=−0.26, *P*=.04) were still significantly correlated with PHQ-8 score at the between-individual level, which indicated these features are relatively robust for reflecting depressive symptom severity in the whole population. Compared with the results of previous phone-based studies, our results showed that population diversity affects correlations between mobility features and the depression score. Most mobility features were significantly correlated with depression scores in student-based studies [16,18], while several features lost their significance in a community-based population with a wide age distribution [19]. These findings indicated that individual differences need to be considered during exploring relationships between depression and mobility.

PHQ-8 score and mobility features both had significant and positive lagged effects (Table 3), indicating that the autoregressive nature of individuals' depressive states [24] and movement habits [25] could be captured by mobile phones. For the direction of relationships over time, we found 3 mobility features significantly correlated with the subsequent PHQ-8 score. Specifically, increases in PHQ-8 score are probably preceded by one or more following changes in the mobility: (1)

lower average time spent at different places (Location Entropy), (2) more time at home (Homestay), and (3) more traveling (Residential Location Count). Conversely, change in PHQ-8 score was significantly and negatively correlated ($\phi 3=-0.07$, $P<.001$) with the subsequent circadian rhythm measured by location data. The findings of a recent study [22] showed changes in several mobility features were associated with subsequent depression changes, but not vice versa. The differences in populations and applied methods could be potential reasons for the slightly inconsistent results. Both our study and that study [22] have shown that the changes in mobility prior to changes in depressive symptom severity can be captured by mobile phones. An interesting finding is that the number of residential locations was positively correlated ($\phi 4=0.05$, $P=.02$) with the subsequent PHQ-8 score (Table 3), which is opposite to their negative correlation ($\rho=-0.09$, $P=.001$) at the within-individual level (Table 2). As the number of temporary residential locations could reflect traveling [36], this finding indicated that traveling may reduce the current depressive symptoms but may worsen some existing depressive feelings. This finding may provide insight into a phenomenon called "post-travel depressed feelings [45,46]." The causes of "post-travel depressed feelings" are fatigue from trips, the shock of re-entry of ordinary life, and jet lag [46,47].

For influences of individual differences on the levels of depressive symptom severity and mobility, we found that PHQ-8 scores tended to be lower in participants who are older or have jobs, which can be expected because previous survey-based studies have shown that depression is negatively correlated with age, and the unemployment rate in the depressed population is high [40-42]. Gender was not significantly correlated with the PHQ-8 score ($\gamma=0.07$, $P=.09$) in our population, possibly due to all participants in our study having at least one diagnosis of depression in recent 2 years [26], which may reduce the link between gender and depressive symptom severity. For the effects of demographic characteristics on mobility features, we found that the mobility in older participants or participants without jobs tended to be lower, which is also expected. For influences of individual differences on the lagged and cross-lagged effects, we found the participants with jobs had lower autocorrelation of the PHQ-8 score, indicating more depressive symptoms severity changes over time in employed participants than unemployed participants. Female participants, older participants, and unemployed participants tended to have lower autocorrelations of some mobility features than male participants, young participants, and employed participants, which indicated that variabilities of mobility over time were larger in these participants. For influences of age on cross-lagged effects, the impact of changes in PHQ-8 score on the subsequent circadian rhythm for older participants was significantly lower than that of young participants ($\gamma=-0.49$, $P=.004$), indicating that the mobility rhythm of the older participants is affected by depressive symptoms for a shorter period than the young participants.

We proposed 3 frequency-domain features to reflect the periodic characteristics of individuals' mobility (Figure 1). They were all significantly correlated with the PHQ-8 score at the within-individual level. Higher values of Long-term Rhythm and Circadian Rhythm represent more regular movement and activity, which were correlated with lower depressive symptom severity. Notably, Circadian Rhythm had the strongest correlation ($\rho=-0.12$, $P<.001$) among these 3 features, and it had significant cross-lagged effect ($\phi 3=-0.07$, $P<.001$) with the preceding PHQ-8 score. These findings demonstrated that the frequency-domain of location data can provide some additional information for evaluating depressive symptom severity in future research.

Limitations

We obfuscated the raw location data due to privacy issues. Therefore, we did not have access to contextual information, which may mean some information was lost. Another limitation is that we only used the Lag-1 vector autoregressive models. We did not use high-order vector autoregressive models because we wanted to make our preliminary model simple to allow easier explanation and to avoid convergence problems in the procedure of coefficient estimations. We will attempt high-order vector autoregressive models in future research when we have more data without the impact of the COVID-19.

We chose to build 11 dynamic structural equation modeling models, one for each mobility feature. Since each mobility feature has a specific meaning, the bivariate model can better explain changes of the feature before and after the changes in PHQ-8 scores indicating the longitudinal relationships. We attempted multivariate dynamic structural equation modeling with all mobility features, but the model failed to converge, possibly due to the multicollinearity between mobility features and complexity of the model. As all mobility features were devised for describing characteristics of individuals' mobility, there were high correlations between mobility features (Figure 3). In future research, we plan to solve the multicollinearity in the multivariate model through further feature engineering and feature selection methods or by using other multivariate time series models which are robust to multicollinearity [48].

Conclusions

This study provides initial evidence of the relationship and the direction of the relationship between depressive symptom severity and phone-measured mobility over time. We found several mobility features affected depressive symptom severity, while changes in the depression score were associated with the subsequent periodic rhythm of mobility. These mobility features have the potential to be used as indicators for assessing depression risk in future clinical applications, which could provide timely suggestions for both people with depression risk (eg, encouraging to attend more activities) and physicians (eg, early interventions). This work may provide support for remote mental health monitoring practice in real-world settings.

Acknowledgments

The Remote Assessment of Disease and Relapse–Major Depressive Disorder project is a component of the Remote Assessment of Disease and Relapse–Central Nervous System (RADAR-CNS) project, which has received funding from the Innovative Medicines Initiative 2 Joint Undertaking (grant 115902), which receives support from European Union Horizon 2020 and the European Federation of Pharmaceutical Industries and Associations. The study was conducted in accordance with the Declaration of Helsinki and Good Clinical Practice, adhering to principles outlined in the National Health Service Research Governance Framework for Health and Social Care (2nd edition).

Participant recruitment in Amsterdam was partially accomplished through Hersenonderzoek.nl [49], which is funded by ZonMw-Memorabel (project 73305095003).

This study was funded in part by the National Institute for Health Research (NIHR) Maudsley Biomedical Research Centre at South London and Maudsley NHS (National Health Service) Foundation Trust and King's College London.

The funding bodies have not been involved in the design of the study, the collection or analysis of data, or the interpretation of data, and the views expressed are those of the authors.

We thank following 4 clinical communities in Spain: *Parc Sanitari Sant Joan de Déu* Network services, *Institut Català de la Salut*, *Institut Pere Mata*, and *Hospital Clínico San Carlos* for recruiting participants in the CIBER site.

We thank all the members of the RADAR-CNS patient advisory board for their contribution to the device selection procedures, and their invaluable advice throughout the study protocol design. We acknowledge that this research was reviewed by the Feasibility and Acceptability Support Team for Researchers (National Institute for Health Research Maudsley Biomedical Research Centre, King's College London and South London and Maudsley NHS Foundation Trust).

We thank all GLAD Study volunteers for their participation and gratefully acknowledge the NIHR BioResource, NIHR BioResource centers, NHS Trusts and staff for their contribution. We also acknowledge NIHR Biomedical Research Centre, King's College London, South London and Maudsley NHS Trust and King's Health Partners. We thank the NIHR, NHS Blood and Transplant, and Health Data Research UK as part of the Digital Innovation Hub program.

MH and TW were supported by NIHR Senior Investigator Awards.

RB is funded in part by the King's College London Medical Research Council Skills Development Fellowship (grant MR/R016372/1), which is funded by the UK Medical Research Council (grant IS-BRC-1215-20018) for the NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London.

RJBD is supported by (1) NIHR Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London (2) Health Data Research UK, which is funded by the UK Medical Research Council, Engineering and Physical Sciences Research Council, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and Wellcome Trust; (3) the BigData@Heart Consortium, which funded by the Innovative Medicines Initiative-2 Joint Undertaking (grant 116074), which receives support from the European Union Horizon 2020 research and innovation program and European Federation of Pharmaceutical Industries and Associations; (4) the NIHR University College London Hospitals Biomedical Research Centre; (5) the UK Research and Innovation London Medical Imaging & Artificial Intelligence Centre for Value Based Health care; and (6) the NIHR Applied Research Collaboration South London at King's College Hospital NHS Foundation Trust.

Authors' Contributions

YZ extracted and integrated the questionnaire and location data for the analysis, planned and performed the analysis, and drafted the manuscript. MH and VAN gained funding and co-led the Remote Assessment of Disease and Relapse–Central Nervous System program. MH is the principal investigator for the Remote Assessment of Disease and Relapse–Major Depressive Disorder study. RJBD, AAF, YR, ZR, PC, HS, and CS have contributed to the development of the RADAR-base platform used for data collection and management across sites, data protection, security, and storage. YZ, AAF, S Sun, NC, SV, RB, PL, SB, DCM, MH, and RJBD contributed to the design of the study. FM, KMW, CO, AI, FL, S Siddi, EV, S Simblett, JMH, BWJHP, MH contributed to data collection. AAF, IMG, AR, VAN, TW, PA, MH, and RJBD contributed to the administrative, technical, and clinical support of the study. All authors were involved in reviewing the manuscript, had access to the study data, and provided direction and comments on the manuscript.

Conflicts of Interest

SV and VAN are employees of Janssen Research and Development LLC. PA is employed by the pharmaceutical company H. Lundbeck A/S. DCM has accepted honoraria and consulting fees from Apple Inc, Otsuka Pharmaceuticals, Pear Therapeutics, and the One Mind Foundation; has received royalties from Oxford Press; and has an ownership interest in Adaptive Health Inc.

MH is the principal investigator of the Remote Assessment of Disease and Relapse–Central Nervous System project, a private public precompetitive consortium that receives funding from Janssen, UCB, Lundbeck, MSD, and Biogen.

References

1. Ferrari AJ, Charlson FJ, Norman RE, Patten SB, Freedman G, Murray CJL, et al. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. PLoS Med 2013 Nov;10(11):e1001547 [FREE Full text] [doi: 10.1371/journal.pmed.1001547] [Medline: 24223526]
2. Beck A, Crain AL, Solberg LI, Unützer J, Glasgow RE, Maciosek MV, et al. Severity of depression and magnitude of productivity loss. Ann Fam Med 2011 Jul 11;9(4):305-311 [FREE Full text] [doi: 10.1370/afm.1260] [Medline: 21747101]
3. Cuijpers P, Smit F. Excess mortality in depression: a meta-analysis of community studies. J Affect Disord 2002 Dec;72(3):227-236. [doi: 10.1016/s0165-0327(01)00413-x] [Medline: 12450639]
4. Katon W, Ciechanowski P. Impact of major depression on chronic medical illness. J Psychosom Res 2002 Oct;53(4):859-863. [doi: 10.1016/s0022-3999(02)00313-6] [Medline: 12377294]
5. Simon GE. Social and economic burden of mood disorders. Biol Psychiatry 2003 Aug 01;54(3):208-215. [doi: 10.1016/s0006-3223(03)00420-7] [Medline: 12893097]
6. Zhang Y, Folarin AA, Sun S, Cummins N, Ranjan Y, Rashid Z, et al. Predicting depressive symptom severity through individuals' nearby bluetooth device count data collected by mobile phones: preliminary longitudinal study. JMIR Mhealth Uhealth 2021 Jul 30;9(7):e29840 [FREE Full text] [doi: 10.2196/29840] [Medline: 34328441]
7. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):617-627 [FREE Full text] [doi: 10.1001/archpsyc.62.6.617] [Medline: 15939839]
8. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]
9. Weyerer S, Kupfer B. Physical exercise and psychological health. Sports Med 1994 Feb;17(2):108-116. [doi: 10.2165/00007256-199417020-00003] [Medline: 8171221]
10. Perrino T, Mason CA, Brown SC, Szapocznik J. The relationship between depressive symptoms and walking among Hispanic older adults: a longitudinal, cross-lagged panel analysis. Aging Ment Health 2010 Mar;14(2):211-219 [FREE Full text] [doi: 10.1080/13607860903191374] [Medline: 20336553]
11. Roshanaei-Moghaddam B, Katon WJ, Russo J. The longitudinal effects of depression on physical activity. Gen Hosp Psychiatry 2009;31(4):306-315. [doi: 10.1016/j.genhosppsych.2009.04.002] [Medline: 19555789]
12. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-generation psychiatric assessment: using smartphone sensors to monitor behavior and mental health. Psychiatr Rehabil J 2015 Sep;38(3):218-226 [FREE Full text] [doi: 10.1037/prj0000130] [Medline: 25844912]
13. Chow PI, Fua K, Huang Y, Bonelli W, Xiong H, Barnes LE, et al. Using mobile sensing to test clinical models of depression, social anxiety, state affect, and social isolation among college students. J Med Internet Res 2017 Mar 03;19(3):e62 [FREE Full text] [doi: 10.2196/jmir.6820] [Medline: 28258049]
14. Farhan AA, Yue C, Morillo R. Behavior vs. introspection: refining prediction of clinical depression via smartphone sensing data. 2016 Presented at: 2016 IEEE Wireless Health; October 25-27, 2016; Bethesda, MD, USA p. 1-8. [doi: 10.1109/WH.2016.7764553]
15. Laiou P, Kaliukhovich DA, Folarin AA, Ranjan Y, Rashid Z, Conde P, RADAR-CNS. The association between home stay and symptom severity in major depressive disorder: preliminary findings from a multicenter observational study using geolocation data from smartphones. JMIR Mhealth Uhealth 2022 Jan 28;10(1):e28095 [FREE Full text] [doi: 10.2196/28095] [Medline: 35089148]
16. Lu J, Shang C, Yue C, Morillo R, Ware S, Kamath J, et al. Joint modeling of heterogeneous sensing data for depression assessment via multitask learning. Proc ACM Interact Mob Wearable Ubiquitous Techno 2018 Mar 26;2(1):1-21. [doi: 10.1145/3191753]
17. Pratap A, Atkins DC, Renn BN, Tanana MJ, Mooney SD, Anguera JA, et al. The accuracy of passive phone sensors in predicting daily mood. Depress Anxiety 2019 Jan 21;36(1):72-81 [FREE Full text] [doi: 10.1002/da.22822] [Medline: 30129691]
18. Saeb S, Lattie EG, Schueller SM, Kording KP, Mohr DC. The relationship between mobile phone location sensor data and depressive symptom severity. PeerJ 2016;4:e2537 [FREE Full text] [doi: 10.7717/peerj.2537] [Medline: 28344895]
19. Saeb S, Zhang M, Karr C, Schueller S, Corden M, Kording K, et al. Mobile phone sensor correlates of depressive symptom severity in daily-life behavior: an exploratory study. J Med Internet Res 2015 Jul 15;17(7):e175 [FREE Full text] [doi: 10.2196/jmir.4273] [Medline: 26180009]
20. Wang R, Chen F, Chen Z. StudentLife: assessing mental health, academic performance and behavioral trends of college students using smartphones. 2014 Presented at: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 13-17, 2014; Seattle, Washington URL: https://dl.acm.org/doi/10.1145/2632048.2632054 [doi: 10.1145/2632048.2632054]

21. Wang R, Wang W, daSilva A, Huckins JF, Kelley WM, Heatherton TF, et al. Tracking depression dynamics in college students using mobile phone and wearable sensing. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Mar 26;2(1):1-26. [doi: 10.1145/3191775]
22. Meyerhoff J, Liu T, Kording KP, Ungar LH, Kaiser SM, Karr CJ, et al. Evaluation of changes in depression, anxiety, and social anxiety using smartphone sensor features: longitudinal cohort study. J Med Internet Res 2021 Sep 03;23(9):e22844 [FREE Full text] [doi: 10.2196/22844] [Medline: 34477562]
23. Gana K, Bailly N, Broc G, Cazauvieilh C, Boudouda NE. The geriatric depression scale: does it measure depressive mood, depressive affect, or both? Int J Geriatr Psychiatry 2017 Oct 15;32(10):1150-1157. [doi: 10.1002/gps.4582] [Medline: 27633329]
24. Wichers M. The dynamic nature of depression: a new micro-level perspective of mental disorder that meets current challenges. Psychol Med 2014 May;44(7):1349-1360. [doi: 10.1017/S0033291713001979] [Medline: 23942140]
25. Rhodes R, Courneya K. Modelling the theory of planned behaviour and past behaviour. Psychol Health Med 2003 Feb 01;8(1):57-69. [doi: 10.1080/1354850021000059269] [Medline: 21888489]
26. Matcham F, Barattieri di San Pietro C, Bulgari V, de Girolamo G, Dobson R, Eriksson H, RADAR-CNS consortium. Remote assessment of disease and relapse in major depressive disorder (RADAR-MDD): a multi-centre prospective cohort study protocol. BMC Psychiatry 2019 Feb 18;19(1):72 [FREE Full text] [doi: 10.1186/s12888-019-2049-z] [Medline: 30777041]
27. Asparouhov T, Hamaker EL, Muthén B. Dynamic structural equation models. Struct Equ Modeling 2017 Dec 27;25(3):359-388. [doi: 10.1080/10705511.2017.1406803]
28. Matcham F, Leightley D, Siddi S, Lamers F, White KM, Annas P, RADAR-CNS consortium. Remote Assessment of Disease and Relapse in Major Depressive Disorder (RADAR-MDD): recruitment, retention, and data availability in a longitudinal remote measurement study. BMC Psychiatry 2022 Feb 21;22(1):136 [FREE Full text] [doi: 10.1186/s12888-022-03753-1] [Medline: 35189842]
29. Ranjan Y, Rashid Z, Stewart C, Conde P, Begale M, Verbeeck D, Hyve, RADAR-CNS Consortium. RADAR-base: Open source mobile health platform for collecting, monitoring, and analyzing data using sensors, wearables, and mobile devices. JMIR Mhealth Uhealth 2019 Aug 01;7(8):e11734 [FREE Full text] [doi: 10.2196/11734] [Medline: 31373275]
30. Kroenke K, Strine TW, Spitzer RL, Williams JBW, Berry JT, Mokdad AH. The PHQ-8 as a measure of current depression in the general population. J Affect Disord 2009 Apr;114(1-3):163-173. [doi: 10.1016/j.jad.2008.06.026] [Medline: 18752852]
31. RADAR-base. URL: https://radar-base.atlassian.net/wiki/spaces/RAD/overview [accessed 2022-02-28]
32. Sun S, Folarin AA, Ranjan Y, Rashid Z, Conde P, Stewart C, RADAR-CNS Consortium. Using smartphones and wearable devices to monitor behavioral changes during COVID-19. J Med Internet Res 2020 Sep 25;22(9):e19992 [FREE Full text] [doi: 10.2196/19992] [Medline: 32877352]
33. Zhang Y, Folarin AA, Sun S, Cummins N, Bendayan R, Ranjan Y, RADAR-CNS Consortium. Relationship between major depression symptom severity and sleep collected using a wristband wearable device: multicenter longitudinal observational study. JMIR Mhealth Uhealth 2021 Apr 12;9(4):e24604 [FREE Full text] [doi: 10.2196/24604] [Medline: 33843591]
34. Depp CA, Bashem J, Moore RC, Holden JL, Mikhael T, Swendsen J, et al. GPS mobility as a digital biomarker of negative symptoms in schizophrenia: a case control study. NPJ Digit Med 2019;2(2):108-302 [FREE Full text] [doi: 10.1038/s41746-019-0182-1] [Medline: 31728415]
35. Ester M, Kriegel H, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. In: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. 1996 Presented at: Second International Conference on Knowledge Discovery and Data Mining; August 2-4, 1996; Portland, Oregon p. 226-231.
36. Isabelle F, Dominique K, Statia E. Home away from home: a longitudinal study of the holiday appropriation process. Tourism Management 2019 Apr;71:327-336. [doi: 10.1016/j.tourman.2018.10.025]
37. Walker WH, Walton JC, DeVries AC, Nelson RJ. Circadian rhythm disruption and mental health. Transl Psychiatry 2020 Jan 23;10(1):28 [FREE Full text] [doi: 10.1038/s41398-020-0694-0] [Medline: 32066704]
38. McNeish D. Two-level dynamic structural equation models with small samples. Struct Equ Modeling 2019 Mar 28;26(6):948-966 [FREE Full text] [doi: 10.1080/10705511.2019.1578657] [Medline: 32863699]
39. McNeish D, Hamaker EL. A primer on two-level dynamic structural equation models for intensive longitudinal data in Mplus. Psychol Methods 2020 Oct;25(5):610-635. [doi: 10.1037/met0000250] [Medline: 31855015]
40. Rizvi SJ, Cyriac A, Grima E, Tan M, Lin P, Gallaugher LA, et al. Depression and employment status in primary and tertiary care settings. Can J Psychiatry 2015 Jan;60(1):14-22 [FREE Full text] [doi: 10.1177/070674371506000105] [Medline: 25886545]
41. Aluoja A, Leinsalu M, Shlik J, Vasar V, Luuk K. Symptoms of depression in the Estonian population: prevalence, sociodemographic correlates and social adjustment. J Affect Disord 2004 Jan;78(1):27-35. [doi: 10.1016/s0165-0327(02)00179-9] [Medline: 14672794]
42. Akhtar-Danesh N, Landeen J. Relation between depression and sociodemographic factors. Int J Ment Health Syst 2007 Sep 04;1(1):4 [FREE Full text] [doi: 10.1186/1752-4458-1-4] [Medline: 18271976]

43. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 2018 Dec 05;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
44. Muthén B, Muthén L, Asparouhov T. Regression and Mediation Analysis Using Mplus. Los Angeles, CA: Muthén & Muthén; 2017.
45. Postvacation blues. Wikipedia. URL: https://en.wikipedia.org/wiki/Postvacation_blues#cite_note-Lillywhite_2017-5 [accessed 2022-02-28]
46. Jafari J. Tourism models: the sociocultural aspects. Tourism Management 1987 Jun;8(2):151-159. [doi: 10.1016/0261-5177(87)90023-9]
47. Katz G, Durst R, Zislin Y, Barel Y, Knobler H. Psychiatric aspects of jet lag: review and hypothesis. Med Hypotheses 2001 Jan;56(1):20-23. [doi: 10.1054/mehy.2000.1094] [Medline: 11133250]
48. Garg A, Tai K. Comparison of statistical and machine learning methods in modelling of data with multicollinearity. Int J Model Identif Control 2013;18(4):295. [doi: 10.1504/ijmic.2013.053535]
49. Hersenonderzoek. URL: https://hersenonderzoek.nl [accessed 2022-02-28]

Abbreviations

NIHR: National Institute for Health Research
NHS: National Health Service
PHQ-8: 8-item Patient Health Questionnaire
RADAR-CNS: Remote Assessment of Disease and Relapse–Central Nervous System

Edited by J Torous; submitted 11.11.21; peer-reviewed by E Sükei; comments to author 02.12.21; revised version received 09.12.21; accepted 12.01.22; published 11.03.22.

Please cite as:
Zhang Y, Folarin AA, Sun S, Cummins N, Vairavan S, Bendayan R, Ranjan Y, Rashid Z, Conde P, Stewart C, Laiou P, Sankesara H, Matcham F, White KM, Oetzmann C, Ivan A, Lamers F, Siddi S, Vilella E, Simblett S, Rintala A, Bruce S, Mohr DC, Myin-Germeys I, Wykes T, Haro JM, Penninx BWJH, Narayan VA, Annas P, Hotopf M, Dobson RJB, RADAR-CNS consortium
Longitudinal Relationships Between Depressive Symptom Severity and Phone-Measured Mobility: Dynamic Structural Equation Modeling Study
JMIR Ment Health 2022;9(3):e34898
URL: https://mental.jmir.org/2022/3/e34898
doi:10.2196/34898
PMID:35275087

Original Paper

Usage Intensity of a Relapse Prevention Program and Its Relation to Symptom Severity in Remitted Patients With Anxiety and Depression: Pre-Post Study

Esther Krijnen-de Bruin[1,2,3], PhD; Anna DT Muntingh[1,2], PhD; Evelien M Bourguignon[2], MD; Adriaan Hoogendoorn[1,2], PhD; Otto R Maarsingh[4], MD, PhD; Anton JLM van Balkom[1,2], MD, PhD; Neeltje M Batelaan[1,2], MD, PhD; Annemieke van Straten[5], PhD; Berno van Meijel[1,3,6], PhD

[1]Department of Psychiatry, Amsterdam Public Health Research Institute, Amsterdam UMC, Amsterdam, Netherlands

[2]Department of Research and Innovation, GGZ inGeest Specialized Mental Health Care, Amsterdam, Netherlands

[3]Department of Health, Sports and Welfare, Research Group Mental Health Nursing, Inholland University of Applied Sciences, Amsterdam, Netherlands

[4]Department of General Practice, Amsterdam Public Health Research Institute, Amsterdam UMC, Amsterdam, Netherlands

[5]Department of Clinical, Neuro and Developmental Psychology, VU University Amsterdam, Amsterdam Public Health Research Institute, Amsterdam, Netherlands

[6]Parnassia Academy, Parnassia Psychiatric Institute, The Hague, Netherlands

Corresponding Author:
Esther Krijnen-de Bruin, PhD
Department of Health, Sports and Welfare
Research Group Mental Health Nursing
Inholland University of Applied Sciences
De Boelelaan 1109
Amsterdam, 1081 HV
Netherlands
Phone: 31 20 495 1111
Email: esther.krijnendebruin@inholland.nl

Abstract

Background: Given that relapse is common in patients in remission from anxiety and depressive disorders, relapse prevention is needed in the maintenance phase. Although existing psychological relapse prevention interventions have proven to be effective, they are not explicitly based on patients' preferences. Hence, we developed a blended relapse prevention program based on patients' preferences, which was delivered in primary care practices by mental health professionals (MHPs). This program comprises contact with MHPs, completion of core and optional online modules (including a relapse prevention plan), and keeping a mood and anxiety diary in which patients can monitor their symptoms.

Objective: The aims of this study were to provide insight into (1) usage intensity of the program (over time), (2) the course of symptoms during the 9 months of the study, and (3) the association between usage intensity and the course of symptoms.

Methods: The Guided E-healTh for RElapse prevention in Anxiety and Depression (GET READY) program was guided by 54 MHPs working in primary care practices. Patients in remission from anxiety and depressive disorders were included. Demographic and clinical characteristics, including anxiety and depressive symptoms, were collected via questionnaires at baseline and after 3, 6, and 9 months. Log data were collected to assess the usage intensity of the program.

Results: A total of 113 patients participated in the study. Twenty-seven patients (23.9%) met the criteria for the minimal usage intensity measure. The core modules were used by ≥70% of the patients, while the optional modules were used by <40% of the patients. Usage decreased quickly over time. Anxiety and depressive symptoms remained stable across the total sample; a minority of 15% (12/79) of patients experienced a relapse in their anxiety symptoms, while 10% (8/79) experienced a relapse in their depressive symptoms. Generalized estimating equations analysis indicated a significant association between more frequent face-to-face contact with the MHPs and an increase in both anxiety symptoms (β=.84, 95% CI .39-1.29) and depressive symptoms (β=1.12, 95% CI 0.45-1.79). Diary entries and the number of completed modules were not significantly associated with the course of symptoms.

Conclusions: Although the core modules of the GET READY program were used by most of the patients and all patients saw an MHP at least once, usage decreased quickly over time. Most patients remained stable while participating in the study. The significant association between the frequency of contact and the course of symptoms most likely indicates that those who received more support had more symptoms, and thus, it is questionable whether the support offered by the program was sufficient to prevent these patients from relapsing.

International Registered Report Identifier (IRRID): RR2-10.1186/s12888-019-2034-6

(JMIR Ment Health 2022;9(3):e25441) doi:10.2196/25441

KEYWORDS

relapse prevention; anxiety disorder; depressive disorder; eHealth; primary care practice; usage intensity; self-management; mobile phone

Introduction

Despite effective treatments for anxiety and depressive disorders [1,2], maintenance phase relapse rates are high. Indeed, up to 57% of remitted patients experience a relapse of either their index disorder or another anxiety or depressive disorder within 4 years of remission [3]. Hence, relapse prevention is crucial in the maintenance phase. Having access to a relapse prevention program could help patients recognize early warning signs of relapse and take appropriate actions to prevent full relapse. There are several relapse prevention programs currently available for patients with remitted anxiety or depressive disorders. Previous research on patients with a depressive disorder showed that psychological relapse prevention programs reduce both residual symptoms [4,5] and relapse rates by 36% compared to treatment-as-usual [6]. Most relapse prevention programs solely involve face-to-face (FTF) contact, but programs using web-based formats are increasingly available [7]. Although web-based programs have the advantage of being easily accessible and flexible [8], the majority of them have low usage and high attrition rates [9-11]. This potentially undermines their effectiveness. A possible limitation of existing relapse prevention programs is that they are not explicitly based on patients' preferences; taking these preferences into account can increase acceptance and adherence, which, in turn, enhances their effectiveness [12].

In the Netherlands, relapse prevention is provided by mental health professionals (MHPs) in primary care practices. However, many MHPs are unfamiliar with relapse prevention interventions, and the tools to support MHPs in providing relapse prevention are lacking [13]. Therefore, we developed the blended relapse prevention program "Guided E-healTh for REapse prevention in Anxiety and Depression" (GET READY), which is explicitly based on patients' preferences. The program aimed to prevent relapsing by promoting self-management skills. Patients' preferences were obtained via a "discrete choice experiment," in which a set of tasks comprising alternative hypothetical treatment options could be chosen by participants [14]. Patients preferred a relapse prevention program that included regular contact with a professional, flexible time investment based on their needs, and a personalized prevention plan. The purpose of the GET READY intervention was to provide a flexible program that could be used over a longer period depending on the symptom level of the patient. Based on these preferences, the GET READY program includes (1) regular FTF contact with an MHP, (2) web-based modules based on evidence-based (cognitive behavioral) interventions, divided into 2 core modules (including a personalized relapse prevention plan) and 12 optional modules, and (3) a mood and anxiety diary to monitor symptoms. Depending on the symptom level and needs of the patient, the program can be used over a longer period. This study examined (1) usage intensity of the program (over time), (2) the course of symptoms during the 9 months of the study, and (3) the association between usage intensity and the course of symptoms.

Methods

Design

The GET READY study was a pre-post study for remitted patients with an anxiety or depressive disorder [15]. This paper presents the results pertaining to the usage intensity of the GET READY program, the course of symptoms (at baseline and after 3, 6, and 9 months), and the association between usage intensity and the course of symptoms.

Setting

This study was conducted in 50 primary care practice settings across the Netherlands. In the Netherlands, most primary care physicians (PCPs) employ an MHP (ie, nurse, psychologist, or social worker) who provides support and treatment for patients with mild mental health problems. These MHPs were involved in the GET READY program. Alternatively, for those patients whose MHP was not participating in the study, the program was offered via an ambulatory mental health care center. Patients began with an FTF meeting with an MHP, whereby they started composing a personalized relapse prevention plan. Next, patients could access web-based modules and a weekly diary via their computer, tablet, or smartphone. They were able to send messages to their MHP, ask for feedback on completed modules, and schedule FTF meetings with their MHP. All MHPs received a 4-hour training course, in which background information on relapse prevention, strategies for relapse prevention, and practical advice on using the program was provided [15]. PCPs did not play an active role in the study, although some MHPs regularly discussed patients with the PCP in their primary care practice.

Participants

Patients were eligible to participate if they had received treatment in specialized mental health care centers for anxiety

or depressive disorder in the previous 2 years. After receiving acute phase treatment, they were referred to primary care services. They had to be in full or partial remission according to their MHP or clinician (clinical judgment), have scored 50 or higher on the Global Assessment of Functioning scale [16], be at least 18 years old, and be sufficiently fluent in Dutch. Patients were excluded if they were participating in another structured psychological intervention, had no access to the internet, or still received specialized treatment for a comorbid psychiatric disorder. Maintenance antidepressant use was allowed.

Procedures

We sought to recruit 50 MHPs and 126 patients for this study. Sample size calculations have been described elsewhere [15]. MHPs and patients were recruited from April 2017 to November 2018. Fifty-four MHPs working in primary care practices throughout the Netherlands were recruited via telephone, letters, advertisements on MHP websites, and through the researchers' professional networks. Informed consent was obtained from MHPs at the start of the training course. PCPs had to agree that MHPs participated in the GET READY study. Patients were recruited either by their MHP or by their clinician at the end of their treatment (N=113), who provided brief information about the study. If patients were interested in participating, then the MHP or clinician asked consent from the patient to provide their contact details to the researchers. Next, consenting patients were contacted by the researchers and received additional information. Informed consent was obtained prior to administering the baseline questionnaire. This questionnaire assessed whether patients met the inclusion criteria pertaining to remission by administering the Inventory of Depressive Symptomatology Self-Report (IDS-SR) and the Beck Anxiety Inventory (BAI). Remission was defined as a score of <39 on the IDS-SR and a score of <30 on the BAI. Scores above these cutoff points indicate severe symptoms that require additional treatment to relapse prevention [17,18]. Therefore, patients with a score of ≥39 on the IDS-SR or ≥30 on the BAI were excluded from the study.

Ethical Approval

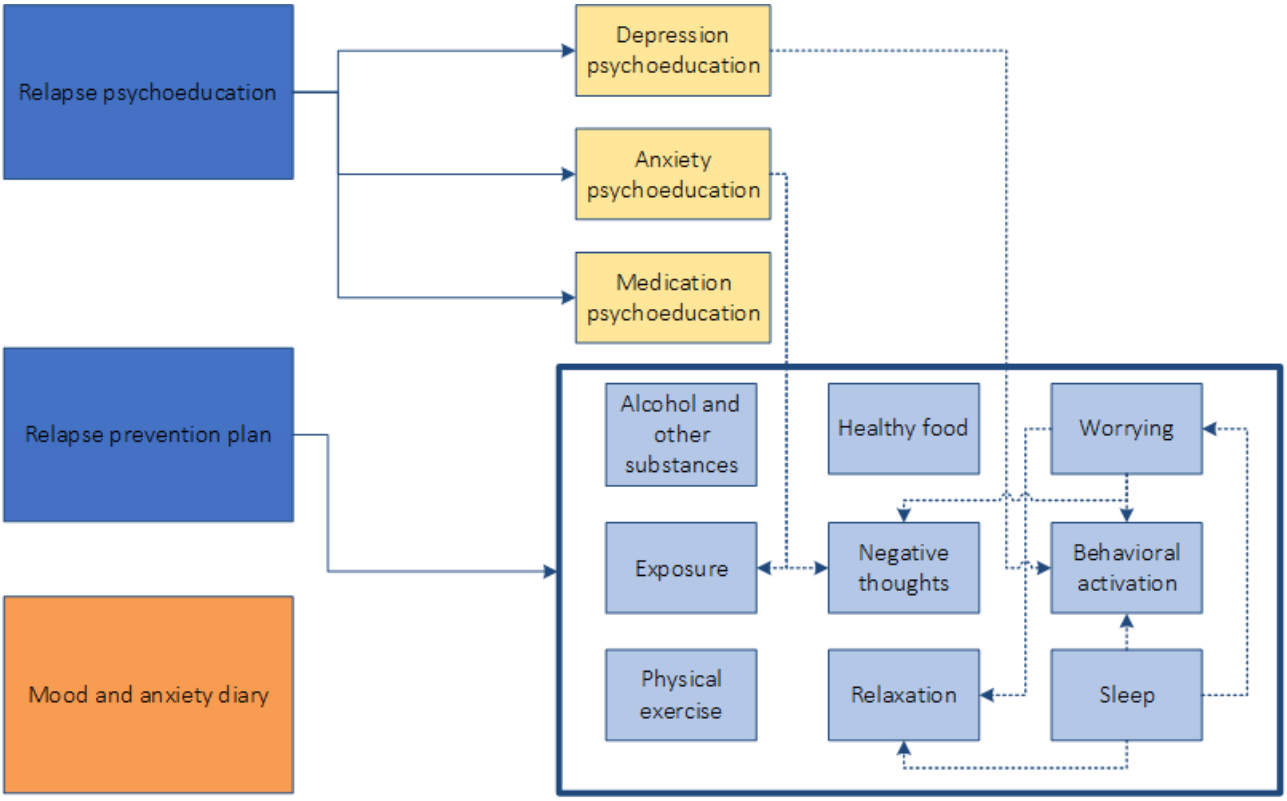
The Medical Ethics Committee of the Vrije Universiteit University Medical Center Amsterdam deemed that ethical approval was not required according to Dutch legislation (registration 2016.280) and thus gave their permission to conduct the study.

GET READY Program

The central aim of the program was to prevent relapse via the promotion of self-management skills. In the field of mental health, strengthening self-management skills is increasingly important, insofar as it allows patients to self-manage their own mental health [19]. More information regarding the content of the GET READY program has been published previously [15]. The program comprised several components. The program offered both FTF and web-based contact with an MHP. Every patient had at least 1 FTF engagement at the start of the study, and patients and MHPs were encouraged by the researchers to have FTF contact every 3 months. In addition, patients were encouraged to contact MHPs if their symptoms increased. In the FTF contact between patients and MHPs, usage of the program was discussed, and patients were encouraged to use the relapse prevention plan and to complete the diary and web-based modules. Patients were able to request feedback from their MHP when using the web-based modules. Besides the feedback on specific modules, patients and MHPs could also send and receive messages via the web-based platform. MHPs had access to their patients' data and could check whether they had logged in or if they had completed modules and the weekly diary. In the event that a patient did not complete a module within a week, they were sent an automatic reminder.

The web-based modules were divided into 2 core modules "relapse psychoeducation" and "relapse prevention plan" (see Multimedia Appendix 1) and 12 optional modules, which included 3 psychoeducation modules with information on depression, anxiety, and medication. The other 9 optional modules contained information on specific topics such as exposure, negative thoughts, and sleep (see Figure 1). These modules also contained exercises, videos, and examples of fictive patients. Some modules had overlapping themes, and patients could easily open these linked modules from the other module (see the dotted lines in Figure 1). Finally, the GET READY program included a "mood and anxiety diary," which allowed patients to monitor their symptoms. Patients received weekly reminders to complete the diary. When patients logged in for the first time, the core components "relapse psychoeducation," "relapse prevention plan," and the "mood and anxiety diary" were available. If patients completed the "relapse psychoeducation" module, the "depression/anxiety/medication psychoeducation" modules were automatically set up. Likewise, if patients completed the "relapse prevention plan," they could choose which optional modules they wish to complete based on their preferences and goals.

Figure 1. Overview of eHealth modules.



Data Collection

Patients were invited to complete questionnaires at baseline (T0) and after 3 (T1), 6 (T2), and 9 months (T3). Completion of the questionnaires took 20-30 minutes. If necessary, patients received an email reminder after 1 week. As part of the treatment protocol, patients were also prompted to complete the mood and anxiety diary once a week for a period of 9 months (39 times). MHPs were requested to complete a case registration form after each FTF contact, in which the clinical status of patients and the duration and content of the FTF contacts were described. In order to assess the usage intensity of the program, log data from the web-based platform were collected.

Measures

Demographic and Clinical Variables

Demographic and clinical variables of patients were assessed at baseline using the questionnaire. Moreover, in the baseline questionnaire, patients were asked to score their own perceived risk of relapse as well as their expectations about the effectiveness of the relapse prevention program (0%-100%). Anxiety severity was measured using the BAI, and symptoms in the past week were assessed. This questionnaire contains 21 items, all of which are rated on a 0- to 3-point scale, with a total score ranging from 0 to 63, with ≥30 indicating severe anxiety symptoms [20]. Severity of depression was measured using the IDS-SR [21]. Depressive symptoms in the past week were assessed. This questionnaire contains 30 items, all of which are rated on a 0- to 3-point scale, and when adding up 28 of the 30 items, the total score ranged from 0 to 84, with ≥39 indicating severe depressive symptoms [18]. To provide insight into the baseline clinical characteristics of patients, anxiety sensitivity and general functioning and disability were also measured. Anxiety sensitivity was measured using the Anxiety Sensitivity Index [22]. This questionnaire has 16 items, all of which are rated on a 0- to 4-point scale, with a total score ranging from 0 (no anxiety sensitivity) to 64 (severe anxiety sensitivity). General functioning and disability were measured using the World Health Organization Disability Assessment Schedule 2.0. This questionnaire has 36 items, all of which are rated on a 0- to 4-point scale, with a total score ranging from 0 (no disability) to 100 (full disability) [23]. Medication and health care use was measured using the Trimbos/iMTA Questionnaire for Costs Associated with Psychiatric Illness (TiC-P) [24].

Primary Outcome

Program Usage Variable

Log data from the web-based platform was used to assess the web-based usage intensity of the program. This included the number of messages from patients to MHPs or vice versa, the number of completed modules, and the number of diary entries. The frequency of FTF contact between patients and MHPs was registered with the TiC-P [24]. Participants were divided into low and regular users based on the median of the separate usage variables, as the data was nonnormally distributed. If participants had completed at least the median amount of FTF contact with the MHP (median 1), modules (median 4), and diary entries (median 4), then they were classified as regular users of these specific usage variables. If they completed less than the median of the separate usage variables, then they were considered to be low users. Furthermore, a “minimal usage intensity” measure was composed. If patients had at least 1 FTF contact during the intervention period, completed the core components of the program (relapse psychoeducation module, relapse prevention

plan, at least 4 mood and anxiety diary entries), and completed at least 1 extra module, then they were classified as regular users. If they did not complete these components, they were considered to be low users.

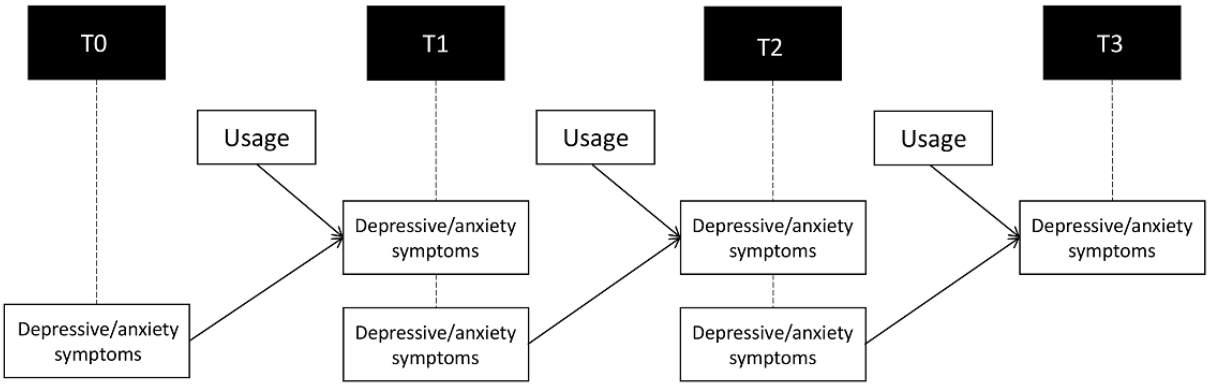
Course of Symptoms

To explore the course of symptoms during the study, the severity of anxiety and depressive symptoms was measured at baseline and 3-month intervals (T1, T2, T3) using the BAI and the IDS-SR. Deterioration/relapse was defined as an increase of at least 1 SD on the IDS-SR or of an increase on the BAI between T0 and T3. If there was an increase of 1 SD on the IDS-SR and the BAI, this was also regarded as deterioration/relapse. Similarly, symptom improvement was defined as a decrease of at least 1 SD. The SD was calculated using data from the Netherlands Study of Depression and Anxiety [25], the study of Kok et al [5], and this study, resulting in an SD of 9.3 on the IDS-SR and an SD of 6.6 on the BAI. By approaching the definition of relapse this way, patients can be regarded as their own controls, and an increase of 1 SD most likely indicates a clinically significant increase in symptoms, and thus indicate relapse.

Statistical Analysis

Descriptive statistics were used to describe the characteristics of the participants to illustrate the extent to which patients used the program (over time) and to explore the course of symptoms. Explorative analyses were conducted to study the association between usage intensity and the course of symptoms. The course of symptoms was determined for both regular and low users in accordance with the "minimal usage intensity" measure as well as for the separate usage intensity measures. Differences in anxiety and depressive symptoms between regular and low users were tested using the Mann-Whitney *U* test (as these were nonnormally distributed), while Bonferroni corrections were applied to correct for multiple testing [26]. Generalized estimating equations (GEE) analyses were carried out to examine the longitudinal association between the different usage intensity variables and the course of symptoms. The usage intensity variables indicated usage of the program between baseline and T1, T1 and T2, and T2 and T3. In this way, the association between usage intensity and the course of symptoms at the point of each follow-up questionnaire was assessed by taking into account the usage intensity in the period immediately prior to the follow-up questionnaire. A time-lag model was used, in which an adjustment was made for the outcome at time point $t-1$, as it assumed that the outcome at time t was predicted by the outcome at time $t-1$ (see Figure 2). All data analyses were performed using SPSS 26 (IBM Corp). Further details regarding the methods employed in this study can be found in the study protocol [15].

Figure 2. Time-lag model. T0: baseline assessment; T1: assessment after 3 months; T2: assessment after 6 months; T3: assessment after 9 months.



Results

Participant Characteristics

The demographic and clinical characteristics of the participants (N=113) are reported in Table 1. The mean age at baseline was 43 (SD 12.9) years. More than half of the participants were females (65/113, 57.5%), while more than half of the participants attended higher professional education or university (64/113, 56.6%). Overall, 36.3% (41/113) of the participants reported being treated for a depressive disorder, 23.9% (27/113) stated they had been treated for an anxiety disorder, and 39.8% (45/113) stated they had been treated for both.

Table 1. Demographic and clinical characteristics of the total sample (N=113).

Variables	Value
Demographic variables	
Age (years), mean (SD)	42.9 (12.9)
Sex (female), n (%)	65 (57.5)
Nationality (Dutch), n (%)	105 (92.9)
Marital status, n (%)	
Single	45 (39.8)
In relationship	68 (60.2)
Highest educational level, n (%)	
High school	23 (20.3)
Secondary vocational education	22 (19.5)
Higher professional education or university	64 (56.6)
Unknown	4 (3.6)
Occupational, n (%)	
Employed	79 (69.9)
Sick leave	18 (15.9)
Other	16 (14.2)
Clinical variables	
Clinical history, n (%)	
Treatment for depressive disorder	41 (36.3)
Treatment for anxiety disorder	27 (23.9)
Treatment for both depressive disorder and anxiety disorder	45 (39.8)
Number of times received treatment for mental health problems, mean (SD)	3.5 (3.3)
Time passed since referral back to the primary care physician from specialized care (months), mean (SD)	5.9 (6.3)
Age of first onset (years), mean (SD)	27.6 (13.8)
Positive family history of anxiety or depressive disorder, n (%)	60.0 (53.1)
Anxiety sensitivity, mean (SD)	10.7 (7.9)
General functioning and disability, mean (SD)	23.6 (15)
Anxiety severity, mean (SD)	10.2 (6.6)
Depression severity, mean (SD)	20.6 (9.5)

Usage Intensity

The use of the program is described in 3 subcategories: (1) contact with MHP, (2) completed modules, and (3) diary entries.

Contact With MHP

The option to correspond with MHPs via the web-based platform was rarely exercised by participants. In total, the 113 patients sent 157 messages to their MHPs (median 0 [IQR 0.0-2.0]) and received 260 messages in return from their MHPs (median 1.0 [IQR 0.0-3.0]). Sixty-five patients (57.5%) never sent a message to their MHP, and 45 patients (39.8%) never received a single message from their MHP. All participants had initial FTF contact with their MHPs. During the 9 months of the study, there were 260 FTF follow-up meetings (median 1.0 [IQR 0.0-4.0]). Forty-nine participants (43.4%) did not have any follow-up meetings with their MHP. Forty-one participants (36.3%) met their MHP at least every 3 months, as prescribed in the research protocol. The number of FTF appointments ranged from 0 to 13.

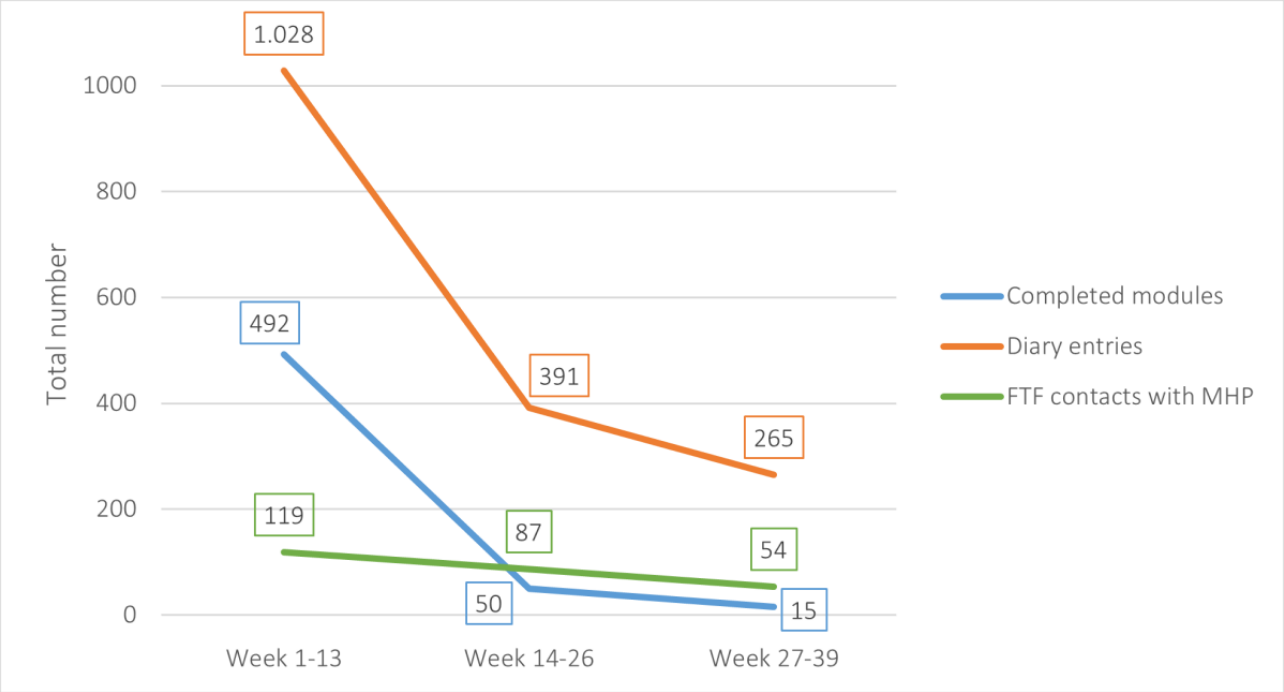
Completed Modules

A median of 4 modules were completed by the participants (IQR 2.0-8.0). Of the 113 participants, 1 (0.01%) completed all the 14 available modules, while 17 participants (15%) failed to complete any module. The 2 core modules were completed the most: 74.3% (84/113) of the participants completed the module “relapse psychoeducation” and 69.9% (79/113) completed the module “relapse prevention plan.” Approximately 46%-54% of the patients completed the other 3 psychoeducation modules, while less than 40% of patients completed the optional modules.

Diary Entries

The number of diary entries varied substantially across the participants, ranging from 0 to 159, with a median of 4 (IQR 1.0-15.0). Seventeen participants (15%) never reported on their mood and anxiety. Only 12 participants (10.6%) completed the diary weekly for the entire duration of the study. Usage of the program decreased considerably over time, as can be seen in Figure 3. In particular, there was a strong decrease in both the number of completed modules and number of diary entries. The median for when participants completed their last module was 31 (IQR 10.0-92.5) days after registering on the web-based platform.

Figure 3. Usage of the modules, diary, and face-to-face contacts over time. FTF: face-to-face; MHP: mental health professional.



Course of Symptoms

In the overall sample, anxiety and depressive symptoms decreased slightly over time. For all 113 participants, the mean BAI score at baseline was 10.2 (SD 6.6). After 9 months, the mean BAI score of the remaining 79 participants was 9.3 (SD 8.2). The differences over time were not significant, as indicated by the overlapping error bars in Figure S1 in Multimedia Appendix 2. A completer analysis (which included only those patients who completed T3) produced similar results. The mean IDS-SR score of all 113 participants decreased from 20.6 (SD 9.5) at baseline to 17.3 (SD 11.8) at 9 months (T3). The differences over time were not significant, as indicated by the overlapping error bars in Figure S2 in Multimedia Appendix 2. A completer analysis produced similar results. Regarding changes in symptomatology (stable/deteriorated/improved), for anxiety symptoms, it was found that the majority of the 79 patients who completed T3 remained stable over time (52/79, 66%), 12 patients (15%) experienced a deterioration of at least 1 SD (defined as a relapse), and 15 patients (19%) saw their anxiety symptoms improve. The numbers were comparable for depression symptoms: 53 patients (67%) remained stable, 8 patients (10%) deteriorated/relapsed, and 18 patients' (23%) depressive symptoms improved. Seven patients (9%) experienced a deterioration in both their anxiety and depressive symptoms. For all 3 categories (stable/deteriorated/improved), antidepressant medication use remained largely stable. For depressive symptoms, most patients in the stable and deterioration group used antidepressant medication: 57% (30/53) and 75% (6/8) respectively. In the improved group, 50% (9/18) used antidepressant medication. For anxiety symptoms, the majority of patients in the stable and deterioration group used antidepressant medication: 64% (33/52) and 58% (7/12) respectively. In the improved group, 33% (5/15) used antidepressant medication.

Association Between Usage Intensity and Course of Symptoms

Minimal Usage Intensity Measure

Of the 113 patients, 27 (23.9%) met the criteria for the "minimal usage intensity" measure, and hence, these patients were defined as regular users. Figure S3 in Multimedia Appendix 2 depicts the course of anxiety symptoms for both regular and low users, as measured by the combined "minimal usage intensity" measure. No significant differences were found in the anxiety symptoms between regular and low users (T0: U=1122.0, P=.79; T1: U=840.0, P=.52; T2: U=738.5, P=.79; T3: U=623.5, P=.59). Figure S4 in Multimedia Appendix 2 depicts the course of depressive symptoms for both regular and low users. Although the mean scores for regular users were higher across all the time points, no statistically significant differences were found between low and regular users (T0: U=1025.0, P=.36; T1: U=700.0, P=.07; T2: U=708.0, P=.57; T3: U=506.5, P=.08). The number of appointments in specialized mental health care facilities during the study did not significantly differ between regular and low users (T1: U=909.0, P=.92; T2: U=702.0, P=.39; T3: U=631.0, P=.55). There was a significant difference

between regular and low users in terms of the number of appointments they had with a psychologist or psychiatrist in a private practice at T3 (U=575.0, P=.04), that is, low users had more appointments than regular users. However, after applying the Bonferroni correction, this difference was no longer significant. No significant differences in medication use between regular and low users were found (T0: U=1148.0, P=.92; T1: U=900.5, P=.87; T2: U=727.5, P=.66; T3: U=645.0, P=.71).

Separate Usage Intensity Measures

The mean BAI scores and IDS-SR scores for regular users (median use of usage variable or higher) and low users (below median of usage variable) on the separate usage intensity measures across all 4 time points are shown in Multimedia Appendix 3. Patients who had 1 or more FTF meetings with their MHP after the initial FTF contact experienced a higher score on the BAI and IDS-SR at T3. For diary entries and the number of completed modules, the BAI and IDS-SR scores did not differ between regular and low users.

GEE Analyses

In the GEE analyses, all of the separate usage variables were used to model the course of anxiety and depressive symptoms in a multivariate analysis. GEE analyses indicated no significant association between module completion and number of diary entries and the course of anxiety or depressive symptoms (Table 2). A significant association was found between the frequency of FTF contact with MHPs and the course of anxiety and depressive symptoms. The coefficient of .84 (95% CI .39-1.29) indicates that each additional FTF meeting with an MHP was associated with an increased BAI score of .84 in the next measurement (corrected for the BAI score one measurement prior). Similarly, the coefficient of 1.12 (95% CI .45-1.79) indicates that each additional FTF meeting with an MHP was associated with an increased IDS-SR score of 1.12 in the next measurement (corrected for the IDS-SR score one measurement prior). Therefore, more FTF contact with MHPs was significantly associated with higher anxiety and depressive scores.

Table 2. Generalized estimating equations analysis of the longitudinal associations of separate usage intensity variables with anxiety and depressive symptoms.

	Anxiety symptoms			Depressive symptoms		
	β (SE)	95% CI	*P* value[a]	β (SE)	95% CI	*P* value
Module completion	.12 (.13)	–.15 to .38	.39	.06 (.12)	–.18 to .30	.65
Face-to-face contact with mental health professional	.84 (.23)	.39 to 1.29	*<.001*	1.12 (.34)	.45 to 1.79	*.001*
Diary completion	–.05 (.03)	–.10 to .003	.07	–.04 (.03)	–.09 to .02	.19

[a]*P* values <.05 were considered to indicate significance and are shown in italics.

Discussion

Principal Results

This study has shown the usage intensity of the GET READY relapse prevention program, explored the course of symptoms of participants across the duration of the study, and examined the association between usage intensity and the course of symptoms. The core modules were used by ≥70% of the patients, while optional modules were regarded as elective and used as such (<40% of the patients). Of the 113 patients, 27 (23.9%) were defined as regular users according to the minimal usage intensity measure. Usage of the self-management components of the program (the web-based modules and web-based mood and anxiety diary) decreased quickly over time. Although no causal effect of the GET READY intervention on the severity of psychopathology could be established owing to its pre-post design, it appeared that most patients remained stable or experienced symptom improvement while they engaged with the GET READY program. Having more FTF contact with MHPs was significantly associated with an increase in anxiety and depressive symptoms. The other usage intensity variables were not significantly associated with the course of symptoms. Overall, the participants were highly educated and employed. These results are consistent with those reported in other studies on web-based or eHealth interventions [27,28], which have shown that this group is more likely to use web-based interventions. Similarly to Kontos et al [28], we also found it difficult to access patients with lower educational levels, which is problematic given that this program might also be beneficial for this group. Therefore, future relapse prevention studies should attempt to access participants with lower educational levels by seeking input from this group during the developmental phase of interventions.

The core components of the program were used fairly well, as the 2 core modules “relapse psychoeducation” and “relapse prevention plan” were completed by 74.3% (84/113) and 69.9% (79/113) of patients, respectively. As expected, optional modules were used less frequently than the core modules, with less than 40% of patients completing them. This result is consistent with data from Hollandäre et al [29], who showed that their basic modules were used more often than optional modules. Nonetheless, the usage of the optional modules in this study was relatively low in comparison to that in other studies [30-33]. The average usage intensity in these other programs was higher, with around 50% of the participants completing all available modules. However, these studies varied in terms of the number of modules (n=3-12), not to mention that the web-based programs were not focused on relapse prevention but rather on treating anxiety and depressive disorders. In this study, patients had already finished treatment for anxiety or depressive disorders, were in remission, and therefore may not have felt the need to actively engage in a relapse prevention program. One explanation for the relatively low usage could be simply that the patients experienced treatment fatigue [34]. Another

explanation might be that since all the patients had received treatment prior to this study, the lessons learned from treatment were very much at the forefront of their mind and, as such, they felt no desire to repeat these lessons in the relapse prevention program. Indeed, it appeared that patients selected the modules that applied to their situation. To conclude, the optional modules in our relapse prevention program were regarded as elective and used as such.

Usage of the program decreased rapidly over time, as most patients used the program for a median of 1 month after registering on the web-based platform. Although this finding has been reported in previous studies on (web-based) guided self-help programs [10,11], it was contrary to the aim of the intervention, which was to provide a flexible program that can be used over a longer period of time depending on patients' symptom levels. Prior studies demonstrated that both the absence of symptoms and an increase of symptoms might hinder patients' capacity to actively use a program. Potentially, patients with fewer symptoms may not need to engage with the entire program to feel well again and cease using the program after obtaining the benefits [35,36]. At the same time, a qualitative study on the GET READY program indicated that increased symptom levels might also limit patients from further using the program [37]. An increase in depressive or anxiety symptoms may result in avoidance behavior, which may also lead to avoiding actively working on the web-based modules. This underlines the importance of the proactive role to be played by MHPs in terms of stimulating patients who are vulnerable to relapsing to continue using the program. Another important way of keeping patients engaged might be the further personalization of the intervention content, for example, by increasing the depth of tailored feedback, providing real-time feedback, and customizing the content based on current symptoms [9,37].

Although all of the patients were in remission, they nevertheless appeared vulnerable to relapse: patients had already received an average of 3.5 treatments in specialized care, 53.1% (60/113) had a family history of anxiety/depression, and 39.8% (45/113) had received treatment for both an anxiety and a depressive disorder, while their baseline mean symptom levels showed mild residual anxiety and depressive symptoms. In the overall sample, anxiety and depressive symptoms decreased slightly over time. Most participants remained stable, while 19%-23% of patients experienced symptom improvement. Only 10%-15% of the patients experienced a relapse. In comparison to other studies, our results show lower relapse rates [38,39]. Hardeveld et al [38] found that after 10 months, 20%-30% of patients experienced a relapse of their depressive symptoms. Similarly, Taylor et al [39] found that around 30% of patients experienced a relapse of their anxiety symptoms within a year. Although no causal pathway could be established in this pre-post study, these results nevertheless indicate that the GET READY program could potentially protect patients from relapse. However, as the definition of relapse differs across studies, comparing the results can be difficult. Therefore, efforts should be made in the field to reach a consensus regarding the definition and assessment of relapse in depression and anxiety disorders. Moreover, the effectiveness of the GET READY program in preventing relapse should be tested in a randomized controlled trial (RCT).

Patients who experienced a deterioration in symptoms more often used antidepressant medication than patients whose symptoms improved. However, this result should be interpreted with caution, as no causal pathway can be established. This study design is not feasible to investigate the influence of medication on the course of symptoms. Patients who had more FTF contact with MHPs had significantly higher anxiety and depressive scores than patients who had less FTF contact with MHPs. It is questionable whether the support they received by their MHP was sufficient to engender a subsequent decrease in symptoms. At the same time, this result might indicate that the web-based program in itself does not provide enough support to patients who experience a deterioration of symptoms. As aforesaid, this is a pre-post study; therefore, no causal pathway could be established [40]. An alternative explanation for this significant association might be that patients adequately responded to early symptoms of relapse by reaching out to their MHP. This explanation is in line with an earlier study, which showed that patients with more severe symptoms were more likely to receive help [41]. Furthermore, no evidence was found for an association between the number of completed modules and diary entries on the one hand and the course of symptoms on the other hand. Although no comparison with other relapse prevention programs is currently possible, other treatment studies have found that more completed web-based modules are associated with better anxiety and depression outcomes [31,42]. The same applies to the number of diary entries [43]. When comparing this study to these studies, it becomes apparent that the sample size of these studies was larger. Therefore, as well as the difference in population (remitted vs present disorder) and aims of the program (relapse prevention vs treatment), these studies may also have had more scope to detect an association.

Limitations

There are several potential limitations in this study. First, attrition from the study was relatively high, with only 79 (69.9%) of the 113 participants completing the last follow-up questionnaire. Despite this, the statistical methods applied in this study, especially GEE analyses, are expedient for handling missing data. Second, self-selection bias and the fact that the patients were highly educated might restrict the generalizability of the results. However, Donkin et al's [44] study showed no indication that these factors actually limit the generalizability, as they found no evidence that these factors were related to study outcomes. Third, this study had a limited follow-up period of 9 months. A longer follow-up period of 2 years would have provided greater insight into the course of anxiety and depressive symptoms over a longer period of time, which, in turn would have facilitated better comparison with other studies [38,45]. However, for pragmatic reasons, it was not feasible to extend the follow-up period. Fourth, owing to methodological considerations, no time-to-event analysis could be performed, as the assumptions for this analysis could not be met. Finally, the definitions of "regular use" that were used in this study should be interpreted with caution. As the median of several usage intensity measures was relatively low (ie, FTF contact median=1), regular use could still indicate relatively low usage when compared to the intended amount of usage (ie, FTF contact

once every 3 months=3 FTF meetings). However, to enhance readability we opted to use the terms “regular use” and “low use.”

Implications for Practice and Research

This study highlights the importance of providing personalized and guided relapse prevention to remitted patients with anxiety and depressive disorders. Usage of the program decreased quickly over time, possibly indicating a rapid decrease in the motivation of patients. As aforesaid, this decrease in motivation can be explained by different causal factors. Therefore, MHPs have the important task of monitoring and motivating patients via personalized intervention strategies, thus ensuring that patients receive guidance when they need it the most. Further research in an RCT with a longer follow-up duration is necessary to establish the effectiveness of blended relapse prevention programs. Within the design of an RCT, greater insight can also be obtained into the association between usage intensity and the course of symptoms.

Conclusions

When relapse prevention was offered, most patients used the core modules, while optional modules were completed by a smaller sample. As indicated in an earlier study [14], the patients showed that they preferred a low level of time investment for relapse prevention programs. Despite the relatively low usage and low time investment, most patients remained stable while participating in the GET READY study. Patients who had more FTF contact with their MHP experienced more anxiety and depressive symptoms. Owing to the pre-post design of this study, no causal pathway could be established. An RCT is needed to provide insight into the effectiveness of the GET READY program and to further explore the causal relationship between usage intensity and the course of symptoms.

Acknowledgments

ADTM, NMB, AvS, and BvM obtained funding for this study. This study is funded by SIA-RAAK: The Taskforce for Applied Research, part of the Netherlands Organization for Scientific Research. The funder had no role in study design, data collection and analysis, interpretation of data, and preparation of the manuscript. The authors gratefully acknowledge the contribution of all patients, mental health professionals, primary care practices, research assistants, data managers, and all others who contributed to the study.

Authors' Contributions

EKdB, ADTM, NMB, AvS, ORM, AJLMvB, and BvM contributed to the design of the study. EKdB, ADTM, and BvM coordinated the recruitment of mental health professionals and patients and the data collection. AH advised in data analysis. EMB participated in data analysis. EKdB analyzed the data and drafted the manuscript. All authors revised and commented on the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Home page for patients with core modules, namely, relapse psychoeducation and relapse prevention plan, and the mood and anxiety diary.
[PDF File (Adobe PDF File), 188 KB - mental_v9i3e25441_app1.pdf]

Multimedia Appendix 2
Course of symptoms specified by usage intensity.
[PDF File (Adobe PDF File), 228 KB - mental_v9i3e25441_app2.pdf]

Multimedia Appendix 3
Low and regular use of separate usage intensity measures.
[PDF File (Adobe PDF File), 149 KB - mental_v9i3e25441_app3.pdf]

References

1. Bandelow B, Sagebiel A, Belz M, Görlich Y, Michaelis S, Wedekind D. Enduring effects of psychological treatments for anxiety disorders: meta-analysis of follow-up studies. Br J Psychiatry 2018 Jun;212(6):333-338. [doi: 10.1192/bjp.2018.49] [Medline: 29706139]
2. Cuijpers P. The Challenges of Improving Treatments for Depression. JAMA 2018 Dec 25;320(24):2529-2530. [doi: 10.1001/jama.2018.17824] [Medline: 30500053]
3. Scholten WD, Batelaan NM, Penninx BW, van Balkom AJLM, Smit JH, Schoevers RA, et al. Diagnostic instability of recurrence and the impact on recurrence rates in depressive and anxiety disorders. J Affect Disord 2016 May;195:185-190. [doi: 10.1016/j.jad.2016.02.025] [Medline: 26896812]

4. Hoorelbeke K, Koster EHW. Internet-delivered cognitive control training as a preventive intervention for remitted depressed patients: Evidence from a double-blind randomized controlled trial study. J Consult Clin Psychol 2017 Feb;85(2):135-146. [doi: 10.1037/ccp0000128] [Medline: 27362792]
5. Kok G, Burger H, Riper H, Cuijpers P, Dekker J, van Marwijk H, et al. The Three-Month Effect of Mobile Internet-Based Cognitive Therapy on the Course of Depressive Symptoms in Remitted Recurrently Depressed Patients: Results of a Randomized Controlled Trial. Psychother Psychosom 2015;84(2):90-99. [doi: 10.1159/000369469] [Medline: 25721915]
6. Biesheuvel-Leliefeld K, Kok G, Bockting C, Cuijpers P, Hollon S, van Marwijk HWJ, et al. Effectiveness of psychological interventions in preventing recurrence of depressive disorder: meta-analysis and meta-regression. J Affect Disord 2015 Mar 15;174:400-410. [doi: 10.1016/j.jad.2014.12.016] [Medline: 25553400]
7. Hennemann S, Farnsteiner S, Sander L. Internet- and mobile-based aftercare and relapse prevention in mental disorders: A systematic review and recommendations for future research. Internet Interv 2018 Dec;14:1-17 [FREE Full text] [doi: 10.1016/j.invent.2018.09.001] [Medline: 30510909]
8. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
9. Beatty L, Binnion C. A Systematic Review of Predictors of, and Reasons for, Adherence to Online Psychological Interventions. Int J Behav Med 2016 Dec;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]
10. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
11. Kelders SM, Bohlmeijer ET, Van Gemert-Pijnen JE. Participants, usage, and use patterns of a web-based intervention for the prevention of depression within a randomized controlled trial. J Med Internet Res 2013 Aug 20;15(8):e172. [doi: 10.2196/jmir.2258] [Medline: 23963284]
12. Cuijpers P, van Straten A, Warmerdam L, van Rooy M. Recruiting participants for interventions to prevent the onset of depressive disorders: possible ways to increase participation rates. BMC Health Serv Res 2010 Jun 25;10:181 [FREE Full text] [doi: 10.1186/1472-6963-10-181] [Medline: 20579332]
13. Hermens MLM, Muntingh A, Franx G, van Splunteren PT, Nuyen J. Stepped care for depression is easy to recommend, but harder to implement: results of an explorative study within primary care in the Netherlands. BMC Fam Pract 2014 Jan 09;15:5 [FREE Full text] [doi: 10.1186/1471-2296-15-5] [Medline: 24400701]
14. Muntingh ADT, Hoogendoorn AW, Van Schaik DJF, Van Straten A, Stolk EA, Van Balkom AJLM, et al. Patient preferences for a guided self-help programme to prevent relapse in anxiety or depression: A discrete choice experiment. PLoS One 2019;14(7):e0219588 [FREE Full text] [doi: 10.1371/journal.pone.0219588] [Medline: 31318918]
15. Krijnen-de Bruin E, Muntingh ADT, Hoogendoorn AW, van Straten A, Batelaan NM, Maarsingh OR, et al. The GET READY relapse prevention programme for anxiety and depression: a mixed-methods study protocol. BMC Psychiatry 2019 Feb 11;19(1):64 [FREE Full text] [doi: 10.1186/s12888-019-2034-6] [Medline: 30744601]
16. American Psychiatric Association. DSM-IV Diagnostic and Statistical Manual of Mental Disorders. Washington DC; 1994.
17. Beck A, Steer R. Beck Anxiety Inventory Manual. San Antonio, TX: The Psychological Corporation Harcourt Brace & Company; 1993.
18. Inventory of Depressive Symptomatology (IDS) and Quick Inventory of Depressive Symptomatology (QIDS) - Interpretation. URL: http://www.ids-qids.org/interpretation.html [accessed 2018-05-22]
19. Kemp V. Use of 'chronic disease self-management strategies' in mental healthcare. Current Opinion in Psychiatry 2011:1. [doi: 10.1097/yco.0b013e3283438014]
20. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. J Consult Clin Psychol 1988 Dec;56(6):893-897. [doi: 10.1037//0022-006x.56.6.893] [Medline: 3204199]
21. Rush AJ, Gullion CM, Basco MR, Jarrett RB, Trivedi MH. The Inventory of Depressive Symptomatology (IDS): psychometric properties. Psychol Med 1996 May;26(3):477-486. [doi: 10.1017/s0033291700035558] [Medline: 8733206]
22. Reiss S, Peterson RA, Gursky DM, McNally RJ. Anxiety sensitivity, anxiety frequency and the prediction of fearfulness. Behav Res Ther 1986;24(1):1-8. [doi: 10.1016/0005-7967(86)90143-9] [Medline: 3947307]
23. WHO disability assessment schedule 2.0. WHO. URL: http://www.who.int/classifications/icf/WHODAS2.0_36itemsSELF.pdf [accessed 2018-08-14]
24. Hakkaart-van Roijen L, van Straten A, Tiemens B, Donker M. Trimbos/iMTA questionnaire for costs associated with psychiatric illness (TiC-P). Research Gate. 2002. URL: https://www.researchgate.net/publication/254758194_Manual_TrimbosiMTA_Questionnaire_for_Costs_Associated_with_Psychiatric_Illness_TIC-P_in_Dutch [accessed 2022-02-18]
25. Penninx BW, Beekman AT, Smit JH, Zitman FG, Nolen WA, Spinhoven P, NESDA Research Consortium. The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. Int J Methods Psychiatr Res 2008;17(3):121-140 [FREE Full text] [doi: 10.1002/mpr.256] [Medline: 18763692]
26. Armstrong RA. When to use the Bonferroni correction. Ophthalmic Physiol Opt 2014 Sep;34(5):502-508. [doi: 10.1111/opo.12131] [Medline: 24697967]

27. Torrent-Sellens J, Díaz-Chao Á, Soler-Ramos I, Saigí-Rubió F. Modelling and Predicting eHealth Usage in Europe: A Multidimensional Approach From an Online Survey of 13,000 European Union Internet Users. J Med Internet Res 2016 Jul 22;18(7):e188 [FREE Full text] [doi: 10.2196/jmir.5605] [Medline: 27450189]
28. Kontos E, Blake KD, Chou WS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014 Jul 16;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
29. Holländare F, Johnsson S, Randestad M, Tillfors M, Carlbring P, Andersson G, et al. Randomized trial of Internet-based relapse prevention for partially remitted depression. Acta Psychiatr Scand 2011 Oct;124(4):285-294. [doi: 10.1111/j.1600-0447.2011.01698.x] [Medline: 21401534]
30. Donkin L, Hickie IB, Christensen H, Naismith SL, Neal B, Cockayne NL, et al. Rethinking the dose-response relationship between usage and outcome in an online intervention for depression: randomized controlled trial. J Med Internet Res 2013 Oct 17;15(10):e231 [FREE Full text] [doi: 10.2196/jmir.2771] [Medline: 24135213]
31. Hilvert-Bruce Z, Rossouw PJ, Wong N, Sunderland M, Andrews G. Adherence as a determinant of effectiveness of internet cognitive behavioural therapy for anxiety and depressive disorders. Behav Res Ther 2012 Aug;50(7-8):463-468. [doi: 10.1016/j.brat.2012.04.001] [Medline: 22659155]
32. Van Gemert-Pijnen JE, Kelders SM, Bohlmeijer ET. Understanding the usage of content in a mental health intervention for depression: an analysis of log data. J Med Internet Res 2014 Jan 31;16(1):e27 [FREE Full text] [doi: 10.2196/jmir.2991] [Medline: 24486914]
33. van Straten A, Cuijpers P, Smits N. Effectiveness of a web-based self-help intervention for symptoms of depression, anxiety, and stress: randomized controlled trial. J Med Internet Res 2008 Mar 25;10(1):e7 [FREE Full text] [doi: 10.2196/jmir.954] [Medline: 18364344]
34. Heckman BW, Mathew AR, Carpenter MJ. Treatment Burden and Treatment Fatigue as Barriers to Health. Curr Opin Psychol 2015 Oct 01;5:31-36 [FREE Full text] [doi: 10.1016/j.copsyc.2015.03.004] [Medline: 26086031]
35. Neil AL, Batterham P, Christensen H, Bennett K, Griffiths KM. Predictors of adherence by adolescents to a cognitive behavior therapy website in school and community-based settings. J Med Internet Res 2009 Feb 23;11(1):e6 [FREE Full text] [doi: 10.2196/jmir.1050] [Medline: 19275982]
36. Postel MG, de Haan HA, ter Huurne ED, Becker ES, de Jong CAJ. Effectiveness of a web-based intervention for problem drinkers and reasons for dropout: randomized controlled trial. J Med Internet Res 2010 Dec 16;12(4):e68 [FREE Full text] [doi: 10.2196/jmir.1642] [Medline: 21163776]
37. Krijnen-de Bruin E, Geerlings JA, Muntingh AD, Scholten WD, Maarsingh OR, van Straten A, et al. Evaluation of a Blended Relapse Prevention Program for Anxiety and Depression in General Practice: Qualitative Study. JMIR Form Res 2021 Feb 16;5(2):e23200 [FREE Full text] [doi: 10.2196/23200] [Medline: 33591277]
38. Hardeveld F, Spijker J, De Graaf R, Hendriks S, Licht C, Nolen W, et al. Recurrence of major depressive disorder across different treatment settings: results from the NESDA study. J Affect Disord 2013 May;147(1-3):225-231 [FREE Full text] [doi: 10.1016/j.jad.2012.11.008] [Medline: 23218899]
39. Taylor JH, Jakubovski E, Bloch MH. Predictors of anxiety recurrence in the Coordinated Anxiety Learning and Management (CALM) trial. J Psychiatr Res 2015 Jun;65:154-165 [FREE Full text] [doi: 10.1016/j.jpsychires.2015.03.020] [Medline: 25896121]
40. Thiese MS. Observational and interventional study design types; an overview. Biochem Med (Zagreb) 2014;24(2):199-210 [FREE Full text] [doi: 10.11613/BM.2014.022] [Medline: 24969913]
41. ten Have M, de Graaf R, Vollebergh W, Beekman A. What depressive symptoms are associated with the use of care services? Journal of Affective Disorders 2004 Jun;80(2-3):239-248. [doi: 10.1016/s0165-0327(03)00132-0]
42. Christensen H, Griffiths KM, Korten A. Web-based cognitive behavior therapy: analysis of site usage and changes in depression and anxiety scores. J Med Internet Res 2002;4(1):e3 [FREE Full text] [doi: 10.2196/jmir.4.1.e3] [Medline: 11956035]
43. de Graaf LE, Huibers MJH, Riper H, Gerhards SAH, Arntz A. Use and acceptability of unsupported online computerized cognitive behavioral therapy for depression and associations with clinical outcome. J Affect Disord 2009 Aug;116(3):227-231. [doi: 10.1016/j.jad.2008.12.009] [Medline: 19167094]
44. Donkin L, Hickie IB, Christensen H, Naismith SL, Neal B, Cockayne NL, et al. Sampling bias in an internet treatment trial for depression. Transl Psychiatry 2012 Oct 23;2:e174 [FREE Full text] [doi: 10.1038/tp.2012.100] [Medline: 23092978]
45. Bockting C, Klein N, Elgersma H, van Rijsbergen G, Slofstra C, Ormel J, et al. Effectiveness of preventive cognitive therapy while tapering antidepressants versus maintenance antidepressant treatment versus their combination in prevention of depressive relapse or recurrence (DRD study): a three-group, multicentre, randomised controlled trial. The Lancet Psychiatry 2018 May;5(5):401-410. [doi: 10.1016/s2215-0366(18)30100-7]

Abbreviations

BAI: Beck Anxiety Inventory
FTF: face-to-face

GEE: generalized estimating equations
GET READY: Guided E-healTh for RElapse prevention in Anxiety and Depression
IDS-SR: Inventory of Depressive Symptomatology Self-Report
MHP: mental health professional
PCP: primary care physician
RCT: randomized controlled trial
TiC-P: Trimbos/iMTA Questionnaire for Costs Associated with Psychiatric Illness

Edited by J Torous; submitted 02.11.20; peer-reviewed by K Mathiasen, CH Cho; comments to author 07.03.21; revised version received 29.04.21; accepted 06.01.22; published 16.03.22.

Please cite as:
Krijnen-de Bruin E, Muntingh ADT, Bourguignon EM, Hoogendoorn A, Maarsingh OR, van Balkom AJLM, Batelaan NM, van Straten A, van Meijel B
Usage Intensity of a Relapse Prevention Program and Its Relation to Symptom Severity in Remitted Patients With Anxiety and Depression: Pre-Post Study
JMIR Ment Health 2022;9(3):e25441
URL: https://mental.jmir.org/2022/3/e25441
doi:10.2196/25441
PMID:35293876

Original Paper

Utilizing Big Data From Google Trends to Map Population Depression in the United States: Exploratory Infodemiology Study

Alex Wang[1], BSc; Robert McCarron[1], DO; Daniel Azzam[1], MD; Annamarie Stehli[1], MPH; Glen Xiong[2], MD; Jeremy DeMartini[2], MD

[1]Department of Psychiatry and Human Behavior, University of California, Irvine, Orange, CA, United States

[2]Department of Psychiatry & Behavioral Sciences, University of California, Davis, Sacramento, CA, United States

Corresponding Author:
Alex Wang, BSc
Department of Psychiatry and Human Behavior
University of California, Irvine
101 The City Dr S
Orange, CA, 92868
United States
Phone: 1 8584059768
Email: wangaj3@uci.edu

Abstract

Background: The epidemiology of mental health disorders has important theoretical and practical implications for health care service and planning. The recent increase in big data storage and subsequent development of analytical tools suggest that mining search databases may yield important trends on mental health, which can be used to support existing population health studies.

Objective: This study aimed to map depression search intent in the United States based on internet-based mental health queries.

Methods: Weekly data on mental health searches were extracted from Google Trends for an 11-year period (2010-2021) and separated by US state for the following terms: “feeling sad,” “depressed,” “depression,” “empty,” “insomnia,” “fatigue,” “guilty,” “feeling guilty,” and “suicide.” Multivariable regression models were created based on geographic and environmental factors and normalized to the following control terms: “sports,” “news,” “google,” “youtube,” “facebook,” and “netflix.” Heat maps of population depression were generated based on search intent.

Results: Depression search intent grew 67% from January 2010 to March 2021. Depression search intent showed significant seasonal patterns with peak intensity during winter (adjusted $P<.001$) and early spring months (adjusted $P<.001$), relative to summer months. Geographic location correlated with depression search intent with states in the Northeast (adjusted $P=.01$) having higher search intent than states in the South.

Conclusions: The trends extrapolated from Google Trends successfully correlate with known risk factors for depression, such as seasonality and increasing latitude. These findings suggest that Google Trends may be a valid novel epidemiological tool to map depression prevalence in the United States.

(JMIR Ment Health 2022;9(3):e35253) doi:10.2196/35253

KEYWORDS

depression; epidemiology; internet; google trends; big data; mental health

Introduction

Over the past few decades, the amount of data stored, transferred, and analyzed has grown extensively, with the big data market reaching a value of US $139 billion in 2020 [1]. The term “big data” was coined in 2005 in reference to a large set of data that was essentially impossible to manage and process using traditional methods and tools [2]. As industries and companies have developed analytic tools targeted toward big data, information that was once inaccessible is now obtainable. One of the most important applications of big data in medicine is extrapolating trends and using them to support health care groups and organizations seeking to understand population health changes and predict the future.

Google Trends is a free online tool developed by Google LLC in 2008 that allows users from anywhere in the world to analyze big data [3]. It tracks search content across various countries and languages and compares relative search intent between 2

or more terms. The usefulness of Google Trends was demonstrated in 2009: Ginsberg et al [4] published a groundbreaking study predicting the spread of influenza earlier than the Centers for Disease Control and Prevention (CDC). Google Trends was subsequently utilized to predict the outbreaks of many viruses, including the West Nile virus, norovirus, varicella, influenza, and HIV [5-8]. More recently, Google Trends has been frequently used to study a variety of health care domains, including the COVID-19 pandemic [9].

Depression is the most common psychiatric disorder in the United States, with 18.5% of adults experiencing symptoms of depression in 2019 [10,11]. Since the start of the COVID-19 pandemic, the prevalence of depression symptoms has increased to 27.8%, affecting an estimated 91.2 million Americans [12]. Epidemiological data for depression have traditionally been collected through surveys. Major organizations such as the National Institute of Mental Health (NIMH), Anxiety & Depression Association of America (ADAA), and CDC provide only limited data specific to the time and population being studied from their surveys [13-15]. In response to the COVID-19 pandemic, the CDC and US Census Bureau collaborated to track mental health in the United States [16].

In this study, we provide estimates of depression search intent across the United States using big data from Google Trends. Our analysis fills the gap in current depression epidemiology, which is mainly derived from voluntary surveys, by extrapolating trends from big data across time and space. We provide an analysis of how internet search intent can be used to map population depression and how this can be compared in relation to depression risk factors. This model serves as a proof of concept that analyzing big data in association with environmental and geographic factors can be used as an epidemiological tool for psychiatric disease surveillance models. In terms of population health, analysis of Google Trends depression search intent represents a digital epidemiological tool that may one day be used for real-time surveillance of high-risk and underserved populations. The trends accessed through internet data may one day guide public policies, workforce supply decisions, and allocation of resources.

Methods

Google Trends

The following methodologies were designed based on published methods [17-19]. All search queries entered into Google's search engine become anonymized and grouped based on both the general query topic and the specific keywords entered. Google Trends interprets the information and normalizes the data into an index between 0 and 100. The numbers represent the search interest relative to the highest point based on the given location and time frame within the query. A value of 100 represents highest search popularity for a term, and a value of 50 represents half the search popularity for a term [20].

To examine the US population's interest in depression, we completed a series of search queries in Google Trends between January 1, 2010 and March 1, 2021. Data sets were downloaded for symptoms and terms listed by the American Psychiatric Association for major depressive disorder: "feeling sad," "depressed," "depression," "empty," "insomnia," "fatigue," "guilty," "feeling guilty," and "suicide" [21]. To account for random variance and overall increases in search queries, data sets were also downloaded across similar time periods for control terms based on previously published studies and popular internet search terms: "sports," "news," "google," "youtube," "facebook," and "netflix" [18,19,22]. The values of depression search intent were summed and normalized relative to the control terms for the given region and time and are represented on a scale of 0 to 100 arbitrary units (AU).

Two separate data sets were extracted from Google Trends. The first data set represents the entire US public interest in depression over time with a data frequency of monthly averages from January 1, 2010 to March 1, 2021. The second data set represents public interest in depression on a statewide level collected as a single value per state averaged from January 1, 2010 to March 1, 2021.

Environmental and Geographic Risk Factors

Given the known phenomenon of seasonal affective disorder, we obtained the annual temperature, humidity, and sunshine percentage from 1971 to 2000 from the National Climatic Center to assess for environmental and geographic risk factors of depression [23]. The sunshine percentage represents the percentage of time between sunrise and sunset that the sun reaches the earth's surface. For the Air Quality Index (AQI), we obtained data from the 2010 to 2014 American Community Survey [24]. Values from the AQI were calculated for 4 major air pollutants regulated by the Clean Air Act [25]. Lastly, data for urban percentage were obtained from the 2010 US Census [26].

Statistical Analysis

Multiple linear regression models were conducted to analyze the relationship between depression search queries and environmental factors and geographic factors. Confounding variables were identified using a correlation matrix and appropriately removed. The *P* values for each variable were adjusted according to the Bonferroni correction for multiple comparisons, with statistical significance determined at an adjusted $P<.05$. For predictive analysis, the multivariable regression models were constructed to generate quadratic forecasts to predict depression search intent and control search intent. The multiple regression models allowed us to account for confounding variables and prevent ecological fallacies according to previously published methods [27,28]. The values for normalized depression search intent were categorized into 4 regions according to the US Census Bureau: Northeast, Midwest, South, and West [29]. Geographic heat maps were generated in Microsoft Excel 2018 (Microsoft Corporation, Redmond, WA) to visualize the relationship between state temperature and state depression search intent.

Results

Multivariable Regression Model and Predictive Analysis in Relation to Time and Seasonality

The Google Trends data from January 2010 to March 2021 demonstrated an upward trend such that depression search intent grew 67% from 58.7 AU to 92.9 AU (n=135), while control search intent grew 24% to 67.1 AU (n=135). Based on the quadratic forecasts, depression search intent is predicted to increase an additional 7.4% to 99.8 AU in 2025 (95% CI 96.6 to 102.9 AU; n=135), while control search intent is predicted to increase 3.5% to 64.7 AU (95% CI 63.7 to 65.7 AU; n=135). A significant pattern of seasonality can be observed in Figure 1 with a peak in depression searches in the spring (March, April, May) and a trough in depression searches during the summer (June, July, August).

Figure 1. Time series plot of search intent for depression and control terms in the United States from 2010 to 2021 with predictive forecasts to 2025; demonstrates significant upward trend and seasonal pattern in depression search intent over time. AU: arbitrary unit.

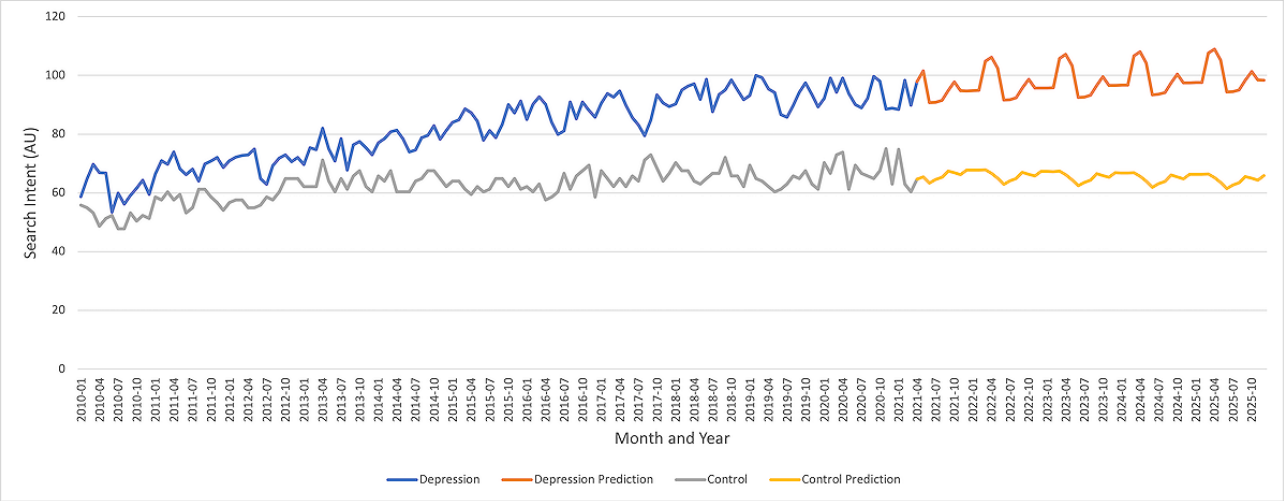


Table 1 presents the multivariable regression model using time and seasonality to predict depression search intent over time. The variables that were significant predictors of search intent were time (r=0.69, adjusted P<.001; n=135), time^2 (r=0.91, adjusted P<.001; n=135), winter (r=0.03, adjusted P<.001; n=135), spring (r=0.12, adjusted P<.001; n=135), and fall (r=0.06, adjusted P<.001; n=135). Applying the regression model, there was a 0.5 AU (95% CI 0.42 to 0.57 AU; n=135) month-over month increase in depression search intent from 2010 to 2021. Depression search intent in the spring, fall, and winter were 7.0 AU (95% CI 5.3 to 8.7 AU; n=135), 4.6 AU (95% CI 2.9 to 6.4 AU; n=135), and 4.5 AU (95% CI 2.8 to 6.2 AU; n=135) higher than in summer, respectively.

Table 1. Multivariable regression model using time variables and season to predict seasonal depression search intent (R^2=0.91).

Variables	Coefficients	Standard error	*t* statistic	*P* value	Adjusted *P* value[a]	*r*
Intercept	56.4	4.7	12.1	<.001	<.001	_[b]
Control	0.0	0.1	–0.4	.70	.99	0.69
Time	0.5	0.0	12.9	<.001	<.001	0.91
Time^2	0.0	0.0	–6.8	<.001	<.001	0.83
Winter[c]	4.5	0.9	5.3	<.001	<.001	0.03
Spring[c]	7.0	0.9	8.2	<.001	<.001	0.12
Fall[c]	4.6	0.9	5.2	<.001	<.001	0.06

[a]Bonferroni correction for 4 independent analyses on the dependent variable (alpha=.05).

[b]Not applicable.

[c]Relative to summer.

Multivariable Regression Model in Relation to Environmental and Geographic Risk Factors

Table 2 presents the multivariable regression model of depression search intent based on state-specific environmental and geographic factors and has a predictive value of R^2=0.57. In this model, variables that were significant predictors of depression search intent were AQI (r=0.30, adjusted P=.01; n=50) and the South (r=–0.2, adjusted P=.01; n=50). Applying the regression model, as AQI increased by 1, the depression search intent increased by 0.4 AU (95% CI 0.14 to 0.61 AU; n=50). Examining the depression search intent relative to US census regions, the South had a decrease of 6.3 AU (95% CI –10.2 to –2.3, adjusted P=.01; n=50) relative to the Northeast. Figure 2 visually demonstrates the regional differences such that states in the South such as Florida and Texas had lower depression search intent in comparison with states in the Northeast such as Maine and Pennsylvania. No relationships existed between depression search intent and temperature (r=–0.5, adjusted P=.99; n=50), humidity (r=0.2, adjusted P=.99; n=50), urban percentage (r=0.3, adjusted P=.06; n=50), or sunshine percentage (r=–0.5, adjusted P=.99; n=50).

Table 2. Multivariable regression model of depression search intent in relation to environmental and geographic risk factors (R^2=0.57).

Variables[a]	Coefficients	Standard error	*t* statistic	*P* value	Adjusted *P* value[b]	*r*
Intercept	94.9	12.2	7.8	<.001	<.001	-[c]
Temperature	0.0	0.1	–0.3	.74	.99	–0.5
Humidity	0.0	0.1	0.1	.89	.99	0.2
Air Quality Index	0.4	0.1	3.2	.002	.01	0.3
Urban %	–0.1	0.0	–2.7	.01	.06	0.3
Sunshine %	–9.0	13.1	–0.7	.50	.99	–0.5
South[d]	–6.3	1.9	–3.2	.002	.01	–0.2
West	–4.4	1.8	–2.5	.02	.11	–0.3
Midwest	–3.8	1.6	–2.4	.02	.12	0.1

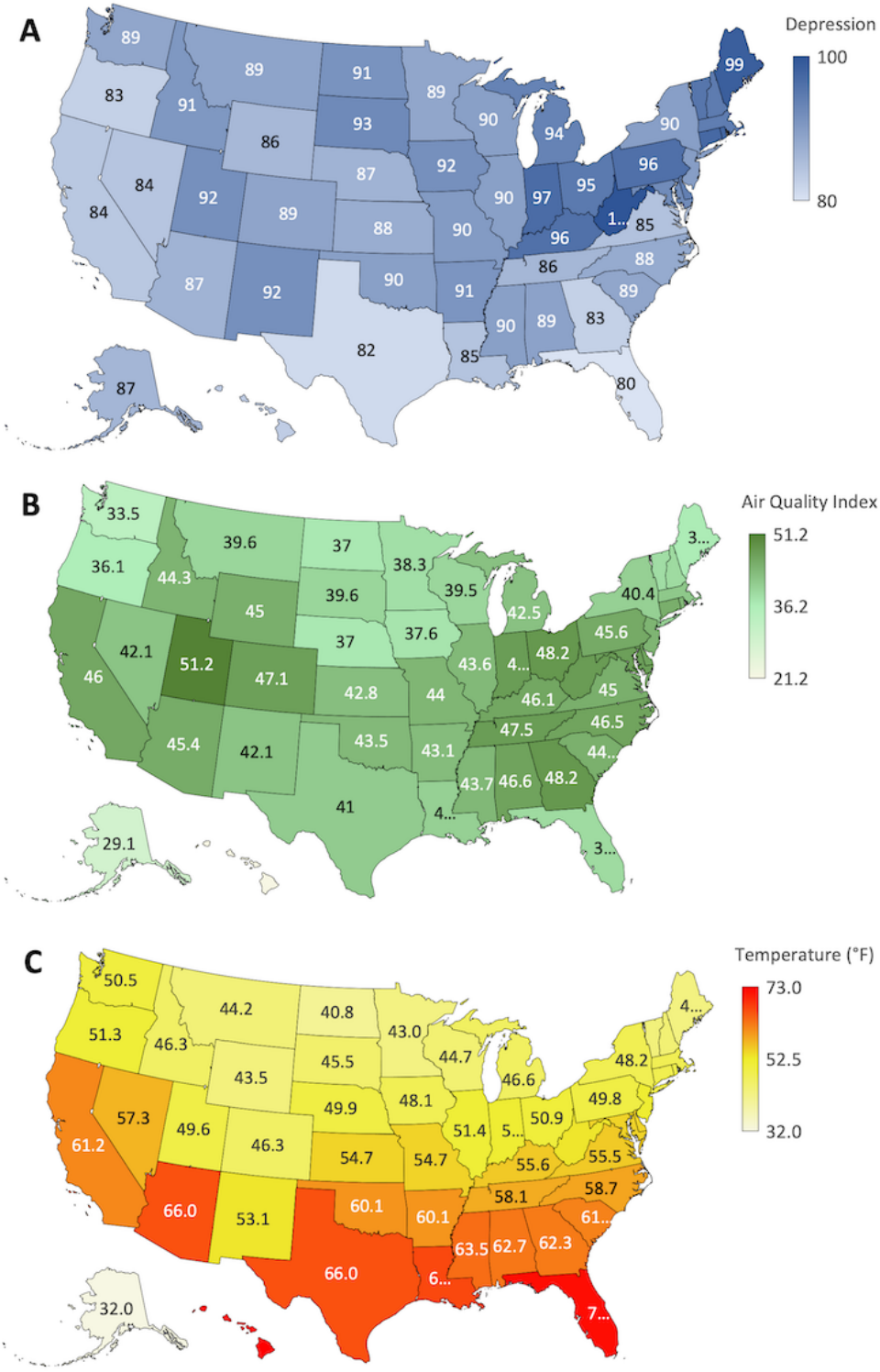
[a]Multivariable regression model using environmental and geographic risk variables to predict depression search intent. Environmental and geographic data sets were collected as an average from 1971 to 2000 and 2008 to 2019, respectively (n=50). This model predicts depression search intent for each state based on the state's average annual temperature, humidity, air quality, urban %, sunshine %, and US census region.

[b]Bonferroni correction for 6 independent analyses on the dependent variable (alpha=.05).

[c]Not applicable.

[d]Relative to the Northeast.

Figure 2. Geographic heat maps of the United States visualizing depression search intent on (A) Google Trends, (B) Air Quality Index, and (C) average annual temperature (° F) by state.



Discussion

Principal Findings

To our knowledge, this is the first study to geographically map depression search intent across the United States in relation to environmental and geographic risk factors by using statistical analysis of big data through Google Trends. Traditionally, prevalence data for mental health and depression have been collected through surveys that require an intensive amount of time and resources to conduct [12-14,30,31]. These surveys are limited not only by human and monetary resources but also by participants' willingness to be included in research. According to the National Survey on Drug Use and Health (SAMHSA), 32.9% of the selected sample did not complete the interview because of refusal to participate, absence from their home, language barriers, or other reasons such as physical or mental

incompetence [32]. Response bias is a known recurring issue with epidemiological surveys and has been difficult to overcome as patients with severe mental health and the homeless population are continuously marginalized by society [33].

The solution to this problem may be utilization of big data found through the internet. In 2020, roughly 86% of the total US population had access to the internet [34]. A US study in greater Los Angeles that examined digital technology use in homeless populations discovered that 94% owned a cell phone [35]. Currently where digital technology is a requirement for survival, internet data can be used to track populations from all over the world over any period. The use of real-time monitoring of internet data to track trends and diseases overcomes the issues of resources, time, and physical location. Analyzing big data through Google Trends is free to researchers and provides information and predictive insight that may one day surpass national or local surveillance systems.

Comparison With Prior Work

The validity of using big data for epidemiology was demonstrated during the influenza outbreak of 2009. At the time, Google Trends was an experimental tool used by researchers for real-time monitoring of influenza outbreaks [36]. By analyzing health care info-seeking behavior on the Google search engine, Google Trends was able to detect regional outbreaks of influenza 7-10 days before the CDC. Google Trends has been successfully used to track viral outbreaks and is currently being used to monitor COVID-19 outbreaks across the world [4-9,37,38].

Depression is a major public health concern and one of the most prevalent mental health illnesses in the United States [10,39]. In 2010, the estimated annual economic consequence of depression was upwards of US $200 billion [40,41]. Considering depression also leads to diminished productivity, poor quality of life, and negative psychological impacts on well-being, the true costs of depression on society are much higher [42]. Worsening mental health and an increasing prevalence of depression, especially during the COVID-19 pandemic, signify the increasing importance of monitoring and treating patients with depression. Based on our analysis, Google search intent for depression in the United States has grown by 67% from 2010 to 2021 and is projected to grow another 7.4% by 2025. This increase reflects the epidemiological trends reported by US national surveys, with an increase in depression prevalence by 61% from 2008 to 2018 (6.6% to 10.4%) [43,44]. This corroborates the concept that, as depression prevalence in the United States continues to grow, so does the information-seeking behavior on Google Trends. Furthermore, depression search intent in the United States demonstrated a significant seasonal pattern, such that depression search intent was lowest in the summer. Relative to the summer, the fall, winter, and spring seasons had an increase in depression search intent by 4.6 AU, 4.5 AU, and 7.0 AU, respectively. This increase in depression search intent reflects the seasonal pattern of seasonal affective disorder (SAD) which has been shown to have higher prevalence in the fall and winter seasons and a decrease in the summer [45,46]. Although SAD has been shown to begin remission in the spring, the increase in depression search intent in the spring may reflect population interest in depression in the early stages of a patient's recovery.

In relation to environmental and geographic risk factors, the state's air quality and geographic location had significant predictive values for depression search intent. States that had a 1-unit higher AQI had an increase in depression search intent by 0.4 AU In other words, states with worse air pollution had higher levels of depression search intent than states with cleaner air. These results reflect previously published findings that air pollution is linked to depression [47-49]. Our results comparing the 4 US census regions demonstrated that the South had less search intent, by 6.3 AU, relative to the Northeast. The West and Midwest also demonstrated decreased levels of depression search intent, by 4.4 AU and 3.8 AU, respectively, though their adjusted *P* values were insignificant. These results reflect the findings that depression is correlated with latitude, with regions further from the equator having a higher prevalence of depression [42,50]. Although the season and location of a state cannot fully predict the depression search intent at a given time, the trends extrapolated from Google Trends have demonstrated their validity in relation to known risks of depression.

Although mining for epidemiological trends within big data is a fascinating prospect, it should not be assumed to replace the work of national and public health organizations. Instead, researchers should consider comparing their results with big data and using big data to support their findings. Our study has demonstrated that depression search intent increased over time following a seasonal pattern and was higher in states with higher air pollution and states with northern latitudes. This supports the trends found in US epidemiological surveys on mental health and supports published results of known risk factors for depression.

Future studies should build upon the results demonstrated here by examining other risk factors for depression such as socioeconomic, demographic, or lifestyle variables. More specifically, whether age, income, marital status, race/ethnicity, or gender are predictive variables of depression search intent, both on national and state levels. Considering the COVID-19 pandemic, future studies should analyze the data based on advanced time series modeling to analyze the effects of the pandemic on mental health. In the future, public organizations such as the CDC or regional hospitals may be able to monitor depression prevalence in real time based on the search intent of their communities through publicly available internet data. The clinical applications of big data in the medical field are limitless and will continue to become more useful as technology software improves.

Limitations

Several limitations are present in our study. First, interpreting the trends extrapolated from Google Trends is challenging without supporting clinical information normally collected by traditional surveys such as medical comorbidities or symptom severity. Second, the data in Google Trends may be influenced by various factors such as trending television shows or bots. For example, in 2017, the internet search intent for suicide queries increased by 19% over a 19-day span after the release of popular Netflix series, *13 Reasons Why*, which elevated

suicide awareness [51]. Third, our data may overrepresent people that search terms in English as Google Trends does not combine search intent of the same word in another language. Fourth, the geographic and environmental data sets were consolidated into a single data point for each state regardless of varying climates and heterogenous landscapes. Lastly, patients with severe or debilitating depression may not have the capacity to search for depression or have the access if they are hospitalized. These limitations illustrate that overreliance on big data, much like on epidemiological studies, may inadvertently exclude certain populations.

Conclusions

Our study is the first to demonstrate that big data in Google Trends can be successfully utilized as a novel epidemiological tool to geographically map out population depression in the United States. This method of mapping allows for easier visualization of areas with higher depression search intent, which were mostly states with higher air pollution and those further from the equator. The interest in depression has grown tremendously in the past decade, with an upward trend that follows a seasonal variation pattern similarly seen in SAD. AQI and geographic location were stronger predictors of depression search intent than temperature, humidity, urban percentage, or sunshine percentage. Further investigation is needed to determine whether the factors significant in our study hold true to depression trends across the world. From a clinical perspective, narrowing the scope of depression search intent to specific cities or high-risk populations should be the next goal of researchers.

Conflicts of Interest

None declared.

References

1. Big Data Market by Component, Deployment Mode, Organization Size, Business Function (Operations, Finance, and Marketing and Sales), Industry Vertical (BFSI, Manufacturing, and Healthcare and Life Sciences), and Region - Global Forecast to 2025. MarketsandMarkets. 2020 Mar 01. URL: https://www.marketresearch.com/MarketsandMarkets-v3719/Big-Data-Component-Deployment-Mode-13117116/ [accessed 2021-04-01]
2. van Rijmenam M. A Short History Of Big Data. Datafloq. 2013 Jan 7. URL: https://datafloq.com/read/big-data-history/ [accessed 2021-03-01]
3. Jun S, Yoo HS, Choi S. Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. Technological Forecasting and Social Change 2018 May;130:69-87. [doi: 10.1016/j.techfore.2017.11.009]
4. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009 Feb 19;457(7232):1012-1014. [doi: 10.1038/nature07634] [Medline: 19020500]
5. Johnson A, Mehta S. A comparison of Internet search trends and sexually transmitted infection rates using Google trends. Sex Transm Dis 2014 Jan;41(1):61-63. [doi: 10.1097/OLQ.0000000000000065] [Medline: 24326584]
6. Jena AB, Karaca-Mandic P, Weaver L, Seabury SA. Predicting new diagnoses of HIV infection using internet search engine data. Clin Infect Dis 2013 May 18;56(9):1352-1353. [doi: 10.1093/cid/cit022] [Medline: 23334812]
7. Pelat C, Turbelin C, Bar-Hen A, Flahault A, Valleron A. More diseases tracked by using Google Trends. Emerg Infect Dis 2009 Aug;15(8):1327-1328 [FREE Full text] [doi: 10.3201/eid1508.090299] [Medline: 19751610]
8. Wilson K, Brownstein JS. Early detection of disease outbreaks using the Internet. Canadian Medical Association Journal 2009 Mar 12;180(8):829-831. [doi: 10.1503/cmaj.1090215]
9. Mavragani A, Gkillas K. COVID-19 predictability in the United States using Google Trends time series. Sci Rep 2020 Nov 26;10(1):20693 [FREE Full text] [doi: 10.1038/s41598-020-77275-9] [Medline: 33244028]
10. Kessler RC, Ormel J, Petukhova M, McLaughlin KA, Green JG, Russo LJ, et al. Development of lifetime comorbidity in the World Health Organization world mental health surveys. Arch Gen Psychiatry 2011 Jan 03;68(1):90-100 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2010.180] [Medline: 21199968]
11. Villarroel M, Terlizzi EP. Symptoms of depression among adults: United States, 2019. NCHS Data Brief 2020 Sep(379):1-8. [Medline: 33054920]
12. Ettman CK, Abdalla SM, Cohen GH, Sampson L, Vivier PM, Galea S. Prevalence of depression symptoms in US adults before and during the COVID-19 pandemic. JAMA Netw Open 2020 Sep 01;3(9):e2019686 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.19686] [Medline: 32876685]
13. Major Depression. National Institute of Mental Health. Bethesda, MD: NIMH; 2020 Jan 1. URL: https://www.nimh.nih.gov/health/statistics/major-depression [accessed 2021-03-31]
14. Anxiety & Depression Association of America. Silver Spring, MD: ADAA; 2020. URL: https://adaa.org/understanding-anxiety/depression [accessed 2021-04-01]
15. FastStats - Depression. Centers for Disease Control and Prevention.: U.S. Department of Health & Human Services; 2020. URL: https://www.cdc.gov/nchs/fastats/depression.htm [accessed 2021-04-01]
16. Mental Health - Household Pulse Survey - COVID-19. Centers for Disease Control and Prevention. 2021. URL: https://www.cdc.gov/nchs/covid19/pulse/mental-health.htm [accessed 2021-04-01]

17. Zitting K, Lammers-van der Holst HM, Yuan RK, Wang W, Quan SF, Duffy JF. Google Trends reveals increases in internet searches for insomnia during the 2019 coronavirus disease (COVID-19) global pandemic. J Clin Sleep Med 2021 Feb 01;17(2):177-184 [FREE Full text] [doi: 10.5664/jcsm.8810] [Medline: 32975191]
18. Azzam D, Nag N, Tran J, Chen L, Visnagra K, Marshall K, et al. A novel epidemiological approach to geographically mapping population dry eye disease in the United States through Google Trends. Cornea 2021 Mar 01;40(3):282-291. [doi: 10.1097/ICO.0000000000002579] [Medline: 33177410]
19. Azzam DB, Cypen SG, Tao JP. Oculofacial plastic surgery-related online search trends including the impact of the COVID-19 pandemic. Orbit 2021 Feb 14;40(1):44-50. [doi: 10.1080/01676830.2020.1852264] [Medline: 33317388]
20. Google Trends. Google. URL: https://support.google.com/trends/?hl=en#topic=6248052 [accessed 2021-03-01]
21. Torres F. What is Depression? American Psychiatric Association. 2020 Oct 1. URL: http://www.psychiatry.org/patients-families/depression/what-is-depression [accessed 2021-04-01]
22. Soulo T. Top Google Searches. Ahrefs. 2021 Jan 1. URL: https://ahrefs.com/blog/top-google-searches/ [accessed 2021-04-01]
23. National Climate Report - Annual 2020. National Centers for Environmental Information. 2021 Jan 1. URL: https://www.ncei.noaa.gov/access/monitoring/monthly-report/national/202013 [accessed 2021-04-01]
24. U.S. Air Quality Index State Rank. USA.com. 2014 Jan 1. URL: http://www.usa.com/rank/us--air-quality-index--state-rank.htm [accessed 2021-04-01]
25. Air Quality by State 2021. World Population Review. 2021 Jan 1. URL: https://worldpopulationreview.com/state-rankings/air-quality-by-state [accessed 2021-04-01]
26. Mesenbourg T, Potok N, Lamas L. United States Summary: 2010: Population and Housing Unit Counts. US Census Bureau. 2012 Sep. URL: https://www2.census.gov/library/publications/decennial/2010/cph-2/cph-2-1.pdf [accessed 2022-03-27]
27. McNamee R. Regression modelling and other methods to control confounding. Occup Environ Med 2005 Jul 01;62(7):500-6, 472 [FREE Full text] [doi: 10.1136/oem.2002.001115] [Medline: 15961628]
28. Portnov BA, Dubnov J, Barchana M. On ecological fallacy, assessment errors stemming from misguided variable selection, and the effect of aggregation on the outcome of epidemiological study. J Expo Sci Environ Epidemiol 2007 Jan 11;17(1):106-121. [doi: 10.1038/sj.jes.7500533] [Medline: 17033679]
29. Census Regions and Divisions of the United States. US Census Bureau. URL: https://www2.census.gov/geo/pdfs/maps-data/maps/reference/us_regdiv.pdf [accessed 2021-04-01]
30. Brody D, Pratt L, Hughes J. Prevalence of Depression Among Adults Aged 20 and Over: United States, 2013-2016. NCHS Data Brief 2018 Feb(303):1-8 [FREE Full text] [Medline: 29638213]
31. 2019 NSDUH Detailed Tables. Substance Abuse and Mental Health Services Administration. 2020 Sep 11. URL: https://www.samhsa.gov/data/report/2019-nsduh-detailed-tables [accessed 2022-03-27]
32. 2017 Methodological Summary And Definitions. Substance Abuse and Mental Health Services Administration. 2018 Sep 14. URL: https://www.samhsa.gov/data/report/2017-methodological-summary-and-definitions [accessed 2021-03-31]
33. Room R. Stigma, social inequality and alcohol and drug use. Drug Alcohol Rev 2005 Mar;24(2):143-155. [doi: 10.1080/09595230500102434] [Medline: 16076584]
34. Johnson J. Internet user penetration in the United States from 2017 to 2025. Statista. 2021 Aug 4. URL: http://www.statista.com/statistics/590800/internet-usage-reach-usa [accessed 2021-04-01]
35. Rhoades H, Wenzel SL, Rice E, Winetrobe H, Henwood B. No digital divide? Technology use among homeless adults. J Soc Distress Homeless 2017;26(1):73-77 [FREE Full text] [doi: 10.1080/10530789.2017.1305140] [Medline: 31097900]
36. Carneiro H, Mylonakis E. Google trends: a web-based tool for real-time surveillance of disease outbreaks. Clin Infect Dis 2009 Nov 15;49(10):1557-1564. [doi: 10.1086/630200] [Medline: 19845471]
37. Prasanth S, Singh U, Kumar A, Tikkiwal VA, Chong PH. Forecasting spread of COVID-19 using google trends: A hybrid GWO-deep learning approach. Chaos Solitons Fractals 2021 Jan;142:110336 [FREE Full text] [doi: 10.1016/j.chaos.2020.110336] [Medline: 33110297]
38. Venkatesh U, Gandhi PA. Prediction of COVID-19 Outbreaks Using Google Trends in India: A Retrospective Analysis. Healthc Inform Res 2020 Jul;26(3):175-184. [doi: 10.4258/hir.2020.26.3.175]
39. Murray CJL, Atkinson C, Bhalla K, Birbeck G, Burstein R, Chou D, U.S. Burden of Disease Collaborators. The state of US health, 1990-2010: burden of diseases, injuries, and risk factors. JAMA 2013 Aug 14;310(6):591-608 [FREE Full text] [doi: 10.1001/jama.2013.13805] [Medline: 23842577]
40. Greenberg PE, Fournier A, Sisitsky T, Pike CT, Kessler RC. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). J. Clin. Psychiatry 2015 Feb 25;76(02):155-162. [doi: 10.4088/jcp.14m09298]
41. Chow W, Doane MJ, Sheehan J, Alphs L, Le H. Economic burden among patients with major depressive disorder: an analysis of healthcare resource use, work productivity, and direct and indirect costs by depression severity. The American Journal of Managed Care 2019 Feb 14:1-3 [FREE Full text]
42. McTernan WP, Dollard MF, LaMontagne AD. Depression in the workplace: An economic cost analysis of depression-related productivity loss attributable to job strain and bullying. Work & Stress 2013 Nov 07;27(4):321-338. [doi: 10.1080/02678373.2013.846948]

43. Gonzalez O, McKnight-Eily LR, Strine T, Edwards VJ, Lu H, Croft JB. QuickStats: Age-Adjusted Death Rates for Parkinson Disease Among Adults Aged ≥65 Years - National Vital Statistics System, United States, 1999-2017. MMWR Morb Mortal Wkly Rep 2019 Sep 06;68(35):773-1235 [FREE Full text] [doi: 10.15585/mmwr.mm6835a6] [Medline: 31487275]
44. Hasin DS, Sarvet AL, Meyers JL, Saha TD, Ruan WJ, Stohl M, et al. Epidemiology of adult major depressive disorder and its specifiers in the United States. JAMA Psychiatry 2018 Apr 01;75(4):336. [doi: 10.1001/jamapsychiatry.2017.4602]
45. Magnusson A. An overview of epidemiological studies on seasonal affective disorder. Acta Psychiatr Scand 2000 Mar;101(3):176-184. [doi: 10.1046/j.0902-4441.2000..x]
46. Rosen LN, Targum SD, Terman M, Bryant MJ, Hoffman H, Kasper SF, et al. Prevalence of seasonal affective disorder at four latitudes. Psychiatry Research 1990 Feb;31(2):131-144. [doi: 10.1016/0165-1781(90)90116-m]
47. Gładka A, Rymaszewska J, Zatoński T. Impact of air pollution on depression and suicide. Int J Occup Med Environ Health 2018 Dec 20;31(6):711-721 [FREE Full text] [doi: 10.13075/ijomeh.1896.01277] [Medline: 30281038]
48. Szyszkowicz M. Air pollution and emergency department visits for depression in Edmonton, Canada. Int J Occup Med Environ Health 2007;20(3):241-245. [doi: 10.2478/v10001-007-0024-2] [Medline: 17932013]
49. Lim Y, Kim H, Kim JH, Bae S, Park HY, Hong Y. Air pollution and symptoms of depression in elderly adults. Environ Health Perspect 2012 Jul;120(7):1023-1028 [FREE Full text] [doi: 10.1289/ehp.1104100] [Medline: 22514209]
50. Levitt AJ, Boyle MH. The impact of latitude on the prevalence of seasonal depression. Can J Psychiatry 2002 May 01;47(4):361-367. [doi: 10.1177/070674370204700407] [Medline: 12025435]
51. Ayers JW, Althouse BM, Leas EC, Dredze M, Allem J. Internet searches for suicide following the release of 13 reasons why. JAMA Intern Med 2017 Oct 01;177(10):1527-1529 [FREE Full text] [doi: 10.1001/jamainternmed.2017.3333] [Medline: 28759671]

Abbreviations

ADAA: Anxiety & Depression Association of America
AQI: Air Quality Index
AU: arbitrary units
CDC: Centers for Disease Control and Prevention
NIMH: National Institute of Mental Health
SAD: seasonal affective disorder
SAMHSA: National Survey on Drug Use and Health

Edited by J Torous; submitted 28.11.21; peer-reviewed by A Sharma, M Nuutinen, A Joseph, N Van; comments to author 03.01.22; revised version received 31.01.22; accepted 18.02.22; published 31.03.22.

Please cite as:
Wang A, McCarron R, Azzam D, Stehli A, Xiong G, DeMartini J
Utilizing Big Data From Google Trends to Map Population Depression in the United States: Exploratory Infodemiology Study
JMIR Ment Health 2022;9(3):e35253
URL: https://mental.jmir.org/2022/3/e35253
doi:10.2196/35253
PMID:35357320

Original Paper

The Behavior Change Techniques Used in Canadian Online Smoking Cessation Programs: Content Analysis

Laura Struik[1], BScN, MSN, PhD; Danielle Rodberg[1]; Ramona H Sharma[2], BSc

[1]School of Nursing, University of British Columbia, Kelowna, BC, Canada

[2]School of Social Work, University of British Columbia, Kelowna, BC, Canada

Corresponding Author:
Laura Struik, BScN, MSN, PhD
School of Nursing
University of British Columbia
1147 Research Road
Kelowna, BC, V1V 1V7
Canada
Phone: 1 2508078000
Email: laura.struik@ubc.ca

Abstract

Background: Smoking rates in Canada remain unacceptably high, and cessation rates have stalled in recent years. Online cessation programs, touted for their ability to reach many different populations anytime, have shown promise in their efficacy. The Government of Canada has therefore funded provincial and national smoking cessation websites countrywide. However, little is known about the behavior change techniques (BCTs) that underpin the content of these websites, which is key to establishing the quality of the websites, as well as a way forward for evaluation.

Objective: The purpose of this study, therefore, is to apply the BCTTv1 taxonomy to Canadian provincial and federal websites, and to determine which BCTs they use.

Methods: A total of 12 government-funded websites across Canada were included for analysis. Using deductive content analysis and through training in applying the BCTTv1 taxonomy, the website content was coded according to the 93 BCTs across the 16 BCT categories.

Results: Of the 16 BCT categories, 14 were present within the websites. The most widely represented BCT categories (used in all 12 websites) included goals and planning, social support, natural consequences, and regulation. Implementation of BCTs within these categories varied across the sites.

Conclusions: Analyzing the content of online smoking cessation websites using the BCTTv1 taxonomy is an appropriate method for identifying the behavior change content of these programs. The findings offer programmers and researchers tangible directions for prioritizing and enhancing provincial and national smoking cessation programs, and an evaluation framework to assess smoking cessation outcomes in relation to the web-based content.

(JMIR Ment Health 2022;9(3):e35234) doi:10.2196/35234

KEYWORDS

content analysis; smoking cessation; internet; behavior change technique; mental health; smoking; online program; website; government; federal; provincial

Introduction

Tobacco use is the leading cause of preventable disease, disability, and death globally. In Canada, 45,000 people die every year due to a tobacco-related illness [1]. Although cigarette smoking has decreased overall in Canada [2], rates of current smoking remain unacceptably high at 15% [3]. In addition, the rise in e-cigarette use has added concerns about nicotine addiction from e-cigarettes that will eventually transfer to smoking [4], and there is some evidence indicating that the increase in e-cigarette use parallels an increase in smoking uptake among younger demographics [5]. Even further, recent evidence indicates that smoking cessation has stalled since the COVID-19 pandemic. According to a cross-sectional research study across Australia, Canada, England, and the United States, although nearly 50% of smokers indicated that they were

thinking about quitting due to the pandemic, changes in smoking status were marginal: only 1.1% attempted to quit, 14.2% reduced smoking, 14.6% increased smoking, and 70.2% reported no change [6]. This is due, in part, to reduced access to smoking cessation treatment and support. For example, a study of the Ontario STOP smoking program, which transitioned from in-person clinic visits to online visits during the pandemic, found that enrollment decreased by 69% and that there was a 42% drop in visits in the spring of 2020 compared to the previous 2 years [7].

Given that one life can be saved for every two people who quit smoking [8], making cessation supports available and accessible to smokers is critical. Online smoking cessation programs are known for their low costs per smoker, accessibility, availability, and their potential to reach a large proportion of smokers [9]. Given that 75% of Canadians 15 years and older reported an increase in internet use since the pandemic [10], the value of online-based cessation support becomes foregrounded. Although there is evidence to support the efficacy of online cessation programs, especially interactive ones, researchers repeatedly emphasize the need to ensure the quality of these interventions to avoid disappointment and failed quit attempts [11-13]. For example, a systematic review that aimed to examine the efficacy and effect modifiers of eHealth interventions for smoking cessation found that, after pooling findings from 67 web-based cessation programs, compared to nonactive control conditions (eg, print materials, assessment only, or no intervention), web-based programs were significantly more effective at 6 months [13]. However, the authors also cautioned that the quality of the web-based programs modified the effect of the program (poorer quality programs yielded fewer positive outcomes) [13].

One way to establish and assess the quality of online smoking cessation interventions is by articulating the behavior change techniques (BCTs) that form the basis of a program's content [14]. The BCT taxonomy provides a list of methods that could be used in a program to yield behavior change [14-16] (Table 1). The BCTTv1 taxonomy consists of 93 BCTs in 16 categories that address the different targets of behavior change, including capability, opportunity, and motivation [17,18]. The benefit of web-based intervention programs is maximized when the BCTs included have been shown to be effective. For example, in England, researchers were able to identify BCTs that were associated with cessation success rates in local smoking cessation services [19]. This subsequently informed guidance on service provision and learning objectives in training courses, which was associated with increased success rates of smokers who were engaged in these services [20]. Therefore, articulating the BCTs used in smoking cessation programming is of utmost importance to understand what works for whom and how.

The Canadian government funds a comprehensive set of online provincial and national smoking cessation programs. However, little is known about how these programs are designed to influence smoking cessation behavior in relation to the BCTs and subsequently how to leverage strengths and address weaknesses. Understanding this is critical, especially in the context of an increasingly complex tobacco use landscape, so that we can reach the national goal of less than 5% tobacco use by 2035 [1]. The purpose of this study, therefore, was to apply the BCTTv1 taxonomy to Canadian provincial and federal websites and determine which BCTs they use.

Table 1. Behavior change techniques (BCTs).

BCT category (n=16)	BCTs (n=93), n	Definition/meaning
1. Goals and planning	9	Develop a goal for a behavior or an outcome of a behavior and determine what factors need to be assessed to work toward that goal. This can include periodically reviewing the goal and making verbal/written commitments to work toward the goal.
2. Feedback and monitoring	7	Monitor the progress made with the behavior or outcome of a behavior either by the individual themselves, by others, or by a device. When monitored by others, feedback may or may not be given to the individual being monitored.
3. Social support	3	This can include social support for three reasons: practical purposes like getting to an appointment, emotional purposes like comforting an individual at an appointment, and unspecified purposes like encouraging an individual to attend their appointment.
4. Shaping knowledge	4	Clarify proper performance of the wanted behavior and determine antecedents associated with the unwanted behavior and causes of the unwanted behavior.
5. Natural consequences	6	Provide information on the consequences associated with the unwanted behavior including health consequences, social and environmental consequences, and emotional consequences, which may also include the individual monitoring their emotional consequences.
6. Comparison of behavior	3	Demonstrate proper performance of the wanted behavior and showcase the performance and opinions of others on the wanted behavior.
7. Associations	8	Increase facilitators for the wanted behavior such as prompts/cues and reduce interest in the unwanted behavior and decrease barriers to the wanted behavior such as nagging and fear.
8. Repetition and substitution	7	Practice performing the wanted behavior to develop a new habit to replace the unwanted behavior. This includes gradually increasing the amount the wanted behavior is performed until it becomes a habit.
9. Comparison of outcomes	3	Identify the pros and cons of continuing the unwanted behavior, including identifying future outcomes that will result from the unwanted behavior. Obtaining information from credible sources like health professionals can help identify this.
10. Reward and threat	11	Reward individuals or give them the incentive that they will be rewarded either when the goal is completed or when effort has been put in toward reaching the goal. This can also include individuals rewarding themselves.
11. Regulation	4	Certain resources can be used to aid in reaching the goal by assisting with maintaining a positive mindset such as medications and stress management techniques.
12. Antecedents	6	Modify the social and physical environment to make it conducive for the wanted behavior and unconducive for the unwanted behavior.
13. Identity	5	Change one's self-identity and beliefs to associate with the wanted behavior rather than the unwanted behavior.
14. Scheduled consequences	10	Eliminate rewards if unwanted behavior occurs and only provide rewards for the wanted behavior in specific circumstances.
15. Self-belief	4	Build the confidence needed to perform the wanted behavior through positive self-talk and persuasion from others as well as focusing on one's past success and envisioning future success.
16. Covert learning	3	Envision the future consequences of the unwanted behavior and the future rewards of the wanted behavior as well as focus on the consequences and rewards others are currently receiving.

Methods

Data Collection

Government-funded websites were found using Google search phrases, such as "smoking cessation Canada provinces and territories" and "smoking cessation federal Canada." We also searched each individual province with the phrase "smoking cessation." The exclusion criteria included websites that only had telephone numbers, websites intended for use outside of Canada, websites with information but not interventions (eg, fact sheets), and websites with only government legislation pages (eg, tobacco control acts). The first 10 pages of the search were scanned for provincial and federal smoking cessation websites. Provincial and federal government health websites were entered to find the hyperlinks for tailored websites on provincial or national smoking cessation. Provinces and territories that did not show up in the Google scan were individually searched. The search yielded 6 provinces and 3 territories with tailored websites, in addition to 1 website designed to provide cessation information for a combination of 4 provinces and 1 territory. Furthermore, 2 federal websites were identified, resulting in a total of 12 websites.

Data Analysis

We used deductive content analysis to determine which BCTs were used in the 12 programs. Deductive content analysis is the process of applying data to a pre-existing framework [21]. All three researchers were familiar with the BCT taxonomy, and one researcher was trained through the BCT online training course, the latter of which did the primary coding of the websites. Through the creation of a document detailing the

decision-making processes around applying the BCT taxonomy to each website, a clear audit trail was generated and frequently consulted by the larger team. Points of confusion were discussed, and consensus was reached through consultation with the BCT taxonomy descriptions and the relevant website section.

Results

Website Details

The 12 websites represented in this paper include QuitNow (BC), Alberta Quits (AB), Tobacco Free Quebec (QC), New Brunswick Anti-Tobacco Coalition (NB), Tobacco Free Nova Scotia (NS), Smokers Help (NL), Nunavut Quits (NU), Quitpath (YT), North West Territories Quitline (NT), Canadian Cancer Society Smokers' Helpline (YT, SK, MB, ON, and PE), Quit Smoking (federal), and Break It Off (federal). All 12 websites addressed combustible cigarettes when discussing nicotine cessation. A total of 10 websites included information on vaping, and 6 websites included information on smokeless tobacco, which encompasses both shisha/hookah and chewing tobacco. Only five websites included informational tabs tailored to specific populations. The specific populations addressed ranged from broader categories, such as women, teens/youth, and older adults, to narrower categories, such as First Nations/Inuit/Metis, pregnant and breastfeeding individuals, individuals with mental illness, and individuals with a cancer diagnosis.

All 12 websites had some method of direct contact support for users: 12 websites offered phone support, 6 offered email support, and 5 offered text support. In addition, the 2 federal websites provided links to the 10 provincial and territorial websites for local support. A total of 9 websites had some type of online community. There were a variety of online communities: Facebook (n=8), Twitter (n=6), Instagram (n=4), YouTube (n=2), website forum (n=4), and 1 website offering a video call online support group. The structure of the websites varied following the number of tabs on the home page. The topics of the tabs included the following: quitting (why quit and how to quit), staying quit, community/support, helping others quit, resources, special concerns, contact, and feedback. The number of tabs on each website included 3 tabs (n=4), 4 tabs (n=3), 6 tabs (n=2), 7 tabs (n=1), 2 tabs (n=1), and 5 tabs (n=1).

BCT Category Representation

The number of BCT categories used in a single website ranged from 5 to 13, with an average of 11 (SD 2.01; see Table 2). All 12 websites include the BCT categories goals and planning, social support, natural consequences, and regulation. This meant that at least one BCT in each of those categories was included in all 12 websites. The least represented categories (in six or fewer websites) included feedback and monitoring, comparison of behavior, and self-belief. The two BCT categories not represented at all include scheduled consequences and covert learning.

Table 2. Behavior change technique (BCT) category representation.

BCT category	Websites (n=12)												
	BC	AB	QC	NB	NS	NL	NU	YT	NWT	CCS[a]	QS[b]	BIO[c]	Total, n
1. Goals and planning	✓[d]	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12
2. Feedback and monitoring			✓		✓	✓	✓		✓			✓	6
3. Social support	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12
4. Shaping knowledge	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	11
5. Natural consequences	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12
6. Comparison of behavior	✓	✓					✓	✓	✓				5
7. Associations	✓	✓	✓		✓	✓				✓	✓	✓	8
8. Repetition and substitution	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	11
9. Comparison of outcomes	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	11
10. Reward and threat	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	11
11. Regulation	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	12
12. Antecedents	✓	✓	✓		✓	✓	✓	✓			✓	✓	10
13. Identity	✓			✓	✓	✓	✓		✓		✓		7
14. Scheduled consequences													0
15. Self-belief		✓			✓	✓		✓		✓	✓		6
16. Covert learning													0

[a]CCS: Canadian Cancer Society.

[b]QS: Quit Smoking.

[c]BIO: Break It Off.

[d]The checkmark indicates that this BCT was used.

BCT Representation and Implementation

The number of BCTs used in a single website ranged from 8 to 33, with an average of 21.91 (SD 5.85). The number of BCTs represented within each category varied. For example, 5 of the 9 BCTs under goals and planning were used. The ways in which each BCT was implemented also varied within the websites. For example, although all 12 websites used goal-setting (behavior), this BCT was implemented in variable manners (eg, setting a quit date, setting goals for cigarette consumption, or taking a readiness quiz and taking an addiction test to shape a quit plan). The complete representation of the BCTs is listed in Textbox 1 along with a list of ways in which the BCTs were implemented.

Textbox 1. Behavior change technique (BCT) representation and implementation.

1. Goals and planning

1.1 Goal-setting (behavior)

- Setting a quit date
- Providing users with an addiction level test
- Providing users with a confidence in quitting test
- Providing users with a readiness to quit test
- Giving users small goals/reduction exercises

1.2 Problem-solving

- Assistance with identifying triggers/roadblocks
- Relapse prevention strategies (eg, activity)
- Tips on managing cravings and withdrawal

1.3 Goal-setting (outcome)

- Encouraging social media engagement around quit goal

1.4 Action planning

- Making “if/then” plans (eg, if I have a craving, then I will go for a walk)

1.9 Commitment

- Making self-commitment statements (eg, “I will...”)

2. Feedback and monitoring

2.3 Self-monitoring of behavior

- Journaling cravings exercise
- Journaling triggers
- Journaling progress to plan self-rewards
- Journaling quit journey

3. Social support

3.1 Social support (unspecified)

- Quit Coaches phone and live chat
- Quitline
- Support groups
- Community forum
- Local help directory/map for local support/in-person clinics
- Social media network (Facebook, Instagram, and Twitter)
- Mental health professional helpline
- Advice for friends and family to support
- Email support
- Text messaging support for up to 3 months
- Quit buddy
- Pregnancy support
- Quit connection referral form
- Quit stories

4. Shaping knowledge

4.1 Instruction on how to perform the behavior

- Visual instructions on use of nicotine replacement therapy (NRT) products
- Instructions on how to navigate the website
- Instructions on how to navigate second-guessing
- Instructions on how to remove smoke from the home
- Instructions on how to prevent weight gain after a quit
- Instructions for families on how to talk to a smoker
- Instructions for youth and young adults on how to quit
- Instructions on how to use the quitline
- Links to self-help resources/apps

4.2 Information about antecedents

- Tracking antecedents to smoking (eg, nicotine cravings)
- Advice on how to manage a craving

5. Natural consequences

5.1 Information about health consequences

- Interactive diagram on the health risks of smoking
- Health risks of nicotine product use
- Risks to pregnancy and breastfeeding
- Risks to cancer recovery
- Links and resources to risks

5.2 Salience of consequences

- Graphic images on the health risks of smoking

5.3 Information about social and environmental consequences

- Second- and thirdhand smoke risks to children and pets
- Economic, environmental, and social effects of smoking
- Increased risk of kids taking up smoking

5.4 Monitoring of emotional consequences

- Recording emotions in a journal while quitting

5.6 Information about emotional consequences

- Negative impacts on quality of life and enjoyment

6. Comparison of behavior

6.1 Demonstration of the behavior

- Stories from ex-smokers (eg, videos on other's quit smoking journey)

7. Associations

7.1 Prompts/cues

- Put up a reminder list with the reasons you quit smoking

7.5 Remove aversive stimuli

- Inform friends/family to not nag

7.6 Satiation

- Mindfulness exercises/videos

8. Repetition and substitution

8.1 Behavior practice/rehearsal

- Complete a practice quit day

8.2 Behavior substitution

- Suggestions for substitutions to smoking (eg, fruit)

8.7 Graded tasks

- Gradual smoking reduction plan

9. Comparison of outcomes

9.1 Credible source

- Quit coaches, doctors, health care providers, pharmacists

9.2 Pros and cons

- Pros and cons list for quitting smoking

10. Reward and threat

10.1 Material incentive (behavior)

- Calculator for money spent on cigarettes

10.2 Material reward (behavior)

- Providing milestone certificates

10.3 Nonspecific reward

- Calculating reward of physical improvements

10.6 Nonspecific incentive

- Incentive of benefitting baby/child/family

10.9 Self-reward

- Encouraging self-reward (eg, purchase special gift)

11. Regulation

11.1 Pharmacology

- Information on pharmacological support (eg, NRT)
- NRT for pregnancy and youth

11.2 Reduce negative emotions

- Stress management techniques
- Information around management of mental health
- Tips on managing cravings and withdrawal

12. Antecedents

12.1 Restructuring the physical environment

- Remove ash trays, remove odor

12.2 Restructuring the social environment

- Hang out with nonsmokers, ask people not to smoke, avoid social situations of temptation

12.3 Avoidance/reducing exposure to behavior cues

- Avoid triggers (eg, change of routine, ensure relaxation)

12.4 Distraction

- Use distractions (eg, keep mind and hands busy)

12.5 Adding objects to the environment

- Prepare by having snacks or gadgets available

13. Identity

13.1 Identification of self as role model

- Identifying self as a role model (eg, to kids, to social groups)

15. Self-belief

15.1 Verbal persuasion about capability

- Encourage family and friends to let smoker know they believe in them

15.2 Mental rehearsal of successful performance

- Imagining life permanently smoke-free

15.3 Focus on past success

- Focus on past successes when relapsing

15.4 Self-talk

- Encourage smoker to use positive self-talk

Discussion

Principal Findings

This is the first study to apply the BCT taxonomy to Canadian government-funded smoking cessation websites. This analysis enables unique comparison and education on a national scale. One major benefit of this analysis is that it provides a framework for understanding which BCTs are used across the nation and how; this provides individual programmers with ideas for strengthening their websites, enables the identification of priority BCTs to include in Canadian cessation programming, and offers a foundation for evaluating strengths and weaknesses in the programs.

Although understanding the overall BCT categories that are being used is helpful and directs programmers to the most widely used BCTs, the granular analysis of BCTs used within each category provides a window into the nuances of how a BCT category and BCT can be executed in an online program. For example, although the BCT information on health consequences under the natural consequences BCT category is used in all 12 websites, its use varied (some provided a text-based list of negative health impacts, while others provided an interactive diagram). This not only cues programmers to the importance of including this BCT in their program but also provides them with ideas for expanding and innovating how to incorporate this BCT in their program (eg, incorporating interactive features).

This analysis also closes a major methodological gap in analyzing provincial and national websites aimed at addressing the same health behavior (smoking). Applying the taxonomy in this context enabled the ability to analyze complicated (ie, nuanced, complex, comprehensive, and tailored) websites, which provides an accessible understanding of the underlying mechanisms that underpin the content to lend to behavior change. In simpler terms, the analysis reveals “how” these websites function.

Further, this analysis provides an evaluation framework for smoking cessation websites. Few websites are evaluated for their effect on behavior change, possibly because websites are more comprehensive and nuanced compared to a single intervention. The first step to investigating the association between the content of the websites and smoking outcomes is the need for a reliable method to describe the content [22]. In this vein, the findings of this study provide researchers with an evaluation framework to investigate the effects of these initiatives. For example, by mapping out the strategies that an individual website uses with regard to the BCT taxonomy and comparing that to the national compilation provided here, researchers could develop survey or interview questions surrounding user experiences with those strategies and if they lent to smoking reduction/cessation. In sum, the findings provide researchers with an evaluation framework that can be used to explore strengths and weaknesses of each technique used within a website with end users, informing what techniques work for whom and how.

Attention to the most widely used (represented in all 12 websites) BCT categories and their associated BCTs is warranted. The most widely used included goals and planning, social support, natural consequences, and regulation. Previous research evidence supports use of most of these BCTs to lend to higher cessation rates. In one study that examined BCT categories used in online interventions and mobile-based interventions and their association with cessation rates, the

authors found that five BCT categories were linked to higher cessation rates [23]. These included goals and planning (eg, advice on coping), reward and threat (eg, self-rewards), regulation (eg, advice on pharmacotherapy), antecedents (eg, advice on changing routines), and identity (eg, supporting identity change of being an ex-smoker) [23]. In a study that included 6 online smoking cessation websites evaluated for their effectiveness, cost-effectiveness, and theoretical underpinnings (use of BCTs), the effects indicated that smokers using an online cessation intervention are 1.15 to 2.84 times more likely to become a former smoker compared to the control condition. The majority of the interventions used most, if not all, of these five BCT categories [11]. Finally, in a recent systematic review and meta-regression analysis of BCT categories and BCTs associated with smoking cessation in smoking cessation trials, the authors found 29 individual BCTs as potentially important predictors of smoking cessation in at least one analysis (controlling for total BCTs in one) [24]. Of these, three consistently predicted higher smoking cessation rates and included goals and planning (prompting commitment), reward and threat (social reward), and identity (identity associated with changed behavior) [24]. These results, in combination with this study's findings, add to a growing body of evidence that supports focusing, expanding, and innovating the use of these BCT categories and their associated BCTs within provincial and national websites.

The findings also highlight areas for improvement and ideas for how improvements could be made by listing the full range of ways in which the BCTs were implemented across the websites. According to the findings and in comparison with previous research findings, there is room to grow with regard to ensuring consistent use of the most effective BCT categories nationwide. Specifically, the BCT categories reward and threat and identity were not included in all websites despite evidence that these categories contribute to higher cessation rates [13,23,24]. In addition, the list of ways in which the BCTs were implemented offers smoking cessation website programmers with the ability to test out and prioritize ideas about how they might want to incorporate and expand on the BCT categories for their individual program. In short, the BCTs ensure the quality of the web-based programs.

It is well-established in the smoking cessation literature that interactive features and tailored content lends to higher user engagement, which subsequently lends to higher cessation success outcomes [25-27]. In this regard, the use of BCTs alone are likely not enough. Instead, making the BCTs as engaging and as tailored as possible is another necessary step forward. For example, sharing health consequences in the form of a text-based list will not likely engage an end user as much as an interactive and graphical map that displays how different parts of the body are impacted by smoking.

Limitations

There are a few limitations to this study that must be noted. First, the BCT taxonomy is vast and complex, requiring training to be well versed in understanding them. Although this research team consisted of a member who engaged in BCT training, interpreting which BCT applies to which web content may result in subjective nuances in representation. In addition, some website sections in Quebec were in the French language and required the use of Google Translate, possibly limiting the appropriateness of the content's translation. Second, some website sections could not be accessed without creating an account and were therefore not included in the analysis, meaning that some BCT categories or BCTs were missed or not represented. Some websites also had accompanying resources (eg, account-based features); the use of BCTs in these resources was not assessed, which adds a layer of complexity. Third, the top BCTs found in Canadian websites may not be appropriate for international contexts due to differences in policies, laws, and tobacco use and cessation attitudes. Fourth, different BCTs may be more prevalent for different topics (eg, nutrition); therefore, BCT analyses should be specific to the topic at hand and should be analyzed and applied according to the needs of the behavior being investigated. Fifth, the success rates of each individual website (eg, based on user traffic) remains unknown; hence, it is difficult to make any assertions about how effective these websites are in relation to the BCTs without further evaluation. Sixth, the websites may have changed since the time of the analysis. Finally, websites were evaluated on just one occasion; for a more comprehensive assessment, each website/service would have to be used across multiple occasions (eg, as a quitter would use it).

Conclusion

Analyzing the BCTs that underpin government-funded smoking cessation websites in Canada is an appropriate method for identifying strengths and weaknesses in these programs for influencing the target behavior of quitting smoking. The findings offer programmers and researchers with tangible directions for prioritizing and enhancing provincial and national smoking cessation programs, and an evaluation framework to assess smoking cessation outcomes in relation to the web-based content. The findings would benefit from being included in national conversations around how to implement and evaluate evidence-based smoking cessation support nationwide.

Acknowledgments

This study was supported by funding from the British Columbia Ministry of Health, via a British Columbia Lung Foundation Research Agreement (grant 018471).

Conflicts of Interest

None declared.

References

1. Canada's tobacco strategy. Canada.ca. 2021. URL: https://www.canada.ca/en/health-canada/services/publications/healthy-living/canada-tobacco-strategy.html [accessed 2021-11-26]
2. Tobacco use in Canada: patterns and trends. University of Waterloo. 2019. URL: https://uwaterloo.ca/tobacco-use-canada/tobacco-use-canada-patterns-and-trends [accessed 2022-02-23]
3. Smoking and tobacco statistics. BREATHE: The Lung Association. 2021. URL: https://www.lung.ca/lung-health/lung-info/lung-statistics/smoking-and-tobacco-statistics [accessed 2021-11-26]
4. National Academies of Sciences, Engineering, and Medicine. Introduction. In: Stratton K, Kwan LY, Eaton DL, editors. Public Health Consequences of E-Cigarettes. Washington, DC: National Academies Press; 2018.
5. Hammond D, Reid JL, Rynard VL, Fong GT, Cummings KM, McNeill A, et al. Prevalence of vaping and smoking among adolescents in Canada, England, and the United States: repeat national cross sectional surveys. BMJ 2019 Jun 20;365:l2219 [FREE Full text] [doi: 10.1136/bmj.l2219] [Medline: 31221636]
6. Gravely S, Craig LV, Cummings KM, Ouimet J, Loewen R, Martin N, et al. Smokers' cognitive and behavioural reactions during the early phase of the COVID-19 pandemic: findings from the 2020 ITC Four Country Smoking and Vaping Survey. PLoS One 2021;16(6):e0252427 [FREE Full text] [doi: 10.1371/journal.pone.0252427] [Medline: 34086706]
7. Minian N, Veldhuizen S, Tanzini E, Duench S, deRuiter WK, Barker M, et al. Changes in the reach of a smoking cessation program in Ontario, Canada, during the COVID-19 pandemic: a cross-sectional study. CMAJ Open 2021;9(4):E957-E965 [FREE Full text] [doi: 10.9778/cmajo.20210053] [Medline: 34667076]
8. West R. The clinical significance of "small" effects of smoking cessation treatments. Addiction 2007 Apr;102(4):506-509. [doi: 10.1111/j.1360-0443.2007.01750.x] [Medline: 17362283]
9. Civljak M, Stead L, Hartmann-Boyce J, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2013 Jul 10(7):CD007078. [doi: 10.1002/14651858.CD007078.pub4] [Medline: 23839868]
10. Bilodeau H, Kehler A, Minnema N. Internet use and COVID-19: Hhow the pandemic increased the amount of time Canadians spend online. Statistics Canada. 2021. URL: https://www150.statcan.gc.ca/n1/pub/45-28-0001/2021001/article/00027-eng.htm [accessed 2021-11-26]
11. Cheung K, Wijnen B, de Vries H. A review of the theoretical basis, effects, and cost effectiveness of online smoking cessation interventions in the Netherlands: a mixed-methods approach. J Med Internet Res 2017 Jun 23;19(6):e230 [FREE Full text] [doi: 10.2196/jmir.7209] [Medline: 28645889]
12. Taylor G, Dalili M, Semwal M, Civljak M, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2017 Sep 04;9:CD007078 [FREE Full text] [doi: 10.1002/14651858.CD007078.pub5] [Medline: 28869775]
13. Do H, Tran B, Le Pham Q, Nguyen L, Tran T, Latkin C, et al. Which eHealth interventions are most effective for smoking cessation? A systematic review. Patient Prefer Adherence 2018;12:2065-2084. [doi: 10.2147/PPA.S169397] [Medline: 30349201]
14. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
15. Michie S, Johnston M. Theories and techniques of behaviour change: developing a cumulative science of behaviour change. Health Psychol Rev 2012 Mar;6(1):1-6. [doi: 10.1080/17437199.2012.654964]
16. Michie S, Wood CE, Johnston M, Abraham C, Francis JJ, Hardeman W. Behaviour change techniques: the development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative data). Health Technol Assess 2015 Nov;19(99):1-188. [doi: 10.3310/hta19990] [Medline: 26616119]
17. Michie S, Atkins L, West R. The Behaviour Change Wheel. London: Silverback; 2014.
18. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
19. West R, Walia A, Hyder N, Shahab L, Michie S. Behavior change techniques used by the English Stop Smoking Services and their associations with short-term quit outcomes. Nicotine Tob Res 2010 Jul;12(7):742-747. [doi: 10.1093/ntr/ntq074] [Medline: 20478957]
20. West R, May S, West M, Croghan E, McEwen A. Performance of English stop smoking services in first 10 years: analysis of service monitoring data. BMJ 2013 Aug 19;347:f4921 [FREE Full text] [doi: 10.1136/bmj.f4921] [Medline: 23963106]
21. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
22. Michie S, Abraham C. Interventions to change health behaviours: evidence-based or evidence-inspired? Psychol Health 2004 Feb;19(1):29-49. [doi: 10.1080/0887044031000141199]
23. Ubhi HK, Michie S, Kotz D, van Schayck OCP, Selladurai A, West R. Characterising smoking cessation smartphone applications in terms of behaviour change techniques, engagement and ease-of-use features. Transl Behav Med 2016 Sep;6(3):410-417 [FREE Full text] [doi: 10.1007/s13142-015-0352-x] [Medline: 27528530]

24. Black N, Johnston M, Michie S, Hartmann-Boyce J, West R, Viechtbauer W, et al. Behaviour change techniques associated with smoking cessation in intervention and comparator groups of randomized controlled trials: a systematic review and meta-regression. Addiction 2020 Nov;115(11):2008-2020. [doi: 10.1111/add.15056] [Medline: 32196796]
25. Civljak M, Sheikh A, Stead L, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2010 Sep 08(9):CD007078. [doi: 10.1002/14651858.CD007078.pub3] [Medline: 20824856]
26. Myung S, McDonnell DD, Kazinets G, Seo HG, Moskowitz JM. Effects of Web- and computer-based smoking cessation programs: meta-analysis of randomized controlled trials. Arch Intern Med 2009 May 25;169(10):929-937. [doi: 10.1001/archinternmed.2009.109] [Medline: 19468084]
27. Shahab L, McEwen A. Online support for smoking cessation: a systematic review of the literature. Addiction 2009 Nov;104(11):1792-1804. [doi: 10.1111/j.1360-0443.2009.02710.x] [Medline: 19832783]

Abbreviations

BCT: behavior change technique

Edited by J Torous; submitted 26.11.21; peer-reviewed by F Velayati, JC Shin; comments to author 12.01.22; revised version received 13.01.22; accepted 18.02.22; published 01.03.22.

Please cite as:
Struik L, Rodberg D, Sharma RH
The Behavior Change Techniques Used in Canadian Online Smoking Cessation Programs: Content Analysis
JMIR Ment Health 2022;9(3):e35234
URL: https://mental.jmir.org/2022/3/e35234
doi:10.2196/35234
PMID:35230253

Original Paper

Quantifying Changes in the Language Used Around Mental Health on Twitter Over 10 Years: Observational Study

Anne Marie Stupinski[1], MS; Thayer Alshaabi[1,2], PhD; Michael V Arnold[1], MS; Jane Lydia Adams[1,3], MS; Joshua R Minot[1], MS; Matthew Price[4], PhD; Peter Sheridan Dodds[1,5], PhD; Christopher M Danforth[1,6], PhD

[1]Computational Story Lab, Vermont Complex Systems Center, University of Vermont, Burlington, VT, United States

[2]Advanced Bioimaging Center, University of California, Berkeley, CA, United States

[3]Data Visualization Lab, Khoury College of Computer Sciences, Northeastern University, Boston, MA, United States

[4]Department of Psychological Science, University of Vermont, Burlington, VT, United States

[5]Department of Computer Science, University of Vermont, Burlington, VT, United States

[6]Department of Mathematics and Statistics, University of Vermont, Burlington, VT, United States

Corresponding Author:
Christopher M Danforth, PhD
Department of Mathematics and Statistics
University of Vermont
Innovation Hall E220
82 University Place
Burlington, VT, 05405
United States
Phone: 1 802 656 3032
Email: chris.danforth@uvm.edu

Abstract

Background: Mental health challenges are thought to affect approximately 10% of the global population each year, with many of those affected going untreated because of the stigma and limited access to services. As social media lowers the barrier for joining difficult conversations and finding supportive groups, Twitter is an open source of language data describing the changing experience of a stigmatized group.

Objective: By measuring changes in the conversation around mental health on Twitter, we aim to quantify the hypothesized increase in discussions and awareness of the topic as well as the corresponding reduction in stigma around mental health.

Methods: We explored trends in words and phrases related to mental health through a collection of 1-, 2-, and 3-grams parsed from a data stream of approximately 10% of all English tweets from 2010 to 2021. We examined temporal dynamics of mental health language and measured levels of positivity of the messages. Finally, we used the ratio of original tweets to retweets to quantify the fraction of appearances of mental health language that was due to social amplification.

Results: We found that the popularity of the phrase *mental health* increased by nearly two orders of magnitude between 2012 and 2018. We observed that mentions of *mental health* spiked annually and reliably because of mental health awareness campaigns as well as unpredictably in response to mass shootings, celebrities dying by suicide, and popular fictional television stories portraying suicide. We found that the level of positivity of messages containing *mental health*, while stable through the growth period, has declined recently. Finally, we observed that since 2015, mentions of mental health have become increasingly due to retweets, suggesting that the stigma associated with the discussion of mental health on Twitter has diminished with time.

Conclusions: These results provide useful texture regarding the growing conversation around mental health on Twitter and suggest that more awareness and acceptance has been brought to the topic compared with past years.

(JMIR Ment Health 2022;9(3):e33685) doi:10.2196/33685

KEYWORDS

mental health; stigma; natural language processing

Introduction

Background

Recent estimates place 1 in 10 people globally as experiencing some form of mental illness [1], with 1 in 30 living with depression [2]. These rates put mental illness among the leading causes of ill health and disability worldwide. Moreover, rates of mental health disorders and deaths by suicide have increased in recent years, especially among young people [3].

Since the beginning of the COVID-19 pandemic and the subsequent social isolation brought on by lockdowns, stay-at-home orders, and the transition to remote work, there have been drastic declines in both physical and social activity, as well as increases in screen time and symptoms of depression [4]. Google searches for mental health–related topics increased in the first weeks of the pandemic, leveling out after more information regarding stay-at-home orders was released [5]. Since March 2020, there has also been a measured increase in suicidal ideation that is associated with increased feelings of social isolation [6]. The Crisis Text Line service reported receiving a higher-than-average volume of messages every day since March 16, 2020, with the main topics being anxiety, depression, grief, and eating disorders [7]. Price et al [8] also found that daily *doomscrolling*—repeatedly consuming negative news and media content on the web—was associated with same-day increases in depression and posttraumatic stress disorder. The pandemic also influenced the type of content that people discussed on social media, with users shifting away from "self-focused" perspectives and toward more "other-focused" topics that used to be taboo to discuss [9].

Historically, the availability of mental health treatment services has been inadequate compared with the demand [10]. Mental health care also experiences a paradox of being overdiagnosed yet undersupported, with patients with some symptoms and disorders being readily medicated despite the symptoms and disorders not being understood and accepted socially [11]. Furthermore, many who would benefit from mental health services do not seek or participate in care because they are either unaware of such services, are unable to afford them, or the stigma associated with seeking treatment proves too great a barrier [12]. In fact, two-thirds of people with a known mental disorder do not seek help from a health professional [13].

Related Work

Many researchers have used social media platforms to explore and understand the dynamics of health care discussions [14]. Several reviews have been carried out on mental health discussion in particular, finding that social media is a viable platform for users to discuss mental health and feel supported, although privacy risks and ethical concerns of the research applications exist as well [15,16]. Previous studies have analyzed the social media content of consenting individuals who have a diagnosed disorder, identifying early markers of depression in Twitter feeds [17,18] and Instagram photographs [19], predicting postpartum depression in Facebook activity [20], and classifying messages from Twitter users self-disclosing various mental illnesses [21,22].

Other studies have analyzed social media feeds of users struggling with mental health more generally, finding that depressed individuals post with higher levels of distorted thinking [23] and identifying markers of suicidal ideation in support threads on Reddit [24] and in messages on Twitter [25]. Several other studies have more directly examined existing social attitudes toward those with mental illnesses, investigating the stigma toward, and treatment of, students with mental illnesses [26,27] and analyzing social media posts that mention various mental illnesses [28-32]. Analysis of text-based crisis-counseling conversations found actionable strategies associated with more effective counseling [33].

Although developments in predicting mental health states provide an opportunity for early detection and treatment, they come with several ethical concerns, such as incorrect predictions, involvement of bad actors, and potential biases [34]. Social media users also hold negative attitudes toward the concept of automated well-being interventions prompted by emotion recognition [35], and they view emotion recognition in general as invasive, scary, and a loss of their control and autonomy [36].

When it comes to using social media as a real-time source of information and opinion, it should be noted that Twitter's user base is limited, skewing younger and more politically left leaning than the US population overall. Mental health discourse is also a sensitive, often personal topic that many individuals will avoid discussing publicly. Although tweets will fail to capture many aspects of human behavior, estimates of public opinion based on the tweets can complement survey-based measures. Twitter is a valuable social ecosystem from which we can sketch a rough portrait of the existing conversation around mental health, and given that social media lowers the barrier for individuals to join difficult conversations, especially with Twitter allowing users to sign up anonymously, it is a promising source of unstructured language data describing the changing experience of a stigmatized group.

Objectives

Although stigma has proven to be a significant barrier to receiving treatment from formal (eg, psychiatrists and counselors) and informal sources (eg, family and friends), the COVID-19 pandemic and the associated isolation, grief, and hardships have spurred awareness of mental illness and discussion on this topic in public forums such as social media. By measuring changes in this conversation, we aim to quantify the hypothesized increase in discussions and awareness and the corresponding reduction in stigma around mental health. Using messages from Twitter, we examine the growth of public attention on mental health, the divergence of language from general messages and their associated happiness shifts, and finally the rise of ambient words or phrases. With these measurements, we can piece together how this topic and its social attention has shifted in the past decade.

Methods

Data

The source of data for this study is the Decahose application programming interface by Twitter, filtered for English messages, from which we collected a 10% random sample of all public tweets between January 2010 and January 2021. This collection was separated into three corpora consisting of (1) all tweets, (2) tweets containing the phrase *mental health*, and (3) tweets containing a small set of phrases related to mental health. Statistics and time series comparisons among the corpora were carried out as described in the following sections.

N-Grams

General Twitter

To explore trends in the appearance of words, we processed messages from January 2010 through January 2021 into 1-, 2-, and 3-grams, where a 1-gram is a 1-word phrase, a 2-gram is a 2-word phrase, and so on, using the n-gram popularity data set StoryWrangler [37].

For each day, we counted the number of times each unique n-gram appeared in tweets and determined use frequencies compared with the appearance of other phrases on Twitter. We ranked n-grams by descending order of count; n-grams with a low rank value assigned to phrases appear on Twitter very often, whereas those with a high rank value appear rarely. For example, the 1-gram *a* has a median rank of 1 because it is typically the most commonly used word in the English language, whereas the 1-gram *America*, which is less common, has a median rank of 990 [38]. To better visualize this concept of descending count in the figures presented in this paper, we plotted rank on an inverted axis.

Mental Health Collection

To explore the specific language used when discussing mental health on Twitter, we compiled a separate collection of n-grams from tweets related to this topic from the same time frame. Restricting the list to messages from 2010 through 2021 that contained the 2-gram *mental health*, we created n-grams in the same fashion as previously described, determining their use frequency within this anchor set and ranking phrases by descending order of counts. We also computed the aggregated frequency and rank of n-grams over each year, using the existing count values for each day, summing them over each year, and ranking them by these counts. Summary statistics for several of the key events in this new data set compared with the general 1-gram data set are shown in Table 1, which highlights the size of the mental health collection over the years. In 2012, roughly 1 in 10,000 messages referenced mental health, whereas in 2018, the rate was roughly 1 in 100 messages. Even so, the mental health collection remains a small subset of messages compared with Twitter as a whole.

Table 1. Summary statistics of the mental health n-gram data set compared with the general Twitter n-gram data set on 3 individual days. The dates shown correspond to several Bell Let's Talk Day events occurring between 2010 and 2021. Bell Let's Talk Day is an annual fundraising and awareness campaign in Canada that coincides with the annual peak in conversation regarding mental health. Unique 1-grams enumerate the set of distinct words found in tweets on these dates, reflecting roughly 10% of all tweets. The Total 1-grams column shows the sum of the counts of each unique 1-gram, and the Total 1-grams (no retweets) column shows the sum of the counts of 1-grams in tweets, not including any messages that were retweeted.

	Unique 1-grams	Total 1-grams	Total 1-grams (no retweets)
February 8, 2012			
Mental health	3.0×10^3	3.0×10^4	9.3×10^3
General	1.7×10^7	3.1×10^8	2.2×10^8
January 21, 2014			
Mental health	1.6×10^3	2.3×10^4	1.5×10^5
General	2.4×10^7	4.9×10^8	2.9×10^8
January 31, 2018			
Mental health	4.9×10^4	4.4×10^6	2.6×10^5
General	2.1×10^7	5.4×10^8	1.6×10^8

Using these data sets, namely counts of phrases in all tweets (general) versus counts of phrases in tweets containing *mental health*, we analyzed changes in the conversation surrounding mental health over time. The dynamics of several other phrases related to mental health were analyzed as well, but we focused primarily on *mental health* as a representative example of such phrases rather than attempting to exhaustively gather all related content.

Results

Growth of Collective Attention

Mental Health Discourse

Public awareness and education regarding an issue is an important step in reducing negative attitudes because a major component of stigma is lack of knowledge [12]. To understand the general public's level of awareness of mental health issues, we quantified the frequency at which people on Twitter have discussions about the topic of mental health. Using Twitter

n-gram data, we constructed a rank time series of the 2-gram *mental health* on a logarithmic axis, which we have presented in Figure 1.

We find that this 2-gram increased in rank by nearly two orders of magnitude between 2012 and 2018, reflecting a dramatic increase in the discussion of mental health on Twitter. For the first 4 years, only a handful of dates resulted in ranks for *mental health* that were more popular than the overall median, whereas for the final 4 years, only a few dates resulted in ranks indicating less attention than the median.

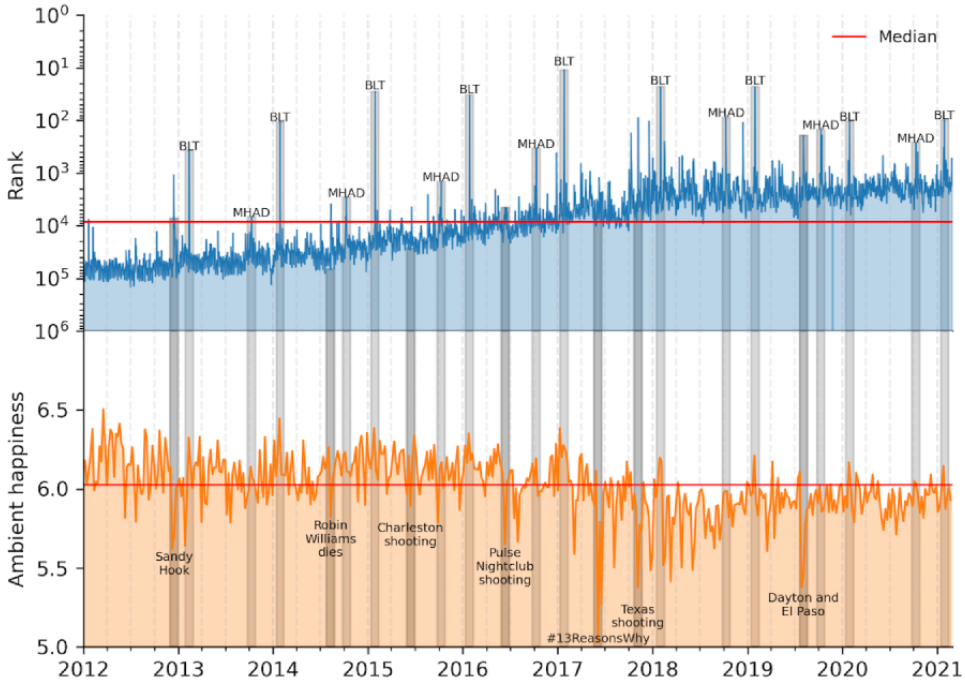
We also examined the positivity of this conversation, calculating the *ambient happiness* score of messages mentioning the phrase *mental health* for each day, which is also shown in Figure 1. Ambient happiness scores for each day were computed by averaging the scores of each word that appeared in a message with *mental health* for a given day, using the Language Assessment by Mechanical Turk dictionary [39]. Although the rank of this 2-gram has increased over the past decade, the ambient happiness of these messages has decreased since 2017.

Examining the daily behavior of these time series, several dates emerged where either the rank or ambient happiness deviated largely from its baseline behavior. In Figure 1, key events associated with large spikes or drops in the time series are highlighted across both panels. Awareness events such as Bell Let's Talk Day and Mental Health Awareness Day contribute to the large, annual spikes in rank beginning in 2013. The 2-gram *mental health* reached its highest rank ever on Bell Let's Talk Day in 2017, peaking at the 18th most popular phrase compared with all other 2-grams on Twitter that day.

Other spikes in rank, and concurrent drops in ambient happiness, occurred on dates with national tragedies such as mass shooting events or celebrity deaths. The largest drop in ambient happiness occurred in 2017 after the deaths of multiple teenagers that were connected to the Netflix series *13 Reasons Why* [40,41].

Looking further into the language used on these specific dates, we show the top 2-grams found in messages containing *mental health* in Multimedia Appendix 1. These co-occurring n-grams are shown with their use rate, rather than rank, so that we can visually see how phrases are being used compared with others in the same list. For example, a popular article shared on December 14, 2012, contained the phrase "It's currently easier for a poor person to get a gun than it is for them to get treatment for mental health issues." This phrase was subsequently quoted by thousands of accounts on Twitter [42]. The resulting phrases (Multimedia Appendix 1) provide more insight into what the broader conversation around mental health looked like after these events.

Figure 1. Timeline of mental health discourse on Twitter. The top panel shows the rank time series of the 2-gram mental health over the past decade on a logarithmic axis. Rank is determined by ordering 2-grams in descending order of counts for each day and is plotted on an inverted axis. The median rank value of the time series is highlighted by a horizontal red line. The bottom panel shows the ambient happiness of all messages containing the 2-gram mental health for each day over the same time period. For clarity, these data are shown as a weekly rolling average, and again the median is highlighted by a red horizontal line. Across both panels, key dates are highlighted in gray and annotated with the associated event. These are dates that led to large spikes or drops in either time series. Annually occurring events such as Bell Let's Talk (BLT) Day or Mental Health Awareness Day (MHAD), are shown in light gray, and unexpected events are highlighted in a darker gray.

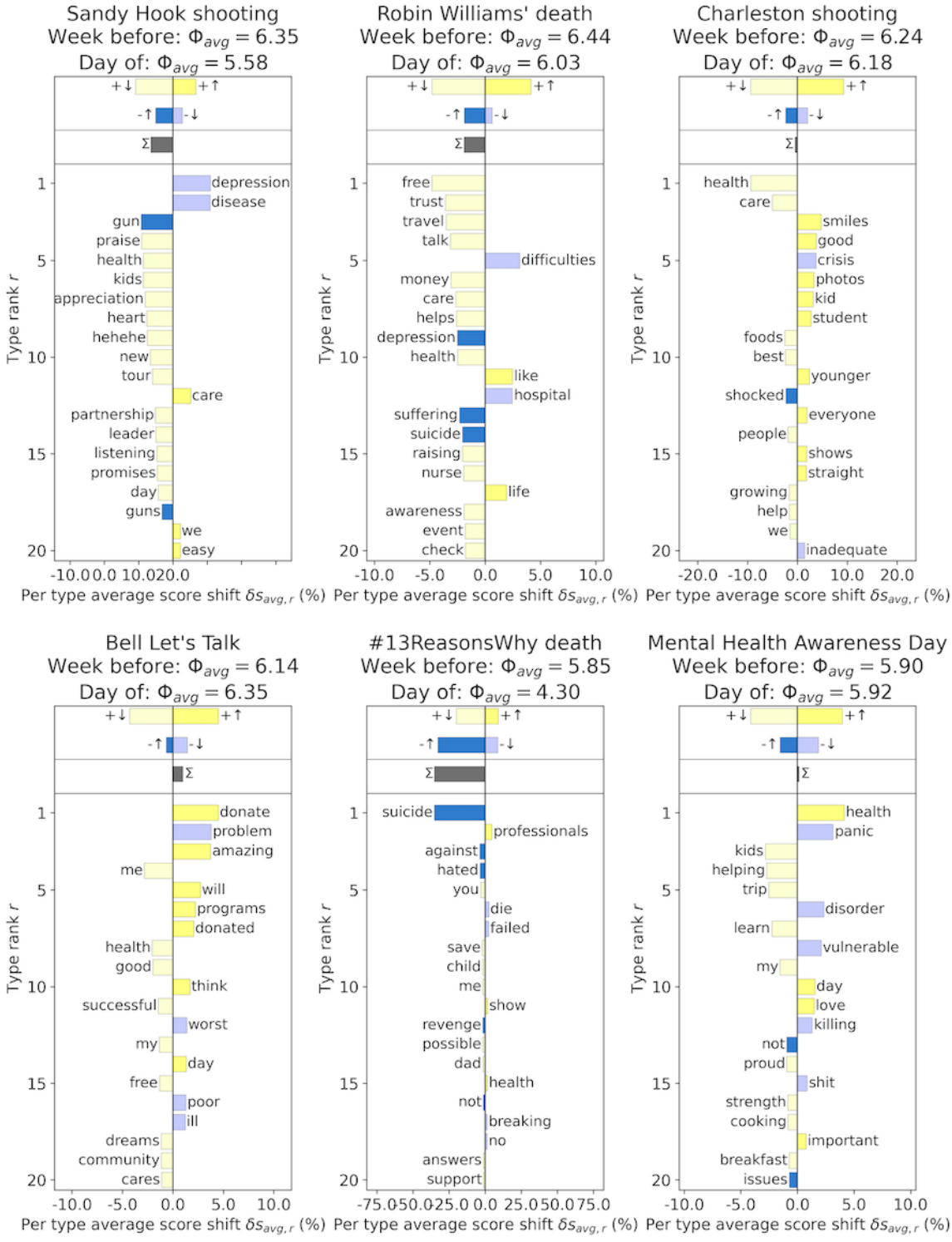


Happiness Word Shifts

To understand the rise and fall of the ambient happiness scores over the time series in Figure 1, we can look at the words that most heavily contributed to these shifts [43]. Figure 2 highlights words associated with the same key events shown in Multimedia Appendix 1, using messages from a week before the event as a reference. Words highlighted with a blue bar are ones that have been coded as negative, and words highlighted with a yellow bar are ones that have been coded as positive. The darker shades

of these 2 colors represent words that have increased in use compared with the reference, whereas lighter shades represent words that have decreased in use. The left side of these panels shows words that are lowering the average score, either through an increase in negative words or a decrease in positive words, and the right side shows words that are raising the score. The average ambient happiness scores for the day of the event and a week before the event are also highlighted at the top of each panel. The 1-grams are also ordered by rank from top to bottom, as shown by the vertical axis.

Figure 2. Happiness word shift graphs. In each of the 6 panels, of the 1-grams, we show the 20 that contribute most to the shift in ambient happiness on key dates, compared with the prior week. The words shown in blue are the ones that have been labeled as relatively negative, whereas the words shown in yellow are the ones that have been labeled as relatively positive [43]. The darker shade of these colors tells us where there is an increase in these words, whereas the lighter shade represents a decrease in use. The happiness score shift is shown on the horizontal axis, representing how positive or negative the language on these days becomes, and the happiness rank of the 1-gram in this subset is shown on the vertical axis. Average ambient happiness scores for the day of the event, as well as a week before the event, are also noted at the top of each subplot.



Looking at Figure 2, we see that mass shooting events have an increase in negative words such as *gun*, *guns*, and *shocked* and a diminishing use of negative words such as *depression*, *disease*, and *crisis*. The day of the Sandy Hook shooting saw fewer positive words such as *praise*, *appreciation*, and *listening*, which would usually be seen in the daily mental health content on Twitter.

Although the Charleston shooting saw a decrease in words such as *health* and *care*, it also saw an increase in positively coded words such as *smiles*, *kid*, and *student*, which likely refer to the shooter in this event. The middle panels in both rows highlight word shifts after death-by-suicide tragedies, and these include an increase in the words *depression*, *suffering*, and *suicide*, which explain the drops in ambient happiness seen on these days.

The awareness events Bell Let's Talk Day and Mental Health Awareness Day, which represent the only increases in ambient happiness on the dates shown in Figure 2, both show an increase in quite a few positive words: *donate*, *amazing*, *programs*, *health*, *love*, and *important*. These days also notably see a decrease in strongly negative words such as *problem*, *disorder*, *vulnerable*, and *killing*.

Narrative and Social Amplifications

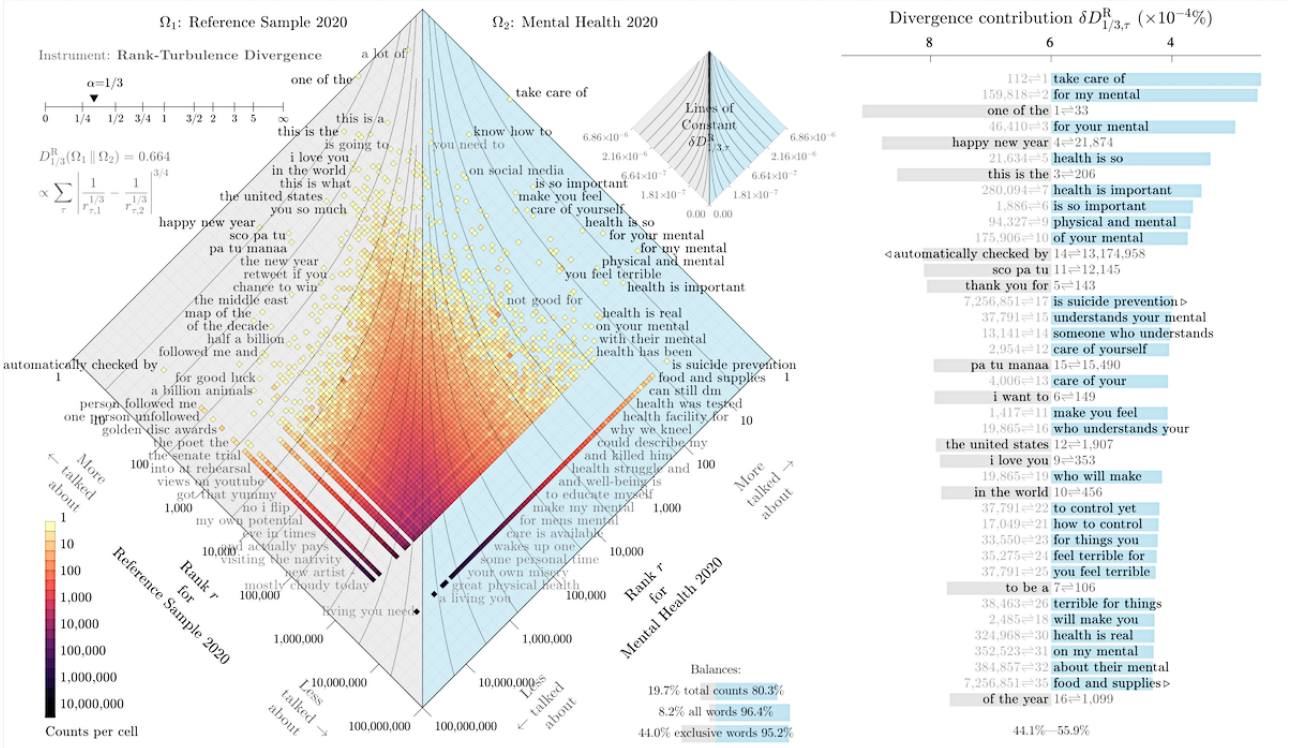
Rank-Turbulence Divergence

The increasing appearance of the phrase *mental health* could be due to several factors. We analyzed the corpus associated with the topic of *mental health* using the n-grams and their relative frequency and rank values for each day and compared the word use in this subset with a random sample of messages on Twitter.

To compare differences in language use, we used rank-turbulence divergence [44]. With this method, we could examine the shift in language between the 2 samples of tweets. We aggregated n-gram counts for phrases found in tweets containing *mental health* over the span of each year, getting annual counts for each of these phrases.

We performed the same aggregation for a smaller random subset of Twitter data, aggregating yearly data for a 1% sample of the Decahose application programming interface. Figure 3 highlights the results of rank divergence comparing the 2 subsets of messages across 2020. When ranking 3-grams from mental health tweets, * *mental health* and *mental health* * phrases were removed for clarity.

Figure 3. Allotaxonograph using rank-turbulence divergence of 1-grams from tweets in 2020 containing the anchor phrase mental health compared with a random sample of tweets in 2020. In the central 2D rank-rank histogram panel, phrases appearing on the right have higher rank in the mental health subset than in random tweets, whereas phrases on the left appeared more frequently in the random sample. The table to the right shows the words that contribute most to the divergence. Note that when ranking 3-grams from mental health tweets, * mental health and mental health * phrases were removed for clarity. The balance of the words in these 2 subsets is also noted in the bottom right corner of the histogram, showing the percentage of total counts, all words, and exclusive words in each set. See Dodds et al [44] for a detailed description of our allotaxonometric instrument.



Each square histogram bin reflects the relative ranks for 3-word phrases in each respective subset. Bins to the right side contain 3-grams with relatively higher rank in the right subset than in the left. The bins down the middle of the plot contain words with a similar rank in both subsets. The bands of bins on the bottom edges of these plots represent words that are exclusive to their respective side's data set.

The color of each bin correlates with the density of words contained in it, and the words appearing on the plot are randomly selected representatives from the bins on the outer edges. The

table on the right shows the words that contribute most to the divergence of the 2 data sets, with small triangles indicating when a word is exclusive to a system. For example, the phrase *take care of* was the 112th most common 3-gram in random tweets posted during 2020, but it was the most common 3-gram in tweets containing *mental health*.

When comparing n-grams from these subsets in Figure 3, we see that the mental health data set, shown on the right side of the figure, includes language related to taking care of one's physical and mental health, suicide prevention, men's mental health, social media, and personal time. Although we would expect to see pandemic-related phrases show up in 2020, these topics were equally mentioned across both samples; therefore, they do not appear on either side of this histogram.

Contagiograms

To better understand the dynamics of phrases related to mental health, we explored ways in which these messages were spreading across Twitter. Tweets can be posted as original content in a new message or a user can retweet a message that another user has posted.

Organic messages show that users are writing their own content related to a topic, whereas retweeted messages show that this topic is being shared and spread to other groups of users; both are important means of contributing to the conversation. Both organic messages and retweeted messages appear in our data set and are included in the previous analyses; therefore, it is important to also examine the proportion of messages that fall into these 2 categories.

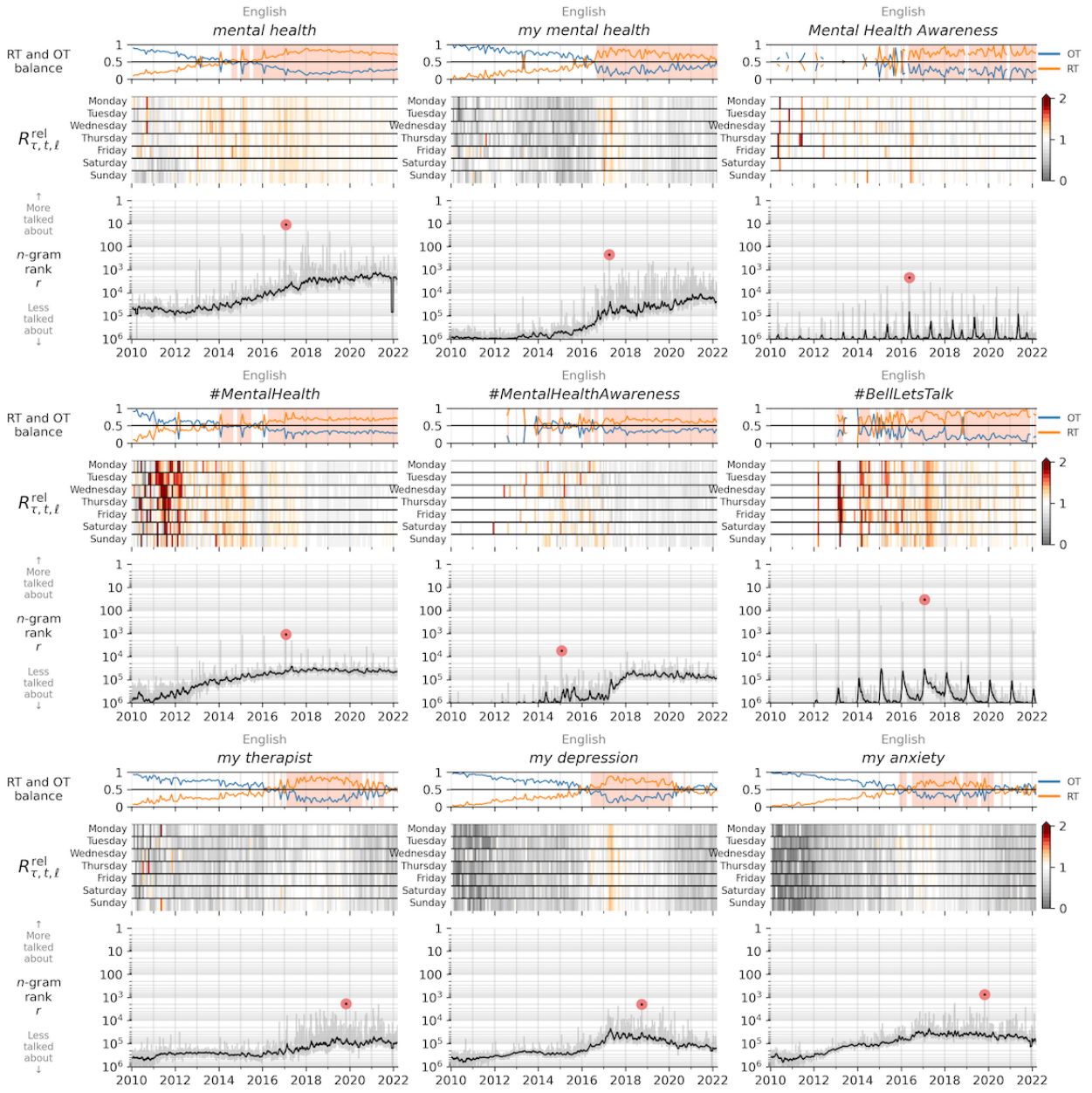
Figure 4 shows *contagiogram* plots, as implemented by Alshaabi et al [45], that highlight the relationship between retweeted and organic content for a given n-gram on Twitter. The top panel of these plots shows the monthly relative use of the specified n-gram, highlighting the use of organic messages in blue and shared retweets in orange. A shaded area in this top panel represents time periods when the number of retweeted messages surpasses that of organic messages, highlighting social amplification.

The middle panel shows retweet use of an n-gram compared with the rate of all retweeting behavior across English Twitter, using a heatmap for each day of the week across the time series. In this heatmap, darker red shades represent a higher relative rate of retweets for the given n-gram compared with a random English n-gram on Twitter and gray shades represent a higher rate of original messages. The bottom panel provides the rank time series of the n-gram, with a month-scale smoothing of the daily values shown in black. In Figure 4, we look at these contagiogram plots for a collection of key n-grams related to the discussion of mental health on Twitter.

Across each of the subplots in Figure 4, we see that phrases and hashtags related to the topic of mental health have grown in volume throughout the time period studied, as reflected by their popularity compared with all tweets. Looking at English Twitter overall, the balance of messages tilted toward primarily organic until mid-2017, when the practice of retweeting messages tipped the balance [45]. Around this same time, retweeted messages reached higher numbers than organic messages for most mental health–related n-grams, as seen in the top panels of these subplots.

Examining the heatmap panels of these subplots, we observe a larger social amplification effect in hashtags related to mental health, highlighted by the darker red shades across the heatmaps. However, in recent years, these hashtags shifted to more organic messages, with the heatmaps becoming more gray after 2018. The hashtag *#BellLetsTalk* sees the most retweeted behavior of these hashtags, as well as an annual spike on the day of the event, followed by a substantial tail of conversation after this date. On Mental Health Awareness Day (October 10) in 2018, organic tweets referencing #BellLetsTalk spiked, leading to the inversion of retweeted messages and organic messages in late 2018 that we see in Figure 4. We also see more original content containing self-disclosure phrases such as *my therapist* or *my depression*, as seen in the third row of n-grams that have largely gray shades across the heatmaps.

Figure 4. Contagiograms for mental health–related n-grams. In each subplot, the top panel displays the monthly relative use of each n-gram, indicating whether they appear organically in new tweets (organic messages [OTs], blue) or in shared retweets (retweeted messages [RTs], orange). The shaded area highlights time frames when the number of RTs is higher than that of OTs, suggesting social amplification [37]. The middle panel of each subplot shows the retweet use of each n-gram compared with the background rate of retweets among all English tweets, with a heatmap for each day of the week. For these heatmaps, the color map is shown on the right, with darker red representing a higher relative rate of RTs among these messages compared with general messages and gray representing a higher rate of OTs. The bottom panel shows the basic n-gram rank time series, with a month-scale smoothing of the daily values shown in black and background shading in gray between the minimum and maximum ranks of each week. Note that phrase counts only reflect tweets that have been identified as messages written in English as discussed by Alshaabi et al [45].



Discussion

Principal Findings

In this project, we explored the conversation around mental health and its appearance on the social media platform Twitter. Using a collection of phrases, we examined how often the topic of mental health was discussed in tweets, finding that the 2-gram *mental health* has increased in rank by nearly two orders of magnitude since 2012. We calculated the associated ambient happiness for the same time series, finding that happiness is largely affected by key dates and has generally decreased over the past decade. After compiling a new data set of n-grams found in the subset of tweets mentioning *mental health*, we analyzed text associated with this specific term, finding the top n-grams related to the topic and their use rates. We examined the language in this conversation across years, finding topics that emerged over the past year since the pandemic began. Comparing the use rates of retweeted content and original content, we found that common *awareness* messages were being amplified on the social media platform, whereas personal self-disclosing statements were being seen more in organic,

originally authored content. These results provide valuable insight into how the discussion of mental health has changed over time and suggest that more awareness and acceptance has been brought to the topic compared with past years.

Growth of Collective Attention

Mental Health Discourse

Our findings suggest that the number of mental health conversations on Twitter has substantially increased in recent years, particularly on dates associated with either awareness campaigns or tragedies. Several dates across the time series emerge where either the rank or ambient happiness deviates largely from its baseline. Awareness events such as Bell Let's Talk Day and Mental Health Awareness Day contribute to the large, annual spikes in rank beginning in 2013. Bell Let's Talk Day, falling on the last Wednesday of January each year, was started by the Canadian company Bell Telephones and aims to bring awareness to the general public about mental health issues by donating CAD $0.05 (US $0.04) for each tweet using its hashtag *#BellLetsTalk*. Other spikes in rank, and concurrent drops in ambient happiness, occurred on dates with national tragedies such as mass shooting events or celebrity deaths. The largest drop in ambient happiness occurred in 2017, immediately after the death of a teenager that was connected to the Netflix series *13 Reasons Why* [39]. Looking at the events that sparked more conversations around the topic of mental health, and their associated levels of ambient happiness, awareness campaigns tended to lead to a rise in ambient happiness, whereas unexpected events, of which all would be considered tragedies, led to a drop in ambient happiness.

Happiness Word Shifts

Looking at the words most heavily contributing to the shifts in sentiment on these dates, we found that although mass shooting events see an increase in negative-coded words related to gun violence, this can sometimes coincide with positive-coded words related to students and children. This example highlights the drawbacks of dictionary-based ambient happiness analysis without context of the words being used because independently positive words can be used to describe a tragic event.

The word shift graphs in Figure 2 also highlight the drop in ambient happiness after death-by-suicide tragedies, which see an increase in words related to depression and suffering. We found that awareness days represent the only increase in ambient happiness on these dates, with an increase in words related to donating, health programs, and love. These awareness days also see a notable decrease in many strongly negative–coded words. These results highlight the shift in language on awareness days, away from phrases with negative connotations and focusing on language relating to community support and aid.

Narrative and Social Amplifications

Rank Divergence

When comparing n-grams from the mental health subset and random Twitter subset, we see that the mental health data set includes language related to taking care of one's physical and mental health, suicide prevention, men's mental health, social media, and personal time. These topics seem to have become more prominent in 2020, with people being at home and isolated during the COVID-19 pandemic and with more awareness being brought to the relationship between social media and mental health.

Studies in 2020 have shown that at the onset of the pandemic, Google searches for terms related to mental health increased initially, followed by a *flattening out* after stay-at-home orders were announced [5]. It has also been recorded that between March 2020 and July 2020, average phone screen time doubled to 5 hours per day and rates of depression increased by 90% [4]. Although these figures cannot tell us everything about how language differs among subsets of conversation, they do provide a sense of the mental health topics individuals discussed in 2020.

Contagiograms

Comparing the use rates of retweeted content and original content, we found that common *awareness* messages are being amplified on the social media platform, whereas personal self-disclosing statements are being seen more in organic, originally authored content. These relationships suggest that users are sharing hashtags to spread awareness and they feel comfortable retweeting hashtags posted by others. The public disclosure of private personal anecdotes, which helps to normalize conversation about personal struggles with mental health, is treated differently. Overall, our results suggest that a larger number of individuals feel comfortable making mental health disclosures publicly, but they are amplified relatively less often than other types of mental health messages.

We also see a substantial increase in the ranks of all phrases and hashtags related to mental health over time, with annual awareness days resulting in spikes corresponding to their given date each year. These findings offer evidence that an understanding of mental health conversations has increased substantially over time, reducing the stigma surrounding mental illness.

Limitations

We acknowledge that using Twitter as a data source for this research has many limitations because its user base is not a broadly representative sample of the human population, and thus these messages will fail to capture many aspects of human behavior. A study by the Pew Research Center [46] shows that as of June 2019, only 22% of all US adults reported using Twitter, smaller than, for example, the 69% who use Facebook. The age breakdown of users is also skewed, with 38% of individuals aged 18-29 years using Twitter, whereas only 17% of those aged 50-64 years use the site. Although demographics of race are fairly uniform (21% of White adults, 24% of Black adults, and 25% of Hispanic adults), the platform is more often used by individuals with a college degree (32%) and living in an urban area (26%) [46].

Mental health discourse is a sensitive and personal topic that many individuals avoid discussing publicly. However, social media has the ability to lower the barrier for individuals to engage in difficult conversations because Twitter allows users to both sign up anonymously and retweet messages in addition to writing their own messages. This being said, we recognize that a portion of Twitter accounts are run by businesses,

institutions, and other organized groups, rather than simply individuals. These corporate accounts, such as *@Bell_LetsTalk*, would have more of a pattern and agenda to their posted tweets, and there is not currently a way to filter out these messages. Because of these complexities of the Twitter user base, care must be taken when interpreting findings based on tweets.

These limitations could be addressed in future studies by expanding the data sources; for example, by looking to other available websites such as Reddit, Instagram, or Facebook, whose user bases differ in some aspects. Turning away from social media, one could examine clinical records for cases of diagnosed mental illness, analyzing the language and positivity of physician notes. Rather than looking at simply the messages of this social media platform, this work could be expanded to address the conversation on a network scale, determining how interactions among users affect the discourse.

This study is also limited to the anchor phrase *mental health* and thus could be leaving out conversations related to the topic. To further enrich these findings, future work could expand the existing mental health data set to include tweets with additional anchor n-grams, although a method for determining these anchors would be necessary.

Our results are also limited to the English language and thus also to events occurring in English-speaking regions. The mass shooting events noted in this study are specific to the United States, and the television show *13 Reasons Why*, although available internationally, led to reports in the United States of an increase in deaths by suicide among teenagers. Although several of the events noted may be specific to the United States, these were events that were discussed heavily across all of English-speaking Twitter and the trends we found relating to awareness campaigns, celebrity deaths, and the pandemic's effect on mental health can be generalized to other regions experiencing these or similar events.

Conclusions

We believe that the results presented here provide useful texture regarding the growing conversation around mental health on Twitter as well as evidence that more people are contributing to this conversation on the social media platform than ever before. Our findings suggest that the number of conversations around this topic have substantially increased on Twitter in recent years and spike especially high on dates coinciding with events such as awareness campaigns, television series releases, mass shootings, and celebrity deaths. These events also drastically shift the ambient happiness associated with the topic of mental health during these time periods. Awareness campaigns positively drive the ambient happiness, as well as shift the focus away from negative connotations and toward the importance of community care, support, and aid, whereas the tragedy events lead to a drop in ambient happiness because they see a focus on suffering, gun violence, and death by suicide. When comparing the mental health data set to a control sample of Twitter users, topics emerge around suicide prevention, taking care of one's mental health, social media, and personal time, all of which became more prominently discussed in 2020. Awareness messages are heavily amplified on the platform through retweets, and personal self-disclosure statements are being posted in more originally authored content.

As mental health becomes talked about more, and awareness campaign efforts seem to be driving a large portion of this increase, public health campaigns aiming to reduce stigma surrounding mental health can leverage this information to improve their messaging. The knowledge that young people on Twitter are participating in these conversations, whether through retweets or personal statements, shows the role social media could have in spreading this conversation to other users in an effort to normalize mental health and reduce the stigma surrounding it.

We also learn from these results that some tragic events, such as mass shootings, bring up interesting conversations related to the link between gun violence and mental health and how much of these horrific events is attributed to the mental illness of the offender. These conversations are complicated and have the potential to not only bring light to the need for better mental health care but also further the stigma around mental illness while avoiding the debate around gun violence as an issue on its own. Knowing that there is a documented link among these conversations after their associated events, perhaps we can inform further debates on, and policy decisions for, these issues.

Finally, we find that television shows can have devastating impacts if their content, portrayal, and significance are not well thought out before creation. Studios, directors, and streaming companies all have a responsibility, especially with projects aimed toward younger audiences, to properly screen their content and think deeply about the impact of each choice that they make. Policies around these safety concerns, if they do not exist already, should be put into place to avoid future tragedies linked to this effect.

As this conversation on the topic of mental health continues to grow, and perhaps becomes more normalized, it will be useful to examine the language associated with future events and how it shifts over time.

Acknowledgments

The authors are grateful for the computing resources provided by the Vermont Advanced Computing Core and financial support from the Massachusetts Mutual Life Insurance Company. The authors thank many of their colleagues at the Computational Story Lab for their feedback on this project.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Top n-grams used in discussions on mental health during spike dates.
[PDF File (Adobe PDF File), 474 KB - mental_v9i3e33685_app1.pdf]

References

1. Dattani S, Ritchie H, Roser M. Mental health. Our World In Data. URL: https://ourworldindata.org/mental-health [accessed 2022-03-06]
2. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1789-1858 [FREE Full text] [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
3. McClure GM. Suicide in children and adolescents in England and Wales 1970-1998. Br J Psychiatry 2001 May;178:469-474. [doi: 10.1192/bjp.178.5.469] [Medline: 11331565]
4. Giuntella O, Hyde K, Saccardo S, Sadoff S. Lifestyle and mental health disruptions during COVID-19. Proc Natl Acad Sci U S A 2021 Mar 02;118(9):e2016632118 [FREE Full text] [doi: 10.1073/pnas.2016632118] [Medline: 33571107]
5. Jacobson NC, Lekkas D, Price G, Heinz MV, Song M, O'Malley AJ, et al. Flattening the mental health curve: COVID-19 stay-at-home orders are associated with alterations in mental health search behavior in the United States. JMIR Ment Health 2020 Jun 01;7(6):e19347 [FREE Full text] [doi: 10.2196/19347] [Medline: 32459186]
6. Fortgang RG, Wang SB, Millner AJ, Reid-Russell A, Beukenhorst AL, Kleiman EM, et al. Increase in suicidal thinking during COVID-19. Clin Psychol Sci 2021 Mar 15;9(3):482-488. [doi: 10.1177/2167702621993857]
7. Everybody hurts. Crisis Text Line. URL: https://www.crisistextline.org/everybody-hurts/ [accessed 2022-03-06]
8. Price M, Legrand A, Brier Z, van Stolk-Cooke K, Peck K, Dodds P, et al. Doomscrolling during COVID-19: the negative association between daily social and traditional media consumption and mental health symptoms during the COVID-19 pandemic. Psychol Trauma 2022 Feb 14 (forthcoming). [doi: 10.1037/tra0001202] [Medline: 35157484]
9. Nabity-Grover T, Cheung CM, Thatcher JB. Inside out and outside in: how the COVID-19 pandemic affects self-disclosure on social media. Int J Inf Manage 2020 Dec;55:102188 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2020.102188] [Medline: 32836645]
10. Detels R, Tan C. The scope and concerns of public health. In: Oxford Textbook of Global Public Health. Oxford, United Kingdom: Oxford University Press; 2015.
11. PLOS Medicine Editors. The paradox of mental health: over-treatment and under-recognition. PLoS Med 2013;10(5):e1001456 [FREE Full text] [doi: 10.1371/journal.pmed.1001456] [Medline: 23723743]
12. Corrigan P, Bink A. The stigma of mental illness. In: Encyclopedia of Mental Health (Second Edition). Amsterdam, Netherlands: Elsevier; 2016.
13. The World Health Report 2001 Mental Health : New Understanding, New Hope. Geneva: World Health Organization; 2001.
14. Gohil S, Vuik S, Darzi A. Sentiment analysis of health care tweets: review of the methods used. JMIR Public Health Surveill 2018 Apr 23;4(2):e43 [FREE Full text] [doi: 10.2196/publichealth.5789] [Medline: 29685871]
15. Conway M, O'Connor D. Social media, big data, and mental health: current advances and ethical implications. Curr Opin Psychol 2016 Jun;9:77-82 [FREE Full text] [doi: 10.1016/j.copsyc.2016.01.004] [Medline: 27042689]
16. Naslund JA, Bondre A, Torous J, Aschbrenner KA. Social media and mental health: benefits, risks, and opportunities for research and practice. J Technol Behav Sci 2020 Sep 20;5(3):245-257 [FREE Full text] [doi: 10.1007/s41347-020-00134-x] [Medline: 33415185]
17. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. In: Proceedings of the International AAAI Conference on Web and Social Media. 2021 Presented at: Proceedings of the International AAAI Conference on Web and Social Media; Jun 7–10, 2021; - URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14432
18. Reece AG, Reagan AJ, Lix KL, Dodds PS, Danforth CM, Langer EJ. Forecasting the onset and course of mental illness with Twitter data. Sci Rep 2017 Oct 11;7(1):13006 [FREE Full text] [doi: 10.1038/s41598-017-12961-9] [Medline: 29021528]
19. Reece AG, Danforth CM. Instagram photos reveal predictive markers of depression. EPJ Data Sci 2017 Aug 8;6(1):15. [doi: 10.1140/epjds/s13688-017-0110-z]
20. De Choudhury M, Counts S, Horvitz E, Hoff A. Characterizing and predicting postpartum depression from shared facebook data. In: Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing. 2014 Presented at: CSCW'14: Computer Supported Cooperative Work; Feb 15 - 19, 2014; Baltimore Maryland USA. [doi: 10.1145/2531602.2531675]
21. Coppersmith G, Dredze M, Harman C. Quantifying mental health signals in Twitter. In: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2014 Presented at: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; Jun, 2014; Baltimore, Maryland, USA. [doi: 10.3115/v1/w14-3207]

22. Coppersmith G, Dredze M, Harman C, Hollingshead K. From ADHD to SAD: analyzing the language of mental health on Twitter through self-reported diagnoses. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2015 Presented at: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; Jun 5, 2015; Denver, Colorado. [doi: 10.3115/v1/w15-1201]
23. Bathina KC, Ten Thij M, Lorenzo-Luaces L, Rutter LA, Bollen J. Individuals with depression express more distorted thinking on social media. Nat Hum Behav 2021 Apr;5(4):458-466 [FREE Full text] [doi: 10.1038/s41562-021-01050-7] [Medline: 33574604]
24. De Choudhury M, Kiciman E, Dredze M, Coppersmith G, Kumar M. Discovering shifts to suicidal ideation from mental health content in social media. Proc SIGCHI Conf Hum Factor Comput Syst 2016 May;2016:2098-2110 [FREE Full text] [doi: 10.1145/2858036.2858207] [Medline: 29082385]
25. Braithwaite SR, Giraud-Carrier C, West J, Barnes MD, Hanson CL. Validating machine learning algorithms for twitter data against established measures of suicidality. JMIR Ment Health 2016 May 16;3(2):e21 [FREE Full text] [doi: 10.2196/mental.4822] [Medline: 27185366]
26. Rose D, Thornicroft G, Pinfold V, Kassam A. 250 labels used to stigmatise people with mental illness. BMC Health Serv Res 2007 Jun 28;7:97 [FREE Full text] [doi: 10.1186/1472-6963-7-97] [Medline: 17598894]
27. Kim T, Hong H. Understanding university students' experiences, perceptions, and attitudes toward peers displaying mental health-related problems on social networking sites: online survey and interview study. JMIR Ment Health 2021 Oct 05;8(10):e23465 [FREE Full text] [doi: 10.2196/23465] [Medline: 34609315]
28. Reavley NJ, Pilkington PD. Use of Twitter to monitor attitudes toward depression and schizophrenia: an exploratory study. PeerJ 2014;2:e647 [FREE Full text] [doi: 10.7717/peerj.647] [Medline: 25374786]
29. Robinson P, Turk D, Jilka S, Cella M. Measuring attitudes towards mental health using social media: investigating stigma and trivialisation. Soc Psychiatry Psychiatr Epidemiol 2019 Jan;54(1):51-58 [FREE Full text] [doi: 10.1007/s00127-018-1571-5] [Medline: 30069754]
30. DeJohn AD, Schulz EE, Pearson AL, Lachmar EM, Wittenborn AK. Identifying and understanding communities using twitter to connect about depression: cross-sectional study. JMIR Ment Health 2018 Nov 05;5(4):e61 [FREE Full text] [doi: 10.2196/mental.9533] [Medline: 30401662]
31. Lachmar EM, Wittenborn AK, Bogen KW, McCauley HL. #MyDepressionLooksLike: examining public discourse about depression on twitter. JMIR Ment Health 2017 Oct 18;4(4):e43 [FREE Full text] [doi: 10.2196/mental.8141] [Medline: 29046270]
32. Li A, Jiao D, Zhu T. Detecting depression stigma on social media: a linguistic analysis. J Affect Disord 2018 May;232:358-362. [doi: 10.1016/j.jad.2018.02.087] [Medline: 29510353]
33. Althoff T, Clark K, Leskovec J. Large-scale analysis of counseling conversations: an application of natural language processing to mental health. Trans Assoc Comput Linguist 2016;4:463-476 [FREE Full text] [Medline: 28344978]
34. Chancellor S, Birnbaum M, Caine E, Silenzio V, De Choudhury M. A taxonomy of ethical tensions in inferring mental health states from social media. In: Proceedings of the Conference on Fairness, Accountability, and Transparency. 2019 Presented at: FAT* '19: Conference on Fairness, Accountability, and Transparency; Jan 29 - 31, 2019; Atlanta GA USA. [doi: 10.1145/3287560.3287587]
35. Roemmich K, Andalibi N. Data subjects' conceptualizations of and attitudes toward automatic emotion recognition-enabled wellbeing interventions on social media. Proc ACM Hum Comput Interact 2021 Oct 13;5(CSCW2):1-34. [doi: 10.1145/3476049]
36. Andalibi N, Buss J. The human in emotion recognition on social media: attitudes, outcomes, risks. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20: CHI Conference on Human Factors in Computing Systems; Apr 25 - 30, 2020; Honolulu HI USA. [doi: 10.1145/3313831.3376680]
37. Alshaabi T, Adams JL, Arnold MV, Minot JR, Dewhurst DR, Reagan AJ, et al. Storywrangler: a massive exploratorium for sociolinguistic, cultural, socioeconomic, and political timelines using Twitter. Sci Adv 2021 Jul;7(29):eabe6534 [FREE Full text] [doi: 10.1126/sciadv.abe6534] [Medline: 34272243]
38. Dodds PS, Minot JR, Arnold MV, Alshaabi T, Adams JL, Dewhurst DR, et al. Fame and Ultrafame: measuring and comparing daily levels of ‘being talked about’ for United States’ presidents, their rivals, God, countries, and K-pop. arXiv 2019:1910 [FREE Full text] [doi: 10.51685/jqd.2022.004]
39. Dodds PS, Harris KD, Kloumann IM, Bliss CA, Danforth CM. Temporal patterns of happiness and information in a global social network: hedonometrics and Twitter. PLoS One 2011;6(12):e26752 [FREE Full text] [doi: 10.1371/journal.pone.0026752] [Medline: 22163266]
40. California families claim ‘13 Reasons Why’ triggered teens’ suicides. ABC News. URL: https://abcnews.go.com/US/california-families-claim-13-reasons-triggered-teens-suicides/story?id=483236 [accessed 2022-03-06]
41. Bridge JA, Greenhouse JB, Ruch D, Stevens J, Ackerman J, Sheftall AH, et al. Association between the release of Netflix's 13 Reasons Why and suicide rates in the United States: an interrupted time series analysis. J Am Acad Child Adolesc Psychiatry 2020 Feb;59(2):236-243 [FREE Full text] [doi: 10.1016/j.jaac.2019.04.020] [Medline: 31042568]

42. Mukherjee S. It's easier for Americans to access guns than mental health services. Think Progress. URL: https://archive.thinkprogress.org/its-easier-for-americans-to-access-guns-than-mental-health-services-13064133675a/ [accessed 2022-03-06]
43. Gallagher RJ, Frank MR, Mitchell L, Schwartz AJ, Reagan AJ, Danforth CM, et al. Generalized word shift graphs: a method for visualizing and explaining pairwise comparisons between texts. EPJ Data Sci 2021 Jan 19;10(1):4. [doi: 10.1140/epjds/s13688-021-00260-3]
44. Dodds PS, Minot JR, Arnold MV, Alshaabi T, Adams JL, Dewhurst DR, et al. Allotaxonometry and rank-turbulence divergence: a universal instrument for comparing complex systems. arXiv 2020 [FREE Full text]
45. Alshaabi T, Dewhurst DR, Minot JR, Arnold MV, Adams JL, Danforth CM, et al. The growing amplification of social media: measuring temporal and social contagion dynamics for over 150 languages on Twitter for 2009-2020. EPJ Data Sci 2021 Mar 31;10(1):15 [FREE Full text] [doi: 10.1140/epjds/s13688-021-00271-0] [Medline: 33816048]
46. Perrin A, Anderson M. Share of US adults using social media, including Facebook, is mostly unchanged since 2018. Pew Research Center. 2019. URL: https://pewrsr.ch/2VxJuJ3 [accessed 2022-03-06]

Edited by J Torous; submitted 19.09.21; peer-reviewed by A Milton, C González-Sanguino; comments to author 20.10.21; revised version received 14.12.21; accepted 26.12.21; published 30.03.22.

Please cite as:
Stupinski AM, Alshaabi T, Arnold MV, Adams JL, Minot JR, Price M, Dodds PS, Danforth CM
Quantifying Changes in the Language Used Around Mental Health on Twitter Over 10 Years: Observational Study
JMIR Ment Health 2022;9(3):e33685
URL: https://mental.jmir.org/2022/3/e33685
doi:10.2196/33685
PMID:35353049

Review

Detecting and Measuring Depression on Social Media Using a Machine Learning Approach: Systematic Review

Danxia Liu[1*], PhD; Xing Lin Feng[2*], PhD; Farooq Ahmed[3,4], PhD; Muhammad Shahid[5], PhD; Jing Guo[2], PhD

[1]School of Sociology, Huazhong University of Science and Technology, Wuhan, China

[2]Department of Health Policy and Management, School of Public Health, Peking University, Beijing, China

[3]Department of Anthropology, University of Washington Seattle, Seattle, WA, United States

[4]Department of Anthropology, Quaid-I-Azam University Islamabad, Islamabad, Pakistan

[5]School of Insurance and Economics, University of International Business and Economics, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Jing Guo, PhD
Department of Health Policy and Management
School of Public Health
Peking University
38 Xueyuan Road
Beijing, 100191
China
Phone: 86 18086471505
Email: jing624218@bjmu.edu.cn

Abstract

Background: Detection of depression gained prominence soon after this troublesome disease emerged as a serious public health concern worldwide.

Objective: This systematic review aims to summarize the findings of previous studies concerning applying machine learning (ML) methods to text data from social media to detect depressive symptoms and to suggest directions for future research in this area.

Methods: A bibliographic search was conducted for the period of January 1990 to December 2020 in Google Scholar, PubMed, Medline, ERIC, PsycINFO, and BioMed. Two reviewers retrieved and independently assessed the 418 studies consisting of 322 articles identified through database searching and 96 articles identified through other sources; 17 of the studies met the criteria for inclusion.

Results: Of the 17 studies, 10 had identified depression based on researcher-inferred mental status, 5 had identified it based on users' own descriptions of their mental status, and 2 were identified based on community membership. The ML approaches of 13 of the 17 studies were supervised learning approaches, while 3 used unsupervised learning approaches; the remaining 1 study did not describe its ML approach. Challenges in areas such as sampling, optimization of approaches to prediction and their features, generalizability, privacy, and other ethical issues call for further research.

Conclusions: ML approaches applied to text data from users on social media can work effectively in depression detection and could serve as complementary tools in public mental health practice.

(JMIR Ment Health 2022;9(3):e27244) doi:10.2196/27244

KEYWORDS

depression; machine learning; social media

Introduction

Over recent decades, depression has increasingly become a matter of global public health concern [1]. The total number of people living with depression globally increased by 18.4% between 2005 and 2015. In 2015, more than 332 million (4.4%) people around the globe were found to be living with depression [2]. Mental disorders like depression rank 9th among global causes of disease burden, following common diseases such as stroke, heart diseases, and AIDS, and it can impair physical, as

well as emotional, and mental health [3]. People with depression experience sleep disorders, lack of energy, low interest in daily activities, feelings of worthlessness, inability to concentrate, and recrudescent suicidality [4]. Detection of depression is critical for helping to relieve these threats.

Traditionally, depression is detected using standardized scales requiring patients' subjective responses or clinical diagnoses given by attending clinicians—methods that have some shortcomings. Firstly, people's responses to standardized scales administered in the traditional way are likely to be affected by context, the patient's mental status at the time, the relationship between the clinician and the patient, the patient's current mood, and the patient's past experiences and memory bias. Traditional diagnostic methods also lack temporal granularity [5]. Secondly, people may be unaware or ashamed of their depressive symptoms and unlikely to consult with professional clinicians, especially in the early stages of depression. A previous study found that more than 70% of the population would not consult with professional clinicians if they were in the early stages of depression, meaning that they would be likely to allow their symptoms to worsen before they would consider seeking help [6]. Finally, detection of depression by traditional methods, being dependent on face-to-face interviews, is costly in terms of both money and time and unaffordable for some people [7]. Therefore, a more cost-effective method for detecting cases of depression, applicable to large populations, is needed.

Fortunately, application of the machine learning (ML) approach to text data from social media can provide an effective solution to this question. Social media such as Twitter, Facebook, discussion forums, and microblogs have long since become popular platforms for expressing and recording individuals' personalities, feelings, moods, thoughts, and behaviors. Social media in this review refers to a cluster of applications that build upon technological and ideological foundations [8]. There were researchers classifying social media according to theories in the field of social processes consisting of self-presentation and self-disclosure. Self-presentation defines that people have the desire to get command of the impressions that other people have of them [9], and it is achieved through self-disclosure. Kaplan and Haenlein [8] classified social media relied on the type of self-presentation and the degree of self-disclosure. Different types of social media can help users conduct different types of self-presentation, such as text-based, video-based, picture-based, etc. And some groups of social media (eg, blogs and social networking sites) have a higher degree of self-disclosure. Hence, data mining of the vast quantities of text through which we can seek out the users found on social media can be of great value for detecting cases of depression [10]. In addition, ML, which has been developing rapidly in recent years, can help text mining and sentiment analysis to become more accurate and intelligent [11]. ML is a subfield of computer science that explores the construction and study of algorithms that can learn from and make predictions on data [12]. With recent and rapid advances in social media technology, mental health researchers have an opportunity to collect vast amounts of online data related to people's mental states, and ML can serve as a robust technique for analyzing these data and detecting trajectories and dimensions of mental disorders (eg, depression and anxiety) [13]. For example, researchers in Australia proposed several ML models for predicting depressive symptoms among users based on text data from Reddit, and the models achieved high predictive precision. As a result, their ML approach was shown to potentially be a useful tool for monitoring social media user populations for early traces of depression and a complementary tool to well-established methods of depression detection [14]. In recent years, researchers have devoted considerable time and effort to developing ML approaches that can make use of words, topics, and other information contained in social media texts for detecting depression [14,15].

As far as we know, there are few existing reviews of ML approaches to depression detection that use text data from social media. Some previous reviews have focused on ML applications that use neuroimaging data to predict depression. For example, Mumtaz et al [16] conducted a detailed review of studies of the use of electroencephalogram and event-related potential data sets to detect major depressive disorder using ML approaches. Orrù et al [17] provided an overview of studies identifying imaging biomarkers of psychiatric diseases, such as major depression, using support vector machines (SVMs). There has also been a review focusing on studies about screening for mental illnesses by applying various methods to social media [18]. None of the existing reviews have focused on the application of ML approaches to texts from social media. However, ML approaches have unique advantages in the detection of depression using text data from social media. With people's memberships in online forums, and their public sharing via the internet, text data from social media records are a treasure trove of psychological data, which can play a vital role in screening for depressive symptoms among users of social media. ML techniques also offer opportunities for identifying hidden patterns in online communication and interaction on social media that may reveal users' mental states such as depression, anxiety, anger, etc [19]. Automatic detection of depressive symptoms through ML algorithms applied to social media data has potential as a way of identifying people at risk of depression through large-scale monitoring of online social networks and could complement traditional screening procedures. Systematic reviews of studies using ML approaches and text data from social media to detect depression can help provide directions for future research in the area, and help guide optimization of data mining, feature extraction, and processing methods so that the limitations of previous studies can be overcome, and prediction accuracy and generalization capability improved. Such reviews should describe the depression identification and classification methods being used.

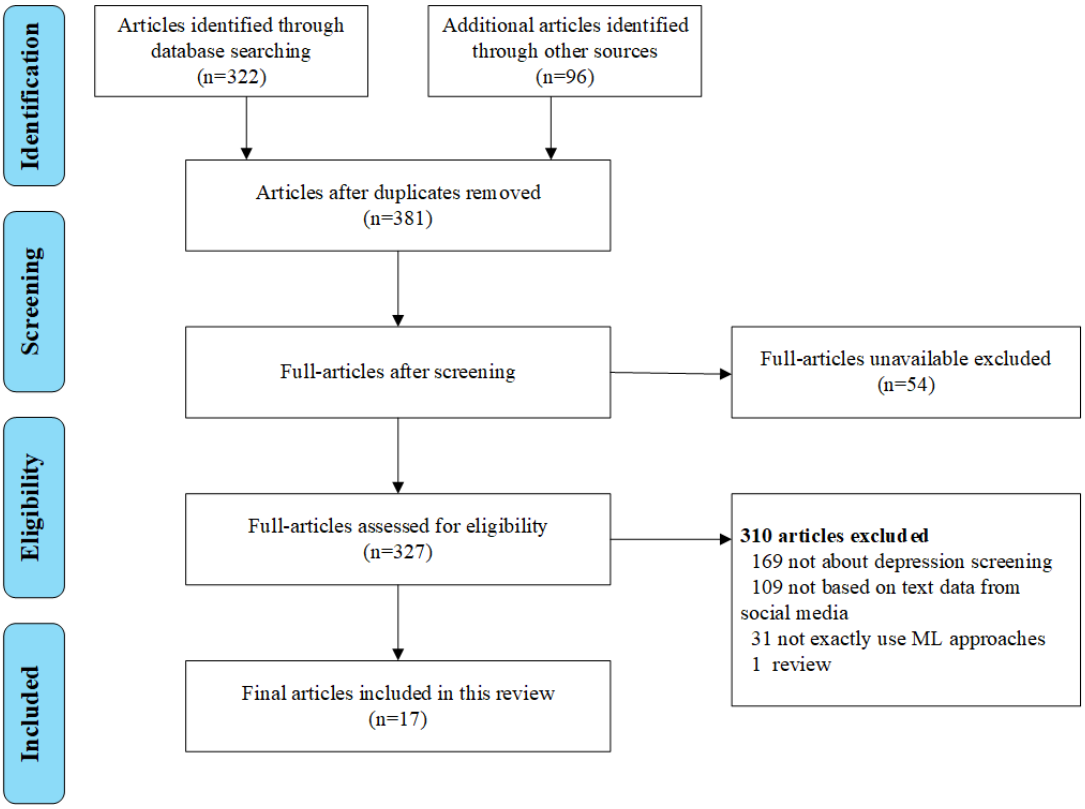
In this paper, we systematically reviewed studies that adopted the ML approach to measure depressive symptoms based on any text mining techniques to identify sentiments using social media data. We specified the ML methods that were used to identify mental status and discuss the evolution of the methods and their pros and cons and provide suggestions for future research in the area.

Methods

Search Strategy

We searched several English- and Chinese-language online bibliographic databases for relevant articles, specifically, Google Scholar, PubMed, Medline, ERIC, PsycINFO, and BioMed, and the Chinese Wangfang, Weipu, and China National Knowledge Infrastructure databases. Our search placed no restrictions on publication type. However, because the age of social media began in the 1990s [8], we did restrict the papers' publication dates to the period between January 1990 and December 2020. Search strings related to ML, depression detection, social media, and text were utilized, that is "Machine Learning" or "Deep Learning" or "Artificial Intelligence" AND "Depression detection" or "screening depression" or "predicting depression" or "recognizing depression" or "major depressive disorder" AND "social media" or "social network" or "online" or "Twitter" or "micro-blog" or "web post" or "Facebook" or "Reddit" or "LiveJournal" or "WeChat" AND "text." We aimed to find studies focusing on the use of ML approaches, such as SVMs, Bayes, latent Dirichlet allocation (LDA), decision tree, and neural networks to detect depression through text mining from social media. "Text mining" refers to mining online textual posts of social media users, including those containing emoticons. And it is worth mentioning that we collect the articles that only screen depression, not those studying several symptoms, which include depression. For example, we first input "machine learning," "depression screening," and "social media" in Google Scholar and obtained more than 20000 articles published during the period between January 1990 and December 2020. We made a preliminary judgment based on the title and abstract to identify the studies we needed. Most of the 20,000 articles do not meet the criterion. Some of the articles do not use text data from social media, but videos, photos, etc. There are also some articles that are depression-related, but they do not detect, screen, or predict depression. This search retrieved 322 articles, all dealing with depression, social media, and ML. We also collected 96 relevant articles that were cited in the 322 articles thus retrieved. After the reviewers screened the retrieved citations according to a set of exclusion criteria, seventeen of them were selected for inclusion in this review. Figure 1 shows the process by which the final set of seventeen studies was selected.

Figure 1. Flowchart for the systematic search of studies in this review.



Study Selection

The article titles and abstracts were screened independently by 2 reviewers (JG and DL). The reviewers then retrieved and assessed the available full texts of the studies and excluded articles that (1) did not discuss ML approaches or detection of depression, (2) were not focused on the use of textual (as opposed to video and image) data from social media, or (3) were themselves reviews of existing research on the use of texts from social media to detect depressive symptoms with ML approaches. The 2 reviewers also recorded important data about the articles such as authors, sample size, platform, study design, assessment tools, outcome of interest, and findings. Disagreements concerning particular articles were resolved through discussions aimed at reaching consensus. Details of the process are shown in Figure 1.

Quality Assessment

The study quality assessment for the 17 studies included was conducted by 2 independent reviewers, using the 14-item NIH

Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies [20]. For each study, they gave each of the checklist items a score of 0 (no) or 1 (yes). The total scores ranged from 0 to 14. Therefore, each reviewer classified each study as low (6), medium (7-10), or high (14) quality and then assigned a quality score to each one. Any discrepancies between the 2 reviewers' ratings were discussed, and a consensus rating was recorded. These consensus ratings are shown in Table 1.

Table 1. Study quality assessment.

Reference	Q1[a]	Q2[b]	Q3[c]	Q4[d]	Q5[e]	Q6[f]	Q7[g]	Q8[h]	Q9[i]	Q10[j]	Q11[k]	Q12[l]	Q13[m]	Q14[n]	Total Score	Rank
Wang et al [21], 2013	1	1	1	1	1	0	1	1	1	1	1	0	1	1	12	high
Burdisso et al [14], 2019	1	1	1	1	1	0	1	1	1	1	1	0	1	1	12	high
Nguyen et al [22], 2014	1	0	0	1	1	0	1	1	1	1	1	0	1	1	10	medium
Fatima et al [23], 2018	1	0	0	1	1	0	1	1	1	1	1	0	1	1	10	medium
Tung & Lu [15], 2016	0	0	1	1	1	0	1	1	1	1	1	0	0	1	9	medium
Husseini Orabi et al [24], 2018	1	0	1	1	1	0	1	1	1	1	1	0	0	1	10	medium
Islam et al [19], 2018	1	1	0	1	1	0	1	1	1	1	1	1	0	1	11	high
Shen et al [6], 2017	0	1	0	1	0	0	1	1	1	1	1	1	0	1	9	medium
De Choudhury, Gamon [25], et al, 2013	0	0	1	1	1	0	1	1	1	1	1	0	1	1	10	medium
Mariñelarena-dondena et al [26], 2017	0	1	1	1	1	0	1	1	1	1	1	0	0	1	10	medium
Tsugawa et al [27], 2015	1	1	0	1	1	1	1	1	1	1	1	1	1	1	13	high
Chen et al [28], 2018	0	0	0	0	0	0	1	1	1	1	1	0	0	1	6	low
De Choudhury, Counts [29], et al, 2013	0	0	0	1	1	0	1	1	1	1	1	1	0	1	9	medium
Dinkel et al [30], 2019	0	0	0	0	0	0	1	1	1	1	1	0	0	1	6	low
Sadeque et al [7], 2017	1	1	1	1	1	0	1	1	1	1	1	1	0	1	12	high
Shatte et al [31], 2020	1	1	1	1	1	1	1	0	1	1	1	0	1	1	12	high
Li et al [32], 2020	1	1	1	1	0	1	1	0	1	1	1	0	1	0	10	medium

[a]Q1: Was the research question or objective in this paper clearly stated?

[b]Q2: Was the study population clearly specified and defined?

[c]Q3: Was the participation rate of eligible persons at least 50%?

[d]Q4: Were all the subjects selected or recruited from the same or similar populations? Were inclusion and exclusion criteria for being in the study prespecified and applied uniformly to all participants?

[e]Q5: Was a sample size justification, power description, or variance and effect estimates provided?

[f]Q6: Were the exposure(s) of interest measured before the outcome(s) were measured?

[g]Q7: Was the timeframe sufficient so that one could reasonably expect to see an association between exposure and outcome if it existed?

[h]Q8: For exposures that can vary in amount or level, did the study examine different levels of the exposure in relation to the outcome?

[i]Q9: Were the exposure measures clearly defined, valid, reliable, and implemented consistently across all study participants?

[j]Q10: Were the exposure(s) assessed more than once over time?

[k]Q11: Were the outcome measures clearly defined, valid, reliable, and implemented consistently across all study participants?

[l]Q12: Were the outcome assessors blinded to the exposure status of participants?

[m]Q13: Was loss to follow-up after baseline 20% or less?

[n]Q14: Were key potential confounding variables measured and their impact on the relationship between exposure(s) and outcome(s) statistically adjusted for?

Results

Depression Identification

The samples, methods, and results of the 17 studies that met the inclusion criteria are summarized in Table 2. This review will further summarize the depression identification and ML methods used in the 17 studies (Table 3). Nine of the studies [15,21,25,27-32] identified depression based on researcher-inferred mental status, while 6 [6,7,14,19,24,26] used user-declared mental status, and 2 [22,23] identified it based on community membership.

Table 2. Summary of machine learning studies of detection of depression using text data from social media.

Reference	Sample	Platform	Outcome	Depression identification method	ML[a] approach type	Features examined	Cross-validation	Type of study
Wang et al [21], 2013	122 depressed and 346 nondepressed subjects, the ages of the samples were not reported	Sina microblog	Bayes: mean absolute error=0.186, ROC[b]=0.908, F-measure=0.85; Trees: mean absolute error=0.239, ROC=0.798, F-measure=0.762; Rules: mean absolute error=0.269, ROC=0.869, F-measure=0.812	Researcher-inferred	3 classification approaches: Bayes, trees and rules	Ten features from three dimensions, including microblog content, interactions, and behaviors. Four of the ten features, (1st person singular, 1st person plural, positive emoticons, and negative emoticons) pertain to microblog content, while three pertain to interactions (mentioning, [being] forwarding, and commenting), and two pertain to behaviors (original blogs and blogs posted between midnight and 6:00 am).	10-fold cross-validation	Observational cohort study
Burdisso et al [14], 2019	486 training subjects (83 depressed/403 nondepressed); 401 test subjects (52 depressed/349 nondepressed), the ages of the samples were not reported	Reddit	SS3[c]: F-measure=0.61, precision=0.63, recall=0.60	User-declared	The proposed model: SS3	Words in users' online text posts on Reddit	4-fold cross-validation	Observational cohort study
Nguyen et al [22], 2014	5000 posts made by users from clinical communities and 5000 posts from control communities, the ages of the samples were not reported	LiveJournal	Lasso to classify communities (Accuracy): ANEW[d]=0.89, mood=0.96, topic=1, LIWC[e]=1; Lasso to classify posts (Accuracy): topic=0.93, LIWC=0.88	Community membership-based	The Lasso model	Affective features, mood tags, features topics from the LIWC, all extracted from posts on LiveJournal.	10-fold cross-validation	Observational cohort study

Reference	Sample	Platform	Outcome	Depression identification method	ML[a] approach type	Features examined	Cross-validation	Type of study
Fatima et al [23], 2018	4026 posts (2019/2007) from depressive and non-depressive communities, the ages of the samples were not reported	LiveJournal	The proposed RF[f]-based model (Accuracy): post=0.898, community=0.950, depression degree=0.923; SVM[g] (Accuracy): post=0.8, community=0.895	Community membership-based	Random forest, SVM	The values of the feature set serve as inputs to the classification algorithm, being extracted from first person singular, positive emotion, negative emotion, anxiety, cognitive process, insight, cause, affiliation health, and informal language of online text.	10-fold cross-validation	Observational cohort study
Tung & Lu [15], 2016	724 posts, the ages of the samples were not reported	PTT[h]	EDDTW[i]: precision=0.593, recall=0.668, F-measure=0.624	Researcher-inferred	EDDTW	Negative emotion lexicon, negative thought lexicon, negative event lexicon, and symptom lexicon.	10-fold cross-validation	Observational cohort study
Husseini Orabi et al [24], 2018	154 subjects (53 labeled as Depressed/101 labeled as Control), the ages of the samples were not reported	Twitter	The optimized CNN[j] model: accuracy=0.880	User-declared	CNN-based models, RNN[k]-based models, SVM	Twitter texts from among which all the @mentions, retweets, nonalphanumeric characters, and URLs were extracted by the researchers.	5-fold cross-validation	Observational cohort study
Islam et al [19], 2018	7145 Facebook comments (58% depressed/42% nondepressed), the ages of the samples were not reported	Facebook	Decision Tree (F-measure): emotional process=0.73, linguistic style=0.73, temporal process=0.73, all features=0.73; SVM (F-measure): emotional process=0.73, linguistic style=0.73, temporal process=0.73, all features=0.73; KNN[l] (F-measure): emotional process=0.71, linguistic style=0.70, temporal process=0.70, all features=0.67; Ensemble (F-measure): emotional process=0.73, linguistic style=0.73, temporal process=0.73, all features=0.73	User-declared	SVM, decision tree, ensemble, KNN	Emotional information (positive, negative, anxiety, anger, and sad), linguistic style (prepositions, articles, personal, conjunctions, auxiliary verbs), temporal process information (past, present, and future)	10-fold cross-validation	Observational cohort study

Reference	Sample	Platform	Outcome	Depression identification method	ML[a] approach type	Features examined	Cross-validation	Type of study
Shen et al [6], 2017	1402 depressed users, 36993 depression-candidate users, and over 300 million non-depressed users, the ages of the samples were not reported	Twitter	Accuracy: NB[m]=0.73, MSNL[n]=0.83, WDL[o]=0.77, MDL[p]=0.85	User-declared	MDL, NB, MSNL, WDL	Features of network interactions (number of tweets, social interactions, and posting behaviors), user profiles (users' personal information in social networks), and visual, emotional, and topic-level features, domain-specific features	5-fold cross-validation	Observational cohort study
De Choudhury, Gamon, et al [25], 2013	476 users (171 depressed/305 nondepressed), with a median age of 25	Twitter	Accuracy: engagement=0.553, egonetwork=0.612, emotion=0.643, linguistic style=0.684, depression language=0.692, demographics=0.513, all features=0.712	Researcher-inferred	SVM	Engagement, egocentric social graph, emotion, linguistic style, depression language, demographics	10-fold cross-validation	Observational cohort study
Mariñelarena-dondena et al [26], 2017	135 articles (20 depressed/115 nondepressed), the ages of the samples were not reported	Reddit	Precision=0.850, recall=0.810, F-measure=0.829, accuracy=0.948	User-declared	SVD, GBM[q], SMOTE[r]	n-grams, use of which can create a large feature space and hold much important information	Not reported	Observational cohort study
Tsugawa et al [27], 2015	209 Japanese users (81 depressed/128 nondepressed), and users were aged 16-55, with a median age of 28.8 years	Twitter	Precision=0.61, recall=0.37, F-measure=0.46, accuracy=0.66	Researcher-inferred	LDA[s], SVM	Frequencies of words used in the tweet, ratio of tweet topics found by LDA, ratio of positive-affect words contained in the tweet, ratio of negative-affect words contained in the tweet, hourly posting frequency, tweets per day, average number of words per tweet, overall retweet rate, overall mention rate, ratio of tweets containing a URL, number of users following, number of users followed	10-fold cross-validation	Observational cohort study
Chen et al [28], 2018	446 perinatal users, the ages of the samples were not reported	WeChat circle of friends	The result of LSTM[w] was similar to EPDS[x]	Researcher-inferred	LSTM	Top 10 emotions in the data set	Not reported	Observational cohort study

Reference	Sample	Platform	Outcome	Depression identification method	ML[a] approach type	Features examined	Cross-validation	Type of study
De Choudhury, Counts, et al [29], 2013	489 users, with a median age of 25 years	Twitter	Accuracy: eng.+ego=0.593, n-grams=0.600, style=0.658, emo.+time=0.686, all features=0.701	Researcher-inferred	PCA[t], SVM	Postcentric features (emotion, time, linguistic style, n-grams), user-centric features (engagement, ego-network)	5-fold cross-validation	Observational cohort study
Dinkel et al [30], 2019	142 speakers (42 depressed/100 nondepressed), the ages of the samples were not reported	Distress Analysis Interview Corpus-Wizard of Oz (WOZ-DAIC) database	Precision=0.93, recall=0.83, F-measure=0.87	Researcher-inferred	LSTM	Words from online posts	10-fold cross-validation	Observational cohort study
Sadeque et al [7], 2017	888 users (136 depressed/752 nondepressed), the ages of the samples were not reported	Reddit	F-measure: LibSVM[u]=0.40, WekaSVM[v]=0.30, RNN=0.34, Ensemble=0.45	User-declared	LibSVM, RNN, Ensemble, WekaSVM	Depression lexicon[y], metamap features[z]	5-fold cross-validation	Observational cohort study
Shatte et al [31], 2020	365 fathers in the perinatal period, the ages of the samples were not reported	Reddit	Precision=0.67, recall=0.68, F-measure=0.67, accuracy=0.66	Researcher-inferred	SVM	Fathers' behaviors, emotions, linguistic style, and discussion topics	10-fold cross-validation	Observational cohort study

Reference	Sample	Platform	Outcome	Depression identification method	ML[a] approach type	Features examined	Cross-validation	Type of study
Li et al [32], 2020	1,410,651 users, the ages of the samples were not reported	Twitter	Accuracy:SVM (radial basis function kernel)=0.82, SVM (linear kernel)=0.87, logistic regression=0.86, naïve Bayes=0.81, simple neural network=0.87	Researcher-inferred	SVM, logistic regression, naïve Bayes Classifier, simple neural network	512 features that were extracted from tweets using a universal sentence encoder	Not reported	Observational cohort study

[a]ML: machine learning.

[b]ROC: receiver operating characteristic.

[c]SS3: sequential S3 (smoothness, significance, and sanction).

[d]ANEW: affective norms for English words.

[e]LIWC: linguistic inquiry and word count.

[f]RF: random forest.

[g]SVM: support vector machine.

[h]PTT: the gossip forum on the Professional Technology Temple.

[i]EDDTW: event-driven depression tendency warning.

[j]CNN: convolutional neural networks.

[k]RNN: recurrent neural network.

[l]KNN: k-nearest neighbor.

[m]NB: naive Bayesian.

[n]MSNL: multiple social networking learning.

[o]WDL: Wasserstein Dictionary Learning.

[p]MDL: multimodal depressive dictionary learning.

[q]GBM: gradient boosting machine.

[r]SMOTE: synthetic minority oversampling technique.

[s]LDA: latent Dirichlet allocation.

[t]PCA: principal component analysis.

[u]LibSVM: library for support vector machines.

[v]WekaSVM: Waikato Environment for Knowledge Analysis for support vector machines.

[w]LSTM: long short-term memory.

[x]EPDS: Edinburgh Postnatal Depression Scale.

[y]A cluster of unigrams that has a great likelihood of appearing in depression-related posts.

[z]The features were extracted using Metamap based on concepts from the Unified Medical Language System Metathesaurus.

Table 3. Summary of the studies' depression identification methods.

Type of depression identification method and reference	Platform	Specific diagnostic method
Researcher-inferred mental status		
De Choudhury, Gamon, et al [25], 2013	Twitter	CES-D[a] questionnaire
De Choudhury, Counts, et al [29], 2013	Twitter	CES-D[a] questionnaire
Tsugawa et al [27], 2015	Twitter	CES-D[a] questionnaire
Chen et al [28], 2018	WeChat	The Edinburgh Postnatal Depression Scale (EPDS) questionnaire
Dinkel et al [30], 2019		The Patient Health Questionnaire (PHQ-8)
Li et al [32], 2020	Twitter	The Patient Health Questionnaire (PHQ)
Wang et al [21], 2013	Sina Microblog	Diagnosis by psychologists using interviews and questionnaires
Tung et al [15], 2016	PTT[b]	Diagnosis by three professional students
Shatte et al [31], 2020	Reddit	ICD-10[c] and diagnosis by a clinical psychologist specializing in perinatal mental health
User-declared mental status		
Burdisso et al [14], 2019	Reddit	Statements specifically indicating depression, such as "I was diagnosed with depression."
Mariñelarena-dondena et al [26], 2017	Reddit	Documents declaring depression diagnoses
Sadeque et al [7], 2017	Reddit	Statements like "I have been diagnosed with depression."
Husseini Orabi et al [24], 2018	Twitter	Documents declaring depression diagnoses
Shen et al [6], 2017	Twitter	Tweets of statements like "I was diagnosed with depression."
Islam et al [19], 2018	Facebook	Indication of depression by ground truth label information on selected posts
Community membership		
Nguyen et al [22], 2014	LiveJournal	Five "clinical" communities and five "control" communities
Fatima et al [23], 2018	LiveJournal	Five depressed and five nondepressed communities

[a]CES-D: Center For Epidemiologic Studies Depression Scale.

[b]PTT: the gossip forum on the professional technology temple.

[c]ICD-10: International Classification Of Diseases, tenth revision.

Identification Based on Researcher-Inferred Mental Status

Researcher-inferred mental status means that researchers identified the users' mental status based on the content of the users' online posts using ML approaches and professional diagnostic scales or expert opinions. Among the studies reviewed, 9 out of 17 studies [15,21,25,27-32] identified depression based on researcher-inferred mental status, while 6 [25,27-30,32] used professional diagnostic scales, and in 2 studies [15,21], depressive tendencies were identified in a traditional way, by clinical professionals; 1 study [31] used both diagnostic scales and expert opinions.

In particular, De Choudhury et al [25] conducted 2 studies using text data from Twitter in 2013. The first study collected data on 476 subjects who had reported depressive symptoms during September 2011-June 2012—among them 171 depressed users and 305 nondepressed users. The second study included 251 male and 238 female users, whose median age was 25 years [29].

In addition, Tsugawa et al [27] implemented data gathered from Japanese-speaking users through the Twitter application programming interface (API). They collected data on 209 participants—among them 121 males and 88 females aged 16 to 55 years, from December 4, 2013, to February 8, 2014, with a depression incidence of about 39%. The authors discuss the fact that Japanese personal pronouns work quite differently from those in Western languages, and subject words are often absent in Japanese texts, which could influence the performance of models being applied across different language contexts. And then, in a Chinese study, Chen et al [28] employed emoticon data from a WeChat circle of friends to detect perinatal depression. They gathered data on 446 perinatal participants, who had posted 1.17 million texts on the WeChat platform, 80% of the group being used as the training set, and the other 20% as the test set.

Finally, Dinkel et al [30] acquired text data from the Distress Analysis Interview Corpus-Wizard of Oz (WOZ-DAIC), a database that is publicly available. They combined data from 107 interviewees from the training set and 35 from the development set. The interviewees from the training set had a depression incidence of 28%, while those in the development set had an incidence of 34%.

In addition, Li et al [32] conducted a study using an ML approach to detect depression during the early COVID-19 outbreaks in the United States, based on the researcher-inferred depressive symptoms on Twitter. They collected tweet posts from 1,410,651 users, which were over 0.4% of the total population.

One of the 2 studies in which depression was identified in a traditional way was Wang et al's [21]. They collected data on users of Sina microblog, which is one of the most popular social network services in China. They collected information from 6013 microblogs dating from August 1-15, 2012, and thus identified 122 depressed and 346 nondepressed subjects from among several hundred who had volunteered for the study. The second study, Tung et al [15], gathered about 18,000 web posts from the Chinese-language online forum PTT (the Gossip Forum on the Professional Technology Temple) from March 2004 to September 2011, of which 724 posts were selected as testing and training data. Next, a Chinese word segmentation and part-of-speech labeling tool was used for sorting and labeling the posts.

One other study combined depression diagnostic criteria with expert opinions. Shatte et al [31] studied the depression-related changes in mood among fathers who reported the births of children on Reddit posts. The study collected social media data on the fathers during the prepartum and postpartum periods and assessed features including behaviors, emotions, linguistic styles, and discussion topics, as well as more basic information.

Identification Based on User-Declared Mental Status

"User-declared mental status" means that users declared, in social media posts, that they had been diagnosed with depression. Six studies used depression identification based on user declarations of mental status in the social media data. Burdisso et al [14] divided the data gathered from users on Reddit into a training set and a test set. The last data set included 135 depressed users and 752 nondepressed ones. Mariñelarena-dondena et al [26] constructed a data set containing 486 submissions, including posts and comments gathered from members of the Reddit community between February 2017 and April 2017. The final data set consisted of 83 depressed users and 403 nondepressed ones. And Sadeque et al [7] used the Reddit API to conduct data collection and constructed a data set of posts by 888 Redditors, among whom 136 were depressed, and 752 were nondepressed.

Two other studies constructed their experimental data sets using Twitter data. Husseini Orabi et al [24] selected 154 users whose Twitter labels were depressed or nondepressed. They also used the users' posts published under the Bell Let's Talk campaign, and the final data collection consisted of 53 depressed users and 101 nondepressed. The second study, Shen et al [6], gathered data from users whose tweets had stated "I was diagnosed with depression" on Twitter. Altogether, they collected 292,564 tweets that had been posted by 1402 depressed subjects over the course of a month.

Islam et al [19] collected text data from Facebook in order to explore ways of detecting depression. Of the total of 7145 posts that they collected, 58% were from depressed users, and 42% were from nondepressed ones.

Identification Based on Community Membership

Two studies that both explored depression identification, using community membership as an identifier, collected their data from LiveJournal. To construct a balanced data set, Nguyen et al [22] selected 5000 posts from five depressed ("clinical") communities and 5000 posts from five nondepressed ("control") communities. Fatima et al [23] also used data from five depressed and five nondepressed communities. Their final data set consisted of a total of 4026 posts, which included 2007 from nondepressed, and 2019 from depressed communities.

The ML Approaches for Depression Detection

The ML approaches used in these studies included supervised learning (SL) and unsupervised learning (UL) approaches. SL methods specify a targeted outcome variable, such as the presence of a mental disorder, and are often used in prediction tasks. UL methods are used to detect relationships among the variables in a data set in the absence of a specified target outcome or response variable to supervise the analyses. UL aims to discover underlying structures such as clusters, components, or dimensions, in the data set [12]. Among the 17 ML-based studies reviewed here, 14 used SL approaches to explore depression detection methods, [7,14,15,19,21-25,27,28,30-32] and 3 used UL approaches [6,26,29] (Table 4).

Table 4. Summary of the machine learning approaches used in the depression detection studies.

Study	Machine learning approaches	Features	Outcomes
Supervised learning approaches			
Nguyen et al [22], 2014	The Lasso model	Affective features, mood tags, thee linguistic inquiry and word count (LIWC) features and topics that were all extracted from posts on LiveJournal.	Community classification of user (Accuracy): ANEW=0.89, mood=0.96, topic=1, LIWC=1; Community classification of post (Accuracy): topic=0.93, LIWC=0.88
Chen et al [28], 2018	LSTM[a]	Top 10 emotions in the data set	Depression, according to the LSTM, and according to the EPDS[b]. The results were similar for both
Dinkel et al [30], 2019	LSTM	Words from online posts	Precision=0.93, recall=0.83, F-measure=0.87
Wang et al [21], 2013	Bayes,Trees, and Rules	Micro-blog content, interactions, and behaviors	Bayes: Mean absolute error=0.186, ROC=0.908, F-measure=0.85; Trees: Mean absolute error=0.239, ROC=0.798, F-measure=0.762; Rules: Mean absolute error=0.269, ROC=0.869,F-measure=0.812
Burdisso et al [14], 2019	The proposed model: SS3	Words in online text users posts on Reddit	SS3: F-measure =0.61, precision=0.63, recall=0.60
De Choudhury, Gamon et al [25], 2013	SVM[c]	Engagement, egocentric social graph, emotion, linguistic style, depression language, demographics	Accuracy: engagement=0.553, ego-network=0.612, emotion=0.643, linguistic style=0.684, depression language=0.692, demographics=0.513, all features=0.712
Tsugawa et al [27], 2015	LDA[d], SVM	Frequencies of words used in the tweet, ratio of tweet topics found by LDA, ratio of positive-affect words contained in the tweet, ratio of negative-affect words contained in the tweet, hourly posting frequency, tweets per day, average number of words per tweet, overall retweet rate, overall mention rate, ratio of tweets containing a URL, number of users following, number of users followed	Precision=0.61, recall=0.37, F-measure=0.46, accuracy=0.66
Islam et al [19], 2018	SVM, decision tree, ensemble, KNN[e]	Emotional information (positive, negative, anxiety, anger, and sad), linguistic style (prepositions, articles, personal, conjunctions, auxiliary verbs), temporal process information (past, present, and future)	Decision Tree (F-measure): emotional process=0.73, linguistic style=0.73, temporal process=0.73, all features=0.73; SVM (F-measure): emotional process=0.73, linguistic style=0.73, temporal process=0.73, all features=0.73; KNN (F-measure): emotional process=0.71, linguistic style=0.70, temporal process=0.70, all features=0.67; Ensemble (F-measure): emotional process=0.73, linguistic style=0.73, temporal process=0.73, all features=0.73
Fatima et al [23], 2018	Random forests, SVM	The feature set values serve as an input to the classification algorithm, which were extracted from first person singular, positive emotion, negative emotion, anxiety, cognitive process, insight, cause, affiliation health, and informal language of text online.	The proposed RF-based model (Accuracy): post=0.898, community=0.950, depression degree=0.923; SVM (Accuracy): post=0.82, community=0.895
Shatte et al [31], 2020	SVM	Fathers' behaviors, emotions, linguistic style, and discussion topics	Precision=0.67, recall=0.68, F-measure=0.67, accuracy=0.66
Husseini Orabi et al [24], 2018	CNN[f]-based models、RNN[g]-based models、SVM	Twitter text, among which all the @mentions, retweets, nonalphanumeric characters and, URLs were removed by researchers.	The optimized CNN model: accuracy=0.880
Sadeque et al [7], 2017	LibSVM, RNN, Ensemble, WekaSVM	Depression lexicon, metamap features	F-measure: LibSVM=0.40, WekaSVM=0.30, RNN=0.34, Ensemble=0.45
Li et al [32], 2020	SVM, logistic regression, naïve Bayes classifier, simple neural network	512 features that were extracted from tweets using a universal sentence encoder	Accuracy: SVM (radial basis function kernel)=0.82, SVM (linear kernel)=0.87, logistic regression=0.86, naïve Bayes=0.81, simple neural network=0.87

Study	Machine learning approaches	Features	Outcomes
Tung et al [15], 2016	EDDTW[j]	Negative emotion lexicon, negative thought lexicon, negative event lexicon, and symptoms lexicon	Precision=0.593 Recall=0.668 F-measure=0.624
Unsupervised learning approaches			
Shen et al [6], 2017	MDL[k], NB[l], MSNL[m], WDL[n]	Social network feature (number of tweets, social interactions, and posting behaviors), user profile feature (users' personal information in social networks), visual feature, emotional feature, topic-level feature、domain-specific feature	Accuracy: NB=0.73, MSNL=0.83, WDL=0.77, MDL=0.85
De Choudhury, Counts et al [29], 2013	PCA[o], SVM	Post-centric features (emotion, time, linguistic style, n-grams), user-centric features (engagement, ego-network)	Accuracy: eng.+ego=0.593, n-grams=0.600, style=0.658, emo.+time=0.686, all features=0.701
Mariñelarena-dondena et al [26], 2017	SVD[p], GBM[q], SMOTE[r]	n-grams that could produce large feature space and hold important information	Precision=0.850, recall=0.810, F-measure=0.829, accuracy=0.948

[a]LSTM: long short-term memory.
[b]EPDS: Edinburgh Postnatal Depression Scale Questionnaire.
[c]SVM: support vector machine.
[d]LDA: latent Dirichlet allocation.
[e]KNN: k-nearest neighbor.
[f]CNN: convolutional neural networks.
[g]RNN: recurrent neural network.
[h]LibSVM: a library for support vector machines.
[i]WekaSVM: Waikato Environment for Knowledge for support vector machines.
[j]EDDTW: event-driven depression tendency warning.
[k]MDL: multimodal depressive dictionary learning.
[l]NB: naive Bayesian.
[m]MSNL: multiple social networking learning.
[n]WDL: Wasserstein Dictionary Learning.
[o]PCA: principal component analysis.
[p]SVD: singular value decomposition.
[q]GBM: gradient boosting machine.
[r]SMOTE: synthetic minority oversampling technique.

Detection With Supervised Learning Approaches

The SL approaches used include regression and classification. Among the 14 studies that employed SL, 3 used regression [22,28,30], 8 adopted classification, [14,15,19,21,23,25,27,31], and 3 combined the two approaches [7,24,32].

Of the 3 that used regression-type SL approaches, Nguyen et al [22] found that the model performed best at community classification (accuracy of 100%) when linguistic inquiry and word count (LIWC) software and topics features were input, while affective feature and mood tags produced accuracies of 89% and 96%, respectively. Moreover, when LIWC and topics were used as feature sets in blog postclassification, performance was effective, with accuracies of 88% and 93%, respectively. Chen et al [28] conducted perinatal depression screening based on data from the WeChat circle of friends with a long short term memory (LSTM) network model. Their results indicated that the prediction power of LSTM was similar to that of the Edinburgh Postnatal Depression Scale, which has been demonstrated to be effective for perinatal depression detection. Similar to Chen et al's study [28], Dinkel et al [30] also adopted LSTM to conduct depression detection based on data from the WOZ-DAIC database. They found that the behavioral aspects of texts were more useful for depression detection than the actual text content, and the proposed bidirectional long short-term memory model obtained the best performance, with the highest F1 score (0.87).

Additionally, 8 studies focused on depression detection using classification-type SL approaches [14,15,19,21,23,25,27,31]. Among the 8 studies, 4 focused on social media users [14,21,25,27], while 4 focused on submissions (comments and posts) [15,19,23,31]. Among the 4 studies focusing on users, Wang et al [21] focused on 468 users of the Sina Microblog, employing three types of approaches: rules, trees, and Bayes, all of which achieved accuracies around 80%. They discovered that users' number of times of mentioning others was highly predictive of depression. Meanwhile, Burdisso et al [14] proposed the sequential S3 (smoothness, significance, and sanction; SS3) model to conduct depression screening using text data from 887 selected Reddit users, acquiring higher

prediction accuracy than other models. The SS3 model would prompt the "large-scale passive monitoring" to be conducted online incrementally. It was not aimed at autonomous diagnosis but intended as a complementary tool to other more well-established methods for diagnosing psychological problems. They also stated that a set of legal and ethical questions related to data ownership and protection were open to debate. In addition, De Choudhury, Gamon et al [25] and Tsugawa et al [27] both used SVM approaches. De Choudhury, Gamon et al [25] conducted depression detection using SVM and obtained good performance, with an accuracy of 70%. Analyzing the behaviors of depressed users, they concluded that depressed users showed decreasing social activity, higher self-attentional focus, more negative emotion, increased expression of religious thoughts, and heightened medicinal and relational concerns. Among those not using SVM, Tsugawa et al [27] applied LDA, which performed with an accuracy of 66%. They found that the research results on effective depression predictors for Japanese users were different from those that were effective for people posting in English. Specifically, the number of times posting and mentioning others had good predictive power for English-based studies [21,25] but were not robust features in the Japanese study.

Of the 4 studies that focused on posting submissions, Tung et al [15] proposed an event-driven depressive tendency warning (EDDTW) model for detecting depressive tendencies based on posts on PPT networks, which showed the highest F-measure score for 0.624 of the EDDTW model, suggesting that EDDTW could be used to track trends or changes in depression among post authors. Fatima et al [23] used random forests and SVM to conduct classifications, achieving post and community classifications based on random forests with accuracies of 90% and 95%, and post and community classifications based on SVM with accuracies of 82% and 90%. Shatte et al [31] collected 3889 submissions and 63,907 comments from fathers reporting birth events over a 6-month period and assessed the data using linear support vector classification. They found that SVM with linear kernel produced the best prediction results. Besides, Islam et al [19] conducted depression prediction based on text data from Facebook, showing that decision trees acquired the highest accuracy in different experiments than other ML approaches.

Finally, 3 studies combined regression and classification [7,24,32]. Husseini Orabi et al [24] used convolutional neural networks (CNN), recurrent neural network (RNN), and SVM to predict depression, obtaining high accuracy of 88% with an optimized CNN model. The experiment indicated that CNN-based models performed better than RNN-based models for depression detection, and user-level classification could perform well in imbalanced or small data sets. Sadeque et al [7] predicted depression with a library for support vector machines, RNN, WekaSVM (Waikato Environment for Knowledge for support vector machines), and Ensemble models. They found the ensemble models performed better than the individual model and more data could improve traditional performance measures. Moreover, Li et al [32] proposed the correlation explanation (CorExQ9) algorithm that integrates with clinical stress measure index (PHQ-9) for depression detection using biweekly COVID-19 related language data from Twitter. And the innovative algorithm predicts depressive symptoms effectively and could be applied to other cases for stress detection.

Detection With Unsupervised Learning Approaches

Three studies combined SL and UL approaches. Shen et al [6] employed four approaches of multimodal depressive dictionary learning (MDL), naive Bayesian, multiple social networking learning, Wasserstein Dictionary Learning, and they demonstrated that the proposed MDL model is effective in depression detecting, obtaining the best performance with an F1-Measure of 85%. The researchers found Twitter users' posting behaviors contributed more to depression detection than posting content. Simultaneously, De Choudhury, Counts et al [29] adopted principal component analysis and SVM as predicting approaches, and the SVM classifier achieved a high accuracy of 73% for depression detection. The researchers pointed out that the study conducted an analysis leveraging people's information and health behaviors, which might involve sensitive privacy and ethical issues about data protection. And the privacy and ethical issues deserved serious consideration in the process of research. Finally, the study of Mariñelarena-dondena et al [26] introduced singular value decomposition, gradient boosting machines, and synthetic minority oversampling techniques as predicting approaches, and the proposed deep learning approach performed better than other classifiers for depression detection, achieving an accuracy of over 94%.

Discussion

Principal Findings

Our review aimed to outline studies that conducted depression detection with ML approaches based on text from social media. According to studies included in this review, researchers would extract features from online text users posted on social media using text analysis strategies such as LIWC and other word-embedding models. Next, the researchers input the features into ML models to conduct depression prediction. The features among the seventeen studies were all produced based on words in the online text, such as emotional information, linguistic style, temporal process information, social network features, etc. As for ML approaches used in depression predicting, SL was adopted more than UL. According to the above-mentioned studies [22,23], ML approaches achieved good accuracies for depression detection using text from social media, such as Facebook, Twitter, mic-blog, etc. Nevertheless, some studies also presented that there were several challenges with ML approaches [14,15,21], and problems of piracy and popularization were ongoing concerns [18].

It is worth noting that there are some common patterns in the studies reviewed here. In terms of depression identification, the existing studies analyzed in this review are consistent in that the researchers, in each case, first identified depressed and nondepressed groups among their subjects, according to either researcher-inferred mental status, user-declared mental status, community membership, or clinicians' judgments and then explored ways of classifying the subjects using ML approaches,

and measured the accuracies of the models' predictions. Furthermore, most of the experiments reviewed here used SL rather than UL models. UL is used to identify unobserved or underlying psychological dimensions and explore how to achieve optimal classification, while SL uses existing information in the feature database for higher-level analyses concerning identification. SL uses classifications established ahead of time to explore ways to forecast a specific outcome of interest, such as the presence of a psychiatric disorder (eg, depression and anxiety). UL explores phenomena such as clustering and compression within sets of unlabeled data [33]. Therefore, for scenarios where prediction of a specific variable, such as depression, is the aim, SL approaches may be more accurate and efficient than UL approaches [12].

Several limitations and challenges of the depression identification and predicting models reviewed here should be acknowledged. Firstly, for depression identification, the fact that some types of information about the individuals, such as sociodemographic characteristics, behaviors behind the scenes, psychological, social, and cultural environment are often lacking in social media data and pose challenges that may be hard to resolve [29,31]. Secondly, the quantities of individual users' posts vary greatly, and posts containing too few of the terms designated as relevant input could lead to bias in depression identification [15]. Moreover, all of the study samples of the 17 studies reviewed were from either China, Japan, the United States, or other English-speaking countries. As a result of cultural and other differences, populations from different countries tend to differ in terms of posting frequencies and content, which may also lead to bias in depression identification. The generalizability of measurement standards for depression is also limited across countries and cultures [27]. It should also be mentioned that the studies reviewed here all explored ML approaches to detecting depression using only text data from social media, which may have limited their predictive efficacy. Given that social media data can also include videos, photos, etc, it may be that including more types of social media data in analyses could make depression identification programs more powerful.

The challenges facing ML approaches for depression detection may, however, be resolvable. For example, existing studies indicate that homophily exists among depressed users; that is to say, friends who interact with depressed users frequently are more likely to have depressive symptoms themselves. Therefore, the interactions and ties between users are significant. But the data used in such prediction models tend to be widely scattered on social media, and it is difficult to analyze the connections among individuals in such a way as to improve the accuracies of the ML approaches [21]. Moreover, only a large-scale data set could facilitate high accuracy in predictive applications. However, due to the characteristics of the data, it is hard to collect a sufficiently large mass of data to optimize the ML approaches applied. Often the studies are conducted based only on several hundred subjects [7,27].

In addition, the approaches and features selected for the analyses are crucial aspects of studies in this area. Wider ranges of possible features, such as specific depression lexicons appropriate for particular cultural populations or groups, and more complex techniques for analyzing posts should be explored with a view to ameliorating experimental processes and improving the accuracies of models [7]. The study conducted by De Choudhury, Gamon, et al [25], for example, in addition to using principal component analysis to perform feature reduction, also employed data abstraction techniques such as entropy, variance, and an average of the features which were significantly helpful in identifying the effects of the methods used in the study. Some approaches, however, tend to have deficiencies in both generalizability and variables selected for measurement. For example, there was a study that ended up not identifying depressed users, but only depressive tendencies, as revealed in posts on social media, because of the methods they applied [17]. Finally, there tends to be bias in the detection of depression when ML approaches are applied to data from social media. We know, for example, that youth and middle-aged people tend to be more active on social media than young children and older adults [32]. It's also true that there is a digital divide between people with higher and lower incomes [34], and people in more developed and richer countries and localities use social media more than those in poor and undeveloped areas, etc. What's more, most older adults seldom go on the internet. For example, according to the Pew Research Center, only 22% of American adults report using Twitter, and 73% of those people are under the age of 50 years [35]. Therefore, we cannot obtain data from social media that will represent all groups, leading to inherent population biases in studies based on social media.

To improve the validity and feasibility of depression detection research based on the application of ML approaches to social media data, increased efforts to reduce research bias will be needed. For depression identification, researchers should employ criteria and tools for depression diagnosis that are both accurate and suitable for different online populations. Moreover, collecting personal information such as sociodemographic characteristics and behaviors behind the scenes should also be considered, where necessary and ethical [31]. Furthermore, on methods used for predicting, first, it is important to refine the prediction results by continually exploring optimal input features, models, and ML approaches through constant training and learning with larger-scale samples. Second, studies should focus on standardizing the measures being used for depression detection with ML approaches and on developing scalable approaches for automated tracking of public psychological health in the future. Third, to avoid estimate biases caused by small sample sizes, researchers should focus on obtaining samples that are as large as possible for their analyses. Finally, discussions about the issues involved in the studies should include computer scientists, psychologists, clinicians, ethicists, lawyers, policymakers, as well as user representatives from various user groups.

Conclusions

In summary, the studies described in this review have demonstrated that ML approaches can be effective for detecting depression using text data from social media and that the objective of developing a highly valid approach for such research may be within reach. Additionally, it seems appropriate and applicable for these methods to function as a complementary

tool to the more traditional, established methods for diagnosing depression. However, further research is still needed in the areas of sample size, optimization of predictive approaches and features, generalizability, privacy issues, and general research ethics.

Acknowledgments

This work was supported by the National Natural Science Fund of China (grant numbers 71761130083 and 82173636).

Authors' Contributions

JG and DL designed the study. DL conducted the analysis, interpreted the collected articles, and wrote the first draft of the manuscript. DL, XLF, JG, FA, and MS contributed to valuable revisions of this manuscript. All authors contributed to further writing and approved the final version.

Conflicts of Interest

None declared.

References

1. Brundtland G. Mental health: a call for action by world health ministers. In: Geneva: World Health Organization. Geneva: World Health Organization; 2014.
2. Depression and other common mental disorders: Global health estimates Internet. Organization WH. 2017. URL: https://apps.who.int/iris/bitstream/handle/10665/254610/WHOMSD?sequence=1 [accessed 2022-02-08]
3. Yu B, Zhang X, Wang C, Sun M, Jin L, Liu X. Trends in depression among Adults in the United States, NHANES 2005-2016. J Affect Disord 2020 Feb 15;263:609-620. [doi: 10.1016/j.jad.2019.11.036] [Medline: 31744739]
4. Diagnostic and statistical manual of mental disorders (DSM-5®) Internet. Association AP. URL: https://www.psychiatry.org/psychiatrists/practice/dsm [accessed 2021-01-11]
5. Kumar M, Dredze M, Coppersmith G, De Choudhury M. Detecting Changes in Suicide Content Manifested in Social Media Following Celebrity Suicides. HT ACM Conf Hypertext Soc Media 2015 Sep;2015:85-94 [FREE Full text] [doi: 10.1145/2700171.2791026] [Medline: 28713876]
6. Shen G, Jia J, Nie L, Feng F, Zhang C, Hu T, et al. Depression detection via harvesting social media: A multimodal dictionary learning solution. 2017 Presented at: the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI-17); 19-25 August 2017; Melbourne, Australia. [doi: 10.24963/ijcai.2017/536]
7. Sadeque F, Xu D, Bethard S. UArizona at the CLEF eRisk 2017 Pilot Task: Linear and Recurrent Models for Early Depression Detection. CEUR Workshop Proc 2017 Sep;1866:1-11 [FREE Full text] [Medline: 29075167]
8. Kaplan AM, Haenlein M. Users of the world, unite! The challenges and opportunities of Social Media. Business Horizons 2010 Jan;53(1):59-68. [doi: 10.1016/j.bushor.2009.09.003]
9. Goffman E. The presentation of self in everyday life. New York: Doubleday Anchor Books; 1958.
10. Schwartz HA, Sap M, Kern ML, Eichstaedt JC, Kapelner A, Agrawal M, et al. Predicting Individual Well-being Through the Language of Social Media. Pac Symp Biocomput 2016;21:516-527 [FREE Full text] [Medline: 26776214]
11. Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. Science 2015 Jul 17;349(6245):255-260. [doi: 10.1126/science.aaa8415] [Medline: 26185243]
12. Jiang T, Gradus JL, Rosellini AJ. Supervised Machine Learning: A Brief Primer. Behav Ther 2020 Sep;51(5):675-687 [FREE Full text] [doi: 10.1016/j.beth.2020.05.002] [Medline: 32800297]
13. Shatte ABR, Hutchinson DM, Teague SJ. Machine learning in mental health: a scoping review of methods and applications. Psychol Med 2019 Jul;49(9):1426-1448. [doi: 10.1017/S0033291719000151] [Medline: 30744717]
14. Burdisso SG, Errecalde M, Montes-y-Gómez M. A text classification framework for simple and effective early depression detection over social media streams. Expert Systems with Applications 2019 Nov;133:182-197. [doi: 10.1016/j.eswa.2019.05.023]
15. Tung C, Lu W. Analyzing depression tendency of web posts using an event-driven depression tendency warning model. Artif Intell Med 2016 Jan;66:53-62. [doi: 10.1016/j.artmed.2015.10.003] [Medline: 26616420]
16. Mumtaz W, Malik AS, Yasin MAM, Xia L. Review on EEG and ERP predictive biomarkers for major depressive disorder. Biomedical Signal Processing and Control 2015 Sep;22:85-98. [doi: 10.1016/j.bspc.2015.07.003]
17. Orrù G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A. Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. Neurosci Biobehav Rev 2012 Apr;36(4):1140-1152. [doi: 10.1016/j.neubiorev.2012.01.004] [Medline: 22305994]
18. Guntuku SC, Yaden DB, Kern ML, Ungar LH, Eichstaedt JC. Detecting depression and mental illness on social media: an integrative review. Current Opinion in Behavioral Sciences 2017 Dec;18:43-49. [doi: 10.1016/j.cobeha.2017.07.005]
19. Islam MR, Kabir MA, Ahmed A, Kamal ARM, Wang H, Ulhaq A. Depression detection from social network data using machine learning techniques. Health Inf Sci Syst 2018 Dec;6(1):8. [doi: 10.1007/s13755-018-0046-0] [Medline: 30186594]

20. Study quality assessment tools Internet. Institute NHL and B. 2019. URL: https://www.nhlbi.nih.gov/health-topics/study-quality-assessmenttools [accessed 2020-02-01]
21. Wang X, Zhang C, Ji Y, Sun L, Wu L. A Depression Detection Model Based on Sentiment Analysis in Micro-blog Social Network. , Berlin, Heidelberg: Springer; 2013 Presented at: Pacific-Asia Conference on Knowledge Discovery and Data Mining 2013; 14-17 April, 2013; Gold Coast, QLD, Australia p. 201-213. [doi: 10.1007/978-3-642-40319-4_18]
22. Nguyen T, Phung D, Dao B, Venkatesh S, Berk M. Affective and Content Analysis of Online Depression Communities. IEEE Trans Affective Comput 2014 Jul 1;5(3):217-226. [doi: 10.1109/Taffc.2014.2315623]
23. Fatima I, Mukhtar H, Ahmad HF, Rajpoot K. Analysis of user-generated content from online social communities to characterise and predict depression degree. Journal of Information Science 2017 Nov 14;44(5):683-695. [doi: 10.1177/0165551517740835]
24. Husseini OA, Buddhitha P, Husseini OM, Inkpen D. Deep Learning for Depression Detection of Twitter Users. In: Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic. 2018 Jan Presented at: the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic; June 2018; New Orleans, LA. [doi: 10.18653/v1/w18-0609]
25. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. : AAAI; 2013 Jul 10 Presented at: the Seventh International AAAI Conference on Weblogs and Social Media; July 8–11, 2013; Cambridge, Massachusetts, USA p. 128-137 URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14432
26. Mariñelarena-dondena L, Sapino M, Errecalde M. Predicting Depression: a comparative study of machine learning approaches based on language usage. Centro de Estudios Académicos en Neuropsicología; Cuadernos de Neuropsicología 2017;11(3):42-54. [doi: 10.7714/CNPS/11.3.201]
27. Tsugawa S, Kikuchi Y, Kishino F, Nakajima K, Itoh Y, Ohsaki H. Recognizing depression from twitter activity. 2015 Aug Presented at: The ACM CHI Conference on Human Factors in Computing Systems 2015; April 18 - 23, 2015; Seoul, Republic of Korea p. 3187-3196. [doi: 10.1145/2702123.2702280]
28. Chen Y, Zhou B, Zhang W, Gong W, Sun G. Sentiment analysis based on deep learning and its application in screening for perinatal depression. : IEEE; 2018 Jul 19 Presented at: 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC); 18-21 June 2018; Guangzhou, China p. 451-456. [doi: 10.1109/DSC.2018.00073]
29. De Choudhury M, Counts S, Horvitz E. Social Media as a Measurement Tool of Depression in Populations. In: Proceedings of the 5th ACM International Conference on Web Science. 2013 May Presented at: the 5th Annual ACM Web Science Conference; May 2-4, 2013; Paris, France p. 47-56. [doi: 10.1145/2464464.2464480]
30. Dinkel H, Wu M, Yu K. Text-based Depression Detection: What Triggers An Alert Heinrich. DeepAI. 2019 Aug 4. URL: http://arxiv.org/abs/1904.05154 [accessed 2020-02-08]
31. Shatte ABR, Hutchinson DM, Fuller-Tyszkiewicz M, Teague SJ. Social Media Markers to Identify Fathers at Risk of Postpartum Depression: A Machine Learning Approach. Cyberpsychol Behav Soc Netw 2020 Sep;23(9):611-618. [doi: 10.1089/cyber.2019.0746] [Medline: 32915660]
32. Li D, Chaudhary H, Zhang Z. Modeling Spatiotemporal Pattern of Depressive Symptoms Caused by COVID-19 Using Social Media Data Mining. Int J Environ Res Public Health 2020 Jul 10;17(14):1-23 [FREE Full text] [doi: 10.3390/ijerph17144988] [Medline: 32664388]
33. Xu L. Optimal reverse prediction: A unified perspective on supervised, unsupervised and semi-supervised learning. In: Proceedings of the 26th Annual International Conference on Machine Learning. New York, the United States: Association for Computing Machinery; 2009 Presented at: the 26th Annual International Conference on Machine Learning; June 14 - 18, 2009; Montreal, Quebec, Canada p. 1137-1144. [doi: 10.1145/1553374.1553519]
34. Pew RC. Digital divide persists even as Americans with lower incomes make gains in tech adoption Internet. Pew Research Center. 2021 Jun 24. URL: https://www.pewresearch.org/fact-tank/2021/06/22/digital-divide-persists-even-as-americans-with-lower-incomes-make-gains-in-tech-adoption/ [accessed 2021-11-25]
35. Sizing up Twitter users Internet. Pew Research Center. 2019 Apr 24. URL: https://www.pewresearch.org/internet/2019/04/24/sizing-up-twitter-users/ [accessed 2021-11-25]

Abbreviations

API: application programming interface
CNN: convolutional neural networks
EDDTW: event-driven depression tendency warning
LDA: latent Dirichlet allocation
LIWC: linguistic inquiry and word count
LSTM: long short term memory
MDL: multimodal depressive dictionary learning
ML: machine learning
PTT: Gossip Forum on the Professional Technology Temple
RNN: recurrent neural network

SL: supervised learning
SS3: sequential S3 (smoothness, significance, and sanction)
SVM: support vector machine
UL: unsupervised learning
WOZ-DAIC: Distress Analysis Interview Corpus-Wizard of Oz

Edited by J Torous; submitted 18.01.21; peer-reviewed by C Biernesser, A Jolliff; comments to author 02.04.21; revised version received 26.05.21; accepted 16.12.21; published 01.03.22.

Please cite as:
Liu D, Feng XL, Ahmed F, Shahid M, Guo J
Detecting and Measuring Depression on Social Media Using a Machine Learning Approach: Systematic Review
JMIR Ment Health 2022;9(3):e27244
URL: https://mental.jmir.org/2022/3/e27244
doi:10.2196/27244
PMID:35230252

Review

Customized Information and Communication Technology for Reducing Social Isolation and Loneliness Among Older Adults: Scoping Review

Gomathi Thangavel[1], MSc; Mevludin Memedi[1], PhD; Karin Hedström[1], PhD

Centre for Empirical Research on Information Systems, Örebro University School of Business, Örebro, Sweden

Corresponding Author:
Gomathi Thangavel, MSc
Centre for Empirical Research on Information Systems
Örebro University School of Business
Fakultetsgatan 1
Örebro, 70281
Sweden
Phone: 46 19 302471
Email: gomathi.thangavel@oru.se

Abstract

Background: Advancements in science and various technologies have resulted in people having access to better health care, a good quality of life, and better economic situations, enabling humans to live longer than ever before. Research shows that the problems of loneliness and social isolation are common among older adults, affecting psychological and physical health. Information and communication technology (ICT) plays an important role in alleviating social isolation and loneliness.

Objective: The aim of this review is to explore ICT solutions for reducing social isolation or loneliness among older adults, the purpose of ICT solutions, and the evaluation focus of these solutions. This study particularly focuses on customized ICT solutions that either are designed from scratch or are modifications of existing off-the-shelf products that cater to the needs of older adults.

Methods: A scoping literature review was conducted. A search across 7 databases, including ScienceDirect, Association for Computing Machinery, PubMed, IEEE Xplore, PsycINFO, Scopus, and Web of Science, was performed, targeting ICT solutions for reducing and managing social isolation and loneliness among older adults. Articles published in English from 2010 to 2020 were extracted and analyzed.

Results: From the review of 39 articles, we identified 5 different purposes of customized ICT solutions focusing on reducing social isolation and loneliness. These were *social communication*, *social participation*, a *sense of belonging*, *companionship*, and *feelings of being seen*. The mapping of purposes of ICT solutions with problems found among older adults indicates that increasing social communication and social participation can help reduce social isolation problems, whereas fulfilling emotional relationships and feeling valued can reduce feelings of loneliness. In terms of customized ICT solution types, we found the following seven different categories: *social network*, *messaging services*, *video chat*, *virtual spaces or classrooms with messaging capabilities*, *robotics*, *games*, and *content creation and management*. Most of the included studies (30/39, 77%) evaluated the usability and acceptance aspects, and few studies (11/39, 28%) focused on loneliness or social isolation outcomes.

Conclusions: This review highlights the importance of discussing and managing social isolation and loneliness as different but related concepts and emphasizes the need for future research to use suitable outcome measures for evaluating ICT solutions based on the problem. Even though a wide range of customized ICT solutions have been developed, future studies need to explore the recent emerging technologies, such as the Internet of Things and augmented or virtual reality, to tackle social isolation and loneliness among older adults. Furthermore, future studies should consider evaluating social isolation or loneliness while developing customized ICT solutions to provide more robust data on the effectiveness of the solutions.

(JMIR Ment Health 2022;9(3):e34221) doi:10.2196/34221

KEYWORDS

social isolation; loneliness; review; ICT; older adults; customization; mobile phone

Introduction

Background

Advancements in science and various technologies have resulted in people having access to better health care, a good quality of life, and better economic situations, enabling humans to live longer than ever before. It is estimated that the number of older adults (aged ≥65 years) in the population, as of 2020, is approximately 727 million, and this number is expected to increase to 1.5 billion, which will be approximately 16% of the world's population by 2050 [1]. Humans are *social beings*; we are biologically and psychologically hardwired to stay connected and be social with other people. If this *socialness* is taken away from us, it can lead to social isolation and loneliness. According to De Jong Gierveld et al [2], loneliness is defined as the subjective feeling of being alone, whereas social isolation is the objective lack of social connections with other people. This review focuses on both social isolation and loneliness.

Studies show that the problems of loneliness and social isolation are much more common among older adults because of various factors such as living alone, the loss of family members or friends, chronic illness, and physical conditions. [3,4]. Furthermore, the recent COVID-19 pandemic has, without doubt, affected people belonging to all age groups. The COVID-19 quarantine restrictions have changed people's daily lives, resulting in reduced social interaction and social participation [5]. As a result, there has been an increased focus on social isolation and loneliness in all ages, especially in older adults [6]. Both social isolation and loneliness affect people psychologically by increasing stress, anxiety, depression, dementia, Alzheimer disease, cognitive decline, and the risk of suicide [7-9]. In addition, the same studies show that social isolation and loneliness affect people biologically by increasing the risk of many health conditions such as high blood pressure, a weakened immune system, obesity, heart disease, and death. Therefore, with this ongoing pandemic and social distancing norms, there is a need for everyone (not least older adults) to stay connected to prevent, reduce, and manage social isolation and loneliness.

Information and communication technology (ICT) can play an important role in alleviating social isolation and loneliness [10-12]. Social networking services such as Facebook and WhatsApp focus on connecting users with their family or friends and enhancing social relationships. However, the existing commercial off-the-shelf (COTS) products or applications such as Facebook mostly cater to the younger generation and do not consider the needs of older adults [13,14]. As a result, older adults find it difficult to adapt and use these technologies [15]. Therefore, there is a necessity to design customized solutions for older adults that are either designed from scratch or modifications of existing COTS products or applications tailored to the needs of individuals or groups. Although there have been some attempts to design and develop customized ICT solutions that are catered to older adults for managing social isolation or loneliness [16], there is a need to summarize current empirical research on these customized ICT solutions to understand what the existing ICT solutions provide and what purpose they have.

Currently, there are some literature reviews summarizing ICT solutions (see the *Related Research* section below) that address social isolation or loneliness for older adults. However, to the best of our knowledge, there is no literature review that summarizes only customized ICT solutions that are designed for older adults for reducing social isolation or loneliness.

In addition, in general, there are different mechanisms and purposes in developing solutions that target social isolation and loneliness [17,18]. For example, messaging applications focus on improving social communication among older adults to reduce their social isolation. In addition, recent technology has developed a long way and introduced social robots for older adults that provide companionship to combat loneliness [19]. Therefore, finding out the purpose of each ICT solution will help in choosing the appropriate solution for managing social isolation or loneliness. Hence, to address this gap, this review summarizes the customized ICT solutions that are designed for older adults for reducing social isolation and loneliness. In addition, this review investigates the purpose of each ICT solution and the evaluation focus of these solutions.

The following research questions have been identified and addressed in this study: (1) What were the purposes of the customized ICT solutions for reducing social isolation or loneliness? (2) What are the customized ICT solutions proposed for reducing social isolation or loneliness among older adults? (3) What aspects of customized ICT solutions have been evaluated?

This study updates the existing literature with the latest evidence on ICT solutions, focusing on social isolation or loneliness among older adults. This review also intends to provide practitioners and researchers in this field with a better insight into how to manage social isolation or loneliness among older adults by distinguishing different types of ICT solutions, the purposes of these solutions, and what has been evaluated.

Related Research

In this section, we will present previous literature reviews that exclusively investigated ICT interventions targeting social isolation or loneliness among older adults to position our literature review and knowledge contribution. Below, Table 1 provides a summary of the current literature reviews, describing the years the literature review covers, the problems investigated, whether the literature review included customized solutions, the purpose of the ICT intervention, and the types of ICT interventions identified in their studies.

As shown in Table 1, 6 reviews provided empirical evidence of ICT solutions for reducing loneliness or social isolation among older adults. However, they mostly covered general ICT use, computer training, and existing social network impact and included less customized solutions or did not include them at all [11,12,20-23]. The review by Baker et al [10] included studies with existing COTS applications as well as small-scale studies that developed prototypes or applications. They addressed social isolation and focused on solutions that increased social participation. In the same review, they found that social networking services (5 COTS and 8 customized) and *touch screen*–based interventions (1 COTS and 8 customized)

were primarily used to combat social isolation. None of the reviews in Table 1 covered only customized solutions that were designed specifically for older adults. In addition, the purposes of the ICT interventions were not examined explicitly in those reviews.

Table 1. Overview of existing literature reviews of information and communication technology (ICT) interventions.

Source	Years included in the review	Problem investigated	Included customized solutions	Purpose of ICT intervention covered (Research question 1)	Types of ICT interventions (part of research question 2)
Baker et al [10]	2000 to August 2016	Social participation and reducing social isolation	Yes (23 out of 36 included studies)	Not covered	• Touch screen technology • Social networking services • Adaptation of existing technology platforms • Use of games • ICT training
Chen and Schulz [11]	2002 to 2015	Social isolation (but included both loneliness and social isolation)	Yes (3 out of 25 included studies)	Not covered	• General ICT use • Social networking services • Telephone befriending • Video games • Virtual pet
Ibarra et al [12]	Until January 2020	Social isolation and loneliness	Yes (10 out of 25 included studies)	Not covered	• General internet use for interaction (eg, discussions in forums) and email • Video chat • Social networks • Virtual spaces or classrooms with messaging capabilities • Messaging services • Virtual companions • Phone calls
Khosravi et al [20]	2000 to 2015	Social isolation and loneliness	Yes (6 out of 34 included studies)	Not covered	• General ICT use • Video game • Robotics • Personal reminder information and social management system • Asynchronous peer support chat room • Social networking sites • Telecare • 3D virtual environment
Noone et al [21]	2004 to April 7, 2020	Social isolation and loneliness	Not included	Not covered	Only 1—videoconferencing
Choi et al [22]	2001 to July 2012	Loneliness	Not included	Not covered	Not categorized but included only computer training and general ICT use studies
Casanova et al [23]	2002 to 2019	Loneliness	Not included	Not covered	Not categorized but included only general ICT use, computer training, and social network studies

Methods

Overview

This paper undertakes a scoping review of the literature to summarize the customized ICT solutions for reducing social isolation or loneliness among older adults. We conducted a scoping literature review in which the mnemonic population, concept, and context guided the focus [24,25]. The population in question was older adults. The concept related to customized ICT solutions proposed for reducing social isolation or loneliness among older adults. The context was the setting; that is, where these types of ICT solutions are being used (eg, in a private home or in a nursing home). The review process started with planning the review protocol and continued with a search process, practical screening of articles, extraction, and analysis of data.

Search Strategy

Electronic searches for this study were conducted in December 2020 using the following seven databases: ScienceDirect, Association for Computing Machinery, PubMed, IEEE Xplore, PsycINFO, Scopus, and Web of Science. In this study, the Association for Computing Machinery and IEEE Xplore were chosen to cover computing and information technology articles, whereas PubMed and PsycINFO were chosen to include medical and psychology-related articles. ScienceDirect, Web of Science, and Scopus were selected to include studies in multidisciplinary areas of interest such as the social sciences. On the basis of the

research questions, keywords were divided into three categories: older adults ("Older" OR "Senior" OR "Elder"), technology intervention ("Information and Communication technology" OR "ICT" OR "Internet" OR "Mobile" OR "Sensor" OR "Social media" OR "Information technology" OR "HCI" OR "Human Computer Interaction" OR "Robot" OR "Computer"), and problem ("Loneliness" OR "Social isolation"). The search was limited to publication years (2010 to 2020); *title, abstract, and keywords*; and articles in English. The initial search process was carried out by the first author (GT). The inclusion and exclusion criteria and data extraction format were drafted by the first author and then reviewed and finalized in coordination with the coauthors. Following that, screening and data extraction were performed by the first author and, in case of any uncertainty, the coauthors were consulted. Further conflicts were resolved through discussion until a consensus among the authors was reached.

Inclusion and Exclusion Criteria

We defined the following inclusion and exclusion criteria (Textbox 1) to retrieve the eligible studies for this review. An article was retained if it met all the inclusion criteria and was rejected if it met any of the exclusion criteria.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- Full paper written in English
- Empirical studies that developed or presented a customized information and communication technology (ICT) solution with a primary focus on loneliness or social isolation
- Focus on older adults (as defined by the articles)

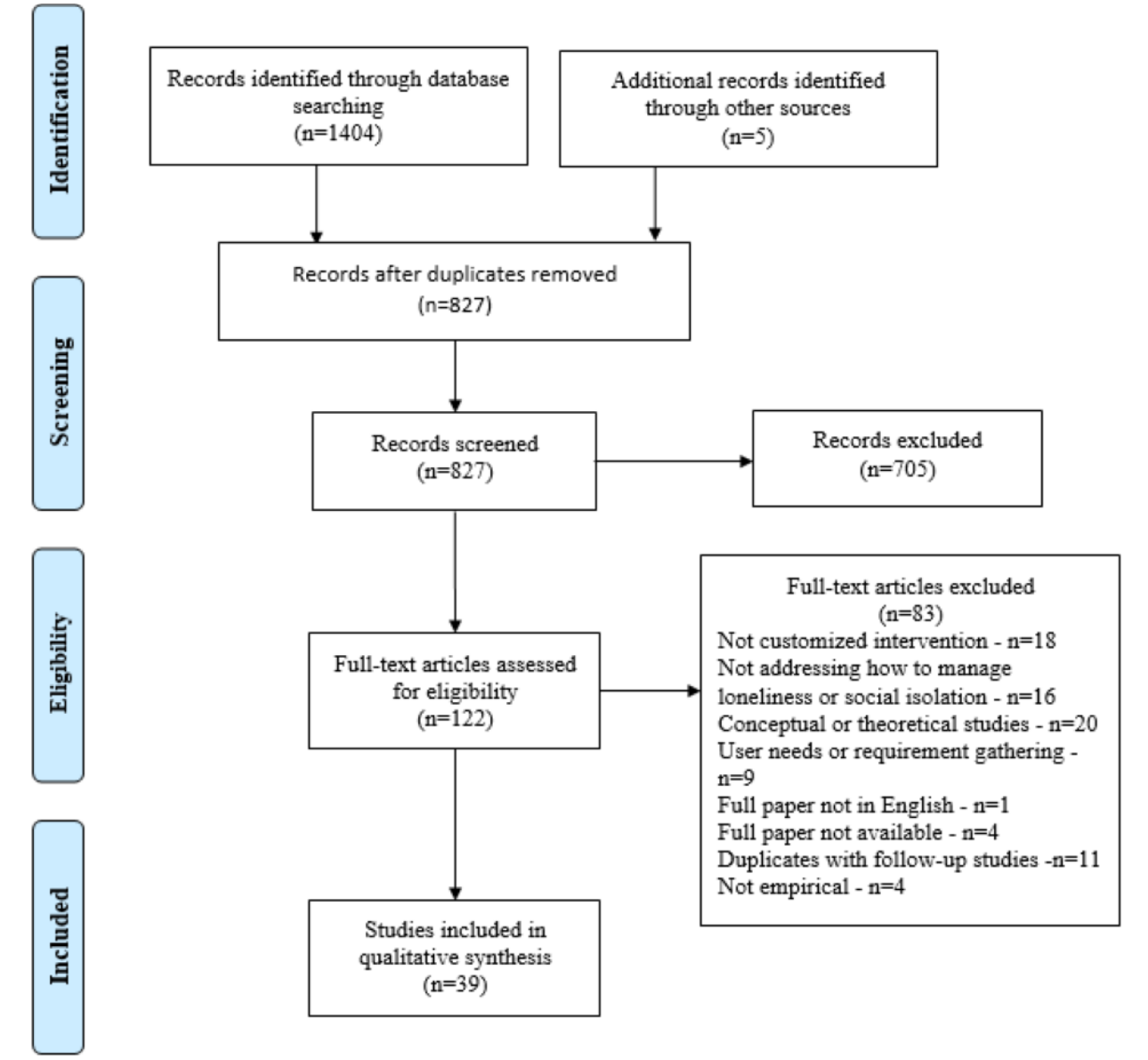
Exclusion criteria

- Nonresearch articles (such as magazines, guest editorial letters, forewords, keynotes, book reviews, posters, and workshop invitations)
- Empirical articles that did not develop or present a customized ICT solution for managing loneliness or social isolation
- Conceptual studies, theoretical studies, or review articles

Selection of Articles

Using the above-mentioned search strategy, 1409 articles were retrieved—584 (41.45%) from Web of Science, 338 (23.99%) from Scopus, 188 (13.34%) from PubMed, 97 (6.88%) from PsycINFO, 83 (5.89%) from the Association for Computing Machinery, 72 (5.11%) from IEEE, and 42 (2.98%) from ScienceDirect, and 5 (0.35%) additional articles were retrieved through a manual search by screening reference lists. Subsequently, 41.31% (582/1409) duplicate articles were removed. The remaining 827 articles were screened by reading their abstracts, which resulted in 122 (14.8%) papers that fit the focus and scope of this study. Full texts of the 122 studies were then screened using the inclusion and exclusion criteria (Textbox 1), and 72 (59%) papers were excluded. Furthermore, if an author published multiple follow-up articles with the same ICT solution, the most recent article or the one with most of the details was selected. In this way, another 9% (11/122) of articles were excluded, which led to 39 articles being retained. The entire selection process and the reasons for exclusion are outlined and reported in Figure 1 based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [26].

Figure 1. Overview of the selection process based on the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [26].



Data Analysis and Synthesis

From the shortlisted articles, the first author extracted the following data based on the research questions. To answer the first research question (ie, the intended purpose of the proposed ICT solution), we extracted the problem area and aim of the proposed solution. If the aim was not mentioned explicitly or was not clear, we analyzed the solution and noted what the solution was set to accomplish. For the second research question (ie, customized ICT solutions), we obtained the types of ICT solutions based on the use or functionality, the devices used, and their features. In addition, adapted from another review study [12], if the solution facilitated social interaction, we noted a different context of interactions and the contacts made by the participants. We used the existing categories (ie, social network, video chat, messaging services, video games, robotics, virtual spaces, or classrooms with messaging capabilities) identified in previous reviews [12,20] as a starting point to categorize the ICT solutions and added if any new type of ICT solution was found. For example, we added *content creation and management system*. For the third research question, we noted how the study was evaluated, its sample size, the age group of the included participants, the evaluation environment, the dependent variables, the scales used for the dependent variables, and the outcome of those measurements. Owing to the heterogeneity of the included studies, narrative synthesis was performed.

Results

Main Characteristics of the Included Studies

Most of the 39 included studies were conducted in Europe (20/39, 51%), followed by North America (9/39, 23%) and Asia (4/39, 10%). Nearly 16% (6/39) of the included studies lacked information about their country of study. The age group of the study population varied among the studies. Most of the studies (16/39, 41%) included older adults with a starting age of 65 years, followed by ≥60 years (8/39, 21%), ≥55 years (4/39, 10%), ≥70 years (3/39, 8%), ≥75 years (2/39, 5%), and ≥50

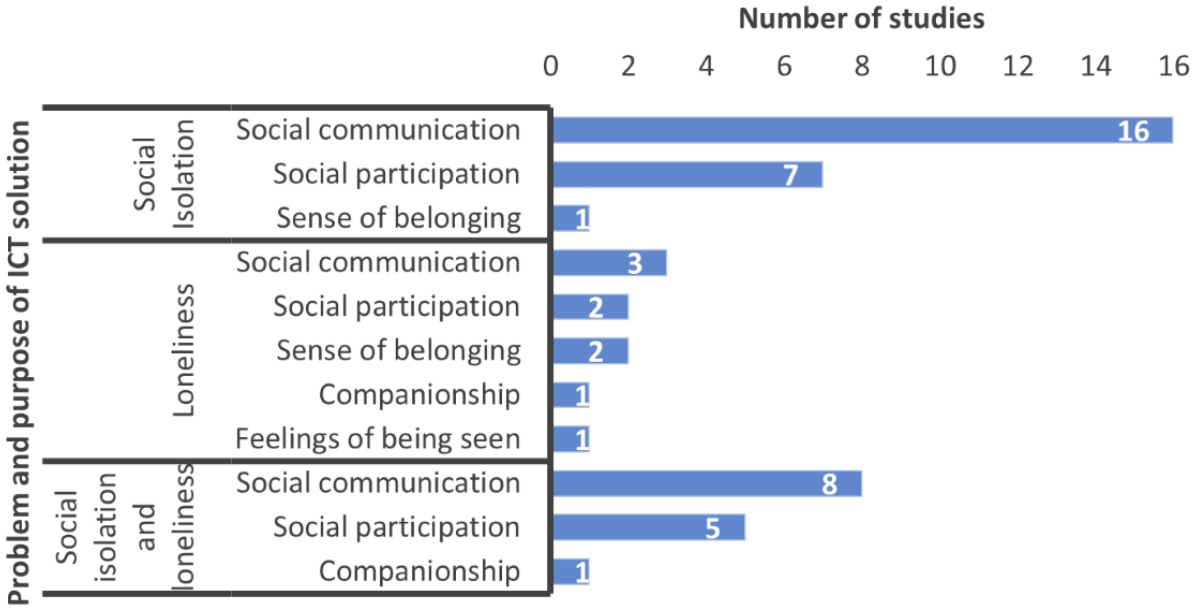
years (2/39, 5%), and the remaining studies did not mention the starting age of their study population. The customized solutions designed by the included studies were mostly tested in regular living environments (28/39, 72%), such as older adults' homes, care centers, retirement homes, and nursing homes. Of the 39 studies, 6 (15%) tested their solutions in their laboratory environments, 1 (3%) tested it in a hospital setting, another one (3%) evaluated it in an exhibition setting, and 3 (8%) did not provide information about the study environment.

From the review of 39 studies, most (20/39, 51%) addressed the problem of social isolation (Multimedia Appendix 1 [12,16,27-63]), whereas 9 (23%) focused on loneliness, and 10 (26%) focused on both social isolation and loneliness. Studies that mentioned both problems considered social isolation and loneliness as similar concepts and referred to them interchangeably. For instance, Sidner et al [27] referred to social isolation in their aim and used *loneliness* in the hypothesis, whereas Goumopoulos, Papa, and Stavrianos [28] used both *social isolation* and *loneliness* in their aim and hypothesis but measured only loneliness. Some studies defined the concept of social isolation as a lack of social relationships, a lack of social support, and reduced participation in social activities [29-31]. Of the 39 studies, 2 (5%) pointed out that, because of the high use of digital communication among younger generations, older adults who were reluctant to use or uncomfortable with existing technologies were not able to have social interactions and felt left out, which contributed to their social isolation [32,33]. The studies addressing loneliness defined it as "feeling invisible," solitude, and living alone [34-36]. Overall, 33% (13/39) of the studies did not have a clear definition of these 2 concepts.

Purposes of Customized ICT Solutions

With respect to the purposes of the proposed solutions, 5 different purposes were identified from the included studies. Most of the studies (31/39, 79%) focused on 1 purpose and designed the solution based on this, whereas 21% (8/39) focused on 2 purposes and provided solutions. The problems addressed in the included studies and the purposes of the proposed solutions are shown in Figure 2. Most of the studies (27/39, 69%; Multimedia Appendix 1) addressed the problem of social isolation or loneliness with the purpose of increasing *social communication* (ie, increasing the older adult's social interaction with their social contacts, mainly family and friends, through web-based chat, videoconferencing, group chat, and email [28,33,37]). Few studies provided the facility to interact with other contacts such as physicians or nurses [38] or with a virtual coach or helper [36,39,64]. One particular study (1/39, 3%) created a virtual coach application that helped and encouraged older adults to interact and make friends with strangers [36].

Figure 2. Problems addressed and purposes of the ICT solutions. ICT: information and communication technology.



The second most common purpose was *social participation* (14/39, 36%). Here, the studies focused on engaging older adults with web-based activities such as gaming [40-42] or participating in web-based exercise [43,64] or art making [44]. Few studies focused on discussing shared interests or hobbies either in a virtual class environment [45] or by forming a small social network group [29,30]. Other studies encouraged social participation by stimulating older adults to visit local social activities [28,46,47] or family events [48] or go shopping [49].

Increasing the *sense of belonging* was emphasized by a few studies to make older adults feel part of broader society, enhance their self-esteem, and feel valued [35,50,51]. For this, those studies created a platform to share user-generated content such as life experiences [35], memories [51], and cooking recipes [50] to pass on to next generations.

Overall, 5% (2/39) of the studies described the purpose as *companionship* to fulfill older adults' emotional relationships by providing virtual support with the help of robots and companionable agents [27] or through a virtual pet application [52].

In total, 3% (1/39) of the studies focused on increasing the feelings of *being seen* to create a sense of care and fulfillment among older adults [34].

Types of ICT Solutions

Overview

We found 7 different types of ICT solutions developed based on the purposes discussed in the previous section to answer the second research question. These were *social networks*, *messaging services*, *video chat*, *virtual spaces or classrooms with messaging capabilities*, *robotics*, *games*, and *content creation and management*. Few studies implemented combinations of these technologies in their solutions. Social networks were mainly proposed by many studies (14/39, 36%), followed by video chat (8/39, 21%), messaging services (8/39, 21%), robotics (6/39, 15%), virtual spaces or classrooms with messaging capabilities (7/39, 18%), games (4/39, 10%), and content creation and management (2/39, 5%). In these different types of solutions, a *tablet* device was the most used for implementation (16/39, 41%). Other devices used in the studies were televisions (8/39, 21%), desktops (6/39, 15%), customized robots or new robotic objects (6/39, 15%), smartphones (4/39, 10%), sensors (2/39, 5%), and a patented communication device named *ippi* [53] (1/39, 3%). Achilleos et al [45] created a customized device using existing devices such as the Mac mini, camera, and microphone, whereas Garattini et al [54] created a novel "building bridges device" using a touch screen computer connected to a custom-made stand along with a phone handset. Similarly, an Android phone was turned into a Raspberry-like board to create a memory music box [55]. Approximately 5% (2/39) of the studies used tangible interfaces such as a flower vase with a microphone [31] or a glass window with a television screen, camera, and microphone [56].

In terms of customization, a few created customized devices as mentioned above. Some created new applications with a customized interface [30,32]. In this regard, most of the studies developed the interface with simple and easy-to-use menus and layouts [16,32,57]. Some studies considered age-appropriate usability and accessibility guidelines based on previous literature or as per older adults' input (eg, voice typing; offering large, nontextual touch icons; audio messages; voice commands; and easy navigations [32,33,35,58,59]). In addition, some studies considered the television as an alternative to the desktop or tablet. For interactions, the user could either use a remote control or speech and gestures [60]. For individual privacy concerns, 3% (1/39) of the studies added avatar features for the older adult group exercise application, which can help reduce physical barriers and get users more engaged in the activities [64]. In addition, 5% (2/39) modified the existing social network interface and made privacy options simple and easy to understand [16,60].

Social Networks

Most of the studies (14/39, 36%) proposed social networks as a solution to overcome social isolation or loneliness problems by connecting older adults with their social contacts. There were 2 different types of social networks designed. Some studies created a *user-friendly tailored interface to access existing social networks* such as Facebook [16,28,32,60-62], Twitter [62], Instagram [33], and YouTube [35]. All the studies presented social media content from friends and family in a way suitable for older adults. In the above studies, except for Romanyk et al [62] and Tapia et al [33], the customized social networks allowed the user to create posts and share them with their network. The interfaces mostly used either the *tablet* or *television* to display the social media content to the user.

Another set of studies *created a new social network* to build a virtual community and participate in social activities [29,30,46,47,49,57]. Buhr et al [30] built a social network specifically for individuals with aphasia to share their personal stories and information about *living with aphasia*. In the same way, FridgeNet, a network created by Lee et al [49], focuses on the topic of a healthy diet, and users can share diet information with their peers. This also encourages users to meet face-to-face by sending a shopping invitation and purchasing food together. Approximately 5% (2/39) of the studies motivated older adults to meet friends or neighbors face-to-face who reside in the same community for participating in social activities [46,47].

Video Chat

The second most used technology by the studies (8/39, 21%) in their solutions was video chat to boost social interaction, mainly with family or friends. Video chat was mostly proposed as an additional social feature along with other solutions such as games and messaging services. Of the 39 studies, 3 (8%) developed video chat as the primary solution. Pereira et al [59] designed a smart remote-control application that can control the television as well as make video calls using voice commands. Approximately 5% (2/39) of the studies proposed a simple interface for video communication that can be used by older adults without much assistance. Angelini et al [56] used glass windows to establish a permanent video connection with a distant relative by opening the blind of a window screen. In a similar way, Kleinberger et al [55] created a memory music box. When older adults open the box, a photo slideshow is played and a notification is sent to their grandchild via email automatically. They can then make a video call, if available.

Messaging Services

Messaging services were used as a solution to enhance the social interactions of older adults, mainly with their family or friends. The different messaging services used in those studies were SMS text messages, voice messages, video messages, email, and photos. The interactions enabled through these messaging services were one-to-one communication in most of the studies except Garattini et al [54], who introduced group chat options. The study by Zaine et al [63] introduced a web-based application in which a human facilitator places a time-based request, collects the media message, and distributes it to a target person with a text commentary to deepen existing relationships.

Virtual Spaces or Classrooms With Messaging Capabilities

Virtual spaces or classrooms are web-based spaces where older adults form a group to discuss their interests and that provide opportunities to participate in voluntary activities such as teaching other older adults [40,45]. Approximately 5% (2/39) of the studies implemented web-based live exercise classes in which older adults could interact with instructors and other members virtually [43,64]. A "virtual coaching" application

was proposed by 5% (2/39) of the studies, providing friendship enrichment lessons to encourage older adults to make new friends by giving them tasks such as "go for a walk with someone" [36,39]. Another study (1/39, 3%) created a *virtual companion* in the form of a web-based pet application mediated by a human helper to provide companionship by monitoring the older adults visually, having deep conversations, and contacting caregivers in case of an emergency [52].

Robotics

Robotics technology provides emotional support by using nonverbal gestures, which increase the feelings of *being seen* to reduce loneliness [34]. Robots talk with older adults about different topics and connect them with their family or friends to provide companionship and increase social interaction [27,38,48]. In addition, robots help older adults engage in activities to keep them occupied, such as listening to music and playing games [27,41,48], or involve them in participatory arts such as reciting Shakespeare sonnets [44].

Games

Of the 39 studies, 2 (5%) primarily proposed games to entertain older adults and reconnect with other persons who participate in the games. Doppler et al [42] created an application for card games with a videoconferencing function to facilitate older adults' interaction while playing. Similarly, Correia et al [41] developed a robot and a touch interface to play card games with older adults as a team player and also as an opponent. Other studies considered games as secondary solutions in their applications.

Content Creation and Management System

Approximately 5% (2/39) of the studies developed simple and easy-to-use web applications, which had built-in templates that helped older adults create content. The NoBits application allowed older adults to capture and upload their memories, local history personified by photos, newspapers, and postcards. [51]. Similarly, the application developed by Tullius and Dogan [50] encouraged older adults to create and share food recipes to help people who were in need.

Evaluation Focus of Customized ICT Solutions

Overview

In terms of evaluations, all the studies except 1 [41] were evaluated with older adults. Correia et al [41] tested their game with younger participants. Most of the included studies (21/39, 54%; Multimedia Appendix 2 [12,16,27-63]) measured the *usability* of the developed solution. In addition, of the 39 studies, 11 (28%) reported the users' acceptance of their solutions; 11 (28%) examined loneliness; and 6 (15%) evaluated social isolation in the form of social support, social engagement, and social connectedness. Few studies also analyzed and reported use, general feedback about the solution, self-perceived health, quality of life, depression, emotional well-being, and self-esteem.

Usability and User Experience

Approximately 14% (3/21) of the studies carried out *heuristic evaluation* with experts initially to determine the usability problems [29,33,45]. The suggestions were implemented and evaluated further with the older adults. Of the 21 studies which measured usability, 6 (29%) measured with the System Usability Scale, 2 (10%) used the User Experience Questionnaire, and 1 (5%) used the Computer System Usability Questionnaire [51]. Almost all the studies that reported using these scales showed positive results and high ratings [29,45,60,64]. Only 5% (1/21) of the studies reported a below-average score of 65.3 on the System Usability Scale [46]. Older adults perceived the ICT interventions as easy to use and found them useful [29,32,40,53]. For some studies, it was initially difficult for older adults to use the services immediately, but training and support or use over time helped them gain confidence, which later improved their usability at the end of the study period [58,64]. The participants also rated the overall *user experience* [46,60,63] as high except for one of the studies [36], where participants were missing some fun in using the system.

Users' Acceptance of the Solutions

Of the 39 studies, 11 (28%) reported the user acceptance and attitudes toward the developed solutions. The results were mostly positive, and the participants were willing to use the system in the future [28,47,53]. Older adults perceived the system as useful and saw potential to improve their social connectedness [55]. Technology satisfaction significantly increased, and there was a significant difference between control and intervention over time [64]. In addition, the participants gave a positive opinion about using new technology solutions [50].

Loneliness and Social Isolation

Loneliness was measured using the University of California, Los Angeles Loneliness Scale in 45% (5/11) of the studies that measured loneliness, and 18% (2/11) used the revised version of the same scale. Jansen et al [46] used the De Jong Gierveld and Kamphuis 11-item loneliness scale, whereas Brandenburgh et al [39] used the short version of the same scale that comprises 6 items. Morganti et al [51] used the Italian Loneliness Scale, which has 18 items that are grouped into three subscales: emotional loneliness, social loneliness, and general loneliness. In total, 3 (27%) studies conducted randomized controlled trials in which there was a significant decrease in loneliness among older adults who used the intervention compared with the control group [40]. There was a reduction in loneliness in other studies [51,64], but there was no significant difference when compared with the control group. This was because there was constant contact with the coach over the phone in the exercise program [64], whereas the older adults in the control groups were doing a reminiscing activity with a children group in Morganti et al [51]. Other 8(73%) pre-post, quasi-experimental, and mixed methods studies had a varied response. Of the 8 studies, 4 (50%) resulted in a reduction in loneliness among older adults [39,43,44,52], whereas Goumopoulos et al [28] showed a moderate improvement in reducing loneliness, and 3 other studies (38%) [27,37,46] reported that there was no significant change in loneliness.

Social isolation was examined using different measurement scales such as the Friendship Scale, Lubben Social Network Scale, Duke Social Support Index, and Norbeck Social Support

Questionnaire. Social interactions and social engagement increased after a period of use [43,54,63]. There was a significant decline in social isolation and an increase in social support [40]. In contrast, in 5% (2/39) of the studies, there were no significant changes in the participants' social relationships or interactions [27,37].

Health-Related Outcomes

Health and health-related quality of life were analyzed in 10% (4/39) of the studies, of which 50% (2/4) showed an improvement in health [40,46], and 50% (2/4) reported no significant changes in health [27,58]. Approximately 5% (2/39) of the studies examined the status of depression and, in both, depression was reduced [43,44]. Emotional well-being improved [43]. The outcome of self-esteem only increased in the control group and did not increase in the intervention group [51].

Discussion

Principal Findings

In relation to previous reviews (Table 1), which mostly examined general ICT use or existing ICT interventions, this study reviewed only customized ICT solutions that were designed and developed for older adults to manage social isolation or loneliness. The analysis of the reviewed studies highlights a growing interest in applying customized ICT solutions for reducing social isolation or loneliness among older adults. The results underline the need to increase the aspects that contribute to reduced social isolation or loneliness among older adults using ICT solutions. Such aspects include *social communication*, *social participation*, a *sense of belonging*, *companionship*, and *feelings of being seen*. The studies mostly focused on increasing social communication followed by social participation to build solutions for managing social isolation. Here, *social communication* helps older adults increase their social connection with their family or friends, and *social participation* facilitates engagement in social activities with others. The studies that focused on loneliness alone (9/39, 23%) included aspects such as sense of belonging, companionship, and feelings of being seen apart from increasing social communication or participation [34,52]. Here, even though the purpose of *social communication* is the same, an emotional parasocial relationship was provided by a virtual coach [36]. Similarly, the other 3 purposes—a sense of belonging, companionship, and feelings of being seen—concentrate on fulfilling emotional relationships.

This mapping of problems (social isolation and loneliness) with the purposes of ICT solutions shows the important differences between the solutions for managing social isolation and loneliness among older adults, which was not highlighted in previous reviews. The results indicate that increasing social communication and social participation can help reduce the social isolation problem, whereas fulfilling emotional relationships and feeling valued can offer support in tackling loneliness. These results also highlight the importance of discussing and managing social isolation and loneliness as different but interrelated concepts and also emphasize the need for future research to use a suitable outcome measure for evaluating the ICTs based on the problem they address.

Seven different types of customized ICT solutions were identified: *social networks*, *messaging services*, *video chat*, *virtual spaces or classrooms with messaging capabilities*, *robotics*, *games*, and *content creation and management*. In contrast to previous reviews [11,12,20], which mostly reported general ICT use such as the use of computers and the internet, it was found that social networks were common in the reviewed studies, which is in line with the review by Baker et al [10]. Previous reviews [11,12,20] have included mostly off-the-shelf solutions and less customized solutions, which can explain this result. In terms of the devices used or suggested in the proposed solutions, *tablet* was the most preferred, followed by *television*. The choice of *tablet* for older adults is because of the portability and usability options it provides, whereas the television medium is widely adopted because users feel more comfortable and familiar as it has been present in almost all houses for several years [65-67]. Even though technology has widely advanced in terms of the Internet of Things and virtual or augmented reality, there were no studies in our review proposing their use. This highlights the need for future research to explore emerging technologies to reduce social isolation and loneliness.

Going further, most of the included studies evaluated the usability and acceptance aspects of the ICT solutions, and fewer studies focused on loneliness or social isolation outcomes. Training and support were shown to be important factors in achieving a greater usability score, which is in line with the results of the review by Ibarra et al [12]. In addition, older adults who perceived the application as useful were more willing to use the system in the future. Studies that evaluated loneliness or social isolation outcomes mostly reported a positive response in reducing loneliness or social isolation, but only 8% (3/39) of the studies were assessed using a randomized controlled trial design. In addition, the included studies in this review examined different outcomes by using several measurement scales or used qualitative methods and reported the outcomes in various ways. Owing to the heterogeneity of the included studies, we were not able to assess the effectiveness of each customized ICT solution. Future studies should consider evaluating social isolation or loneliness along with other outcomes while testing their customized ICT solutions.

Strengths and Weaknesses of This Review

This is the first review that had a clear focus on customized ICT solutions targeting social isolation or loneliness. This review highlighted the differences between the solutions for managing social isolation and loneliness. Although a significant effort was made to ensure the rigor of the search strategy, potentially relevant studies may not have been identified if the authors did not use the search keywords that were included in this review. In addition, limiting the search to include articles published in the English language may also have omitted additional relevant studies in other languages.

Conclusions

This scoping review summarized the customized ICT solutions designed for older adults for managing social isolation or loneliness. In addition, this review investigated the purpose of each ICT solution and the evaluation focus of these solutions. The mapping of social isolation and loneliness problems and

the purpose of ICT solutions shows the important differences between the solutions targeting social isolation and loneliness. In terms of ICT solutions, we found 7 different categories, of which social networks were the most proposed. Furthermore, this review highlights the importance of discussing and managing social isolation and loneliness as different but related concepts and emphasizes the need for future research to use suitable outcome measures for evaluating the interventions based on the problem. Even though a wide range of customized ICT solutions have been developed, future studies need to explore recent emerging technologies such as the Internet of Things and augmented or virtual reality to tackle social isolation and loneliness among older adults. Finally, future studies should consider evaluating social isolation or loneliness while testing customized ICT solutions.

Acknowledgments

This project received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement number 754285

Conflicts of Interest

None declared.

Multimedia Appendix 1
Details of the included studies.
[PDF File (Adobe PDF File), 102 KB - mental_v9i3e34221_app1.pdf]

Multimedia Appendix 2
Details of the interventions' dependent variables and their outcomes.
[PDF File (Adobe PDF File), 140 KB - mental_v9i3e34221_app2.pdf]

References

1. United Nations Department of Economic and Social Affairs. World Population Ageing 2020 Highlights: Living Arrangements of Older Persons. New York City: United Nations; 2021. URL: https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/undesa_pd-2020_world_population_ageing_highlights.pdf [accessed 2021-06-20]
2. de Jong J, Van Tilburg T, Dykstra P. Loneliness and social isolation. In: Cambridge Handbook of Personal Relationships. Cambridge, MA: Cambridge University Press; 2006:485-500.
3. Savikko N, Routasalo P, Tilvis RS, Strandberg TE, Pitkälä KH. Predictors and subjective causes of loneliness in an aged population. Arch Gerontol Geriatr 2005;41(3):223-233. [doi: 10.1016/j.archger.2005.03.002] [Medline: 15908025]
4. Grenade L, Boldy D. Social isolation and loneliness among older people: issues and future challenges in community and residential settings. Aust Health Rev 2008 Aug;32(3):468-478. [doi: 10.1071/ah080468] [Medline: 18666874]
5. Plagg B, Engl A, Piccoliori G, Eisendle K. Prolonged social isolation of the elderly during COVID-19: between benefit and damage. Arch Gerontol Geriatr 2020;89:104086 [FREE Full text] [doi: 10.1016/j.archger.2020.104086] [Medline: 32388336]
6. Berg-Weger M, Morley JE. Editorial: loneliness and social isolation in older adults during the COVID-19 pandemic: implications for gerontological social work. J Nutr Health Aging 2020;24(5):456-458 [FREE Full text] [doi: 10.1007/s12603-020-1366-8] [Medline: 32346678]
7. Hawthorne G. Measuring social isolation in older adults: development and initial validation of the friendship scale. Soc Indic Res 2006 Feb 13;77(3):521-548. [doi: 10.1007/s11205-005-7746-y]
8. Crewdson J. The effect of loneliness in the elderly population: a review. Healthy Aging Clin Care Elderly 2016;8. [doi: 10.4137/hacce.s35890]
9. Holt-Lunstad J, Smith TB, Baker M, Harris T, Stephenson D. Loneliness and social isolation as risk factors for mortality: a meta-analytic review. Perspect Psychol Sci 2015 Mar;10(2):227-237. [doi: 10.1177/1745691614568352] [Medline: 25910392]
10. Baker S, Warburton J, Waycott J, Batchelor F, Hoang T, Dow B, et al. Combatting social isolation and increasing social participation of older adults through the use of technology: a systematic review of existing evidence. Australas J Ageing 2018 Sep;37(3):184-193. [doi: 10.1111/ajag.12572] [Medline: 30022583]
11. Chen YR, Schulz PJ. The effect of information communication technology interventions on reducing social isolation in the elderly: a systematic review. J Med Internet Res 2016 Jan 28;18(1):e18 [FREE Full text] [doi: 10.2196/jmir.4596] [Medline: 26822073]
12. Ibarra F, Baez M, Cernuzzi L, Casati F. A systematic review on technology-supported interventions to improve old-age social wellbeing: loneliness, social isolation, and connectedness. J Healthc Eng 2020 Jul 13;2020 [FREE Full text] [doi: 10.1155/2020/2036842] [Medline: 32765823]

13. Eisma R, Dickinson A, Goodman J, Syme A, Tiwari L, Newell AF. Early user involvement in the development of information technology-related products for older people. Univers Access Inform Society 2004 Jun 1;3(2):131-140. [doi: 10.1007/s10209-004-0092-z]
14. McMurtrey M, McGaughey R, Downey J. Seniors and information technology: are we shrinking the digital divide? J Int Technol Inform Manag 2008;17(2):1-17 [FREE Full text]
15. Lee C, Coughlin JF. PERSPECTIVE: older adults' adoption of technology: an integrated approach to identifying determinants and barriers. J Prod Innov Manag 2014 Jun 03;32(5):747-759. [doi: 10.1111/jpim.12176]
16. Gomes G, Duarte C, Coelho J, Matos E. Designing a Facebook interface for senior users. Scient World J 2014;2014:741567 [FREE Full text] [doi: 10.1155/2014/741567] [Medline: 24672361]
17. Gardiner C, Geldenhuys G, Gott M. Interventions to reduce social isolation and loneliness among older people: an integrative review. Health Soc Care Community 2018 Mar;26(2):147-157. [doi: 10.1111/hsc.12367] [Medline: 27413007]
18. Fakoya OA, McCorry NK, Donnelly M. Loneliness and social isolation interventions for older adults: a scoping review of reviews. BMC Public Health 2020 Feb 14;20(1):129 [FREE Full text] [doi: 10.1186/s12889-020-8251-6] [Medline: 32054474]
19. Pirhonen J, Tiilikainen E, Pekkarinen S, Lemivaara M, Melkas H. Can robots tackle late-life loneliness? Scanning of future opportunities and challenges in assisted living facilities. Futures 2020 Dec;124:102640 [FREE Full text] [doi: 10.1016/j.futures.2020.102640] [Medline: 33041358]
20. Khosravi P, Rezvani A, Wiewiora A. The impact of technology on older adults' social isolation. Comput Hum Behav 2016 Oct;63:594-603. [doi: 10.1016/j.chb.2016.05.092]
21. Noone C, McSharry J, Smalle M, Burns A, Dwan K, Devane D, et al. Video calls for reducing social isolation and loneliness in older people: a rapid review. Cochrane Database Syst Rev 2020 May 21;5:CD013632 [FREE Full text] [doi: 10.1002/14651858.CD013632] [Medline: 32441330]
22. Choi M, Kong S, Jung D. Computer and internet interventions for loneliness and depression in older adults: a meta-analysis. Healthc Inform Res 2012 Sep;18(3):191-198 [FREE Full text] [doi: 10.4258/hir.2012.18.3.191] [Medline: 23115742]
23. Casanova G, Zaccaria D, Rolandi E, Guaita A. The effect of information and communication technology and social networking site use on older people's well-being in relation to loneliness: review of experimental studies. J Med Internet Res 2021 Mar 01;23(3):e23588 [FREE Full text] [doi: 10.2196/23588] [Medline: 33439127]
24. Lockwood C, Dos Santos KB, Pap R. Practical guidance for knowledge synthesis: scoping review methods. Asian Nurs Res (Korean Soc Nurs Sci) 2019 Dec;13(5):287-294 [FREE Full text] [doi: 10.1016/j.anr.2019.11.002] [Medline: 31756513]
25. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Nov 19;18(1):143 [FREE Full text] [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
26. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
27. Sidner CL, Bickmore T, Nooraie B, Rich C, Ring L, Shayganfar M, et al. Creating new technologies for companionable agents to support isolated older adults. ACM Trans Interact Intell Syst 2018 Aug 08;8(3):1-27. [doi: 10.1145/3213050]
28. Goumopoulos C, Papa I, Stavrianos A. Development and evaluation of a mobile application suite for enhancing the social inclusion and well-being of seniors. Informatics 2017 Jun 22;4(3):15. [doi: 10.3390/informatics4030015]
29. Alaoui M, Lewkowicz M. A livinglab approach to involve elderly in the design of smart TV applications offering communication services. In: Proceedings of the International Conference on Online Communities and Social Computing. 2013 Presented at: International Conference on Online Communities and Social Computing; July 21-26, 2013; Las Vegas, NV, USA p. 325-334. [doi: 10.1007/978-3-642-39371-6_37]
30. Buhr HR, Hoepner JK, Miller H, Johnson C. AphasiaWeb: development and evaluation of an aphasia-friendly social networking application. Aphasiology 2016 Sep 22;31(9):999-1020. [doi: 10.1080/02687038.2016.1232361]
31. Zhao M, Chen Z, Lu K, Li C, Qu H, Ma X. Blossom: design of a tangible interface for improving intergenerational communication for the elderly. In: Proceedings of the International Symposium on Interactive Technology and Ageing Populations. 2016 Presented at: ITAP '16: International Symposium on Interactive Technology and Ageing Populations; October 20 - 22, 2016; Kochi Japan p. 87-98. [doi: 10.1145/2996267.2996276]
32. Muñoz D, Gutierrez F, Ochoa S, Baloian N. SocialConnector: a ubiquitous system to ease the social interaction among family community members. Comput Syst Sci Eng 2015;30(1):57-68 [FREE Full text]
33. Tapia J, Gutierrez F, Ochoa S. Using Smart TV applications for providing interactive ambient assisted living services to older adults. In: Proceedings of the International Conference on Ubiquitous Computing and Ambient Intelligence. 2016 Presented at: International Conference on Ubiquitous Computing and Ambient Intelligence; November 29 - December 2, 2016; Las Palmas de Gran Canaria, Spain p. 514-524. [doi: 10.1007/978-3-319-48746-5_53]
34. Zuckerman O, Walker D, Grishko A, Moran T, Levy C, Lisak B, et al. Companionship is not a function: the effect of a novel robotic object on healthy older adults' feelings of "Being-seen". In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20: CHI Conference on Human Factors in Computing Systems; April 25 - 30, 2020; Honolulu HI USA p. 1-14. [doi: 10.1145/3313831.3376411]

35. Marcelino I, Laza R, Pereira A. SSN: senior social network for improving quality of life. Int J Distrib Sensor Networks 2016 Jul 22;12(7):2150734. [doi: 10.1177/155014772150734]
36. Muuraiskangas S, Leist A, Braun A, Klauß K, Roelofsma P, Wichert R, et al. V2me: evaluating the first steps in mobile friendship coaching. J Ambient Intellig Smart Environ 2012;4(6):517-534. [doi: 10.3233/ais-2012-0177]
37. Barbosa NB, Franz R, Judges R, Beermann C, Baecker R. Can digital technology enhance social connectedness among older adults? A feasibility study. J Appl Gerontol 2017 Nov 01:733464817741369. [doi: 10.1177/0733464817741369] [Medline: 29166818]
38. Koceski S, Koceska N. Evaluation of an assistive telepresence robot for elderly healthcare. J Med Syst 2016 May;40(5):121. [doi: 10.1007/s10916-016-0481-x] [Medline: 27037685]
39. Brandenburgh A, van Breda W, van der Ham W, Klein M, Moeskops L, Roelofsma P. Increasing physical and social activity through virtual coaching in an ambient environment. In: Proceedings of the International Conference on Active Media Technology. 2014 Presented at: International Conference on Active Media Technology; August 11-14, 2014; Warsaw, Poland p. 487-500. [doi: 10.1007/978-3-319-09912-5_41]
40. Czaja SJ, Boot WR, Charness N, Rogers WA, Sharit J. Improving social support for older adults through technology: findings from the prism randomized controlled trial. Gerontologist 2017 Feb 15:467-477 [FREE Full text] [doi: 10.1093/geront/gnw249] [Medline: 28201730]
41. Correia F, Alves-Oliveira P, Maia N, Ribeiro T, Petisca S, Melo F, et al. Just follow the suit! trust in human-robot interactions during card game playing. In: Proceedings of the 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). 2016 Presented at: 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN); Aug. 26-31, 2016; New York, NY, USA p. 507-512. [doi: 10.1109/roman.2016.7745165]
42. Doppler J, Gradl C, Sommer S, Rottermanner G. Improving user engagement and social participation of elderly people through a TV and tablet-based communication and entertainment platform. In: Proceedings of the International Conference on Computers Helping People with Special Needs. 2018 Presented at: International Conference on Computers Helping People with Special Needs; July 11-13, 2018; Linz, Austria p. 365-373. [doi: 10.1007/978-3-319-94274-2_51]
43. Isaacson M, Cohen I, Shpigelman C. Leveraging emotional wellbeing and social engagement of the oldest old by using advanced communication technologies: a pilot study using Uniper-Care's technology. In: Proceedings of the 3rd IET International Conference on Technologies for Active and Assisted Living (TechAAL 2019). 2019 Presented at: 3rd IET International Conference on Technologies for Active and Assisted Living (TechAAL 2019); March 25, 2019; London, UK URL: https://tinyurl.com/2vcs9ebw [doi: 10.1049/cp.2019.0102]
44. Fields N, Xu L, Greer J, Murphy E. Shall I compare thee…to a robot? An exploratory pilot study using participatory arts and social robotics to improve psychological well-being in later life. Aging Ment Health 2021 Mar;25(3):575-584. [doi: 10.1080/13607863.2019.1699016] [Medline: 31851830]
45. Achilleos A, Mettouris C, Papadopoulos G, Neureiter K, Rappold C, Moser C, et al. The connected vitality system: enhancing social presence for older adults. In: Proceedings of the 12th International Conference on Telecommunications. 2013 Presented at: 12th International Conference on Telecommunications; June 26-28, 2013; Zagreb, Croatia p. 199-206 URL: https://ieeexplore.ieee.org/document/6578290
46. Jansen-Kosterink SM, Bergsma J, Francissen A, Naafs A. The first evaluation of a mobile application to encourage social participation for community-dwelling older adults. Health Technol 2020 May 02;10(5):1107-1113. [doi: 10.1007/s12553-020-00430-9]
47. Gao Q, Ebert D, Chen X, Ding Y. Design of a mobile social community platform for older Chinese people in urban areas. Hum Factors Man 2012 Jun 27:66-89. [doi: 10.1002/hfm.20523]
48. Casey D, Barrett E, Kovacic T, Sancarlo D, Ricciardi F, Murphy K, et al. The perceptions of people with dementia and key stakeholders regarding the use and impact of the social robot MARIO. Int J Environ Res Public Health 2020 Nov 20;17(22):8621 [FREE Full text] [doi: 10.3390/ijerph17228621] [Medline: 33233605]
49. Lee Y, Huang M, Zhang X, Xu W. Fridgenet: a nutrition and social activity promotion platform for aging populations. IEEE Intell Syst 2015 Jul;30(4):23-30. [doi: 10.1109/mis.2015.55]
50. Tullius G, Dogan G. PrecRec: supporting older adults sharing recipes. In: Proceedings of the Conference on Mensch und Computer. 2020 Presented at: MuC'20: Mensch und Computer 2020; September 6 - 9, 2020; Magdeburg Germany p. 455-459. [doi: 10.1145/3404983]
51. Morganti L, Scaratti C, Cipresso P, Gaggioli A, Bonfiglio S, Riva G. How can technology help intergenerational reminiscence? A pilot study. Int J Web Based Commun 2016;12(1):35. [doi: 10.1504/ijwbc.2016.074275]
52. Machesney D, Wexler S, Chen T, Coppola J. Gerontechnology companion: virutal pets for dementia patients. In: Proceedings of the IEEE Long Island Systems, Applications and Technology (LISAT) Conference 2014. 2014 Presented at: IEEE Long Island Systems, Applications and Technology (LISAT) Conference 2014; May 2, 2014; Farmingdale, NY, USA p. 1-3. [doi: 10.1109/lisat.2014.6845226]
53. Scandurra I, Sjölinder M. Participatory design with seniors: design of future services and iterative refinements of interactive ehealth services for old citizens. Med 2 0 2013 Dec;2(2):e12 [FREE Full text] [doi: 10.2196/med20.2729] [Medline: 25075235]

54. Garattini C, Wherton J, Prendergast D. Linking the lonely: an exploration of a communication technology designed to support social interaction among older adults. Univ Access Inf Soc 2011 Jun 7;11(2):211-222. [doi: 10.1007/s10209-011-0235-y]
55. Kleinberger R, Rieger A, Sands J, Baker J. Supporting elder connectedness through cognitively sustainable design interactions with the memory music box. In: Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology. 2019 Presented at: UIST '19: The 32nd Annual ACM Symposium on User Interface Software and Technology; October 20 - 23, 2019; New Orleans LA USA p. 355-369. [doi: 10.1145/3332165.3347877]
56. Angelini L, Carrino F, Caon M, Lemaréchal F, Couture N, Khaled OA, et al. Testing the tangible interactive window with older adults. GeroPsych 2016 Dec;29(4):215-224. [doi: 10.1024/1662-9647/a000159]
57. Restyandito, Febryandi, Nugraha K, Sebastian D. Mobile social media interface design for elderly in Indonesia. In: Proceedings of the International Conference on Human-Computer Interaction. 2020 Presented at: International Conference on Human-Computer Interaction; July 19-24, 2020; Copenhagen, Denmark p. 79-85. [doi: 10.1007/978-3-030-60703-6_10]
58. Caroux L, Consel C, Sauzéon H, Dupuy L. Designing an accessible and engaging email application for aging in place. In: Proceedings of the IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI) Conference. 2017 Presented at: IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computed, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI) Conference; Aug. 4-8, 2017; San Francisco, CA, USA p. 1-8. [doi: 10.1109/uic-atc.2017.8397527]
59. Pereira A, Silva F, Ribeiro J, Marcelino I, Barroso J. Smart remote control design for seniors. In: Proceedings of the International Conference on Universal Access in Human-Computer Interaction. 2015 Presented at: International Conference on Universal Access in Human-Computer Interaction; August 2-7, 2015; Los Angeles, CA, USA p. 484-495. [doi: 10.1007/978-3-319-20681-3_46]
60. Coelho J, Rito F, Duarte C. "You, me and TV" — Fighting social isolation of older adults with Facebook, TV and multimodality. Int J Hum-Comput Stud 2017 Feb;98:38-50. [doi: 10.1016/j.ijhcs.2016.09.015]
61. Boyd K, Bond R, Nugent C, Donnelly M. EasiSocial: recommendations in the development and training of social media tools for older people. In: Proceedings of the 32nd International BCS Human Computer Interaction Conference (HCI). 2018 Presented at: Human Computer Interaction Conference; July 4 - 6, 2018; Belfast, Northern Ireland. [doi: 10.14236/ewic/hci2018.112]
62. Romanyk C, Salehi P, Sant J, Tsotsos L, Chavez R. Family channel: accessible social media for older adults. In: Proceedings of the International Conference on Human-Computer Interaction. 2015 Presented at: International Conference on Human-Computer Interaction; August 2-7, 2015; Los Angeles, CA, USA p. 321-326. [doi: 10.1007/978-3-319-21380-4_55]
63. Zaine I, Frohlich DM, Rodrigues KR, Cunha BC, Orlando AF, Scalco LF, et al. Promoting social connection and deepening relations among older adults: design and qualitative evaluation of media parcels. J Med Internet Res 2019 Oct 03;21(10):e14112 [FREE Full text] [doi: 10.2196/14112] [Medline: 31584001]
64. Báez M, Ibarra F, Far I, Ferron M, Casati F. Online group-exercises for older adults of different physical abilities. In: Proceedings of the International Conference on Collaboration Technologies and Systems (CTS).: SAGE Publications Ltd; 2016 Presented at: International Conference on Collaboration Technologies and Systems (CTS); Oct. 31- Nov. 4, 2016; Orlando, FL, USA p. 524-533. [doi: 10.1109/cts.2016.0098]
65. Coelho J, Duarte C. Socially networked or isolated? Differentiating older adults and the role of tablets and television. In: Proceedings of the IFIP Conference on Human-Computer Interaction. 2015 Presented at: IFIP Conference on Human-Computer Interaction; September 14-18, 2015; Bamberg, Germany p. 129-146. [doi: 10.1007/978-3-319-22701-6_10]
66. Andreadis A, Zambon R, Parlangeli O. TV as an experience conveyer for better acceptance of ICT services by older adults. Univ Access Inf Soc 2020 Jul 03;20(2):359-374. [doi: 10.1007/s10209-020-00731-w]
67. Vaportzis E, Clausen MG, Gow AJ. Older adults perceptions of technology and barriers to interacting with tablet computers: a focus group study. Front Psychol 2017 Oct 04;8:1687 [FREE Full text] [doi: 10.3389/fpsyg.2017.01687] [Medline: 29071004]

Abbreviations

COTS: commercial off-the-shelf
ICT: information and communication technology
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by J Torous; submitted 12.10.21; peer-reviewed by P Schulz, M Donnelly; comments to author 19.11.21; revised version received 14.01.22; accepted 16.01.22; published 07.03.22.

Please cite as:
Thangavel G, Memedi M, Hedström K
Customized Information and Communication Technology for Reducing Social Isolation and Loneliness Among Older Adults: Scoping Review
JMIR Ment Health 2022;9(3):e34221
URL: https://mental.jmir.org/2022/3/e34221
doi:10.2196/34221
PMID:35254273

Review

Telehealth Versus Face-to-face Psychotherapy for Less Common Mental Health Conditions: Systematic Review and Meta-analysis of Randomized Controlled Trials

Hannah Greenwood[1], BPsycSci (Hons); Natalia Krzyzaniak[1,2], BPharmSci, MPharm, PhD; Ruwani Peiris[1], BBiomedSci (Hons), MPH, MD; Justin Clark[1], BA; Anna Mae Scott[1], BA, MA, MSc, PhD; Magnolia Cardona[1,3], MBBS, MPH, PhD; Rebecca Griffith[4], BSocSci, PostGrad Dip (Psych), MPsych (Clin); Paul Glasziou[1], MBBS, PhD

[1]Institute for Evidence-Based Healthcare, Bond University, Robina, Australia

[2]School of Pharmacy, University of Queensland, Brisbane, Australia

[3]Gold Coast University Hospital Evidence-Based Practice Professorial Unit, Southport, Australia

[4]Private Practice Clinical Psychology, Gold Coast, Australia

Corresponding Author:
Hannah Greenwood, BPsycSci (Hons)
Institute for Evidence-Based Healthcare
Bond University
14 University Drive
Robina, 4226
Australia
Phone: 61 755955816 ext 55816
Email: hgreenwo@bond.edu.au

Abstract

Background: Mental disorders are a leading cause of distress and disability worldwide. To meet patient demand, there is a need for increased access to high-quality, evidence-based mental health care. Telehealth has become well established in the treatment of illnesses, including mental health conditions.

Objective: This study aims to conduct a robust evidence synthesis to assess whether there is evidence of differences between telehealth and face-to-face care for the management of less common mental and physical health conditions requiring psychotherapy.

Methods: In this systematic review, we included randomized controlled trials comparing telehealth (telephone, video, or both) versus the face-to-face delivery of psychotherapy for less common mental health conditions and physical health conditions requiring psychotherapy. The psychotherapy delivered had to be comparable between the telehealth and face-to-face groups, and it had to be delivered by general practitioners, primary care nurses, or allied health staff (such as psychologists and counselors). Patient (symptom severity, overall improvement in psychological symptoms, and function), process (working alliance and client satisfaction), and financial (cost) outcomes were included.

Results: A total of 12 randomized controlled trials were included, with 931 patients in aggregate; therapies included cognitive behavioral and family therapies delivered in populations encompassing addiction disorders, eating disorders, childhood mental health problems, and chronic conditions. Telehealth was delivered by video in 7 trials, by telephone in 3 trials, and by both in 1 trial, and the delivery mode was unclear in 1 trial. The risk of bias for the 12 trials was low or unclear for most domains, except for the lack of the blinding of participants, owing to the nature of the comparison. There were no significant differences in symptom severity between telehealth and face-to-face therapy immediately after treatment (standardized mean difference [SMD] 0.05, 95% CI −0.17 to 0.27) or at any other follow-up time point. Similarly, there were no significant differences immediately after treatment between telehealth and face-to-face care delivery on any of the other outcomes meta-analyzed, including overall improvement (SMD 0.00, 95% CI −0.40 to 0.39), function (SMD 0.13, 95% CI −0.16 to 0.42), working alliance client (SMD 0.11, 95% CI −0.34 to 0.57), working alliance therapist (SMD −0.16, 95% CI −0.91 to 0.59), and client satisfaction (SMD 0.12, 95% CI −0.30 to 0.53), or at any other time point (3, 6, and 12 months).

Conclusions: With regard to effectively treating less common mental health conditions and physical conditions requiring psychological support, there is insufficient evidence of a difference between psychotherapy delivered via telehealth and the same therapy delivered face-to-face. However, there was no includable evidence in this review for some serious mental health conditions,

such as schizophrenia and bipolar disorders, and further high-quality research is needed to determine whether telehealth is a viable, equivalent treatment option for these conditions.

(JMIR Ment Health 2022;9(3):e31780) doi:10.2196/31780

KEYWORDS

telemedicine; psychology; mental health; psychotherapy; primary health care; behavioral sciences; systematic review

Introduction

Background

Worldwide, mental health disorders are a leading cause of distress and disability, with 1 in every 4 people expected to be personally impacted throughout their lifetime [1]. Some evidence suggests that mental health difficulties may be increasing; a previous systematic review found a small but significant increase in mental illness prevalence rates from 1978 to 2015, although the authors note that this may have been driven by demographic changes across this period [2]. In addition, the emergence of COVID-19 has seen mental health adversely impacted worldwide [3,4]. This seems to indicate that this already debilitating problem may become a further global burden in the future. Thus, it seems crucial for quality mental health support to be widely available to the public in a safe and accessible way.

Although telehealth was available and suggested to be effective for psychotherapy before the COVID-19 pandemic [5], its uptake was somewhat limited within the delivery of psychological services [6]. A study in the United States found that before the COVID-19 pandemic, psychologists were hesitant to use telehealth owing to lack of training, concerns for client safety, and privacy, among other concerns [7]. In addition, a qualitative study of mental health professionals highlighted concerns around the quality of the patient-therapist relationship [8]. Given the health risks posed by face-to-face meetings, especially for older people or otherwise vulnerable, there was a rapid shift to remote delivery in health care services worldwide [9-12].

Although the pandemic was the catalyst that thrust telehealth to the forefront of health care delivery, there are many advantages to telehealth service provision for mental health. Telehealth extends care to patients with limited access to in-person therapy, including those in rural and remote areas. A narrative review examining telehealth access in rural communities in the United States found telehealth to be a convenient and efficient way to treat patients, and participants reported acceptability and satisfaction with telehealth services [13]. Furthermore, telehealth also offers a safe and effective option for those who may have access issues or face stigmatization [14]. For some conditions, such as substance use disorder, access to therapy delivered remotely may increase engagement with treatment services among groups who would not otherwise attend therapy [15]. For patients being treated for substance abuse, video-delivered treatment was preferred to face-to-face treatment, mostly because of convenience and increased confidentiality [16]. Taken together, the availability of telehealth facilitates increased access of care to those unable or unwilling to engage in face-to-face therapy and promotes continued therapeutic engagement owing to its flexibility and privacy.

Telehealth may also enhance care accessibility for those requiring specialized therapies or those with less common mental health conditions that may not be treated by all clinicians. The skills needed to effectively treat those with less common or more complex mental health conditions or to adequately deliver less common therapy types may require additional training, guided supervision, professional development, or years of clinical experience. This is further compounded in rural and remote areas, where health care disparity is well documented [17-19]. Telehealth presents a potentially effective medium to connect patients requiring specialized forms of care with relevant, qualified therapists.

Objectives

Evidence supports the use of telehealth for application in some psychotherapies [5,20,21] and the management of common mental health conditions, including reviews in this series for depression (Scott AM et al, PhD, unpublished data, February 2022), anxiety [22], and posttraumatic stress disorder (PTSD) [23]. It is important to rigorously assess whether its effectiveness is generalizable beyond these groups. The aim of this systematic review is to assess whether there are any differences between telehealth-based psychotherapy and face-to-face psychotherapy across outcomes (patient, process, and cost) for less common mental health conditions (eg, substance use disorder, eating disorders, or childhood disorders) and physical conditions requiring psychological support (eg, cancer or chronic fatigue syndrome).

Methods

Overview

We aim to find, appraise, and synthesize studies that compared psychotherapy delivered via telehealth (video, telephone, or both) versus face-to-face for patients of any age in the primary health care setting. This systematic review is reported following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2009 statement [24], and the review protocol was developed prospectively.

Inclusion and Exclusion Criteria

Study Design

We included randomized controlled trials (RCTs) of any design (eg, parallel, cluster, crossover, factorial, or mixed), which included >10 patients. We excluded all other study designs, such as controlled nonrandomized trials, qualitative studies, and observational studies (cohort, case-control, cross-sectional, case series, and case reports).

Participants

We included studies with people of any age or gender, who were receiving psychotherapy for less common mental health conditions, such as bulimia nervosa and substance use disorder, or any conditions where psychotherapy was used, such as cognitive behavioral therapy (CBT) for patients with cancer with high psychological needs. Although anxiety [22], depression (Scott AM et al, PhD, unpublished data, February 2022), and PTSD [23] were within the scope of this review, there was enough literature to conduct separate systematic reviews by condition, and hence, these were excluded. Studies involving hospital patients (eg, explicitly identified as taking place in hospital wards, or with patients shortly after discharge) or those consulting a secondary or tertiary specialist (ie, a psychiatrist) were excluded. Studies in hospital-discharged patient populations that explicitly identified the provision of therapy by a psychologist, therapist, psychotherapist, or counselor, however, were included.

Interventions

We included studies of interventions involving standard care psychological therapies for mental health conditions or physical conditions where psychological therapy was required, including but not limited to CBT, parent-child interaction therapy, cognitive behavioral intervention for tics, and parent training. Studies examining novel treatments for mental health were excluded.

Comparators

We included studies with an equivalent face-to-face comparator or other telehealth comparators (ie, video intervention with a telephone comparator). The intervention and comparator had to deliver a similar or identical level of care (ie, care similar in intensity, frequency, and duration). Studies with a comparator that included a wait-list control or clinically inequivalent active comparator were excluded.

Outcomes (Primary and Secondary)

The primary patient outcome was global or symptom severity. The secondary patient outcomes included improvement in psychological symptoms and functioning. The tertiary process (working alliance and satisfaction) or financial (cost) outcomes are included but reported in Multimedia Appendix 1. Studies that met other inclusion criteria but did not report on one of the primary or secondary outcomes were included and reviewed. This is important to distinguish, as we either meta-analyzed outcomes or summarized them narratively if meta-analysis was not possible.

Search Strategy to Identify Studies

Database Search for Primary Studies

The following databases were searched from inception until November 18, 2020: PubMed (MEDLINE), Embase, and CENTRAL via the Cochrane Library. The original search string (Multimedia Appendix 2) was designed in PubMed and translated for use in other databases using the Institute for Evidence-Based Healthcare's Polyglot Search Translator, an automation tool designed to translate search strings between databases [25]. This included a number of concepts and variants, such as *Telemedicine* AND *Primary healthcare* AND *face-to-face* AND *randomised*. On January 11, 2021, we conducted a backward (cited) and forward (citing) citation analysis using the web-based citation database Scopus [26] on included studies identified during previous searches. These were screened against the inclusion criteria.

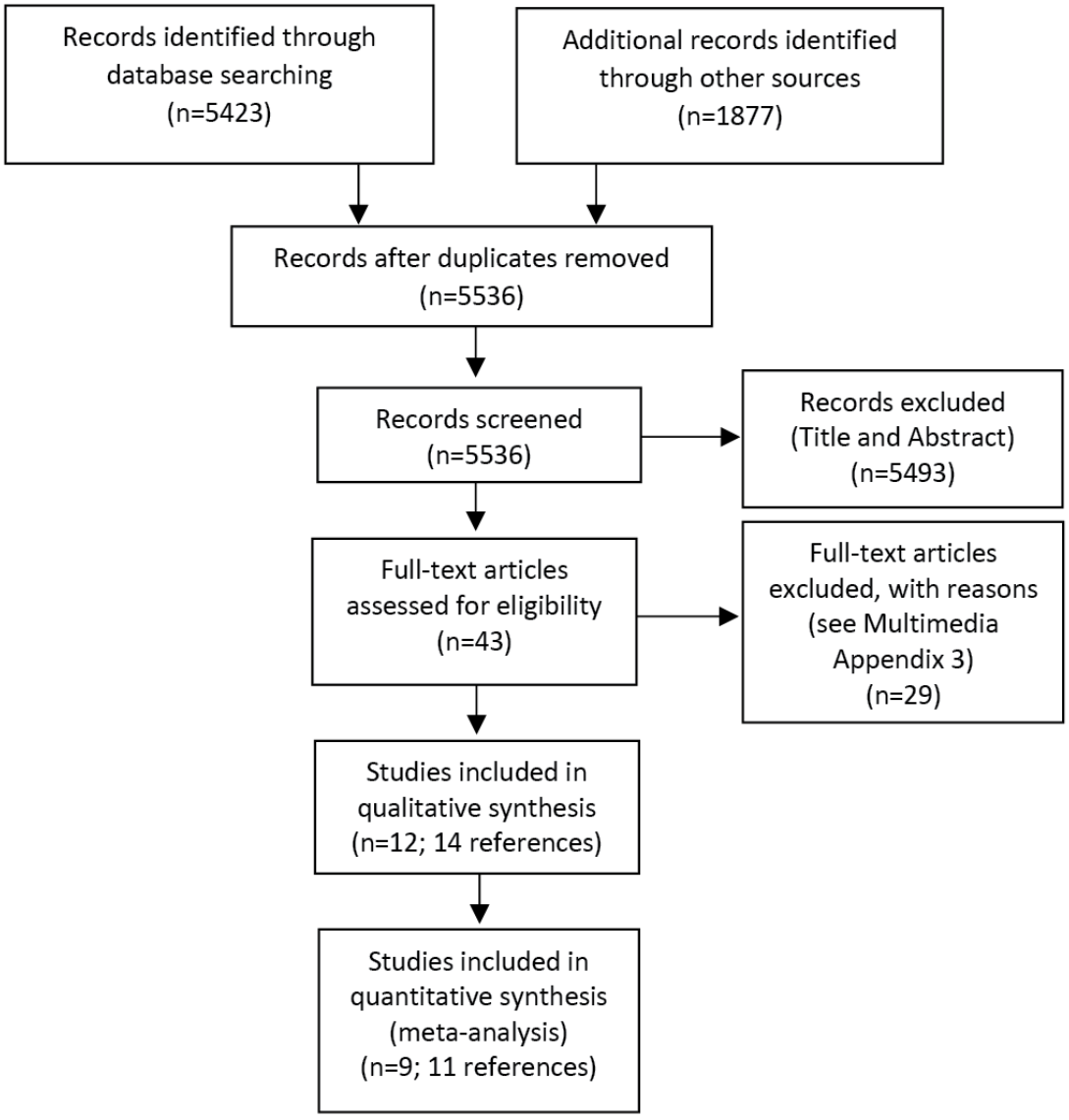
Restriction on Publication Type

No restrictions by language or publication date were imposed. We included only those publications from RCTs that were published in full. We excluded publications available as abstract only (eg, conference abstract) or with no additional results information available (eg, from a clinical trial registry record).

Study Selection and Screening

Titles and abstracts were screened independently by author pairs (AMS, RP, MC, JC, NK, HG, and PG) against the inclusion criteria. In addition, 1 author (JC) retrieved full texts, and 2 authors (HG and NK) screened the full texts for inclusion. Any disagreements were resolved by discussion or reference to the third screener. The forward backward citation analysis was conducted by 1 author (JC) and screened by 3 authors (HG, NK, and RP), and full text was obtained by HG. The study selection process for includable studies is reported in the PRISMA flow diagram (Figure 1), and studies excluded at the full-text screening stage are in Multimedia Appendix 3 with reasons for exclusion.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.



Data Extraction

We used 2 data extraction forms for study characteristics and outcome data, which were piloted on 2 studies in the review. Data from the included studies were extracted independently by 2 authors (NK and HG) into the data extraction forms (Textbox 1), and discrepancies were resolved by discussion or by reference to a third author.

Textbox 1. List of extracted information.

Extracted information

- Methods: Study authors, country, design, and duration of follow-up—as reported
- Participants: n, condition needing psychotherapy, randomization, age (years), mean (SD)
- Interventions: telehealth—provider, therapy, and dose
- Comparators: face-to-face—provider, therapy, and dose
- Outcomes: n, mean (SD), and *P* value (or as reported by authors)—patient (global or symptom severity, improvement in psychological symptoms, and functioning), process (working alliance and satisfaction), and financial (cost)

Risk of Bias in Included Studies

A total of 2 review authors (HG and NK) independently assessed the risk of bias for the included studies using the Cochrane Collaboration Risk of Bias Tool 1, as outlined in the Cochrane Handbook [27], and all disagreements were resolved by discussion.

The following domains were assessed for possible bias: (1) random sequence generation, (2) allocation concealment, (3)

the blinding of participants and personnel, (4) the blinding of the outcome assessment, (5) incomplete outcome data, (6) selective outcome reporting, and (7) other bias (focusing on potential biases due to funding or conflict of interest).

Each domain was graded as low, high, or unclear, including quote or summary from the relevant trial, which summarized why the grading was applied.

Data Synthesis and Analysis

Review Manager 5.4, the Cochrane Collaboration tool for conducting meta-analyses and creating forest plots, was used to calculate the treatment effect [28]. As all outcome measures were continuous, we used mean difference or standardized mean difference (SMD). We performed meta-analyses only when possible (when ≥2 studies or comparisons reported the same or similar outcome) and where appropriate data were available that allowed us to calculate the SMD. Where these data were not available and thus meta-analysis was not possible, we narratively report the results. We anticipated a considerable heterogeneity between studies and used a random-effects model.

The unit of analysis was the individual, which was available for every study in this review. We did not contact study authors to provide missing data. We used the I^2 statistic to examine the heterogeneity of the included studies. Subgroup analyses were conducted according to the duration of follow-up: posttreatment and 3, 6, and 12 months.

As <10 trials were included in any data synthesis, we did not create a funnel plot, and sensitivity analyses were not conducted. We planned to conduct a subgroup analysis of gender, setting, age, and sensitivity by including or excluding studies at high risk of bias; however, the low number of included studies did not allow for this.

Results

Search Results

The primary study search found 5423 references, and 1877 additional references were found in the forward and backward citation search and clinical trial registries. After deduplication, 5536 records were screened in title and abstract. A total of 5493 references were excluded, and 43 full texts were assessed for inclusion. Moreover, 12 RCTs (across 14 articles) were included in this systematic review, and 9 were able to be meta-analyzed (Figure 1). We found 2 potentially relevant but still in-progress clinical trials (Multimedia Appendix 4 [29,30]).

Characteristics of Included Studies

Of the 12 included RCTs, 10 (83%) were conducted in the United States, and the other 2 (17%) studies were conducted in the United Kingdom. A total of 931 patients were included in aggregate. Studies have examined psychotherapy delivered for a variety of less common mental health conditions and other conditions requiring psychotherapy. Of the 12 studies, 2 (17%) included patients with type 1 diabetes mellitus, 2 (17%) included patients with addiction disorders, 1 (8%) (reported in 4 articles) treated patients with bulimia nervosa or eating disorder not otherwise specified, 3 (25%) studies included participants with children's disorders (including disruptive behavior disorder, tic disorders, and attention-deficit hyperactivity disorder), 2 (17%) included patients with chronic illness (chronic fatigue syndrome and chronic multisymptom illness), 1 (8%) study included patients with a range of mental health conditions, and 1 (8%) included patients with cancer who had high psychological needs. The types of therapies varied by target condition: of the 12 studies, 5 (42%) used CBT, 4 (33%) used a family therapy (parent-child interaction therapy, parent training, and behavioral family systems therapy for diabetes), 2 (17%) used addiction therapies (opioid treatment program and acute therapy service), and 1 (8%) used a cognitive behavioral intervention for tics. Finally, of the 12 studies, 3 (25%) used the telephone to deliver telehealth, 7 (58%) used video, and 1 (8%) had included video and telephone groups, and in 1 (8%) study, it was unclear whether video or telephone was used. All studies compared the telehealth intervention to face-to-face intervention (Table 1).

Table 1. Characteristics of included studies.

Reference	Country	RCT[a] design	Follow-up (months)	Study participants, total N (n TH[b], n F2F[c])	Participants	Age (years), mean (SD)	Intervention	Telehealth: modality dose	Comparator: modality dose
Burgess et al [31]	United Kingdom	Parallel, 2 arm	12	80 (45, 35)	Adults (aged 18-65 years) with chronic fatigue syndrome (comorbidities excluded)	37.4 (10.1).	CBT[d]	Telephone, 3-hour 1 × F2F; 30 minutes, 13 sessions, fortnightly	F2F, 3-hour 1 x F2F; 50-60 minutes, 13 sessions
Comer et al [32]	United States	Parallel, 2 arm	6	40 (20, 20)	Children (aged 3-5 years) with principal diagnosis disruptive behavior disorder (serious comorbidities excluded) and their parents or caregivers	4.0 (0.9)	Parent-child interaction therapy	Video, until mastery was achieved, mean sessions 21.7	F2F until mastery was achieved, mean sessions 20.8
Day and Schneider [33]	United States	Parallel, 3 arm	None	91 (completers only reported—26 video, 27 telephone, and 27 F2F)	Adults (aged 19-75 years) presenting with any mental health issue to a community counseling center	39.3 (15.9)	CBT	Video and 2-way audio (telephone analogous), 5 sessions	F2F, 5 sessions
Duke et al [34]	United States	Parallel, 2 arm	3	90 (46, 44)	Adolescents (aged 12-19 years) with type 1 diabetes (uncontrolled comorbidities excluded) and their caregivers	15.0 (1.75)	Behavioral family systems therapy for diabetes	Video, 60-90 minutes, up to 10× sessions, 12 weeks	F2F, 60-90 minutes, up to 10× sessions, 12 weeks
Freeman et al [35]	United States	Parallel, 2 arm	None	92 (47, 45)	Adolescents (aged 12-19 years) with poorly controlled type 1 diabetes (no comorbidity exclusion) and 1 parent or legal guardian	TH 14.9 (1.9); F2F 15.2 (1.8)	Behavioral family systems therapy for diabetes	Video, 60-90 minutes, up to 10× sessions, 12 weeks	F2F, 60-90 minutes, up to 10× sessions, 12 weeks
Himle et al [36]	United States	Parallel, 2 arm	4	20 (10, 10)	Children (aged 8-17 years) who met DSM[e] criteria for Tourette or chronic tic disorder with or without comorbidities	TH 11.3 (2.3); F2F 12 (3.3)	Cognitive behavioral intervention for tics	Video, 6× weekly sessions+2× biweekly sessions, 10 weeks	F2F, 6× weekly session+2× biweekly sessions, 10 weeks
King et al [37]	United States	Parallel, 2 arm	3	85 (50, 35)	Adult outpatients receiving opioid dependence treatment (no comorbidity exclusion)	TH 40.5 (11.2); F2F 41.1 (10.5)	Opioid treatment program	Video, 30-40 minutes, 12× weekly sessions, 12 weeks	F2F, 30-40 minutes, 12× weekly session, 12 weeks
King et al [16]	United States	Parallel, 2 arm	None	37 (20, 17)	Adult outpatients with a partial response to methadone maintenance treatment (no comorbidity exclusion)	TH 42.7; F2F 41.4	Acute therapy service	Video, 1 hour, 2× sessions, 6 weeks	F2F, 1 hour, 2× sessions, 6 weeks
McAndrew et al [38]	United States	Parallel, 3 arm	12	128 (42, 43; 43 UC[f])	Adult veterans with chronic multisymptom illness (serious psychiatric and medical comorbidities excluded)	TH 57.6 (6.6); F2F 55.4 (8.2)	CBT	Telephone, up to 10 sessions	F2F, up to 10 sessions
Crow et al [39], Ertelt et al [40], Mitchell et al [41]	United States	Parallel, 2 arm	12	128 (62, 66)	Adults (aged >18 years) with bulimia nervosa (including comorbidities but excluding suicidal ideation, psychosis, schizophrenia and bipolar)	TH 28.4 (10.4); F2F 29.6 (10.9)	CBT	Unclear, 20 sessions, 16 weeks	F2F, 20 sessions, 16 weeks

Reference	Country	RCT[a] design	Follow-up (months)	Study participants, total N (n TH[b], n F2F[c])	Participants	Age (years), mean (SD)	Intervention	Telehealth: modality dose	Comparator: modality dose
Watson et al [42]	United Kingdom	Parallel, 2 arm	None	118 (60, 58)	Adults (aged 18-79 years) with a cancer diagnosis and comorbid high psychological needs	TH 48.5 (13.3); F2F 52.4 (13.1)	CBT	Telephone, 8 sessions, 12 weeks	F2F, 8 sessions, 12 weeks
Xie et al [43]	United States	Parallel, 2 arm	None	22 (9, 13)	Children (aged 6-14) with primary diagnosis ADHD[g] (excluding unstable medical conditions and other serious psychiatric disorders) and their parents	10.4 (NR[h])	Parent training	Video, 10 weekly session, 10 weeks	F2F, 10 weekly sessions, 10 weeks

[a]RCT: randomized controlled trial.

[b]TH: telehealth.

[c]F2F: face-to-face.

[d]CBT: cognitive behavioral therapy.

[e]DSM: Diagnostic and Statistical Manual of Mental Disorders.

[f]UC: usual care.

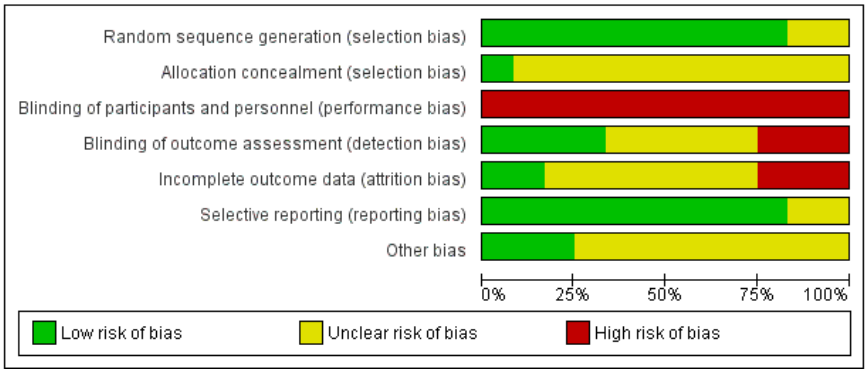
[g]ADHD: attention-deficit/hyperactivity disorder.

[h]NR: not reported.

Risk of Bias

Overall, of the 12 studies, 10 (83%) adequately reported on random sequence generation and selective reporting. Declarations of conflicts of interest and funding (reported under other bias) were adequately reported for only 25% (3/12) of the studies, with the remaining 75% (9/12) not reporting this clearly. Allocation concealment was not clearly reported in most studies, with only 8% (1/12) of the studies reporting this satisfactorily. The blinding of the outcome assessment and incomplete outcome data were at high risk of bias for 25% (3/12) of the studies, with the remaining 75% (9/12) of the studies rated at either unclear or low risk of bias. Notably, the blinding of participants and personnel was a high bias risk for all 12/12 (100%) studies, as the telehealth versus face-to-face nature of the interventions was incompatible with blinding (Figure 2).

Figure 2. Risk of bias graph: review authors' judgments about each risk of bias item presented as percentages across all included studies.



Primary Outcome: Global or Symptom Severity

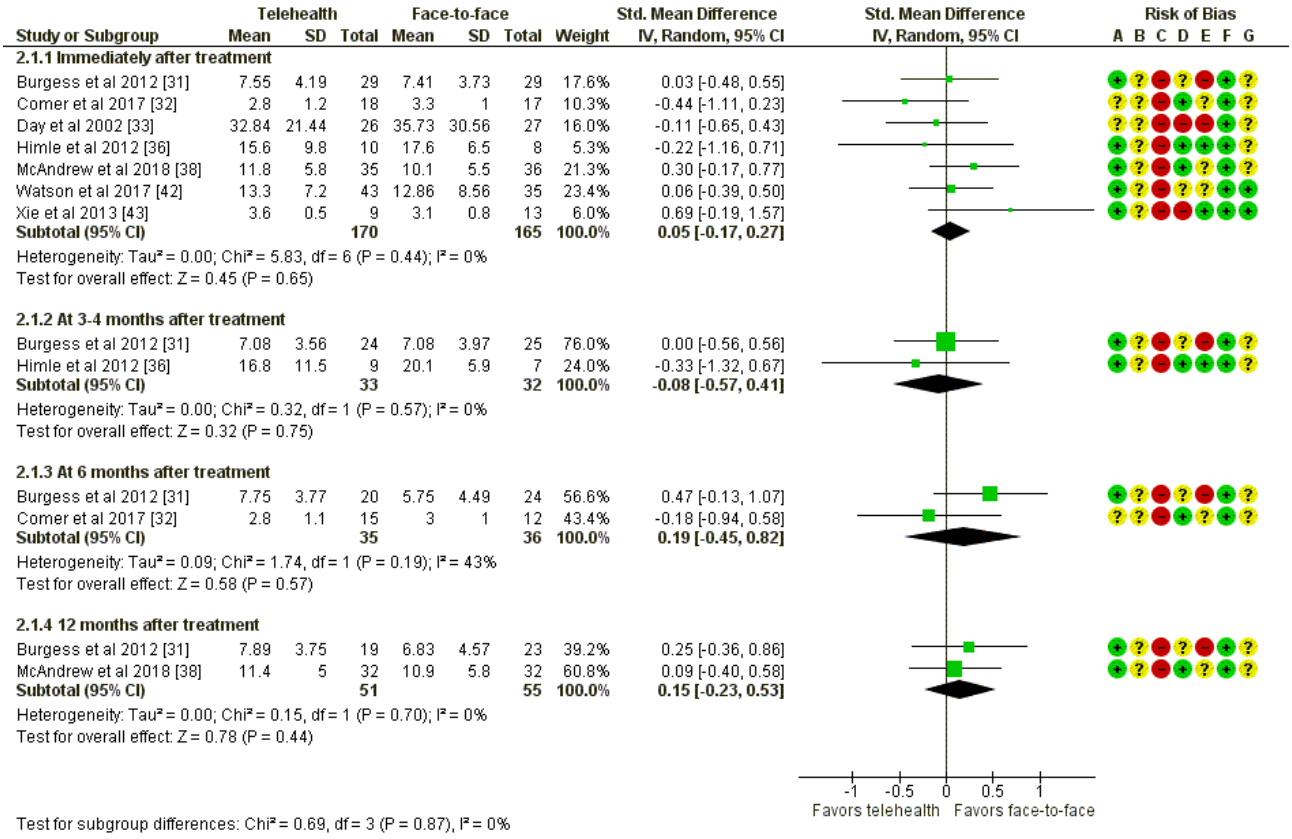
A total of 6 scales across 7 studies were used to report outcomes related to symptom severity (see Multimedia Appendix 5 for a summary of scales used).

In addition, 7 studies reported sufficient data for this outcome and were able to be pooled and meta-analyzed (Figure 3). Data were available for four time point subgroups—immediately after treatment and 3- to 4-, 6-, and 12-month follow-ups.

There were no significant differences in severity outcomes between telehealth and face-to-face therapy immediately after treatment (335 participants; mean difference 0.05, 95% CI −0.17 to 0.27; P=.65) or at any of the follow-up time points, including 3 to 4 months (65 participants; SMD −0.08, 95% CI −0.57 to 0.41; P=.75), 6 months (71 participants; SMD 0.19, 95% CI −0.45 to 0.82; P=.57), and 12 months (106 participants; SMD 0.15, 95% CI −0.23 to 0.53; P=.44).

There was moderately high heterogeneity reported for the 6-month follow-up subgroup (I^2=43%).

Figure 3. Telehealth versus face-to-face for mental conditions: assessment of symptom severity. Std: standard. [31-33, 36, 38, 42, 43].



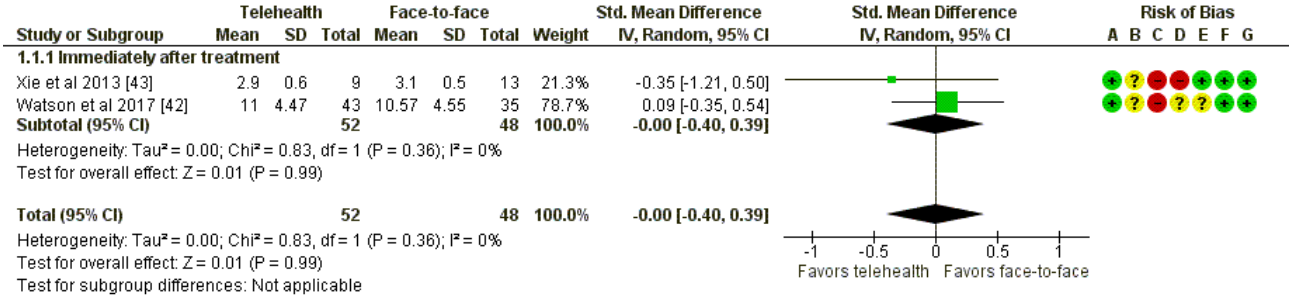
Secondary Outcomes

Improvement

A total of 3 different scales were used to describe patients' overall improvement across the studies (see Multimedia Appendix 5 for a summary of scales used).

In addition, 2 studies were able to be meta-analyzed; the remaining 3 studies are reported narratively. These 2 meta-analyzed studies involved a total of 100 participants (Figure 4). Data were available at one time point; that is, immediately after treatment. There was no evidence of difference between the 2 groups in this comparison, with an SMD of −0 (95% CI −0.4 to 0.39; *P*=.99).

Figure 4. Telehealth versus face-to-face for mental conditions: assessment of improvement of psychological symptoms. Std: standard. [42, 43].



Burgess et al [31] reported global improvement on a self-rated 6-item scale, ranging from very much better to very much worse. Among 29 telehealth participants immediately after treatment, 14 (48%) rated their improvement as very much or much better, whereas 15 (52%) rated their improvement as a little better to very much worse. For 28 face-to-face participants immediately after treatment, 15 (54%) rated themselves as improved, whereas 13 (46%) rated themselves as only a little better or worse. Although this is variable at the 6- and 12-month follow-up time points, there were no differences between groups at any time point (at 6 months, 8/20, 40% telehealth participants and 15/25, 60% face-to-face participants rated themselves as very much or much better, and at 12 months, 11/20, 55% telehealth and 13/23, 57% face-to-face participants rated themselves as very much or much better).

Comer et al [32] and Himle et al [36] both reported using the Clinical Global Impression-Improvement scale, reporting the percentage of participants who received a score of 1 or 2 (very much improved or much improved). Among participants in the study by Comer et al [32], of the 14 participants in the telehealth group, 12 (86%) had improvement, whereas of the 14 participants in the face-to-face group, 11 (79%) improved. Furthermore, at 6 months after treatment, 83% (10/12) of the telehealth participants and 73% (8/11) of the face-to-face participants still reported very much or much improvement. It

is unclear whether the differences between groups were significant, and the outcomes were reported only for treatment completers, not all participants. The findings from Himle et al [36] are similar at immediately after treatment: 80% (8/10) of the telehealth participants were very much or much improved, whereas 75% (6/8) of the face-to-face participants were improved. However, at follow-up, 56% (5/9) of the telehealth participants and 44% (3/7) of the face-to-face participants were very much or much improved.

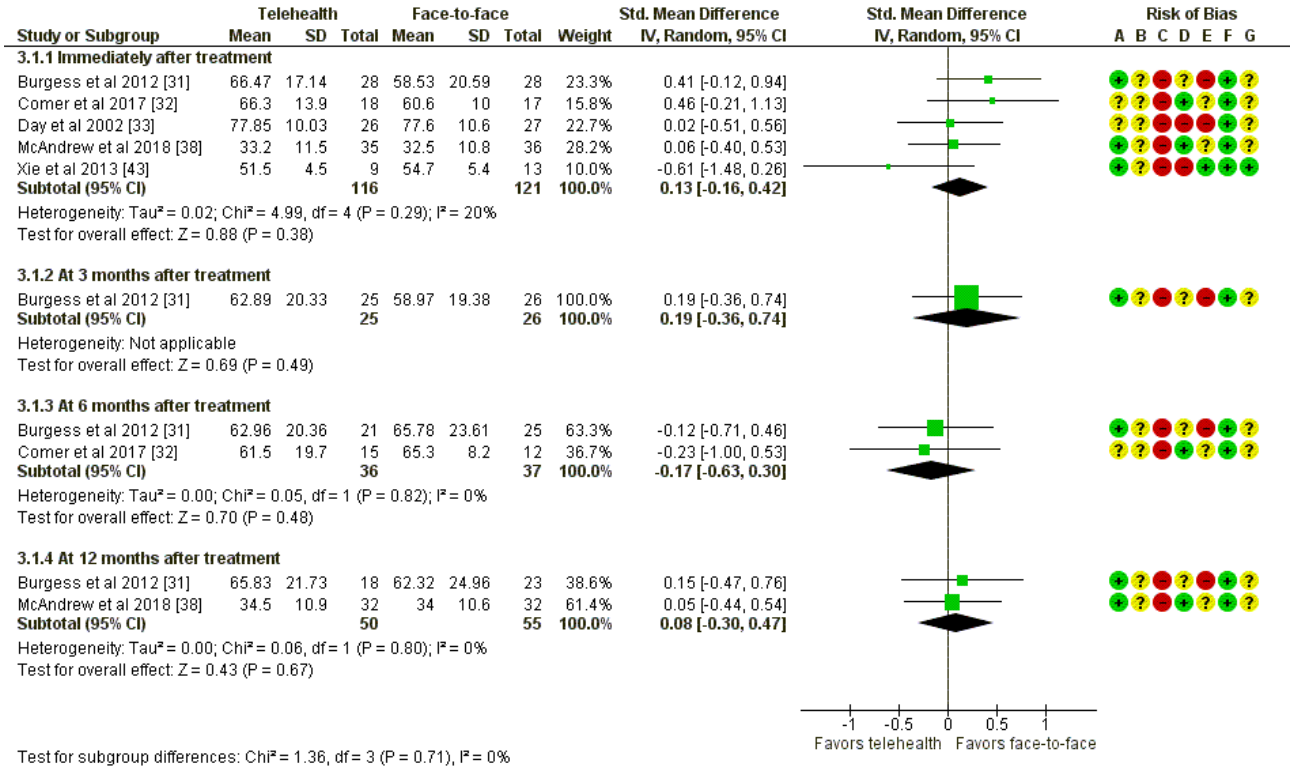
Function

The outcome was assessed using 4 different scales (see Multimedia Appendix 5 for a summary of scales used).

In addition, 6 studies reported sufficient data for this outcome; 5 were able to be meta-analyzed (Figure 5). There were no significant differences in functioning outcomes between telehealth and face-to-face therapy immediately after treatment (237 participants; mean difference 0.13 (95% CI −0.16 to 0.42; *P*=.38) or at any of the follow-up time points, including 3 months (51 participants; SMD 0.19, 95% CI −0.36 to 0.74; *P*=.49), 6 months (73 participants; SMD −0.17, 95% CI −0.63 to 0.3; *P*=.48), and 12 months (105 participants; SMD 0.08, 95% CI −0.3 to 0.47; *P*=.67).

Mitchell et al [41] also reported a function measure using the 36-Item Short-Form Health Survey, reporting on both the mental and physical subscales. At immediately after treatment, there was no difference on the physical subscale between telehealth (41 participants; 54.1, SD 7.9) and face-to-face groups (39 participants; 56.2, SD 5.7). For the mental health subscale, there was no difference between groups for telehealth (41 participants; 42.9, SD 12.6) and face-to-face treatment (39 participants; 45.5, SD 11.9). These results were similar for 3 and 6 months after treatment.

Figure 5. Telehealth versus face-to-face for mental conditions: assessment of functioning. Std: standard. [31-33, 38, 43].



Tertiary Outcomes

Process

A total of 5 studies reported client working alliance outcomes (3 meta-analyzed, n=223, immediately after treatment). There was no difference between telehealth and face-to-face therapy; the SMD was 0.11 (95% CI −0.34 to 0.57; *P*=.63). This subgroup had moderate to high levels of heterogeneity (I^2=63%; see Figure S1 in Multimedia Appendix 1). In addition, 2 studies also reported therapist working alliance outcomes (2 meta-analyzed, n=104, immediately after treatment). There was no evidence of difference between telehealth and face-to-face therapy (SMD −0.16, 95% CI −0.91 to 0.59; *P*=.67), and heterogeneity was high (I^2=72%; see Figure S2 in Multimedia Appendix 1).

A total of 7 studies reported client satisfaction outcomes (3 meta-analyzed, n=131, immediately after treatment), and we found no evidence of difference in satisfaction between groups (SMD 0.12, 95% CI −0.3 to 0.53; *P*=.58; see Figure S3 in Multimedia Appendix 1).

More detailed data for process outcomes (working alliance and client satisfaction) are available in Multimedia Appendix 1, including figures and a narrative analysis of included studies that could not be meta-analyzed.

Financial

A total of 3 studies reported cost, but no outcomes were able to be meta-analyzed. Please see Multimedia Appendix 1 for a narrative review of financial outcomes.

Discussion

Principal Findings

This systematic review of 12 trials shows insufficient evidence of a difference between psychotherapy delivered via telehealth (telephone or video) and face-to-face therapy, when treating less common mental health conditions or physical conditions requiring psychological support. There were no significant differences between telehealth and face-to-face delivery for patient outcomes (symptom severity, symptom improvement, or global function), immediately after treatment, or at any follow-up time point. For process outcomes (working alliance or therapeutic quality and client satisfaction), there was no significant difference between telehealth and face-to-face care delivery for either clients or therapists, although 1 study reported no difference between groups for therapist satisfaction. Although financial outcome data on costs were not meta-analyzable, patients with substance abuse disorder valued telehealth therapy more highly than face-to-face therapy, treatment costs were lower for telehealth than for face-to-face therapy for patients with bulimia nervosa (especially over large geographical areas), and the cost of therapists' time was equivalent, regardless of delivery mode, for patients with cancer receiving CBT. This suggests that telehealth is at least as cost-effective as face-to-face care and potentially perceived as more valuable by the client. Overall, the risk of bias of included studies was unclear, owing to unclear reporting, and blinding of participants was not possible because of the nature of the interventions.

Although we found no significant differences between telehealth and face-to-face delivery of psychotherapy across any outcome, to assess equivalence between telehealth and face-to-face psychotherapy, CIs around the effect estimate should be examined to determine whether they exclude the minimally important difference [44]. In the absence of a prespecified minimally important difference, we accept Cohen cutoff for a small effect (0.2), whereby a CI between (−0.20 and 0.20) suggests equivalence between telehealth and face-to-face therapy, and a CI outside these bounds indicates that the minimally important difference cannot be excluded and there is the possibility of a small effect favoring one or the other intervention. For the primary outcome, symptom severity (Figure 3), immediately after treatment, the upper-bound CI is >0.2, so it is possible that the true effect favors face-to-face therapy. For 12 months after treatment, the CI ranges from a possible small effect favoring telehealth to a possible medium effect favoring face-to-face therapy. The same could be applied to all other time points and outcomes to assess the evidence for equivalence. Although we can demonstrate that there is insufficient evidence of a difference between telehealth- and face-to-face–delivered psychotherapy, we cannot conclude whether they are equivalent, given that the CIs around the effect size are rarely narrow enough to exclude the minimally important difference. For common mental health conditions, there is evidence that telehealth is an effective modality in the provision of psychological therapy as face-to-face therapy. There is some evidence of equivalence between videoconferencing and face-to-face care for depression [20], anxiety [21,22], PTSD [23], and psychotherapy broadly [5]. Furthermore, there is evidence of telephone-delivered therapy being effective for depression and anxiety [45]. Although these reviews suggest comparability between telehealth and face-to-face psychotherapy delivery, they all included nonrandomized and noncontrolled studies, which may introduce bias. This review shows no evidence of difference in patient, process, or cost outcomes between telehealth and face-to-face psychotherapy across more diverse patient groups.

Comparison With Prior Work

This review and meta-analysis shows that telehealth psychotherapy may be similar to face-to-face psychotherapy in treating populations with less common mental health disorders or physical conditions that require psychological support. These synthesized findings support previous primary research suggesting that psychotherapy delivered via telehealth for the treatment of mental health conditions may be comparable with conventional face-to-face therapy. A previous review examined the use of video therapy across a range of mental health conditions, including some of the less common conditions reviewed here, and found video-delivered therapy was equivalent to face-to-face care for outcomes of clinical effectiveness, treatment adherence, and patient satisfaction [14]. In line with our narrative findings for financial outcomes, they also found video therapy to be less costly than face-to-face care. This contrasts with a recent scoping review, finding that telehealth service provision across health care in Australia does not routinely reduce the cost of care delivery [46]. Our findings also support the results of a single-arm study conducted in Japan examining video-delivered CBT, which found that this delivery mode is feasible for the treatment of bulimia nervosa and binge-eating disorders. Previous evidence regarding the impact of telehealth on working alliance is mixed. An RCT examining psychologists' perceptions of therapeutic alliance in videoconferencing found that therapeutic alliance was rated significantly lower for telehealth than for face-to-face care [47]. A survey conducted during the COVID-19 pandemic of psychotherapists' experience with remote care found that it was "better than expected" but that telehealth care could not be compared with face-to-face care [48]. In contradiction, results from a more recent survey study found that telehealth was widely accepted by primary mental health care providers [49], although this was not specific to delivery of psychotherapy via telehealth. A recent study examining working alliance via telehealth for anxiety disorders found that these clients had a stronger working alliance with their clinician when treated via telehealth [50]. Our findings, using only data from RCTs, support previous research suggesting that working alliance is as strong in telehealth as it is in face-to-face care. However, further research is needed to fully understand how telehealth changes the client and clinician relationship dynamic and how this may change circumstantially based on clinician and client perceptions of telehealth and the patient's specific treatment needs.

Strengths and Limitations

This review has many strengths, which add weight to our findings and conclusions. We applied rigorous methodology to find includable studies by establishing a prospective protocol

and following PRISMA guidelines. Clear, strict inclusion and exclusion criteria allowed for studies in a variety of different health conditions to be synthesized and systematically reviewed. Further, we only included RCTs, and bias was reviewed for all included studies.

However, there are some limitations to our findings. First, although includable, there were no eligible randomized studies available for telehealth treatment of some less common mental health conditions, such as schizophrenia, bipolar disorders, and personality disorders. This limits the generalizability of these findings across these serious mental health concerns. To assess whether treatment of these conditions is feasible by telehealth, evidence beyond randomized trials should be examined or further high-quality research primary conducted. Second, we only included studies of therapies delivered verbally via telephone or video, as this is most similar to the face-to-face nature of primary care, and we intentionally excluded chat-based or self-guided internet therapy modalities. There is emerging evidence to support the efficacy of chat for mental health treatment services [51,52]. There is also a growing body of work on internet-based therapies for the treatment of psychological conditions such as addictions [53], eating disorders [54], and depression [55] and for the delivery of specific therapies such as CBT [56]. Therefore, although these therapeutic approaches are outside the scope of this review, the role of chat-based or self-guided internet therapy cannot be discounted for remote management of mental health difficulties. Third, most included trials were conducted in the United States, with 2 from the United Kingdom. These health care systems may not be comparable in other countries or regions [57], which limits the generalizability of our findings across medical systems internationally. Fourth, the risk of bias in included studies was largely unclear. We were unable to conduct prespecified subgroup analysis excluding studies at high risk of bias, owing to the small number of studies eligible for inclusion. The possibility of risk of bias in included studies should be considered when interpreting these results. Fifth, we included both telephone and video modalities as telehealth and did not conduct a sensitivity analysis to test any differences between these modalities owing to the small number of included studies. It is possible that there may be differences between telephone and video telehealth care, and future studies may explore this. Sixth, although we anticipated heterogeneity and used a random-effects model, some measures of heterogeneity are high. In each of these cases, the maximum number of studies available at the time point was 3, and it is thought that even when appropriate, meta-analysis with a small number of included studies can lead to fluctuations in the I^2 statistic and should be interpreted with caution [58]. The small number of included studies precluded explorations into heterogeneity, so it is unclear whether heterogeneity observed is solely due to variation between included studies or whether instability of the I^2 statistic due to the small number of included studies inflated the estimate. Regardless, the presence of heterogeneity highlights differences between included studies and reinforces the need for large, high-quality studies exploring psychotherapy delivered via telehealth versus face-to-face care for less common mental illnesses. Seventh, although the outcomes selected are appropriate for the study question and aims, they are all clinician or patient self-report measures, which are subject to measurement and other biases. Finally, the maximum follow-up time for included studies was 12 months, and there was variability in the follow-up periods among studies. The management of mental illness can be chronic or lifelong, so our results do not speak to the effectiveness of telehealth for longer-term management of these conditions.

Clinical and Research Implications

There are some important clinical implications of this research. To date, there has been some reported hesitancy from clinicians to use telehealth in their practice [6]. This appears driven by care providers rather than care receivers; patients report equal satisfaction and experience of therapeutic alliance when receiving individual care via telehealth versus face-to-face [59]. Therapist hesitancy may be due to lack of training, concerns about the quality of the therapeutic alliance including rapport building, ethical concerns around risk management, and technological limitations [7,60-62]. Given the increasing body of evidence demonstrating the similarity of mental health care delivered via telehealth compared with face-to-face, it is critical that therapist barriers toward telehealth modalities be addressed. This may take various potential forms, including the provision of training for the delivery of specific therapies via telehealth, which could be incorporated into professional development or tertiary training. Furthermore, regulatory bodies (eg, Australian Health Practitioner Regulation Agency in Australia) could also provide support and advice for the implementation of telehealth infrastructure such as billing processes or technical logistics.

In addition to these clinical implications, there are several possible directions for further research. Given the diverse range of patient populations, therapies, and psychological conditions that may be treated using telehealth, and the multiple modes of care delivery (ie, telephone, video, or blended), it may be beneficial to investigate how to optimize telehealth therapy for various patient groups. Future considerations could include understanding whether certain conditions are better suited to video or telephone delivery and whether telehealth is as effective when treating complex or comorbid mental illnesses and identifying whether there are any groups for which telehealth is not recommended. Developing specific and structured protocols or guidelines for the delivery of psychotherapy via telehealth to diverse patient groups will help ensure the consistent provision of best-practice telehealth care.

Conclusions

The COVID-19 pandemic has pushed telehealth to the forefront of mental health care out of necessity. This review shows that there is insufficient evidence of difference between psychotherapy delivered via telehealth and psychotherapy delivered via face-to-face care for the management of less common mental and physical health conditions requiring psychological support. There was insufficient evidence of difference between groups across patient, process, and cost outcomes, including symptom severity, improvement, function, therapeutic working alliance, satisfaction, and cost. However, CIs often included the minimally important difference, so we cannot conclude whether psychotherapy delivered via telehealth

versus via face-to-face are equivalent. Further research is needed to assess the efficacy of telehealth for some conditions for which this review found no evidence (such as schizophrenia and bipolar disorders) and to optimize the delivery of telehealth interventions across diverse patient groups. The current evidence indicates that psychotherapy delivered via telehealth may be an alternative to face-to-face psychotherapy for the treatment of less common mental health conditions and physical conditions requiring psychological care.

Acknowledgments

We wish to acknowledge and thank Dr Mark Jones for his statistical guidance and assistance in the interpretation of the results. This systematic review was commissioned by the Department of Health, Canberra, Australia, as part of a series of systematic reviews on the effectiveness of telehealth. The funder was involved in the refining of the study question in Population, Intervention, Comparator, Outcome (PICO) format. The funder was not involved in the conduct, analysis, or interpretation of the systematic review or in the decision to submit the manuscript for publication.

Authors' Contributions

The conception or design of work was the responsibility of AMS and PG. Data collection was conducted by HG, NK, RP, JC, AMS, and MC. Data analysis and interpretation were conducted by HG, NK, AMS, and PG. The manuscript was drafted by HG and NK. The critical revision of the manuscript was conducted by HG, NK, AMS, RG, and PG. The final approval of the version of the manuscript to be published was provided by HG, NK, RP, JC, AMS, MC, RG, and PG.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Process and financial outcomes.
[DOCX File , 79 KB - mental_v9i3e31780_app1.docx]

Multimedia Appendix 2
Search strings.
[DOCX File , 17 KB - mental_v9i3e31780_app2.docx]

Multimedia Appendix 3
Table of excluded studies.
[DOCX File , 21 KB - mental_v9i3e31780_app3.docx]

Multimedia Appendix 4
Potentially relevant in-progress clinical trials.
[DOCX File , 13 KB - mental_v9i3e31780_app4.docx]

Multimedia Appendix 5
Table of included scales.
[DOCX File , 19 KB - mental_v9i3e31780_app5.docx]

References

1. The World Health Report 2001: mental disorders affect one in four people. World Health Organisation. 2001. URL: https://www.who.int/news/item/28-09-2001-the-world-health-report-2001-mental-disorders-affect-one-in-four-people [accessed 2021-06-01]
2. Richter D, Wall A, Bruen A, Whittington R. Is the global prevalence rate of adult mental illness increasing? Systematic review and meta-analysis. Acta Psychiatr Scand 2019 Nov;140(5):393-407 [FREE Full text] [doi: 10.1111/acps.13083] [Medline: 31393996]
3. Torales J, O'Higgins M, Castaldelli-Maia JM, Ventriglio A. The outbreak of COVID-19 coronavirus and its impact on global mental health. Int J Soc Psychiatry 2020 Jun;66(4):317-320. [doi: 10.1177/0020764020915212] [Medline: 32233719]
4. Vindegaard N, Benros ME. COVID-19 pandemic and mental health consequences: systematic review of the current evidence. Brain Behav Immun 2020 May 30:531-543 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.048] [Medline: 32485289]
5. Backhaus A, Agha Z, Maglione ML, Repp A, Ross B, Zuest D, et al. Videoconferencing psychotherapy: a systematic review. Psychol Serv 2012 May;9(2):111-131. [doi: 10.1037/a0027924] [Medline: 22662727]

6. Hale M, Leah BM. Why have psychologists been slow to adopt telehealth? InPsych. 2020. URL: https://psychology.org.au/for-members/publications/inpsych/2020/june-july-issue-3/why-have-psychologists-been-slow-to-adopt-teleheal [accessed 2021-06-01]
7. Pierce BS, Perrin PB, McDonald SD. Pre-COVID-19 deterrents to practicing with videoconferencing telepsychology among psychologists who didn't. Psychol Serv 2020 Nov 30:157-166. [doi: 10.1037/ser0000435] [Medline: 33252923]
8. Hoffmann M, Wensing M, Peters-Klimm F, Szecsenyi J, Hartmann M, Friederich H, et al. Perspectives of psychotherapists and psychiatrists on mental health care integration within primary care via video consultations: qualitative preimplementation study. J Med Internet Res 2020 Jun 18;22(6):e17569 [FREE Full text] [doi: 10.2196/17569] [Medline: 32554369]
9. Chang JE, Lai AY, Gupta A, Nguyen AM, Berry CA, Shelley DR. Rapid transition to telehealth and the digital divide: implications for primary care access and equity in a post-covid era. Milbank Q 2021 Jun;99(2):340-368 [FREE Full text] [doi: 10.1111/1468-0009.12509] [Medline: 34075622]
10. Monaghesh E, Hajizadeh A. The role of telehealth during COVID-19 outbreak: a systematic review based on current evidence. BMC Public Health 2020 Aug 01;20(1):1193 [FREE Full text] [doi: 10.1186/s12889-020-09301-4] [Medline: 32738884]
11. Punia V, Nasr G, Zagorski V, Lawrence G, Fesler J, Nair D, et al. Evidence of a rapid shift in outpatient practice during the COVID-19 pandemic using telemedicine. Telemed J E Health 2020 Oct;26(10):1301-1303. [doi: 10.1089/tmj.2020.0150] [Medline: 32429769]
12. Wosik J, Fudim M, Cameron B, Gellad ZF, Cho A, Phinney D, et al. Telehealth transformation: COVID-19 and the rise of virtual care. J Am Med Inform Assoc 2020 Apr 20:957-962 [FREE Full text] [doi: 10.1093/jamia/ocaa067] [Medline: 32311034]
13. Butzner M, Cuffee Y. Telehealth interventions and outcomes across rural communities in the United States: narrative review. J Med Internet Res 2021 Aug 26;23(8):e29575 [FREE Full text] [doi: 10.2196/29575] [Medline: 34435965]
14. Fletcher TL, Hogan JB, Keegan F, Davis ML, Wassef M, Day S, et al. Recent advances in delivering mental health treatment via video to home. Curr Psychiatry Rep 2018 Jul 21;20(8):56. [doi: 10.1007/s11920-018-0922-y] [Medline: 30032337]
15. Postel MG, de Jong CA, de Haan HA. Does e-therapy for problem drinking reach hidden populations? Am J Psychiatry 2005 Dec;162(12):2393. [doi: 10.1176/appi.ajp.162.12.2393] [Medline: 16330613]
16. King VL, Stoller KB, Kidorf M, Kindbom K, Hursh S, Brady T, et al. Assessing the effectiveness of an Internet-based videoconferencing platform for delivering intensified substance abuse counseling. J Subst Abuse Treat 2009 Apr;36(3):331-338. [doi: 10.1016/j.jsat.2008.06.011] [Medline: 18775625]
17. ABS national survey of mental health and wellbeing: summary of results 2007. Australian Bureau of Statistics. 2008. URL: https://www.abs.gov.au/statistics/health/mental-health/national-survey-mental-health-and-wellbeing-summary-results/latest-release [accessed 2021-06-01]
18. Pong RW, Desmeules M, Lagacé C. Rural-urban disparities in health: how does Canada fare and how does Canada compare with Australia? Aust J Rural Health 2009 Feb;17(1):58-64 [FREE Full text] [doi: 10.1111/j.1440-1584.2008.01039.x] [Medline: 19161503]
19. Smith KB, Humphreys JS, Wilson MG. Addressing the health disadvantage of rural populations: how does epidemiological evidence inform rural health policies and research? Aust J Rural Health 2008 Apr;16(2):56-66. [doi: 10.1111/j.1440-1584.2008.00953.x] [Medline: 18318846]
20. Berryhill MB, Culmer N, Williams N, Halli-Tierney A, Betancourt A, Roberts H, et al. Videoconferencing Psychotherapy and Depression: A Systematic Review. Telemed J E Health 2019 Jun;25(6):435-446. [doi: 10.1089/tmj.2018.0058] [Medline: 30048211]
21. Berryhill MB, Halli-Tierney A, Culmer N, Williams N, Betancourt A, King M, et al. Videoconferencing psychological therapy and anxiety: a systematic review. Fam Pract 2019 Jan 25;36(1):53-63. [doi: 10.1093/fampra/cmy072] [Medline: 30188992]
22. Krzyzaniak N, Greenwood H, Scott AM, Peiris R, Cardona M, Clark J, et al. The effectiveness of telehealth versus face-to face interventions for anxiety disorders: a systematic review and meta-analysis. J Telemed Telecare 2021 Dec 03:1357633X211053738. [doi: 10.1177/1357633X211053738] [Medline: 34860613]
23. Scott AM, Bakhit M, Greenwood H, Cardona M, Clark J, Krzyzankiak N. Real-time telehealth versus face-to-face management for patients with PTSD in primary care: a systematic review and meta-analysis (forthcoming). J Clin Psychiatry 2022;Forthcoming.
24. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
25. Clark JM, Sanders S, Carter M, Honeyman D, Cleo G, Auld Y, et al. Improving the translation of search strategies using the Polyglot Search Translator: a randomized controlled trial. J Med Libr Assoc 2020 Apr;108(2):195-207 [FREE Full text] [doi: 10.5195/jmla.2020.834] [Medline: 32256231]
26. Scopus. Elsevier B.V. 2021. URL: https://www.scopus.com/search/form.uri?display=basic#basic [accessed 2021-12-01]
27. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M. Cochrane handbook for systematic reviews of interventions. Hoboken, New Jersey, United States: John Wiley & Sons; 2019.

28. Review Manager (RevMan) [Computer program]. Version 5.4. The Cochrane Collaboration. 2020. URL: https://training.cochrane.org/online-learning/core-software-cochrane-reviews/revman/revman-non-cochrane-reviews [accessed 2022-02-11]
29. TIA: a feasibility study of telemedicine in addictions. ISRCTN Registry Identifier: ISRCTN36756455. 2021. URL: https://www.isrctn.com/ISRCTN36756455 [accessed 2022-02-21]
30. OT intervention via telehealth for children with ASD and ADHD. ClinicalTrials.gov Identifier: NCT03957993. URL: https://clinicaltrials.gov/ct2/show/NCT03957993 [accessed 2022-02-21]
31. Burgess M, Andiappan M, Chalder T. Cognitive behaviour therapy for chronic fatigue syndrome in adults: face to face versus telephone treatment: a randomized controlled trial. Behav Cogn Psychother 2012 Mar;40(2):175-191. [doi: 10.1017/S1352465811000543] [Medline: 21929831]
32. Comer JS, Furr JM, Miguel EM, Cooper-Vince CE, Carpenter AL, Elkins RM, et al. Remotely delivering real-time parent training to the home: an initial randomized trial of Internet-delivered parent-child interaction therapy (I-PCIT). J Consult Clin Psychol 2017 Sep;85(9):909-917. [doi: 10.1037/ccp0000230] [Medline: 28650194]
33. Day SX, Schneider PL. Psychotherapy using distance technology: a comparison of face-to-face, video, and audio treatment. J Counsel Psychol 2002;49(4):499-503. [doi: 10.1037/0022-0167.49.4.499]
34. Duke DC, Wagner DV, Ulrich J, Freeman KA, Harris MA. Videoconferencing for teens with diabetes: family matters. J Diabetes Sci Technol 2016 Jul;10(4):816-823 [FREE Full text] [doi: 10.1177/1932296816642577] [Medline: 27075708]
35. Freeman KA, Duke DC, Harris MA. Behavioral health care for adolescents with poorly controlled diabetes via Skype: does working alliance remain intact? J Diabetes Sci Technol 2013 May 01;7(3):727-735 [FREE Full text] [doi: 10.1177/193229681300700318] [Medline: 23759406]
36. Himle MB, Freitag M, Walther M, Franklin SA, Ely L, Woods DW. A randomized pilot trial comparing videoconference versus face-to-face delivery of behavior therapy for childhood tic disorders. Behav Res Ther 2012 Sep;50(9):565-570. [doi: 10.1016/j.brat.2012.05.009] [Medline: 22743661]
37. King VL, Brooner RK, Peirce JM, Kolodner K, Kidorf MS. A randomized trial of Web-based videoconferencing for substance abuse counseling. J Subst Abuse Treat 2014 Jan;46(1):36-42 [FREE Full text] [doi: 10.1016/j.jsat.2013.08.009] [Medline: 24035556]
38. McAndrew LM, Greenberg LM, Ciccone DS, Helmer DA, Chandler HK. Telephone-based versus in-person delivery of cognitive behavioral treatment for veterans with chronic multisymptom illness: a controlled, randomized trial. Mil Behav Health 2018;6(1):56-65 [FREE Full text] [doi: 10.1080/21635781.2017.1337594] [Medline: 31192051]
39. Crow SJ, Mitchell JE, Crosby RD, Swanson SA, Wonderlich S, Lancanster K. The cost effectiveness of cognitive behavioral therapy for bulimia nervosa delivered via telemedicine versus face-to-face. Behav Res Ther 2009 Jun;47(6):451-453 [FREE Full text] [doi: 10.1016/j.brat.2009.02.006] [Medline: 19356743]
40. Ertelt TW, Crosby RD, Marino JM, Mitchell JE, Lancaster K, Crow SJ. Therapeutic factors affecting the cognitive behavioral treatment of bulimia nervosa via telemedicine versus face-to-face delivery. Int J Eat Disord 2011 Dec;44(8):687-691. [doi: 10.1002/eat.20874] [Medline: 22072405]
41. Mitchell JE, Crosby RD, Wonderlich SA, Crow S, Lancaster K, Simonich H, et al. A randomized trial comparing the efficacy of cognitive-behavioral therapy for bulimia nervosa delivered via telemedicine versus face-to-face. Behav Res Ther 2008 May;46(5):581-592 [FREE Full text] [doi: 10.1016/j.brat.2008.02.004] [Medline: 18374304]
42. Watson M, White C, Lynch A, Mohammed K. Telephone-delivered individual cognitive behavioural therapy for cancer patients: an equivalence randomised trial. Psychooncology 2017 Mar;26(3):301-308. [doi: 10.1002/pon.4338] [Medline: 27943570]
43. Xie Y, Dixon JF, Yee OM, Zhang J, Chen YA, Deangelo S, et al. A study on the effectiveness of videoconferencing on teaching parent training skills to parents of children with ADHD. Telemed J E Health 2013 Mar;19(3):192-199. [doi: 10.1089/tmj.2012.0108] [Medline: 23405952]
44. Treadwell JR, Uhl S, Tipton K, Shamliyan T, Viswanathan M, Berkman ND, et al. Assessing equivalence and noninferiority. J Clin Epidemiol 2012 Nov;65(11):1144-1149. [doi: 10.1016/j.jclinepi.2012.05.001] [Medline: 22732455]
45. Coughtrey AE, Pistrang N. The effectiveness of telephone-delivered psychological therapies for depression and anxiety: a systematic review. J Telemed Telecare 2016 Dec 30;24(2):65-74. [doi: 10.1177/1357633x16686547]
46. Snoswell CL, Taylor ML, Comans TA, Smith AC, Gray LC, Caffery LJ. Determining if telehealth can reduce health system costs: scoping review. J Med Internet Res 2020 Oct 19;22(10):e17298 [FREE Full text] [doi: 10.2196/17298] [Medline: 33074157]
47. Rees CS, Stone S. Therapeutic Alliance in Face-to-Face Versus Videoconferenced Psychotherapy. Professional Psychology: Research and Practice 2005 Dec;36(6):649-653. [doi: 10.1037/0735-7028.36.6.649]
48. Humer E, Stippl P, Pieh C, Pryss R, Probst T. Experiences of Psychotherapists With Remote Psychotherapy During the COVID-19 Pandemic: Cross-sectional Web-Based Survey Study. J Med Internet Res 2020 Nov 27;22(11):e20246 [FREE Full text] [doi: 10.2196/20246] [Medline: 33151896]
49. Malas N, Klein E, Tengelitsch E, Kramer A, Marcus S, Quigley J. Exploring the Telepsychiatry Experience: Primary Care Provider Perception of the Michigan Child Collaborative Care (MC3) Program. Psychosomatics 2019 Mar;60(2):179-189. [doi: 10.1016/j.psym.2018.06.005] [Medline: 30097288]

50. Watts S, Marchand A, Bouchard S, Gosselin P, Langlois F, Belleville G, et al. Telepsychotherapy for generalized anxiety disorder: impact on the working alliance. J Psychother Integrat 2020 Jun;30(2):208-225. [doi: 10.1037/int0000223]
51. Dowling M, Rickwood D. Online counseling and therapy for mental health problems: a systematic review of individual synchronous interventions using chat. J Technol Hum Serv 2013 Jan;31(1):1-21. [doi: 10.1080/15228835.2012.728508]
52. Hoermann S, McCabe KL, Milne DN, Calvo RA. Application of synchronous text-based dialogue systems in mental health interventions: systematic review. J Med Internet Res 2017 Jul 21;19(8):e267 [FREE Full text] [doi: 10.2196/jmir.7023] [Medline: 28784594]
53. Gainsbury S, Blaszczynski A. A systematic review of internet-based therapy for the treatment of addictions. Clin Psychol Rev 2011 Apr;31(3):490-498. [doi: 10.1016/j.cpr.2010.11.007] [Medline: 21146272]
54. Loucas CE, Fairburn CG, Whittington C, Pennant ME, Stockton S, Kendall T. E-therapy in the treatment and prevention of eating disorders: a systematic review and meta-analysis. Behav Res Ther 2014 Dec;63:122-131 [FREE Full text] [doi: 10.1016/j.brat.2014.09.011] [Medline: 25461787]
55. Josephine K, Josefine L, Philipp D, David E, Harald B. Internet- and mobile-based depression interventions for people with diagnosed depression: a systematic review and meta-analysis. J Affect Disord 2017 Dec 01;223:28-40. [doi: 10.1016/j.jad.2017.07.021] [Medline: 28715726]
56. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
57. Schütte S, Acevedo PN, Flahault A. Health systems around the world - a comparison of existing health system rankings. J Glob Health 2018 Jun;8(1):010407 [FREE Full text] [doi: 10.7189/jogh.08.010407] [Medline: 29564084]
58. Thorlund K, Imberger G, Johnston BC, Walsh M, Awad T, Thabane L, et al. Evolution of heterogeneity (I2) estimates and their 95% confidence intervals in large meta-analyses. PLoS One 2012;7(7):e39471 [FREE Full text] [doi: 10.1371/journal.pone.0039471] [Medline: 22848355]
59. Jenkins-Guarnieri MA, Pruitt LD, Luxton DD, Johnson K. Patient perceptions of telemental health: systematic review of direct comparisons to in-person psychotherapeutic treatments. Telemed J E Health 2015 Aug;21(8):652-660. [doi: 10.1089/tmj.2014.0165] [Medline: 25885491]
60. Brenes GA, Ingram CW, Danhauer SC. Benefits and challenges of conducting psychotherapy by telephone. Prof Psychol Res Pract 2011 Dec;42(6):543-549 [FREE Full text] [doi: 10.1037/a0026135] [Medline: 22247588]
61. Brooks E, Turvey C, Augusterfer EF. Provider barriers to telemental health: obstacles overcome, obstacles remaining. Telemed J E Health 2013 Jun;19(6):433-437. [doi: 10.1089/tmj.2013.0068] [Medline: 23590176]
62. Lopez A, Schwenk S, Schneck CD, Griffin RJ, Mishkind MC. Technology-based mental health treatment and the impact on the therapeutic alliance. Curr Psychiatry Rep 2019 Jul 08;21(8):76. [doi: 10.1007/s11920-019-1055-7] [Medline: 31286280]

Abbreviations

CBT: cognitive behavioral therapy
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PTSD: posttraumatic stress disorder
RCT: randomized controlled trial
SMD: standardized mean difference

Edited by J Torous; submitted 28.07.21; peer-reviewed by J Lindsay, M Keijsers, KP Wong; comments to author 01.09.21; revised version received 16.12.21; accepted 12.01.22; published 11.03.22.

Please cite as:
Greenwood H, Krzyzaniak N, Peiris R, Clark J, Scott AM, Cardona M, Griffith R, Glasziou P
Telehealth Versus Face-to-face Psychotherapy for Less Common Mental Health Conditions: Systematic Review and Meta-analysis of Randomized Controlled Trials
JMIR Ment Health 2022;9(3):e31780
URL: https://mental.jmir.org/2022/3/e31780
doi:10.2196/31780
PMID:35275081

Review

Development of a Framework for the Implementation of Synchronous Digital Mental Health: Realist Synthesis of Systematic Reviews

David Villarreal-Zegarra[1,2], BSc, MPH; Christoper A Alarcon-Ruiz[3], MD; GJ Melendez-Torres[4], DPhil; Roberto Torres-Puente[2], BSc; Alba Navarro-Flores[2,5], MD; Victoria Cavero[6], BSc, MSc; Juan Ambrosio-Melgarejo[2], BSc; Jefferson Rojas-Vargas[7], BSc; Guillermo Almeida[2], BSc; Leonardo Albitres-Flores[6,8], MD; Alejandra B Romero-Cabrera[9], MD; Jeff Huarcaya-Victoria[10,11], MD, MSc

[1]Escuela de Medicina, Universidad César Vallejo, Trujillo, Peru

[2]Dirección de Investigación, Desarrollo e Innovación, Instituto Peruano de Orientación Psicológica, Lima, Peru

[3]Unidad de Investigación para la Generación y Síntesis de Evidencias en Salud, Universidad San Ignacio de Loyola, Lima, Peru

[4]Peninsula Technology Assessment Group, College of Medicine and Health, University of Exeter, Devon, United Kingdom

[5]Georg-August-University Göttingen, International Max Planck Research School for Neurosciences, Göttingen, Germany

[6]Center of Excellence in Chronic Diseases, Universidad Peruana Cayetano Heredia, Lima, Peru

[7]Facultad de Psicología, Universidad Nacional Mayor de San Marcos, Lima, Peru

[8]Facultad de Medicina, Universidad Nacional de Trujillo, Trujillo, Peru

[9]Carrera Profesional de Medicina Humana, Facultad de Ciencias de la Salud, Universidad Científica del Sur, Lima, Peru

[10]Unidad de Psiquiatría de Enlace, Departamento de Psiquiatría, Hospital Nacional Guillermo Almenara Irigoyen, Lima, Peru

[11]Escuela Profesional de Medicina Humana, Universidad Privada San Juan Bautista, Filial Ica, Peru

Corresponding Author:
Jeff Huarcaya-Victoria, MD, MSc
Unidad de Psiquiatría de Enlace
Departamento de Psiquiatría
Hospital Nacional Guillermo Almenara Irigoyen
Av Grau 800
Lima, 15001
Peru
Phone: 51 987382101
Email: jeff.huarcaya@upsjb.edu.pe

Abstract

Background: The use of technologies has served to reduce gaps in access to treatment, and digital health interventions show promise in the care of mental health problems. However, to understand what and how these interventions work, it is imperative to document the aspects related to their challenging implementation.

Objective: The aim of this study was to determine what evidence is available for synchronous digital mental health implementation and to develop a framework, informed by a realist review, to explain what makes digital mental health interventions work for people with mental health problems.

Methods: The SPIDER (Sample, Phenomenon of Interest, Design, Evaluation, and Research type) framework was used to develop the following review question: What makes digital mental health interventions with a synchronous component work on people with mental health problems, including depression, anxiety, or stress, based on implementation, economic, quantitative, qualitative, and mixed methods studies? The MEDLINE, EBM Reviews, PsycINFO, EMBASE, SCOPUS, CINAHL Complete, and Web of Science databases were searched from January 1, 2015, to September 2020 with no language restriction. A Measurement Tool to Assess Systematic Reviews-2 (AMSTAR-2) was used to assess the risk of bias and Confidence in Evidence from Reviews of Qualitative Research (CERQual) was used to assess the confidence in cumulative evidence. Realist synthesis analysis allowed for developing a framework on the implementation of synchronous digital mental health using a grounded-theory approach with an emergent approach.

Results: A total of 21 systematic reviews were included in the study. Among these, 90% (n=19) presented a critically low confidence level as assessed with AMSTAR-2. The realist synthesis allowed for the development of three hypotheses to identify the context and mechanisms in which these interventions achieve these outcomes: (1) these interventions reach populations otherwise unable to have access because they do not require the physical presence of the therapist nor the patient, thereby tackling geographic barriers posed by in-person therapy; (2) these interventions reach populations otherwise unable to have access because they can be successfully delivered by nonspecialists, which makes them more cost-effective to implement in health services; and (3) these interventions are acceptable and show good results in satisfaction because they require less need of disclosure and provide more privacy, comfortability, and participation, enabling the establishment of rapport with the therapist.

Conclusions: We developed a framework with three hypotheses that explain what makes digital mental health interventions with a synchronous component work on people with mental health problems. Each hypothesis represents essential outcomes in the implementation process.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42020203811; https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020203811

International Registered Report Identifier (IRRID): RR2-10.12688/f1000research.27150.2

(JMIR Ment Health 2022;9(3):e34760) doi:10.2196/34760

KEYWORDS

telemedicine; digital health; internet-based intervention; mental health; mental disorders; systematic reviews; qualitative research; realist review; mHealth; eHealth; telehealth

Introduction

Mental health is in crisis globally and the COVID-19 pandemic has suddenly revealed the magnitude of this problem [1,2]. To minimize health care gaps, the use of digital technologies has been proposed to be able to provide specialized treatment to a greater number of people in places with limited resources and to those with difficult access [3-7]. These technologies have been very well received and served to complement or improve the effectiveness of treatments for various chronic diseases [6]. In addition, these digital interventions show great promise in the care of mental health problems [8-10].

With the undeniable contribution of technologies in mental health care, it is important to document the aspects related to their challenging implementation [11], such as adaptability, cost, complexity, external policies and incentives, compatibility, or general fit between the digital health intervention and the organization, among others [12]. These features provide understanding about how and what works in these interventions, and considering the complexity as challenges in the implementation of telemedicine can help to reveal the deficiencies and inequalities of health care systems worldwide [13].

Currently, there are different frameworks to guide the implementation process, including Expert Recommendations for Implementing Change (ERIC), Promoting Action on Research Implementation in Health Services (PARIHS), or Consolidated Framework for Implementation Research (CFIR) [14,15]. However, we have not been able to find studies that developed frameworks to explain what makes digital mental health interventions work, specify in which contexts these digital interventions can be implemented, identify the mechanisms that facilitate or hinder their implementation, and elucidate the most important outcomes within the implementation process.

Despite not developing such a framework, previous studies have identified critical aspects to consider within the implementation process, such as the effectiveness of digital mental health interventions [16,17], barriers and facilitators to the implementation of electronic mental health interventions [18], or long-term cost-effectiveness studies [19]. However, this evidence alone is not sufficient to warrant the implementation of these interventions or their adoption by health systems [4].

The problem with not having a specific framework for the implementation of interventions focused on digital mental health is that this type of intervention has particular nuances compared with other types of health interventions [4], especially in low- or middle-income countries. More qualitative and flexible approaches are needed to understand the complexity of these interventions and what key elements could help their implementation [4]. Thus, the aim of this study was to determine what evidence is available for synchronous digital mental health implementation and to develop a framework, informed by a realist review, to explain what makes digital mental health interventions work for people with mental health problems.

Methods

Research Question

This systematic review adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [20]; a completed PRISMA checklist can be found in Multimedia Appendix 1. The detailed methodology is available in the published study protocol [21], and the study was registered in PROSPERO (CRD420203811). The SPIDER framework was used to develop the review question, which is based on describing the Sample (S), Phenomenon of Interest (PI), Design (D), Evaluation (E), and Research type (R) [22] (see Textbox 1).

Textbox 1. Research question development based on the SPIDER framework.

- Sample

Adults with depression (or major depressive disorder), anxiety (or generalized anxiety disorder), stress (or trauma-related disorders), and/or general mental health problems (unspecified). Participants may be diagnosed through clinical interviews or categorized based on screening assessments (self-reported scales).

- Phenomenon of Interest

Any digital mental health intervention that includes a synchronous component, namely communication with a mental health professional (eg, psychiatrist, psychologist) or a health professional trained in mental health. These interventions included, among others, remote consultation, interactive application, video chats, and calls.

- Design

Systematic review.

- Evaluation

We included all types of outcomes of interest assessed by implementation studies, economic, qualitative, quantitative, and other study designs, including (1) health effectiveness outcomes (eg, depression, anxiety and/or stress symptoms, adherence to treatment), (2) patient outcomes (eg, quality of life, satisfaction), (3) economic outcomes, and (4) damage or adverse effects.

- Research type

Quantitative, qualitative, and mixed methods.

Eligibility Criteria

Inclusion Criteria

Systematic reviews were selected that reported on inclusion/exclusion criteria for their included studies; conducted an adequate systematic literature search using at least two databases; and synthesized, assessed the quality of, and presented sufficient detail on their individual primary included studies [23]. The reviews had to include primary studies as a unit of analysis focused on a research question. We selected a publication start date of January 1, 2015, without language restrictions. We selected this time frame to include only the latest systematic reviews, since in the field of digital health, the launch of new technologies makes scientific development dynamic. Articles were also included if the primary studies in the review focused on adults with common mental health problems, defined as (1) adults with depression (or major depressive disorder), anxiety (or generalized anxiety disorder), stress (or trauma-related disorders), and/or general mental health problems (unspecified); or (2) adults attending an outpatient mental health consultation. The final inclusion criterion was that at least 90% of the primary studies assessed synchronous digital mental health or the results only for synchronous digital mental health are presented separately.

Exclusion Criteria

Narrative reviews, scoping reviews, primary studies, opinion/editorial manuscripts, letters to the editor, and reviews of mobile health intervention repositories (ie, app stores) were excluded. In addition, reviews that included primary studies of (1) adult participants with some other specific mental health condition outside of those listed above, (2) healthy adult participants (without mental health conditions), (3) adult participants receiving emergency/crisis psychiatric care, (4) interventions that lack a synchronic component (real-time information exchange between the user and mental health professional using technologies) or were not sufficiently clear of having a synchronic component, or (5) women with depressive postpartum symptoms were also excluded from the analysis.

Information Sources

We searched the MEDLINE (Ovid), EBM Reviews (Ovid), PsycINFO (Ovid), EMBASE (Elsevier), SCOPUS, CINAHL Complete (EBSCOhost), and Web of Science databases, including Science Citation Index Expanded, Social Sciences Citation Index, and Conference Proceedings Citation Index (Clarivate Analytics). Articles published in the last 5 years (January 1, 2015, to April 30, 2020) were included with no language restrictions. The search of the databases was performed on April 30, 2020.

Search Strategy

The search formula was created using thesaurus and entry terms for the following syntaxis: “telemedicine” AND “mental health, anxiety, depression or stress” AND “systematic reviews.” The full search strategy for each database is available in Multimedia Appendix 2.

Study Records

Data Management

The records retrieved after the search were managed using the Rayyan QCRI free online application (eliminate duplicates, and review titles and abstracts) [24]. Full-text review and data extraction were performed in an Excel template.

Selection Process

The records were screened by title and abstract and then by full-text assessment. The records were divided into three groups with each consisting of a pair of independent reviewers (six people in total). Before conducting the review of the records, a

calibration process was carried out, which was based on a pilot review of 30 registries and identifying that there was a discrepancy of less than 5% in the decision of whether or not to include the studies. During the review, in case of discrepancies between decisions within groups, peers discussed the discrepancies to reach an agreement. When it was not possible to reach an agreement among the peers, a third reviewer was included if necessary.

Data Collection Process

For each eligible study, data were extracted independently and duplicated on predesigned extraction forms. Reviewers solved discrepancies and a third reviewer evaluated any unresolved disagreement.

Data Items

An extraction form was created for the included systematic reviews. We collected the following information: first author and publication date of the study, characteristics of the participants, main objective, research questions, inclusion criteria for the systematic review, search date, study selection process, quality assessment (if any), main findings, and limitations. The full text of the included articles, tables, and supplementary material were also gathered to perform the qualitative analysis of the text.

Outcomes and Prioritization

The aim of our study was to perform a realist review of systematic reviews using a qualitative strategy to synthesize the information and answer our research question. Therefore, we did not look for a specific result such as effectiveness, cost-effectiveness, or similar. Instead, we were interested in identifying the full text of all studies that answered our research question to perform a grounded-theory analysis with an emergent approach [25]. Priority was given in the analysis to studies with the lowest risk of bias assessed.

Risk of Bias in Individual Studies

To assess the quality of the included systematic reviews, we used A Measurement Tool to Assess Systematic Reviews-2 (AMSTAR-2), which has 16 domains. Seven of these domains are considered critical: (1) protocol registered before the start of the review, (2) adequacy of the literature search, (3) justification for the exclusion of individual studies, (4) risk of bias of individual studies included in the review, (5) adequacy of meta-analytic methods, (6) consideration of the risk of bias in interpreting the results of the review, and (7) assessment of the presence and likely impact of publication bias [26].

AMSTAR-2 classifies the quality of systematic reviews into four categories: high (none or one noncritical weakness), moderate (more than one noncritical weakness), low (one critical weakness with or without noncritical weaknesses), and very low (more than one critical weakness with or without noncritical weaknesses). The quality assessment was rated by two trained researchers independently. In case of difference in the overall quality assessment of the systematic reviews, the AMSTAR-2 criteria were discussed between the two researchers to reach a consensus.

Data Synthesis

We developed a framework informed by a realist analysis of synchronous digital mental health interventions using a grounded-theory approach with an emergent approach [27]. The realist synthesis was based on interpreting, integrating, and inferring the evaluation elements to better understand the implementation of synchronous digital mental health interventions from all of the included studies [28]. To answer the question "what makes the implementation of these interventions work?", hypotheses supported by the included studies' results were developed and generated through discussion and consensus among the researchers [28]. Since our study was designed to perform a realist synthesis of the evidence, we focused on different outcomes to use them as input for assessing the implementation of synchronous digital mental health interventions. Therefore, we did not perform a quantitative synthesis in any case (ie, a meta-analysis of effectiveness).

Three researchers followed the three steps established by Thomas and Harden [28] for qualitative syntheses [29]. First, the extracted data were freely coded. The researchers read the full texts of the included articles and coded each text fragment that provided information to answer the research question. Second, the codified data were organized and then grouped based on descriptive aspects using a context-linked causality approach represented as "context+mechanism=outcome" [25]. Finally, the analytical concepts generated in the previous step were grouped so that they were related to each other. The elements that were related to each other were assumed to be part of a hypothesis that would help to answer the research aim.

The selection of the studies for the realist review was based on the AMSTAR-2 score, with the highest-quality studies being assessed first. We assessed all included studies, down to the criterion of theoretical saturation [30]. All qualitative analyses were performed with NVivo software (version 12, QSR International).

Confidence in Cumulative Evidence

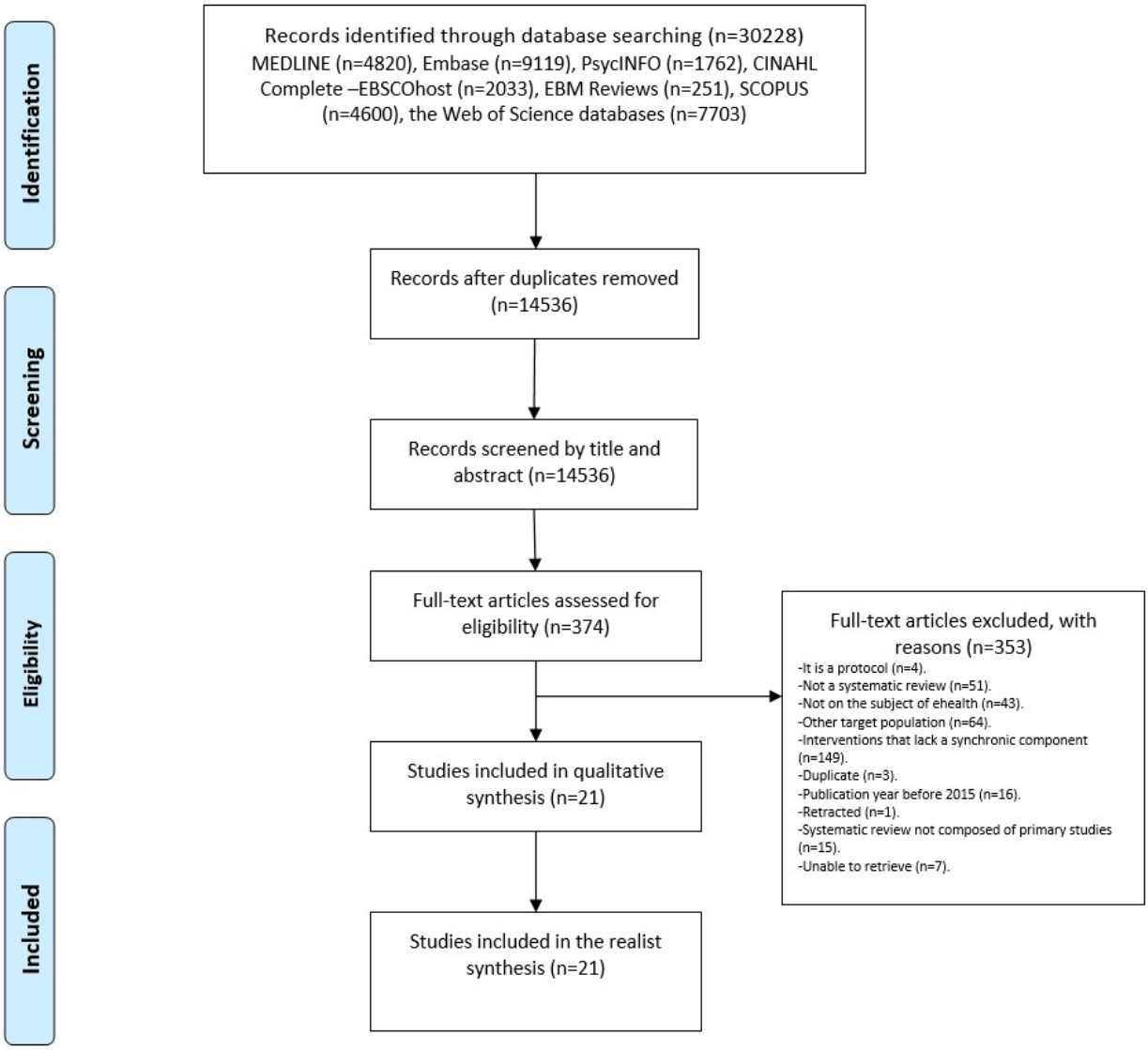
The Confidence in Evidence from Reviews of Qualitative Research (CERQual) approach, which has four components (Methodological Limitations, Relevance, Coherence, and Appropriateness Data), was assessed by a researcher and then reviewed by another independent researcher. The CERQual was evaluated to contribute to an overall assessment of each hypothesis resulting from the realist synthesis to determine the level of confidence (high, moderate, low, or very low) and to present the overall assessment in a Summary of Qualitative Findings table [31,32].

Results

Study Selection

The search strategy retrieved 30,228 records, and after duplicated cleaning, we obtained 14,536 unique records. The evaluation by title and abstract identified 374 results that were evaluated at the full-text level. Among those, 353 were excluded. The reasons for exclusion are listed in Multimedia Appendix 3. Finally, 21 systematic reviews were included in this study (see Figure 1).

Figure 1. Flowchart of the study selection process.



Study Characteristics

The included systematic reviews analyzed a median of 27 studies (range 9-155). Eleven studies reported some form of synchronous digital mental health intervention based on internet, telephone, or online cognitive behavioral therapy (CBT) as the primary intervention [19,33-42]. The remaining studies reported a mix of digital mental health interventions based on synchronous components (ie, telephone, videoconferencing) and asynchronous components (ie, text messages, email, chats, instructional videos, podcasts). Most of the systematic reviews included exclusively randomized controlled trials (RCTs) as primary studies, two included only non-RCTs, and five studies included both. Only six studies did not include a meta-analysis. With respect to the type of therapy, nine reviews stated CBT as the target therapy, one review used the transdiagnosis method, and one included mindfulness-based interventions. The individual characteristics of the included studies are presented in Multimedia Appendix 4. It is important to mention that despite having no language restrictions, all of the included articles were published in English and the systematic reviews did not include qualitative studies.

Risk of Bias Within Studies

Most of the studies (19/21, 90%) of the included systematic reviews performed a risk of bias assessment. The most commonly used instrument was the Risk of Bias Cochrane Collaboration tool (12/24, 57%) [33-35,37,38,40,43-48]. Seven studies used other tools to assess the risk of bias such as the Effective Public Health Practice Project Quality Assessment Tool (2/21, 10%); Grading of Recommendations Assessment, Development and Evaluation (1/21, 5%); and others. Only two studies did not report using any risk of bias tool [41,42]. Ten studies did not appropriately account for the risk of bias of the individual studies included when interpreting the results of their review.

Olthuis et al [47] presented a medium level of confidence and Lewis et al [38] presented low confidence. The rest of the included systematic reviews presented a critically low level of confidence (see Figure 2). On average, the included reviews only met 40% of the AMSTAR-2 risk of bias items. The reviews included in Rees et al [41] failed to accomplish any of the AMSTAR-2 items and those in the study by Turgoose et al [49] only passed one AMSTAR-2 item.

The AMSTAR-2 items that were the most fulfilled (if applicable) were item 15 (critical) assessing the presence and likely impact of publication bias (14/21, 93%) and item 12 (noncritical) assessing the potential impact of risk of bias in individual studies (11/21, 73%) in the case of meta-analysis. The AMSTAR-2 items that were the least fulfilled were item 10 (noncritical) on whether the review reported the funding sources of the included studies. Only the study by Irvine et al [36] achieved compliance. Two other items that had a low compliance rate (3/21, 14%) were item 4 (critical) on the adequate literature search and item 3 (noncritical) on the justification for the decision on the study designs to be included in the review, and only one study met each of these criteria [46].

Figure 2. Risk of bias assessment of individual studies, according to AMSTAR-2. 1: Did the research questions and inclusion criteria for the review include the components of PICO (Population, Intervention, Control, Outcomes)? *2: Did the report of the review contain an explicit statement that the review methods were established prior to conduct of the review and did the report justify any significant deviations from the protocol? (critical item); 3: Did the review authors explain their selection of the study designs for inclusion in the review? *4: Did the review authors use a comprehensive literature search strategy? (critical item); 5: Did the review authors perform study selection in duplicate? 6: Did the review authors perform data extraction in duplicate? *7: Did the review authors provide a list of excluded studies and justify the exclusions? (critical item); 8: Did the review authors describe the included studies in adequate detail? *9: Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review? (critical item); 10: Did the review authors report on the sources of funding for the studies included in the review? *11: If meta-analysis was justified, did the review authors use appropriate methods for statistical combination of results? (critical item); 12: If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis? *13: Did the review authors account for RoB in individual studies when interpreting/discussing the results of the review? (critical item); 14: Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review? *15: If they performed quantitative synthesis, did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review? (critical item); 16: Conflict of interest declaration.

First author and year /Criterial	1	2*	3	4*	5	6	7*	8	9*	10	11*	12	13*	14	15*	16	Confidence
Ahern, E. et al. (2018)	+	-	-	+	-	-	-	+	-	-	+	-	-	+	+	+	Critically low
Carlbring, P. et al. (2017)	+	-	-	-	-	-	-	-	+	-	+	+	+	+	+	+	Critically low
Castro, A. et al. (2020)	+	+	-	+	+	+	-	+	+	-	-	+	+	+	+	+	Critically low
Coughtrey, A. et al. (2016)	-	-	-	+	-	-	+	+	+	+	NA	NA	-	-	NA	+	Critically low
Cuijpers, P. et al. (2019)	+	+	+	+	+	+	-	+	+	-	+	+	+	+	+	+	Critically low
Domhardt, M. et al (2018)	+	+	-	+	+	+	-	+	+	-	-	+	+	+	+	+	Critically low
Drago, A. et al. (2016)	-	-	-	-	-	+	-	-	-	-	-	-	-	+	+	+	Critically low
Finley, B. et al. (2020)	NA	-	-	+	+	-	-	NA	-	-	NA	NA	-	-	NA	+	Critically low
Irvine, A. et al. (2020)	-	-	-	+	-	-	-	+	-	+	+	-	-	-	-	+	Critically low
Josephine, K. et al (2017)	+	+	-	+	+	+	-	+	+	-	-	+	+	+	+	-	Critically low
Kampmann, I. et al (2016)	-	-	-	+	-	-	-	-	+	-	-	-	-	-	+	+	Critically low
Lewis, C. et al (2017)	+	+	-	+	+	+	+	+	+	-	-	+	+	+	+	-	Low
Linde, K. et al (2015)	+	+	-	+	+	+	-	+	+	-	-	+	+	+	+	+	Critically low
Moulton-Perkins, A. et al (2020)	-	+	-	+	+	+	-	+	+	-	NA	NA	-	-	NA	-	Critically low
Olthuis, J. et al (2015)	+	+	+	+	+	+	+	+	+	-	-	+	+	+	+	+	Medium
Olthuis, J. V. et al (2016)	+	-	+	+	+	+	-	+	+	-	+	+	+	+	+	-	Critically low
Pasarelu, CR. et al (2016)	-	-	-	+	+	-	-	+	+	-	-	+	+	+	+	+	Critically low
Proctor, B. J. et al (2018)	-	-	-	+	-	-	-	-	+	-	-	+	+	+	+	-	Critically low
Rees, C. S. et al (2015)	-	-	-	-	-	-	-	-	-	-	NA	NA	-	-	NA	-	Critically low
Sunjaya, A. et al (2020)	+	-	-	+	-	-	-	-	-	-	NA	NA	-	-	NA	+	Critically low
Turgoose, D. et al (2017)	-	-	-	-	-	-	-	-	+	-	NA	NA	-	-	NA	+	Critically low

+ Yes - No + Partial Yes NA Not applicable

Realist Synthesis

Overview

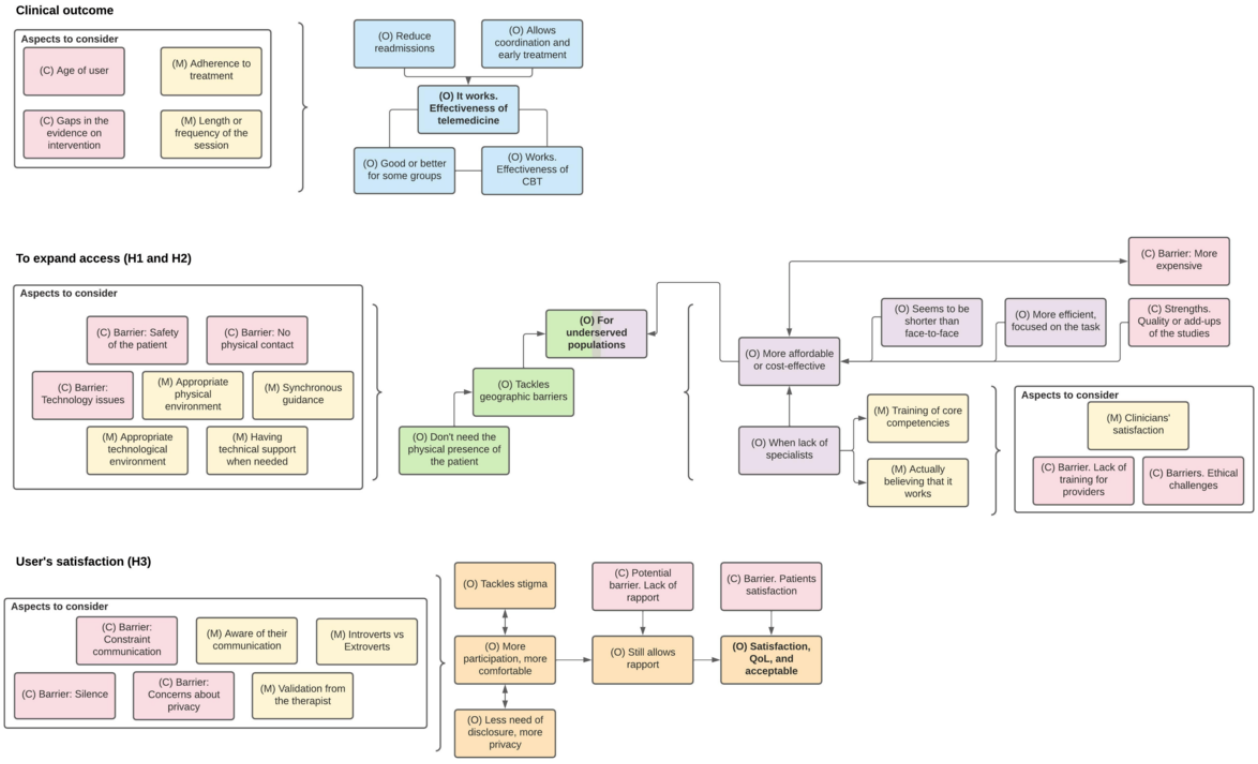
Synchronous digital mental health interventions provide effective clinical outcomes (see Figure 3). Some systematic reviews identified that digital mental health interventions based on CBT (ie, telephone, internet-based, videoconferencing, online) were equally effective as face-to-face CBT in the treatment of specific mental health conditions (eg, social anxiety disorder, posttraumatic stress disorder [PTSD], panic, depressive symptoms, body dissatisfaction, insomnia, specific phobias) [33-36,42,47,49,50]. In addition, the different theoretical models used in CBT-based digital mental health interventions (ie, classical, mindfulness, transdiagnostic, nonspecific) and nonspecific digital mental health interventions had a moderate to large effect in reducing depressive, anxious, and PTSD symptoms compared to control situations [35,37-40,44-46,48,49,51]. Furthermore, different formats of individual and group electronic interventions (ie, telephone, videoconferencing) and guided self-help treatment had comparable effectiveness in depression and anxiety treatment [35,45,46]. In addition, digital interventions have shown to be effective in different population groups such as adults and elder people [33,34,40,45], veterans [47,49], and people with multiple sclerosis [48].

The advantages of interventions using technology are allowing the inclusion of add-ups to the therapy (eg, written, audio or visual materials to access online or download, diary-keeping, chats [19], emails [19,47], online forums [19,43,46], new or existing platforms such as Skype or Zoom) [49]. These interventions also promote better coordination of care and early treatment [42,49].

Guided synchronous components are essential elements in digital interventions to reduce anxiety. They are more effective and significantly improve adherence compared to unguided interventions or those with only asynchronous components [43]. It is also unclear which guided synchronous components are the most effective or whether there are cumulative effects when combining them [43]. Of note, CBT-based and heterogeneous digital mental health interventions (not CBT-based) showed no difference in their effectiveness in reducing PTSD symptoms [38].

Three main hypotheses were derived from this analysis, which are summarized in Textbox 2 and described in detail in the following sections.

Figure 3. Results of the three hypotheses (H1-H3) of the realist synthesis. C: context (pink); M: mechanism (yellow); O: outcome (different colors for each hypothesis); CBT: cognitive behavioral therapy; QoL: quality of life.



Textbox 2. Hypotheses for why digital mental health interventions work for people with mental health problems based on the realist synthesis of reviews.

Hypothesis 1

Synchronous digital mental health interventions reach populations otherwise unable to have access through face-to-face interventions, since they do not require the physical presence of the therapist nor the patient, thereby tackling geographic barriers posed by in-person therapy (to expand access).

Hypothesis 2

Synchronous digital mental health interventions reach populations otherwise unable to have access via face-to-face interventions because they can be successfully delivered by nonspecialists, which makes them more cost-effective to implement in health services (to expand access).

Hypothesis 3

Synchronous digital mental health interventions are acceptable by patients and show good results in satisfaction, because they require less need of disclosure and provide more privacy, comfortability, and participation, enabling the establishment of rapport with the therapist (user satisfaction).

Hypothesis 1

Synchronous digital interventions in mental health reach populations that would not have access through face-to-face interventions, such as children, veterans, refugees, and people living in rural areas [50,52]. This is because these interventions do not require the physical presence of both the patient and the therapist (see Figure 3). We also found that these interventions can reduce geographical barriers to access (eg, mobilization for several hours). In addition, they can interact in real time [41] and tackle the geographic barriers of travel required to receive care, thereby being accessible even from remote areas [37,41,42,46-48,50].

Some aspects need to be taken into consideration for the delivery of successful therapy through synchronous digital mental health interventions. The first is to find a quiet area in the home or at the usual environment of the patient to receive the session, which could represent a challenge for many [49]. The second aspect is that the platform should be as stable as possible since ineffective internet service could lead to withdrawing the therapy [49], and the quality of the image and sound could be associated

with satisfaction [39]. Third, the possibility to expand the use of telepsychiatry will require the development or improvement of a software specially designed for that purpose [42]. Finally, the presence of technical support when needed should be considered, as one systematic review found that scheduled guidance showed better outcomes on anxiety symptom severity at postintervention and follow-up [43].

The presence of synchronous human support seems to improve the delivery of digital mental health sessions, although the evidence is not conclusive [19,43]. Guided interventions were superior to completely unguided interventions for symptom severity across mental disorders and presented higher treatment adherence [43]. In studies that used local clinics rather than home-based teletherapy, it was recommended to have local sta on hand to assist, such as to receive homework and other materials via fax machine and disseminate them to participants [42]. However, in the future, artificial intelligence could replace human support to generate computer responses [33].

Additionally, we found some barriers. The first barrier is the absence of physical contact. One review identified that patients receiving in-person treatment were more likely to complete the home assessments and tasks given [49]. The second barrier is that the safety of the patient could be compromised. It is worth noting a potential issue with interventions using technology. The distance between the patient and therapist could put patients' safety at risk, as they may not receive the necessary care in the event of a crisis or emergency [43]. Some studies also suggested the presence of an extra person to provide in-person support in case of emergencies [43,49], although not all studies showed favorable results [19,33,43]. Finally, the presence of technical issues could impose a potentially modifiable barrier. Some flaws found during the therapy delivery were limited connectivity, the lack of human resources and telepsychiatry equipment [42], low image resolution, difficulties for establishing the connection, slight audio delays, and problems with the internet connection [42]. Moreover, a systematic review assessing mindfulness-based cognitive therapy for stress reduction found that the users' dissatisfaction was linked to technical issues [39].

Hypothesis 2

A second reason for why these interventions reach populations that otherwise would not have access to face-to-face interventions is that they are an accessible and cost-effective treatment in the short term [19]. This may lead to reductions in mental health costs, at least in depression [19]. It should be noted that CBT-based digital interventions tend to be slightly more expensive compared to usual treatment at baseline. This is because their cost-effectiveness improves when considering their positive effect on quality-adjusted life years [19] and their costs in the long-term, since they require limited interaction between the patient and therapist [34,42].

This higher cost-effectiveness is associated with different components. Regarding phone sessions, they adhere to a more structured format and focus on problem-solving and tasks, resulting in more efficient and direct sessions [36] with shorter durations than in-person therapies [34,36]. It should be noted that the session duration of these interventions was not associated with better outcomes in cases of anxiety and depression, although the therapy duration varied from 19 to 150 minutes [40].

Evidence suggests that physicians, psychiatrists, psychologists, or nurses trained for various mental health problems could perform digital interventions such as telepsychiatry or teleconsultations [52]. This enables optimization for using available human resources when there is a reduced number of specialists for large populations, since nonspecialists with adequate training and supervision are as effective as specialists for this purpose [41,43]. For this outcome, it is important to consider some barriers. A potential barrier was the provision of care by nonspecialists, highlighting the importance of having appropriate training and supervision to provide long-distance care. Training for therapists providing interventions using technology should include content on good clinical practices [39,52], the use of technology [49] and telepsychiatry [52], the management of risk or crises [43], as well as potential ethical and/or legal conflicts [50]. Another potential barrier is distrust of the health personnel. One study pointed out that therapists showed greater preference for face-to-face interventions compared to online interventions [38], while another found that some professionals may be reluctant to apply electronic interventions using telephones to treat mental health problems, arguing that it could harm the interactions with the user [36]. However, evidence suggests that the use of electronic interventions with telephones does not change interaction patterns in consultations (duration, alliance, disclosure, empathy, attention, and participation) [36].

Some relevant aspects to consider are clinicians' satisfaction, the lack of training for providers, and ethical challenges. For example, a systematic review of teletherapy for veterans with PTSD found high fidelity to the intervention and good therapist competence, as well high levels of satisfaction among clinicians in terms of their confidence for the delivery of these forms of therapies [49]. However, as mentioned above, proper training is needed for successful delivery [39,49,52], and the ethical and legal aspects should be established [50].

Hypothesis 3

Telepsychiatry for patients with PTSD shows the advantage of diminishing the risk of stigmatization. Since patients are treated from their own homes and are no longer required to visit a psychiatric facility, they feel more motivated to seek mental health care [42]. One systematic review found that patients exhibited more active participation at distance-delivered therapies compared to face-to-face interviews. This may be due to the feeling of "safety" that being at a different location from the therapist could produce. They found that neither empathy, attention, nor participation diminished when using telephone interventions [49]. Additionally, telephonic interventions offer the patient a potentially immediate, anonymous, and easy-to-access option [34]. Other authors pointed out that patients felt that the therapist could understand them better during face-to-face therapies. However, there were no differences in the ability of the therapist to guide the patients to "open themselves" between modalities [36]. It was reported that the efficacy of interventions was similar across modalities

and although the interaction between patient and therapist was lower [39], the therapeutic alliance was able to be achieved without limitations [42,47], except for the difficulties at reading corporal language [49].

Telephone and video call interventions were usually acceptable and efficient for digital mental health [41]. This is probably because more access to care was allowed for children and adults with comorbid psychiatric and complex medical illnesses in various settings, age spans, and demographic characteristics, including rural areas [52]. Although there is greater satisfaction on the users' side (and therefore an improvement in mood state), this does not imply that there are improvements in the quality of life, since recovery (the relief of depressive symptoms) does not necessarily amount to parallel improvements in quality-of-life measures [19]. In addition, it should be considered that these two outcomes do not follow the same recovery rate.

It is also worth noting that during telephone therapies, the patients could develop an awareness of their own emotional and affective changes by listening to their own voice. Moreover, since there is no difference in the measure of how "closely" the therapist could be listening as in usual face-to-face communication, the patients could more easily feel the "connection" with their therapist and enhance disclosure of feelings and emotions [36]. It was found that the use of technology did not influence the therapeutic alliance with their patients [39,47,49]. This could be explained by the fact that, in this context, the therapist's validation is not based on nonverbal communication but rather by their listening capacity, their verbal clarity, their tone of voice, and how the patient experiences it [43]. Indeed, telephone therapy could work better for introverted patients because it provides more anonymity, creating a sense of safety [34,43].

Some aspects to consider include barriers such as awkward silence, concerns about privacy, and constrained communication. Some patients had expressed their privacy concerns. For instance, veterans with PTSD mentioned questions about the confidentiality of the video transmissions and the data they shared during the consultation [49]. In that same review, constrained communication for detecting body language and nonverbal communication by clinicians when conducting teletherapy for veterans with PTSD was reported. However, they could still develop rapport [49]. Finally, during communications where there is no video of the patient, as in telephone therapy, silences during the patients' speech were more challenging to interpret [36].

One review noted that only two studies reported providing ongoing technical support during interventions [39]. In addition, none of the studies included in their review mentioned videoconferencing-specific good practice guidelines, training of facilitators to conduct online psychological interventions, or contingency plans to support remote participants [39]. Moreover, few studies reported on the frequency of technical problems [39].

Gaps: Limitations of Digital Mental Health Reported in Reviews

Lastly, even though technology interventions have proven to be as effective as in-person sessions and have a 2.13-times higher probability of achieving an appointment once a month [52], some limitations should be noted. First, their effectiveness will depend on treatment adherence [40]. Second, there is limited information on whether CBT-based electronic interventions maintain their beneficial effects over time; two systematic reviews did not identify sufficient evidence to support the benefits of this therapy at 3 or 6 months posttreatment for PTSD cases [38,47]. Third, most of these studies did not use randomization and their sample sizes were small; therefore, more research is needed [19,35-37,39,41,44-46,48,49,51]. Finally, most of the available evidence comes from high-resource countries with integrated health systems and larger research budgets [42]. Hence, some results may not be extrapolated to low- or middle-income countries.

Confidence in Cumulative Evidence

An overall analysis of the CERQual assessment showed that the hypotheses presented have low or very low confidence in the evidence (see Multimedia Appendix 5). The main methodological limitations are that the studies come from research with a low or very low confidence level. In terms of coherence, the baseline assumption and hypothesis 1 showed adequate coherence between the different findings, whereas hypotheses 2 and 3 showed moderate concern, since some reviews have heterogeneous results. Finally, all hypotheses showed the adequacy of the data and relevance of the results.

Discussion

Main Findings and Interpretation

Our study developed a framework based on three hypotheses and a baseline assumption to understand/explain the implementation of synchronous digital mental health interventions. From the 21 systematic reviews included, studies showed that synchronous digital mental health interventions provide effective clinical outcomes and are as effective as face-to-face therapies that address mental health conditions [33-36,42,47,49,50]. These digital interventions reach populations such as children, veterans, refugees, and people living in rural areas [50,52], thereby reducing geographical barriers to access. Moreover, since patients are treated from their own homes and are no longer required to visit a psychiatric facility, this can reduce the fear of mental health stigma [39]. Nevertheless, there are few considerations to achieve successful therapy, such as a quiet environment for the patient to receive the session, a stable platform [49], the development or improvement of a software specifically designed for that purpose [42], and the presence of technical support when needed [43]. Some limitations should be noted due to the critically low level of confidence presented in the studies and the fact that most of the available evidence comes from high-resource countries with integrated health systems and larger research budgets [42]. Hence, some results may not be extrapolated to low- or middle-income countries.

Comparison With Other Studies

Implementation science is an emerging and rapidly growing field that has established frameworks, methods, and strategies to improve the adoption and sustainability of interventions within the real world [53]; it has also identified different barriers and facilitators to the implementation of digital mental health interventions [53]. However, strategies specifically designed for implementing digital mental health interventions within the health care system are still limited [53-55].

The implementation of digital mental health interventions allows for overcoming many barriers in health access, such as geographic, human resources, and stigma barriers. These types of interventions allow patients and therapists to remain in their usual, more comfortable, or safer locations. Another advantage is that our framework supports that other mental health providers with lower degrees could deliver digital mental health interventions after appropriate training, which would increase the available human resources pool of therapists [41,43]. In addition, digital mental health interventions could be more attractive than face-to-face therapies, as they present the opportunity to increase privacy and minimize the risk of stigmatization since they can take place outside of mental health institutions, which is especially relevant for populations in which the presence of potential social stigma interferes with the decision to attend mental health facilities [42].

Our study provides hypotheses based on systematic reviews, which allow for obtaining a better understanding for the implementation of synchronous interventions in digital mental health. However, our framework does not provide specific steps or strategies to carry out the implementation process. Therefore, to fill this gap, other researchers could use the ERIC project framework, which presents four general phases for implementing digital interventions in the health system: an implementation strategy exploration phase, preparation phase, implementation phase, and sustainability phase [53,56]. It should be noted that other frameworks that systematize the implementation steps could be used to perform the implementation task, as long as they are adapted to the particularities of the context, health system, resources, and willingness of the actors involved. An alternative that has proven to be useful in favoring the implementation of interventions from heterogeneous contexts is a formative study design that allows for the contextualization of these interventions while evaluating their acceptability, efficiency, and safety within the health system or community [57]. However, this requires greater investment in research by low- and medium-resource countries.

There are currently no frameworks to explain the implementation of digital interventions as the main component in mental health care. Although we have not identified any studies directly comparable to ours, there are related studies. For example, a systematic review of barriers and facilitators to the implementation of electronic mental health interventions identified that the acceptability of electronic interventions depends on (1) patients' and professionals' expectations, (2) preferences about what they would receive and what they provide during care, and (3) the appropriateness of the electronic intervention to address patients' mental health conditions [18]. One study proposed an ethical framework for the development, use, and implementation of digital mental health interventions such as chatbots, based on the principles of beneficence, nonmaleficence, autonomy, justice, and explicability [58]. Although chatbots are not synchronous interventions, they can be used as additional components in synchronous interventions. In the absence of an integrative framework, our study proposes a technical underpinning of available evidence to enable decision-makers to implement electronic interventions to address mental health. We identified different reviews supported by electronic interventions for anxiety, depression, and PTSD, which are equivalent to face-to-face interventions [33-36,42,47,49,50] and are cost-effective in the long term [19].

Despite evidence in favor of digital mental health interventions, there is a considerable difference between the reports from high-income and low-income countries. Some high-income countries had sufficient evidence to conduct country-focused effectiveness evaluations. For example, a systematic review from the United Kingdom identified 7 out of 48 digital interventions promoted by their health system for depression and anxiety as having a small but consistent effect and recommended their use [59]. In addition, the disparity in the amount of evidence remains in economic research, where a systematic review of economic studies identified that internet-based digital interventions for anxiety and depression are cost-effective and recommended their use; however, only studies from high-income countries were identified [60].

In contrast, no reviews of effectiveness, cost-effectiveness, or acceptability of electronic interventions were identified for low- and middle-income countries. The limited evidence from low- and middle-income countries suggests that their health systems made decisions based on minimal local evidence, low-quality evidence (ie, expert review), and evidence from only high-income countries (ie, different contexts). Additionally, material and economic resources and internet access are limited in low- and middle-income countries. Thus, sufficient internet access for health care providers and users should be assured for implementing these technologies. Other problems that could generate inequity, such as limited access to smartphones in rural and low-income areas, low internet speed, and network instability, could generate gaps for adequate implementation of these technologies.

An additional element to highlight, apart from the effectiveness or cost-effectiveness of electronic interventions, is the positive effects they could have on patients' quality of life. Although quality of life was not an outcome in our study, we found evidence that electronic interventions to treat mental health positively affect quality of life [38,40,46]. These results are consistent with other systematic reviews showing that CBT-based interventions (eg, face-to-face, internet, or group) improve participants' quality of life [61,62]. Furthermore, this secondary benefit of electronic mental health interventions on quality of life appeared to affect years of life lost due to disability [63]. This explains why this outcome is key for understanding the cost-effectiveness of this type of intervention since its long-term effect is to reduce costs within the health system [19].

Implementation and Public Health Implications

Decision-makers and researchers could use this relevant information to support the implementation of electronic mental health interventions within their health systems (ie, teleconsultation network). There is evidence to support digital interventions due to their effectiveness in depression, anxiety, and PTSD; their feasibility and acceptability; their safety; and the additional effect on the quality of life of patients [35,37-40,44-46,48,49,51]. The treatment models that have the most empirical support are those based on CBT, which could be the first type of interventions to be implemented. In addition, evidence supports that models of CBT electronic interventions are cost-effective, making their implementation within health systems feasible in the long term.

Health systems must develop legislation and basic technological conditions to achieve the implementation of synchronous digital mental health interventions. First, legislation such as privacy policies, terms of use, and technological requirements of teleconsultation platforms should be established [4]. All of these issues should be covered and regulated by national policies and there should be an entity to enable their regulation. Consequently, health care systems should develop an integrated digital health/digital mental health system that is user-friendly for all literacy levels.

Second, there is a need for quality internet and cell phone services to increase the likelihood of adherence [4,39,42]. Collaboration among public and private sectors is needed in this regard. Technical support and access to therapies should be flexible in terms of schedules, since participants would adjust the delivery to their own timetables. Hence, night schedules should be considered. In addition, training for personnel with minor degrees must be guaranteed in a standardized and systematic way [41,43,52].

Third, for the implementation and use of electronic interventions, it is necessary to identify the barriers within each health system to achieve the acceptance of the different actors. Lack of access to technology (especially in low-resource countries), limited training in teleconsultation or reluctance of health personnel to use the technology, problems related to patient safety or privacy, and limited legislation on teleconsultation at the country level are necessary elements to evaluate during the planning of electronic interventions in mental health [64].

Fourth, the context of the COVID-19 pandemic has enhanced the use of technologies to provide health care and reduce health care access gaps, and decision-makers need to take advantage of this context to enhance the implementation and adoption of these types of interventions [3-7]. It should be noted that digital interventions are not only a short-term solution, as the trend is to incorporate them as a key part of cost-effective health care systems [19,34,42].

Strengths and Limitations

One of the strengths of our study is that we collected information from systematic reviews in a large number of databases, assuring the comprehensiveness of the evidence included. However, our study has limitations. First, the quality of the systematic reviews included was critically low for the most part, which could limit the confidence in the conclusions of the study. Other studies have already reported the low quality of systematic reviews and clinical practice guidelines in mental health [65-67]. Second, the electronic interventions evaluated are very heterogeneous both in the form of delivery (ie, telephone, internet-based, videoconferencing, online) and in the theoretical models used (classical CBT, mindfulness-based CBT, transdiagnostic CBT, nonspecific). Therefore, there may be variations in effect, safety, and acceptability in the way of delivery and the theoretical model used. Third, most of the research has been conducted in high-income countries, and therefore the results may not be comparable in low- and-middle income countries. Fourth, although a realist review analysis was rigorously carried out, the evidence evaluated has methodological limitations, resulting in overall low certainty of the evidence.

Conclusions

Our study assessed all available evidence for the implementation of synchronous digital mental health interventions and developed a framework for the implementation of synchronous digital mental health based on three hypotheses. Since it is known that digital mental health interventions are clinically effective, we hypothesized that those interventions reach otherwise inaccessible populations since they abolish the need of physical presence and mobilization (hypothesis 1) or because a nonspecialist could deliver it with the additional advantage of reducing expenses (hypothesis 2), and that digital interventions are acceptable for those receiving them and maintain the establishment of rapport (hypothesis 3). Each hypothesis represents important outcomes in the implementation process. In addition, we analyzed the barriers and facilitators for these outcomes and identified gaps in the body of evidence that require attention from future researchers.

Our study provides a framework to understand the implementation of synchronous digital mental health interventions, suggests elements to consider at the time of implementation, and establishes gaps. This information will guide decision-makers, researchers, health system managers, and implementation teams.

Acknowledgments

The authors acknowledge Jackeline Garcia, Juan Barrera, and Ana Lindo-Cavero for helping with the initial phases of the information search. JHV received funding for this study from EsSalud (Premio Kaelin 2020). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

DVZ: Conceptualization, methodology, investigation, writing-original draft, writing-review and editing, supervision, project administration, funding acquisition. CAR: Conceptualization, methodology, investigation, writing-review and editing, supervision, funding acquisition. GMT: Conceptualization, methodology, investigation, writing-review and editing, Supervision. RTP: Investigation, writing-original draft, writing-review and editing. ANF: Investigation, writing-review and editing. VC: Methodology, formal analysis, writing-review and editing. JAM: Investigation, writing-original draft, writing-review and editing. JRV: Formal analysis, writing-review and editing. GA: Formal analysis, writing-review and editing. LAF: Investigation, writing-review and editing. ARC: Investigation, writing-review and editing. JHV: Conceptualization, methodology, investigation, writing-review and editing, supervision, funding acquisition.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA checklist.
[DOCX File , 32 KB - mental_v9i3e34760_app1.docx]

Multimedia Appendix 2
Search strategy.
[DOCX File , 34 KB - mental_v9i3e34760_app2.docx]

Multimedia Appendix 3
Reasons for articles excluded.
[DOCX File , 48 KB - mental_v9i3e34760_app3.docx]

Multimedia Appendix 4
Characteristics of the included studies.
[DOCX File , 61 KB - mental_v9i3e34760_app4.docx]

Multimedia Appendix 5
Confidence in cumulative evidence assessed with the Confidence in Evidence from Reviews of Qualitative Research (CERQual) approach.
[DOCX File , 15 KB - mental_v9i3e34760_app5.docx]

References

1. Vindegaard N, Benros ME. COVID-19 pandemic and mental health consequences: systematic review of the current evidence. Brain Behav Immun 2020 Oct;89:531-542 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.048] [Medline: 32485289]
2. Talevi D, Socci V, Carai M, Carnaghi G, Faleri S, Trebbi E, et al. Mental health outcomes of the CoViD-19 pandemic. Riv Psichiatr 2020;55(3):137-144. [doi: 10.1708/3382.33569] [Medline: 32489190]
3. Sullivan AB, Kane A, Roth AJ, Davis BE, Drerup ML, Heinberg LJ. The COVID-19 crisis: a mental health perspective and response using telemedicine. J Patient Exp 2020 Jun;7(3):295-301 [FREE Full text] [doi: 10.1177/2374373520922747] [Medline: 32821785]
4. Torous J, Bucci S, Bell I, Kessing L, Faurholt-Jepsen M, Whelan P, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021 Oct;20(3):318-335. [doi: 10.1002/wps.20883] [Medline: 34505369]
5. Haxhihamza K, Arsova S, Bajraktarov S, Kalpak G, Stefanovski B, Novotni A, et al. Patient satisfaction with use of telemedicine in University Clinic of Psychiatry: Skopje, North Macedonia during COVID-19 pandemic. Telemed J E Health 2021 Apr 01;27(4):464-467. [doi: 10.1089/tmj.2020.0256] [Medline: 32809916]
6. Ruiz EF, Proaño Á, Ponce OJ, Curioso WH. Tecnologías móviles para la salud pública en el Perú: lecciones aprendidas. Rev Peru Med Exp Salud Publica 2015 Jun 19;32(2):364. [doi: 10.17843/rpmesp.2015.322.1634]
7. Kaonga N, Morgan J. Common themes and emerging trends for the use of technology to support mental health and psychosocial well-being in limited resource settings: A review of the literature. Psychiatry Res 2019 Nov;281:112594. [doi: 10.1016/j.psychres.2019.112594] [Medline: 31605874]
8. Miranda JJ, Moscoso MG, Toyama M, Cavero V, Diez-Canseco F, Ovbiagele B. Role of mHealth in overcoming the occurrence of post-stroke depression. Acta Neurol Scand 2018 Jan 13;137(1):12-19 [FREE Full text] [doi: 10.1111/ane.12832] [Medline: 28901543]

9. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010 Oct 13;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
10. Turvey C, Fortney J. The use of telemedicine and mobile technology to promote population health and population management for Psychiatric Disorders. Curr Psychiatry Rep 2017 Oct 16;19(11):88. [doi: 10.1007/s11920-017-0844-0] [Medline: 29035422]
11. Meurk C, Leung J, Hall W, Head BW, Whiteford H. Establishing and governing e-Mental health care in Australia: a systematic review of challenges and a call for policy-focussed research. J Med Internet Res 2016 Jan 13;18(1):e10 [FREE Full text] [doi: 10.2196/jmir.4827] [Medline: 26764181]
12. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
13. Luo J, Tong L, Crotty BH, Somai M, Taylor B, Osinski K, et al. Telemedicine adoption during the COVID-19 pandemic: gaps and inequalities. Appl Clin Inform 2021 Aug;12(4):836-844. [doi: 10.1055/s-0041-1733848] [Medline: 34496419]
14. Nilsen P. Making sense of implementation theories, models and frameworks. Implement Sci 2015 Apr 21;10:53 [FREE Full text] [doi: 10.1186/s13012-015-0242-0] [Medline: 25895742]
15. Tabak RG, Khoong EC, Chambers DA, Brownson RC. Bridging research and practice: models for dissemination and implementation research. Am J Prev Med 2012 Sep;43(3):337-350 [FREE Full text] [doi: 10.1016/j.amepre.2012.05.024] [Medline: 22898128]
16. Griffiths KM, Farrer L, Christensen H. The efficacy of internet interventions for depression and anxiety disorders: a review of randomised controlled trials. Med J Aust 2010 Jun 07;192(S11):S4-S11. [doi: 10.5694/j.1326-5377.2010.tb03685.x] [Medline: 20528707]
17. Senanayake B, Wickramasinghe SI, Chatfield MD, Hansen J, Edirippulige S, Smith AC. Effectiveness of text messaging interventions for the management of depression: A systematic review and meta-analysis. J Telemed Telecare 2019 Oct 20;25(9):513-523. [doi: 10.1177/1357633x19875852]
18. Vis C, Mol M, Kleiboer A, Bührmann L, Finch T, Smit J, et al. Improving implementation of eMental health for mood disorders in routine practice: systematic review of barriers and facilitating factors. JMIR Ment Health 2018 Mar 16;5(1):e20 [FREE Full text] [doi: 10.2196/mental.9769] [Medline: 29549072]
19. Ahern E, Kinsella S, Semkovska M. Clinical efficacy and economic evaluation of online cognitive behavioral therapy for major depressive disorder: a systematic review and meta-analysis. Expert Rev Pharmacoecon Outcomes Res 2018 Feb;18(1):25-41. [doi: 10.1080/14737167.2018.1407245] [Medline: 29145746]
20. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
21. Villarreal-Zegarra D, Alarcon-Ruiz CA, Melendez-Torres GJ, Torres-Puente R, Ambrosio-Melgarejo J, Romero-Cabrera AB, et al. Development of a framework for the implementation of synchronous e-mental health: a protocol for a realist synthesis of systematic reviews. F1000Res 2020;9:1282 [FREE Full text] [doi: 10.12688/f1000research.27150.2] [Medline: 34540205]
22. Cooke A, Smith D, Booth A. Beyond PICO: the SPIDER tool for qualitative evidence synthesis. Qual Health Res 2012 Oct 24;22(10):1435-1443. [doi: 10.1177/1049732312452938] [Medline: 22829486]
23. Centre for Reviews and Dissemination. Systematic Reviews: CRD's guidance for undertaking reviews in health care. York: CRD, University of York; 2009. URL: https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf [accessed 2021-10-01]
24. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
25. Rycroft-Malone J, McCormack B, Hutchinson AM, DeCorby K, Bucknall TK, Kent B, et al. Realist synthesis: illustrating the method for implementation research. Implement Sci 2012 Apr 19;7(1):33 [FREE Full text] [doi: 10.1186/1748-5908-7-33] [Medline: 22515663]
26. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 2017 Sep 21;358:j4008 [FREE Full text] [doi: 10.1136/bmj.j4008] [Medline: 28935701]
27. Guetterman TC, Babchuk WA, Howell Smith MC, Stevens J. Contemporary approaches to mixed methods–grounded theory research: a field-based analysis. J Mixed Methods Res 2017 Jun 01;13(2):179-195. [doi: 10.1177/1558689817710877]
28. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008 Jul 10;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
29. Thorne S, Jensen L, Kearney MH, Noblit G, Sandelowski M. Qualitative metasynthesis: reflections on methodological orientation and ideological agenda. Qual Health Res 2004 Dec;14(10):1342-1365. [doi: 10.1177/1049732304269888] [Medline: 15538004]

30. Saunders B, Sim J, Kingstone T, Baker S, Waterfield J, Bartlam B, et al. Saturation in qualitative research: exploring its conceptualization and operationalization. Qual Quant 2018;52(4):1893-1907 [FREE Full text] [doi: 10.1007/s11135-017-0574-8] [Medline: 29937585]
31. The GRADE-CERQual Project Group. A practical guide for decision makers in health and social care and other users. GRADE-CERQual. 2018 Mar. URL: https://drive.google.com/file/d/1tF0rhDmihpzd3B6BDFHcD8msDu1rqYjR/view [accessed 2021-10-01]
32. Lewin S, Booth A, Glenton C, Munthe-Kaas H, Rashidian A, Wainwright M, et al. Applying GRADE-CERQual to qualitative evidence synthesis findings: introduction to the series. Implement Sci 2018 Jan 25;13(Suppl 1):2 [FREE Full text] [doi: 10.1186/s13012-017-0688-3] [Medline: 29384079]
33. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
34. Castro A, Gili M, Ricci-Cabello I, Roca M, Gilbody S, Perez-Ara M, et al. Effectiveness and adherence of telephone-administered psychotherapy for depression: a systematic review and meta-analysis. J Affect Disord 2020 Jan 01;260:514-526. [doi: 10.1016/j.jad.2019.09.023] [Medline: 31539688]
35. Cuijpers P, Noma H, Karyotaki E, Cipriani A, Furukawa TA. Effectiveness and acceptability of cognitive behavior therapy delivery formats in adults with depression: a network meta-analysis. JAMA Psychiatry 2019 Jul 01;76(7):700-707 [FREE Full text] [doi: 10.1001/jamapsychiatry.2019.0268] [Medline: 30994877]
36. Irvine A, Drew P, Bower P, Brooks H, Gellatly J, Armitage C, et al. Are there interactional differences between telephone and face-to-face psychological therapy? A systematic review of comparative studies. J Affect Disord 2020 Mar 15;265:120-131 [FREE Full text] [doi: 10.1016/j.jad.2020.01.057] [Medline: 32090733]
37. Kampmann IL, Emmelkamp PMG, Morina N. Meta-analysis of technology-assisted interventions for social anxiety disorder. J Anxiety Disord 2016 Aug;42:71-84. [doi: 10.1016/j.janxdis.2016.06.007] [Medline: 27376634]
38. Lewis C, Roberts NP, Simon N, Bethell A, Bisson JI. Internet-delivered cognitive behavioural therapy for post-traumatic stress disorder: systematic review and meta-analysis. Acta Psychiatr Scand 2019 Dec;140(6):508-521. [doi: 10.1111/acps.13079] [Medline: 31359407]
39. Moulton-Perkins A, Moulton D, Cavanagh K, Jozavi A, Strauss C. Systematic review of mindfulness-based cognitive therapy and mindfulness-based stress reduction via group videoconferencing: Feasibility, acceptability, safety, and efficacy. J Psychother Integr 2020 Sep 14:online ahead of print. [doi: 10.1037/int0000216]
40. Păsărelu CR, Andersson G, Bergman Nordgren L, Dobrean A. Internet-delivered transdiagnostic and tailored cognitive behavioral therapy for anxiety and depression: a systematic review and meta-analysis of randomized controlled trials. Cogn Behav Ther 2017 Jan;46(1):1-28. [doi: 10.1080/16506073.2016.1231219] [Medline: 27712544]
41. Rees CS, Maclaine E. A systematic review of videoconference ‐ delivered psychological treatment for anxiety disorders. Austral Psychol 2020 Nov 12;50(4):259-264. [doi: 10.1111/ap.12122]
42. Sunjaya AP, Chris A, Novianti D. Efficacy, patient-doctor relationship, costs and benefits of utilizing telepsychiatry for the management of post-traumatic stress disorder (PTSD): a systematic review. Trends Psychiatry Psychother 2020 Mar;42(1):102-110. [doi: 10.1590/2237-6089-2019-0024] [Medline: 32321088]
43. Domhardt M, Geßlein H, von Rezori RE, Baumeister H. Internet- and mobile-based interventions for anxiety disorders: A meta-analytic review of intervention components. Depress Anxiety 2019 Mar;36(3):213-224. [doi: 10.1002/da.22860] [Medline: 30450811]
44. Josephine K, Josefine L, Philipp D, David E, Harald B. Internet- and mobile-based depression interventions for people with diagnosed depression: a systematic review and meta-analysis. J Affect Disord 2017 Dec 01;223:28-40. [doi: 10.1016/j.jad.2017.07.021] [Medline: 28715726]
45. Linde K, Sigterman K, Kriston L, Rücker G, Jamil S, Meissner K, et al. Effectiveness of psychological treatments for depressive disorders in primary care: systematic review and meta-analysis. Ann Fam Med 2015;13(1):56-68 [FREE Full text] [doi: 10.1370/afm.1719] [Medline: 25583894]
46. Olthuis JV, Watt MC, Bailey K, Hayden JA, Stewart SH. Therapist-supported Internet cognitive behavioural therapy for anxiety disorders in adults. Cochrane Database Syst Rev 2015 Mar 05(3):CD011565. [doi: 10.1002/14651858.CD011565] [Medline: 25742186]
47. Olthuis JV, Wozney L, Asmundson GJG, Cramm H, Lingley-Pottie P, McGrath PJ. Distance-delivered interventions for PTSD: a systematic review and meta-analysis. J Anxiety Disord 2016 Dec;44:9-26. [doi: 10.1016/j.janxdis.2016.09.010] [Medline: 27697658]
48. Proctor BJ, Moghaddam N, Vogt W, das Nair R. Telephone psychotherapy in multiple sclerosis: a systematic review and meta-analysis. Rehabil Psychol 2018 Feb;63(1):16-28. [doi: 10.1037/rep0000182] [Medline: 29553780]
49. Turgoose D, Ashwick R, Murphy D. Systematic review of lessons learned from delivering tele-therapy to veterans with post-traumatic stress disorder. J Telemed Telecare 2018 Oct;24(9):575-585. [doi: 10.1177/1357633X17730443] [Medline: 28958211]
50. Drago A, Winding T, Antypa N. Videoconferencing in psychiatry, a meta-analysis of assessment and treatment. Eur Psychiatry 2016 Aug 23;36:29-37. [doi: 10.1016/j.eurpsy.2016.03.007] [Medline: 27311105]

51. Coughtrey AE, Pistrang N. The effectiveness of telephone-delivered psychological therapies for depression and anxiety: a systematic review. J Telemed Telecare 2016 Dec 30;24(2):65-74. [doi: 10.1177/1357633x16686547]
52. Finley BA, Shea KD, Maixner R, Slebodnik M. Advanced practice registered nurses using synchronous telepsychiatry: an integrative systematic review. J Am Psychiatr Nurses Assoc 2021;27(4):271-282. [doi: 10.1177/1078390320939159] [Medline: 32648509]
53. Graham AK, Lattie EG, Powell BJ, Lyon AR, Smith JD, Schueller SM, et al. Implementation strategies for digital mental health interventions in health care settings. Am Psychol 2020 Nov;75(8):1080-1092 [FREE Full text] [doi: 10.1037/amp0000686] [Medline: 33252946]
54. Anton MT, Jones DJ. Adoption of technology-enhanced treatments: conceptual and practical considerations. Clin Psychol 2017 Sep;24(3):223-240 [FREE Full text] [doi: 10.1111/cpsp.12197] [Medline: 28966479]
55. LaMonica HM, Davenport TA, Braunstein K, Ottavio A, Piper S, Martin C, et al. Technology-enabled person-centered mental health services reform: strategy for implementation science. JMIR Ment Health 2019 Sep 19;6(9):e14719 [FREE Full text] [doi: 10.2196/14719] [Medline: 31538938]
56. Powell BJ, Waltz TJ, Chinman MJ, Damschroder LJ, Smith JL, Matthieu MM, et al. A refined compilation of implementation strategies: results from the Expert Recommendations for Implementing Change (ERIC) project. Implement Sci 2015 Feb 12;10:21 [FREE Full text] [doi: 10.1186/s13012-015-0209-1] [Medline: 25889199]
57. Peters D, Tran N, Adam T. Implementation research in health: a practical guide. Geneva: World Health Organization; 2013.
58. Vilaza GN, McCashin D. Is the automation of digital mental health ethical? Applying an ethical framework to chatbots for cognitive behaviour therapy. Front Digit Health 2021;3:689736 [FREE Full text] [doi: 10.3389/fdgth.2021.689736] [Medline: 34713163]
59. Simmonds-Buckley M, Bennion MR, Kellett S, Millings A, Hardy GE, Moore RK. Acceptability and effectiveness of NHS-recommended e-Therapies for depression, anxiety, and stress: meta-analysis. J Med Internet Res 2020 Oct 28;22(10):e17049 [FREE Full text] [doi: 10.2196/17049] [Medline: 33112238]
60. Mitchell LM, Joshi U, Patel V, Lu C, Naslund JA. Economic evaluations of internet-based psychological interventions for anxiety disorders and depression: a systematic review. J Affect Disord 2021 Apr 01;284:157-182. [doi: 10.1016/j.jad.2021.01.092] [Medline: 33601245]
61. Getu MA, Chen C, Panpan W, Mboineki JF, Dhakal K, Du R. The effect of cognitive behavioral therapy on the quality of life of breast cancer patients: a systematic review and meta-analysis of randomized controlled trials. Qual Life Res 2021 Feb;30(2):367-384. [doi: 10.1007/s11136-020-02665-5] [Medline: 33068239]
62. Hajihasani A, Rouhani M, Salavati M, Hedayati R, Kahlaee AH. The influence of cognitive behavioral therapy on pain, quality of life, and depression in patients receiving physical therapy for chronic low back pain: a systematic review. PM R 2019 Feb;11(2):167-176. [doi: 10.1016/j.pmrj.2018.09.029] [Medline: 30266349]
63. Ayuso-Mateos JL, Salvador-Carulla L, Chisholm D. Use of quality of life measures in mental health economics and care planning. Actas Esp Psiquiatr 2006;34(1):1-6. [Medline: 16525899]
64. Cowan KE, McKean AJ, Gentry MT, Hilty DM. Barriers to use of telepsychiatry: clinicians as gatekeepers. Mayo Clin Proc 2019 Dec;94(12):2510-2523. [doi: 10.1016/j.mayocp.2019.04.018] [Medline: 31806104]
65. Zafra-Tanaka JH, Goicochea-Lugo S, Villarreal-Zegarra D, Taype-Rondan A. Characteristics and quality of clinical practice guidelines for depression in adults: a scoping review. BMC Psychiatry 2019 Feb 20;19(1):76 [FREE Full text] [doi: 10.1186/s12888-019-2057-z] [Medline: 30786870]
66. Banks K, Newman E, Saleem J. An overview of the research on mindfulness-based interventions for treating symptoms of posttraumatic stress disorder: a systematic review. J Clin Psychol 2015 Oct;71(10):935-963. [doi: 10.1002/jclp.22200] [Medline: 26192755]
67. Matthias K, Rissling O, Pieper D, Morche J, Nocon M, Jacobs A, et al. The methodological quality of systematic reviews on the treatment of adult major depression needs improvement according to AMSTAR 2: A cross-sectional study. Heliyon 2020 Sep;6(9):e04776 [FREE Full text] [doi: 10.1016/j.heliyon.2020.e04776] [Medline: 32939412]

Abbreviations

AMSTAR-2: A Measurement Tool to Assess Systematic Reviews-2
CBT: cognitive behavioral therapy
CERQual: Confidence in Evidence from Reviews of Qualitative Research
CFIR: Consolidated Framework for Implementation Research
ERIC: Expert Recommendations for Implementing Change
PARIHS: Promoting Action on Research Implementation in Health Services
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PTSD: posttraumatic stress disorder
RCT: randomized controlled trial
SPIDER: Sample, Phenomenon of Interest, Design, Evaluation, and Research type

Edited by J Torous; submitted 06.11.21; peer-reviewed by H Mehdizadeh, I Wilson; comments to author 30.12.21; revised version received 18.01.22; accepted 20.01.22; published 29.03.22.

Please cite as:
Villarreal-Zegarra D, Alarcon-Ruiz CA, Melendez-Torres GJ, Torres-Puente R, Navarro-Flores A, Cavero V, Ambrosio-Melgarejo J, Rojas-Vargas J, Almeida G, Albitres-Flores L, Romero-Cabrera AB, Huarcaya-Victoria J
Development of a Framework for the Implementation of Synchronous Digital Mental Health: Realist Synthesis of Systematic Reviews
JMIR Ment Health 2022;9(3):e34760
URL: https://mental.jmir.org/2022/3/e34760
doi:10.2196/34760
PMID:35348469

Review

The Current State and Validity of Digital Assessment Tools for Psychiatry: Systematic Review

Nayra A Martin-Key[1*], PhD; Benedetta Spadaro[1*], MPhil; Erin Funnell[1], BSc; Eleanor Jane Barker[2], MSc; Thea Sofie Schei[3], PhD; Jakub Tomasik[1], PhD; Sabine Bahn[1,3], MD, PhD, MRCPsych

[1]Cambridge Centre for Neuropsychiatric Research, Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge, United Kingdom

[2]University of Cambridge Medical Library, University of Cambridge, Cambridge, United Kingdom

[3]Psyomics Ltd, Cambridge, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Sabine Bahn, MD, PhD, MRCPsych
Cambridge Centre for Neuropsychiatric Research
Department of Chemical Engineering and Biotechnology
University of Cambridge
Philippa Fawcett Drive
Cambridge, CB3 0AS
United Kingdom
Phone: 44 1223 334151
Email: sb209@cam.ac.uk

Abstract

Background: Given the role digital technologies are likely to play in the future of mental health care, there is a need for a comprehensive appraisal of the current state and validity (ie, screening or diagnostic accuracy) of digital mental health assessments.

Objective: The aim of this review is to explore the current state and validity of question-and-answer–based digital tools for diagnosing and screening psychiatric conditions in adults.

Methods: This systematic review was based on the Population, Intervention, Comparison, and Outcome framework and was carried out in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. MEDLINE, Embase, Cochrane Library, ASSIA, Web of Science Core Collection, CINAHL, and PsycINFO were systematically searched for articles published between 2005 and 2021. A descriptive evaluation of the study characteristics and digital solutions and a quantitative appraisal of the screening or diagnostic accuracy of the included tools were conducted. Risk of bias and applicability were assessed using the revised tool for the Quality Assessment of Diagnostic Accuracy Studies 2.

Results: A total of 28 studies met the inclusion criteria, with the most frequently evaluated conditions encompassing generalized anxiety disorder, major depressive disorder, and any depressive disorder. Most of the studies used digitized versions of existing pen-and-paper questionnaires, with findings revealing poor to excellent screening or diagnostic accuracy (sensitivity=0.32-1.00, specificity=0.37-1.00, area under the receiver operating characteristic curve=0.57-0.98) and a high risk of bias for most of the included studies.

Conclusions: The field of digital mental health tools is in its early stages, and high-quality evidence is lacking.

International Registered Report Identifier (IRRID): RR2-10.2196/25382

(JMIR Ment Health 2022;9(3):e32824) doi:10.2196/32824

KEYWORDS

diagnostic accuracy; digital mental health; digital questionnaire; psychiatry; systematic review

Introduction

Background

Mental health disorders are highly prevalent [1] and represent the main source of health-related economic burden worldwide [2-4], with barriers to ensuring adequate mental health care provision being complex and multifaceted. For instance, in addition to the lack of available mental health care professionals worldwide [5], short primary care consultation times coupled with the complexity and subjectivity of diagnosing mental health disorders mean that many patients are not receiving adequate support. Furthermore, attitudinal factors, including a low perceived treatment need and a fear of stigmatization, contribute significantly to non–help-seeking behavior [6]. Moving forward, there is a need for innovative, cost-effective, and highly scalable solutions for the assessment, diagnosis, and management of mental health disorders.

To this end, digital technologies for psychiatry may offer attractive *add-ons* or alternatives to conventional mental health care services. Clinical decision support tools may range from simple digitized versions of existing pen-and-paper mental health screening instruments to more sophisticated question-and-answer–based digital solutions for psychiatry such as adaptive questionnaires. Given the ubiquitous nature of technology, these tools can be used on patients' personal devices, such as via a website, thereby offering private and convenient mental health care provision from the comfort of one's home.

Critically, although there exists a plethora of research evaluating digital psychotherapeutic technologies such as internet-delivered cognitive behavioral therapy [7,8], to our knowledge, little effort has been put into evaluating diagnostic decision support technologies. The limited number of studies on diagnostic and screening tools for mental health have mainly focused on establishing the psychometric properties of digitized versions of existing pen-and-paper questionnaires (see van Ballegooijen et al [9] for a systematic review) and have often compared these tools to existing scales such as the 9-item Patient Health Questionnaire (PHQ–9) [10] as opposed to a *gold standard* assessment by a psychiatrist or a diagnostic interview based on the Diagnostic and Statistical Manual of Mental Disorders (DSM; now in its fifth edition [DSM–5]) [11] or the International Statistical Classification of Diseases and Related Health Problems (ICD; now in its 11th edition [ICD–11]) [12,13]. In fact, despite the rapidly growing number of digital assessment tools for screening and diagnosing mental health disorders, little is known about their accuracy.

Objectives

To this end, the key objectives of this systematic review are to summarize available digital mental health assessment tools as well as evaluate their accuracy among studies using a *gold standard* reference test. We will first examine the types of available digital mental health assessment tools (eg, digitized versions of existing psychiatric pen-and-paper questionnaires vs more sophisticated digital tools). Second, we will evaluate the screening or diagnostic accuracy of the identified digital mental health assessment tools for each mental health condition of interest. Finally, we will assess the risk of bias and applicability of all the included studies. Given the rapid pace of technological development and the role digital technologies are likely to play in the future of mental health care, this comprehensive systematic review is timely and has important implications for clinical practice and the development of digital solutions for psychiatry.

Methods

Database Search

The methods are described in detail in a previously published protocol [14], which has been registered with the International Prospective Register of Systematic Reviews (PROSPERO CRD42020214724). The search strategy was developed using the Population, Intervention, Comparison, and Outcome framework and performed following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses [15]) guidelines. Keywords and subject headings were extracted from a preliminary scan of the literature and the DSM–5 and ICD–11 (or DSM–IV and ICD–10 for older publications) diagnostic manuals and were decided in consultation with a medical librarian (EJB) and a practicing psychiatrist (SB). The following electronic databases were searched: MEDLINE, Embase, Cochrane Library, ASSIA, Web of Science Core Collection, CINAHL, and PsycINFO. Search terms were grouped into four themes and combined using the following structure: "digital technology" *AND* "assessment tool" *AND* "mental health" *AND* "accuracy." The search was completed on October 12, 2021. Gray literature (eg, clinical trial databases, unpublished theses, reports, and conference presentations) was identified by hand searching. Other potentially eligible publications were identified by hand searching the reference lists of relevant systematic reviews and meta-analyses. Hand searching was completed on October 21, 2021. A complete list of the search strategies, including keywords and subject headings, can be found in Multimedia Appendix 1.

Inclusion and Exclusion Criteria

Owing to ongoing developments in the digitization of existing psychiatric questionnaires and the rapid growth in digital assessment tools for the screening and diagnosing of mental health conditions, the initial search was limited to studies published between January 1, 2005, and October 12, 2021, with hand searching completed by October 21, 2021. Studies published in any language were included. The study design was not limited to ensure that no relevant studies were missed.

The population included adults with a mean age of 18 to 65 years who had been assessed for the presence of any of the following mental health conditions: bipolar disorder (BD), major depressive disorder (MDD), anxiety disorders, obsessive-compulsive disorder (OCD), insomnia, schizophrenia, attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorders, eating disorders, personality disorders, alcohol use disorder (AUD), substance use disorder (SUD), posttraumatic stress disorder (PTSD), acute stress disorder, and adjustment disorder. In addition to these conditions, notable symptom domains such as self-harm, suicidality, and psychosis were included based on their relevance in psychiatric

assessments. The population included any gender, severity of mental health concern, ethnicity, and geographical location.

As the review focused on the screening or diagnostic accuracy of digital mental health assessments for use in the primary care or general and psychiatric populations, specific subgroups such as pregnant individuals, refugee or asylum seekers, prisoners, and those in acute crisis or admitted to emergency services were excluded. In consultation with a practicing psychiatrist (SB), we also excluded studies on somatoform disorders and specific phobias as these are less frequently diagnosed in primary care and rarely present in secondary care. Studies on tools used to identify neuropsychiatric disorders (eg, dementias) or any disorders that are due to clinically confirmed temporary or permanent dysfunction of the brain were outside the scope of the review. In addition, studies on tools used to identify mental health disorders in physical illnesses (eg, cancer) were excluded.

The interventions targeted in this review included question-and-answer–based digital mental health screening or diagnostic tools completed by the patient. Studies of digital assessment tools that were not *exclusively* question-and-answer–based, such as blood tests, imaging techniques, monitoring tools, genome analyses, accelerometer devices, and wearables, were excluded. Furthermore, studies on digital assessment tools used to predict future risk of developing a mental health disorder were also excluded, except in the case of suicidality.

Only studies that evaluated the accuracy of a digital mental health assessment tool against a *gold standard* reference test, such as an assessment by a psychiatrist or a standardized structured or semistructured interview based on the DSM–5 and ICD–11 criteria (or DSM–IV and ICD–10 for older publications), were included. Studies that did not include an outcome measure of accuracy (eg, sensitivity and specificity or area under the receiver operating characteristic curve [AUC]) were not included.

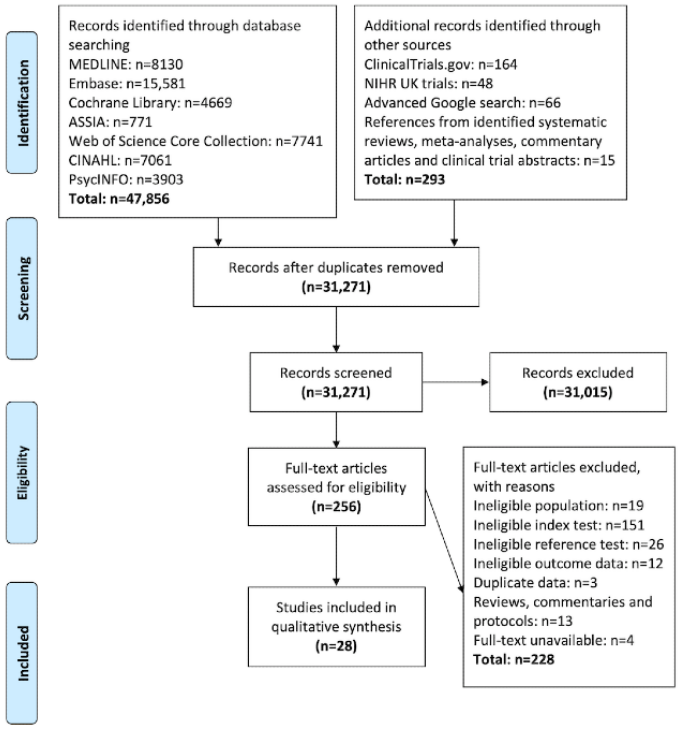
Outcomes Measured

The primary outcome was to examine the current state of digital mental health assessment tools, including the type of tools being used (eg, digitized versions of existing psychiatric pen-and-paper questionnaires) and targeted conditions. The secondary outcome was the validity (ie, screening or diagnostic accuracy) of the identified digital mental health assessment tools.

Screening and Study Selection

Articles identified from the database searches were first stored in the reference management software package EndNote (Clarivate Analytics), which was used to eliminate any duplicates. Once duplicates had been eliminated, all identified articles were transferred to the systematic review software Rayyan (Rayyan Systems Inc). In total, 2 independent reviewers (BS and EF) screened the titles and abstracts of all the studies. Any disagreements were discussed and resolved with a third reviewer (NAM-K). Full texts were then retrieved for the included studies and subsequently assessed for relevance against the eligibility criteria by the 2 independent reviewers. In addition, the full texts of any studies that did not specify in the title or abstract whether the tools used were digital or pen-and-paper versions were examined by the 2 independent reviewers. Once again, any disagreements were discussed and resolved with the third reviewer. Reasons for inclusion and exclusion were recorded at the full-text screening stage and are shown in Figure 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of included studies. NIHR: National Institute for Health Research.



Study Characteristics

A descriptive evaluation of the study characteristics, including conditions of interest, sample type and size, proportion of women, mean age, and country, was extracted by the 2 independent reviewers and summarized.

Digital Mental Health Assessments and Their Validity Per Condition

Information regarding the digital mental health assessments (ie, index tests), including the type and number of questions, reference tests, time flow, and blinding, was extracted by the 2 independent reviewers and summarized. In addition, a descriptive appraisal of the screening or diagnostic accuracy of the included digital mental health assessment tools separated by condition of interest was conducted. The following values were extracted or calculated based on the available data for each digital tool separated by condition of interest:

- Sensitivity: the capacity of the digital tool to correctly classify those with the condition
- Specificity: the capacity of the digital tool to correctly classify those without the condition
- Youden index: a single statistic that measures the performance of a dichotomous diagnostic test at a given cutoff and can be used for maximizing sensitivity and specificity, with scores ranging from 0 (not useful) to 1 (perfect)
- AUC: shows the degree of separability between 2 conditions and represents the probability that a randomly selected individual with the condition is rated or ranked as more likely to have the condition than a randomly selected individual without the condition (≥0.9=excellent, ≥0.8=good, ≥0.7=fair, ≥0.6=poor, ≥0.5=fail [16])

Given the wide range of digital mental health assessment tools and cutoffs used and the differences in methodology and patient populations, as well as the lack of available raw data (after having contacted the authors for further details), a meta-analysis was not deemed clinically informative at this stage.

Risk of Bias and Applicability Assessment

The 2 independent reviewers assessed the risk of bias and applicability of all the included studies using the revised tool for the Quality Assessment of Diagnostic Accuracy Studies 2 (QUADAS–2 [17]), which is recommended for use in systematic reviews of diagnostic accuracy by the United Kingdom National Institute for Health and Clinical Excellence and the Agency for Healthcare Research and Quality, Cochrane Collaboration [18]. Any disagreements were discussed and resolved with the third reviewer. The developers of the QUADAS–2 tool recommend that the tool be tailored for each specific review by adding or omitting signaling questions, which are included to assist in judgments about risk of bias. To this end, the following question was omitted: *Did all patients receive a reference standard?* The reason for removing this question was based on the fact that screening and diagnostic test accuracy studies in the field of mental health ordinarily provide the reference standard to a subset of the original sample, primarily because of missing data by study design or clinical practice [19]. It was agreed that this question was overly conservative for this review. In light of this amendment, we rephrased the following question—*Were all patients included in the analysis?*—to *Did the data analysis only include patients who received both the index test and the reference standard?*

Results

Included Studies

In total, 31,271 articles were retrieved, of which 256 (0.82%) were selected for full-text review. Of these 256 articles, 28 (10.9%) were identified for inclusion. The reasons for exclusion at the full-text review stage are outlined in Figure 1.

Study Characteristics

The characteristics of the 28 included studies are summarized in Table 1 (refer to Multimedia Appendix 2 [20-47] for a checklist summary of the mental health disorders investigated in the included studies). Notably, a large proportion of studies did not meet the inclusion criteria. This was primarily due to the studies not using a *digital* index test or appropriate reference test (ie, an assessment by a psychiatrist or a diagnostic interview based on the DSM or ICD criteria). Other exclusions regarded studies focusing on ineligible populations (eg, children or specific subgroups such as pregnant individuals, refugee or asylum seekers, prisoners, and those in acute crisis or admitted to emergency services) as well as studies that did not include an outcome measure of accuracy (eg, sensitivity and specificity or AUC).

Most of the studies included participants from primary care services or the general population (18/28, 64% [20,22-25,28,32,35,37-45,47]). This was followed by the inclusion of participants from secondary care or specialist services, including psychiatric outpatients (12/28, 43% [20,27,29-31,33-35,38,45-47]). Of the 28 studies, 6 (21%) included university students [21,23-26,36], whereas 4 (14%) purposely recruited nonpsychiatric controls [29-31,33].

Sample sizes ranged from 100 [44] to 6361 [45], with all but 3 studies [26,27,33] including a larger proportion of women. The mean age across studies ranged from 20 [26] to 53 years [44], although not all studies provided this information. Most of the included studies were conducted in the United States (12/28, 43% [20,27-34,37,43,44]). Of the 28 studies, 6 (21%) were conducted in the Netherlands [23-25,38,45,46], and 4 (14%) took place in Spain [21,22,39,42]. The remaining 6 studies (6/28, 21%) were conducted in Australia (1/28, 4%) [40], China (1/28, 4%) [26], Denmark (1/28, 4%) [41], South Korea (2/28, 7%) [35,47], and Thailand (1/28, 4%) [36].

Table 1. Characteristics of the included studies, including conditions of interest, sample type and size, proportion of women, mean age, and country.

Study	Conditions	Occurrence of conditions	Sample				Sample size, N	Women, n	Age (years)	Country
			Primary care or general population	Secondary care	Nonpsychiatric controls	University students				
Achtyes et al [20][a]	MDD[b]	Current and lifetime	✓	✓			145	79	__[c]	United States
Ballester et al [21]	Any mood disorder[d], any anxiety disorder[e], any depressive disorder[f], panic disorder, GAD[g]	Current and lifetime				✓	Total: 575; interviewed: 287	Total: 55	—	Spain
Cano-Vindel et al [22]	MDD, GAD	Current	✓				Total: 1052; interviewed: 178	Total: 77; interviewed: 70	—	Spain
Donker et al [23]	Any depressive disorder[h], GAD, social phobia, panic disorder, agoraphobia, OCD[i], PTSD[j], AUD[k]	Current	✓			✓	Total: 502; interviewed: 157	Total: 57	Total: mean 43 (SD 13)	Netherlands
Donker et al [24]	Any depressive disorder[h]	Current	✓			✓	Total: 502; interviewed: 157	Total: 57	Total: mean 43 (SD 13)	Netherlands
Donker et al [25]	Any depressive disorder[h], any anxiety disorder[l], GAD, panic disorder, social phobia, PTSD	Current	✓			✓	Total: 502; interviewed: 157	Total: 57	Total: mean 43 (SD 13)	Netherlands
Du et al [26]	MDD	Current				✓	Total: 230; interviewed: 150	Total: 44	Total: mean 20 (SD 3)	China
Fowler et al [27]	EUPD[m]	Current		✓			Sample 1: 653; sample 2: 1000	Sample 1: 51; sample 2: 46	Sample 1: mean 36 (SD 15); sample 2: mean 34 (SD 15)	United States
Gaynes et al [28]	Any mood or anxiety disorder[n], any anxiety disorder[o], any depressive disorder[p], bipolar spectrum disorder, PTSD	Current; lifetime only for bipolar spectrum disorder	✓				723	60	Mean 46	United States
Gibbons et al [29]	Any depressive disorder[q], MDD	Current		✓	✓		Total: 1605; interviewed: 292	Total: 70	Total: median 40-49	United States

Study	Conditions	Occurrence of conditions	Sample				Sample size, N	Women, n	Age (years)	Country
			Primary care or general population	Secondary care	Nonpsychiatric controls	University students				
Gibbons et al [30]	MDD	Current		✓	✓		Total: 657; interviewed: 259	Total: 65	—	United States
Gibbons et al [31]	MDD, GAD	Current		✓	✓		Total: 1614; interviewed: 387	Total: 70	Total: median 40-49	United States
Graham et al [32]	MDD, GAD	Current	✓				269	71	Mean 57	United States
Guinart et al [33]	Psychosis	Current		✓	✓		Total: 200; interviewed: 79	Total: 44	Total: median 30	United States
Kertz et al [34]	GAD	Current		✓			Total: 232; interviewed: 218	Total: 60	Total: mean 35 (SD 13)	United States
Kim et al [35]	GAD	Current	✓	✓			527	65	Mean 39 (SD 15)	South Korea
Lohanan et al [36]	EUPD	Current				✓	Total: 342; interviewed: 68	Total: 81	Total: mean 20 (SD 1)	Thailand
McNeely et al [37]	AUD, SUD[r]	Current	✓				Total: 462; interviewed: 459	Total: 52	Total: mean 46 (SD 12)	United States
Meuldijk et al [38]	Any depressive disorder[s], GAD, panic disorder, social phobia, OCD, PTSD, agoraphobia, AUD	Current	✓	✓			1292	61	Mean 40 (SD 13)	Netherlands
Munoz-Navarro et al [39]	GAD	Current	✓				Total: 260; interviewed: 178	Total: 72; interviewed: 70	—	Spain
Nguyen et al [40]	MDD, GAD, social phobia, panic disorder, PTSD, OCD, BN[t], AUD	Current	✓				Total: 616; interviewed: 158	Total: 72; interviewed: 73	Total: mean 40 (SD 12)	Australia
Nielsen et al [41]	MDD	Current	✓				Total: 246; interviewed: 152	Total: 60; interviewed: 59	Total: mean 37 (SD 13); interviewed: mean 34 (SD 13)	Denmark
Oromendia et al [42]	Panic disorder	Current	✓				171	61	Mean 36 (SD 9)	Spain

Study	Conditions	Occurrence of conditions	Sample				Sample size, N	Women, n	Age (years)	Country
			Primary care or general population	Secondary care	Nonpsychiatric controls	University students				
Rogers et al [43]	Any depressive disorder[p], GAD, social phobia, panic disorder, BD[u], ADHD[v], SUD, suicidality	Current	✓				234	64	Mean 47 (SD 16)	United States
Sanchez et al [44]	AUD	Current	✓				100	66	Mean 53 (SD 12)	United States
Schulte-van Maaren et al [45]	Any anxiety disorder[w]	Current	✓	✓			Psychiatric outpatients: 5066; general population: 1295	Psychiatric outpatients: 64; general population: 63	Psychiatric outpatients: mean 37 (SD 12); general population: mean 40 (SD 13)	Netherlands
Ter Huurne et al [46]	AN[x], BN, BED[y], EDNOS[z]	Current		✓			134	88	Mean 31 (SD 11)	Netherlands
Yoon et al [47]	Suicidality	Current	✓	✓			528	65	No risk group: mean 39 (SD 15); risk-positive group: mean 38 (SD 15)	South Korea

[a]The authors also looked at generalized anxiety disorder and bipolar disorder, but no accuracy data were reported.

[b]MDD: major depressive disorder.

[c]Missing data.

[d]Major depressive episode or mania or hypomania.

[e]Panic disorder or generalized anxiety disorder.

[f]Major depressive episode (unspecified).

[g]GAD: generalized anxiety disorder.

[h]MDD, dysthymia, or minor depression.

[i]OCD: obsessive-compulsive disorder.

[j]PTSD: posttraumatic stress disorder.

[k]AUD: alcohol use disorder.

[l]GAD, panic disorder, social phobia, or PTSD.

[m]EUPD: emotionally unstable personality disorder (also known as borderline personality disorder).

[n]MDD, bipolar depression, bipolar spectrum disorder, GAD, agoraphobia, panic disorder, social phobia, PTSD, or OCD.

[o]GAD, agoraphobia, panic disorder, social phobia, PTSD, or OCD.

[p]MDD or bipolar depression.

[q]MDD or minor depression.

[r]SUD: substance use disorder.

[s]Depression (unspecified) or dysthymia.

[t]BN: bulimia nervosa.

[u]BD: bipolar disorder.

[v]ADHD: attention-deficit/hyperactivity disorder.

[w]Anxiety disorder (unspecified).

[x]AN: anorexia nervosa.

[y]BED: binge eating disorder.

[z]EDNOS: eating disorder not otherwise specified.

Digital Mental Health Assessments and Their Validity Per Condition

Overview

The characteristics of the 28 included studies are summarized in Table 2. None of the included studies targeted schizophrenia, autism spectrum disorders, acute stress disorder, adjustment disorder, or self-harm. Insomnia was considered by Nguyen et al [40], but the reference standard used did not meet our eligibility criteria as it did not comprise an assessment by a psychiatrist or a diagnostic interview based on the DSM or ICD criteria. Regarding screening or diagnostic accuracy, below we summarize sensitivity, specificity, and AUCs per tool by condition of interest, where available. For simplicity, where multiple cutoffs were provided for a particular tool, only sensitivity and specificity scores that resulted in the highest Youden index were presented. In the event of multiple sensitivity and specificity values being associated with an equivalent (and highest) Youden index, the values resulting in the smallest difference (ie, sensitivity-specificity) were reported (see Multimedia Appendix 3 [20-47] for sensitivity and specificity values per cutoff score as well as Youden index values and AUCs).

Table 2. Characteristics of the included studies, including conditions of interest, index tests, type and number of questions, reference tests, time flow, and blinding.

Study	Conditions	Index tests	Type of questions	Questions, N	Reference tests	Time flow	Blinded to index test
Achtyes et al [20][a]	MDD[b]	CAD–MDD[c,d]	Based on existing questionnaires, DSM–IV[e], and an expert panel	389	SCID–I[f], DSM–IV–TR[g]	?[h]	?
Ballester et al [21]	Any mood disorder[i], any anxiety disorder[j], any depressive disorder[k], panic disorder, GAD[l]	WMH–ICS[m] surveys	Based on existing questionnaires	291	Spanish MINI[n] (version 5.0 and 6.0), DSM–IV–TR	Within 4 weeks	✓[o]
Cano-Vindel et al [22]	MDD, GAD	PHQ–2[p], GAD–2[q]	Digital versions of existing questionnaires	PHQ–2=2; GAD–2=2	CIDI[r] GAD module, SCID–I, DSM–IV	?	?
Donker et al [23]	Any depressive disorder[s], GAD, social phobia, panic disorder, agoraphobia, OCD[t], PTSD[u], AUD[v]	WSQ[w], GAD–7[x], CES–D[y], PDSS[z], FQ[aa], IES–R[ab], YBOCS[ac], AUDIT[ad]	Based on existing questionnaires, MINI, and AUDIT; digital versions of existing questionnaires	WSQ=15; GAD–7=7; CES–D=20; PDSS=7; FQ=15; IES–R=15; YBOCS=10; AUDIT=10	Lifetime version 2.1 of the CIDI Dutch version, DSM–IV	Mean of 13 days	✓
Donker et al [24]	Any depressive disorder[s]	SID[ae], CES–D, and K10[af]	Digital versions of existing questionnaires	SID=1; CES–D=20; K10=10	Lifetime version 2.1 of the CIDI Dutch version, DSM–IV	Mean of 13 days	?
Donker et al [25]	Any depressive disorder[s], any anxiety disorder[ag], GAD, panic disorder, social phobia, PTSD	GAD–7, GAD–2, GAD–SI[ah], CES–D	Digital versions of existing questionnaires	GAD–7=7; GAD–2=2; GAD–SI=1; CES–D=20	Lifetime version 2.1 of the CIDI Dutch version, DSM–IV	Mean of 13 days	✓
Du et al [26]	MDD	PHQ–9[ai]	Digital version of existing questionnaire	9	MINI (version 5.0, Chinese depression modules), DSM–IV	Within 48 hours	✓
Fowler et al [27]	EUPD[aj,ak]	PID–5[al], FFM[am], SCID–II–PQ[an]	Digital versions of existing questionnaires	PID–5=220; FFM=44; SCID–II–PQ=15	SCID–II[ao], DSM–IV	Within 72 hours	?
Gaynes et al [28]	Any mood or anxiety disorder[ap], any anxiety disorder[aq], any depressive disorder[ar], bipolar spectrum disorder, PTSD	M-3[as]	Questions generated by a panel of mental health clinicians and researchers	27	MINI (version 5.0), DSM–IV	Same day or within 30 days	✓
Gibbons et al [29]	Any depressive disorder[at], MDD	CAT–DI[c,au]	Based on existing questionnaires, DSM–IV, and an expert panel	389	SCID–I, DSM–IV, DSM–IV appendix B (for minor depression)	?	×[av]
Gibbons et al [30]	MDD	CAD–MDD[c]	Based on existing questionnaires, DSM–IV, and an expert panel	88	SCID–I, DSM–IV–TR	?	?
Gibbons et al [31]	MDD, GAD	CAT–ANX[c,aw], CAT–DI[c]	Based on existing questionnaires, DSM–IV, and an expert panel	CAT–ANX=431; CAT–DI=389	SCID–I, DSM–IV	?	?
Graham et al [32][ax]	MDD, GAD	CAD–MDD[c], CAT–ANX[c]	Based on existing questionnaires, DSM–IV, and an expert panel	CAD–MDD=389; CAT–ANX=431	SCID–I, DSM–5	Same day	✓

Study	Conditions	Index tests	Type of questions	Questions, N	Reference tests	Time flow	Blinded to index test
Guinart et al [33]	Psychosis	CAT–Psychosis[c,ay]	Based on existing questionnaires and clinician-rated measures	144	SCID–I, DSM–5	Same day if not completed within last 12 months	?
Kertz et al [34]	GAD	GAD–7	Digital version of existing questionnaire	7	MINI (version 6.0), DSM–IV	?	?
Kim et al [35]	GAD	MHS: A[az]	Based on existing questionnaires and diagnostic criteria, focus group interviews with patients with GAD, and an expert panel	11	MINI (version 5.0), DSM–IV	?	✓
Lohanan et al [36]	EUPD	SI–Bord[ba]	Based on SCID–II criteria	5	SCID–II, DSM–IV	?	?
McNeely et al [37]	AUD, SUD[bb]	SISQs[bc] for alcohol and drugs	Digital version of existing interviewer-administered SISQs	SISQ−alcohol=1; SISQ−drugs=1	MINI–Plus (version 6.0), DSM–IV	Same day	?
Meuldijk et al [38]	Any depressive disorder[bd], GAD, panic disorder, social phobia, OCD, PTSD, agoraphobia, AUD	WSQ	Based on existing questionnaire, MINI, and AUDIT	15	MINI–Plus (version 5.0), DSM–IV–TR	?	?
Munoz-Navarro et al [39]	GAD	GAD–7	Digital version of existing questionnaire	7	CIDI GAD module Spanish version, DSM–IV	?	✓
Nguyen et al [40]	MDD, GAD, social phobia, panic disorder, PTSD, OCD, BN[be], AUD	e-PASS[c,bf]	Based on the DSM–IV–TR criteria; includes a variety of demographic questions	>540	MINI–Plus (version 5.0), DSM–IV, ADIS–IV[bg] (if anxiety symptoms present), DSM–IV–TR	Mean of 10.5 (range 1-34) days	✓
Nielsen et al [41]	MDD	MDI[bh]	Digital version of existing questionnaire	13	M–CIDI[bi] computerized Norwegian version, DSM–IV	Within 2 weeks	✓
Oromendia et al [42]	Panic disorder	WSQ	Based on existing questionnaire	1	SCID–I, DSM–IV	Mean of 14 days	?
Rogers et al [43]	Any depressive disorder[ar], GAD, social phobia, panic disorder, BD[bj], ADHD[bk], SUD, suicidality	CMFC[bl] (initial screener and SAMs[bm])	Expert panel	Initial screener=8; SAMs=11-27	SCID–5–RV[bn], DSM–5	Same day	✓
Sanchez et al [44]	AUD	TAPS–1[bo]	Based on the NIDA[bp] Quick Screen version 1.0	4	CIDI Spanish version, DSM–5	Same day	?
Schulte-van Maaren et al [45]	Any anxiety disorder[bq]	BSA[br], PI–R[bs], PAI[bt], PSWQ[bu], WDQ[bv], SIAS[bw], SPS[bx], IES–R	Digital versions of existing questionnaires	BSA=10; PI–R=41; PAI=15; PSWQ=16; WDQ=30; SIAS=20; SPS=20; IES–R=22	MINI–Plus (version 5.0), DSM–IV	?	?
Ter Huurne et al [46]	AN[by], BN, BED[bz], EDNOS[ca]	EDQ–O[cb]	Based on MINI–Plus and DSM–IV–TR criteria	26	Clinical interview based on the DSM–IV–TR criteria	Mean of 9 days (range of several hours to 48 days)	✓

Study	Conditions	Index tests	Type of questions	Questions, N	Reference tests	Time flow	Blinded to index test
Yoon et al [47]	Suicidality	UBCS[cc]	Literature review and expert panel	12	MINI (version 5.0), DSM–IV	Same day	✓

[a]The authors also used the Computerized Adaptive Test–Depression Inventory, Computerized Adaptive Test–Anxiety, and Computerized Adaptive Test–Mania, but no accuracy data were reported.

[b]MDD: major depressive disorder.

[c]Adaptive in nature, meaning that participants would only answer questions based on their answers to previous items.

[d]CAD–MDD: Computerized Adaptive Diagnosis for Major Depressive Disorder.

[e]DSM–IV: Diagnostic and Statistical Manual of Mental Disorders (fourth edition).

[f]SCID–I: Structured Clinical Interview for DSM Axis I Disorders.

[g]DSM–IV–TR: DSM–IV (text revision).

[h]Unclear.

[i]Major depressive episode or mania or hypomania.

[j]Panic disorder or generalized anxiety disorder.

[k]Major depressive episode (unspecified).

[l]GAD: generalized anxiety disorder.

[m]WMH–ICS: World Health Organization World Mental Health International College Student.

[n]MINI: Mini-International Neuropsychiatric Interview.

[o]Yes.

[p]PHQ-2: 2-item Patient Health Questionnaire.

[q]GAD-2: 2-item Generalized Anxiety Disorder Scale.

[r]CIDI: Composite International Diagnostic Interview.

[s]MDD, dysthymia, or minor depression.

[t]OCD: obsessive-compulsive disorder.

[u]PTSD: posttraumatic stress disorder.

[v]AUD: alcohol use disorder.

[w]WSQ: Web-Based Screening Questionnaire.

[x]GAD–7: 7-item Generalized Anxiety Disorder Scale.

[y]CES–D: Center for Epidemiological Studies–Depression Scale.

[z]PDSS: Panic Disorder Severity Scale.

[aa]FQ: Fear Questionnaire.

[ab]IES–R: Impact of Event Scale–Revised.

[ac]YBOCS: Yale–Brown Obsessive Compulsive Scale.

[ad]AUDIT: Alcohol Use Disorders Identification Test.

[ae]SID: single-item depression scale.

[af]K10: Kessler Psychological Distress Scale.

[ag]GAD, panic disorder, social phobia, or PTSD.

[ah]GAD–SI: single-item Generalized Anxiety Disorder Scale.

[ai]PHQ–9: 9-item Patient Health Questionnaire.

[aj]EUPD: emotionally unstable personality disorder.

[ak]Also known as borderline personality disorder.

[al]PID–5: Personality Inventory for the DSM–5.

[am]FFM: Five Factor Model questionnaire.

[an]SCID–II–PQ: Structured Clinical Interview for DSM Axis II Disorders Personality Questionnaire.

[ao]SCID–II: Structured Clinical Interview for DSM Axis II Disorders.

[ap]MDD, bipolar depression, bipolar spectrum disorder, GAD, agoraphobia, panic disorder, social phobia, PTSD, or OCD.

[aq]GAD, agoraphobia, panic disorder, social phobia, PTSD, or OCD.

[ar]MDD or bipolar depression.

[as]M-3: My Mood Monitor.

[at]MDD or minor depression.

[au]CAT–DI: Computerized Adaptive Test–Depression Inventory.

[av]No.

[aw]CAT–ANX: Computerized Adaptive Test–Anxiety.

[ax]The authors also used the CAT–DI, but no accuracy data were reported.
[ay]CAT–Psychosis: Computerized Adaptive Test–Psychosis.
[az]MHS: A: Mental Health Screening Tool for Anxiety Disorders.
[ba]SI–Bord: screening instrument for borderline personality disorder.
[bb]SUD: substance use disorder.
[bc]SISQ: single-item screening question.
[bd]Depression (unspecified) or dysthymia.
[be]BN: bulimia nervosa.
[bf]e-PASS: electronic psychological assessment screening system.
[bg]ADIS–IV: Anxiety Disorders Interview Schedule (fourth edition).
[bh]MDI: Major Depression Inventory.
[bi]M–CIDI: Munich–Composite International Diagnostic Interview.
[bj]BD: bipolar disorder.
[bk]ADHD: attention-deficit/hyperactivity disorder.
[bl]CMFC: Connected Mind Fast Check.
[bm]SAM: standardized assessment module.
[bn]SCID–V–RV: Structured Clinical Interview for the DSM–5 Research Version.
[bo]TAPS–1: Tobacco, Alcohol, Prescription Medication, and Other Substance Use scale.
[bp]NIDA: National Institute on Drug Abuse.
[bq]Anxiety disorder (unspecified).
[br]BSA: Brief Scale for Anxiety.
[bs]PI–R: Padua Inventory–Revised.
[bt]PAI: Panic Appraisal Inventory.
[bu]PSWQ: Penn State Worry Questionnaire.
[bv]WDQ: Worry Domains Questionnaire.
[bw]SIAS: Social Interaction and Anxiety Scale.
[bx]SPS: Social Phobia Scale.
[by]AN: anorexia nervosa.
[bz]BED: binge eating disorder.
[ca]EDNOS: eating disorder not otherwise specified.
[cb]EDQ–O: Eating Disorder Questionnaire–Online.
[cc]UBCS: Ultra Brief Checklist for Suicidality.

Any Mood or Anxiety Disorder Identification

A total of 1 study (1/28, 4%) targeted the identification of *any* mood or anxiety disorder [28]. To do this, the authors used the My Mood Monitor (M-3) checklist, which is a commercially available test developed by a panel of mental health clinicians and researchers and intended for use in primary care. The tool consists of a total of 27 items focusing on the presence of psychiatric symptoms over the past 2 weeks and covers the following disorders: MDD (7 questions), generalized anxiety disorder (GAD; 2 questions), panic disorder (2 questions), social phobia (1 question), PTSD (4 questions), and OCD (3 questions). In addition, the M-3 inquires about lifetime symptoms of BD (4 questions) and includes a set of 4 functional impairment questions. The authors assessed whether a positive screen on any of the diagnostic categories could be used to identify any mood or anxiety disorder. The sensitivity and specificity of the M-3 were 0.83 and 0.76, respectively.

Any Mood Disorder Identification

The study by Ballester et al [21] targeted the identification of *any* mood disorder. To this end, the authors used the World Health Organization World Mental Health International College Student (WMH–ICS) surveys, which are based on existing questionnaires and include a total of 291 questions. These surveys were designed to generate epidemiological data on mental health disorders among college students worldwide. For current mood disorders, the sensitivity and specificity of the WMH–ICS surveys were 0.76 and 0.80, respectively (AUC=0.78). Lifetime/past mood disorders were identified with a sensitivity of 0.95 and a specificity of 0.60 (AUC=0.77). Overall, discrimination ability was fair for both current and lifetime prevalence of mood disorders.

Any Anxiety Disorder Identification

A total of 4 studies (4/28, 14%) targeted *any* anxiety disorder [21,25,28,45], resulting in a total of 13 unique tools. The study by Ballester et al [21] used the WMH–ICS surveys, which had a sensitivity of 0.79 and a specificity of 0.89 (AUC=0.84) for current anxiety disorders. Lifetime anxiety disorders were identified with a sensitivity of 0.92 and a specificity of 0.71 (AUC=0.81). Accuracy was good for both current and lifetime prevalence of any anxiety disorder.

Digitized versions of the well-validated 7-item Generalized Anxiety Disorder Scale (GAD–7) and its more succinct versions, the 2-item (GAD–2) and single-item (GAD–SI) scales, were used by Donker et al [25]. For cutoff scores with the highest Youden indexes, the sensitivity and specificity of these tools

were 0.36 and 0.78 (GAD–7), 0.47 and 0.72 (GAD–2), and 0.72 and 0.41 (GAD–SI), respectively.

The Brief Scale for Anxiety, Padua Inventory–Revised, Panic Appraisal Inventory, Penn State Worry Questionnaire, Worry Domains Questionnaire, Social Interaction and Anxiety Scale, Social Phobia Scale, and Impact of Event Scale–Revised were used in their digitized versions by Schulte-van Maaren et al [45]. The total number of questions varied from 15 to 21, with excellent discrimination ability (AUC=0.92-0.96). The sensitivity and specificity values for these tools ranged from 0.86 to 0.91 and 0.85 to 0.91, respectively.

Finally, the study by Gaynes et al [28] used the anxiety items of the M-3 (ie, GAD, panic disorder, social phobia, PTSD, and OCD), comprising a total of 12 questions. The sensitivity and specificity of the M-3 were 0.82 and 0.78, respectively.

Any Depressive Disorder Identification

Among the 8 studies (8/28, 29%) targeting the recognition of *any* depressive disorder [21,23-25,28,29,38,43], 11 unique digital mental health assessments were used. These comprised a combination of digitized versions of existing questionnaires, including the single-item depression scale, Center for Epidemiological Studies–Depression Scale, and Kessler Psychological Distress Scale as well as the GAD–7, GAD–2, and GAD–SI, with the total number of questions ranging from 1 to 20. For cutoff scores with the highest Youden indexes, the sensitivity and specificity of these tools were 0.87 and 0.51 (single-item depression scale [24]), 0.94 and 0.69 (Center for Epidemiological Studies–Depression Scale [23,24]), 0.71 and 0.77 (Kessler Psychological Distress Scale [24]), 0.94 and 0.37 (GAD–7 [25]), 0.61 and 0.75 (GAD–2 [25]), and 0.82 and 0.43 (GAD–SI [25]), respectively.

In addition, tools based on existing questionnaires included the WMH–ICS–Major Depressive Episode survey (current: sensitivity=0.93, specificity=0.83, AUC=0.88; lifetime: sensitivity=0.96, specificity=0.65, AUC=0.80), which demonstrated good accuracy [21], and the 2 MDD items of the 15-item Web-Based Screening Questionnaire (WSQ; sensitivity=0.85 [23] and 0.58 [38], specificity=0.59 [23] and 0.94 [37]), which showed fair to good discrimination ability (AUC=0.72 [23] and 0.83 [38]). The WSQ is based on an existing questionnaire, the Mini-International Neuropsychiatric Interview, and the Alcohol Use Disorders Identification Test and can be used to assess depression, GAD, panic disorder, panic disorder with agoraphobia, agoraphobia, specific phobia, social phobia, PTSD, OCD, alcohol abuse and dependence, and suicide.

Furthermore, 1 study (1/28, 4%) [28] used the 7 MDD questions of the M-3 (sensitivity=0.84, specificity=0.80), whereas another study (1/28, 4%) [29] used the Computerized Adaptive Test–Depression Inventory (CAT–DI), which includes a total of 389 items and comprises one of the modules of the commercially available Computerized Adaptive Test–Mental Health (CAT–MH). These modules are based on existing questionnaires, DSM–IV criteria, and an expert panel. Notably, the tests can be fully integrated into routine care and are adaptive in nature, meaning that participants only answer questions based on their answers to previous items. The accuracy of the CAT–DI varied depending on the comparison group (nonpsychiatric comparator: sensitivity=0.90, specificity=0.88; psychiatric comparator: sensitivity=0.90, specificity=0.64). Finally, the study by Rogers et al [43] used the Connected Mind Fast Check (CMFC), which was developed by an expert panel that included psychologists. The tool screens and assesses for several psychiatric disorders using initial screeners and standardized assessment modules (SAMs). The number of questions ranges from 1 to 2 for the initial screeners, resulting in a total of 8 screening questions, and between 11 and 27 for the SAMs. The SAMs are adaptive in nature, meaning that individuals only answer questions based on their answers to previous items. Notably, the CMFC is eligible for reimbursement for primary care practices in the United States. In terms of diagnostic accuracy, the sensitivity and specificity of the CMFC initial screener were 0.94 and 0.65, respectively. In contrast, the SAM had a sensitivity of 0.45 and a specificity of 0.93. Importantly, when reviewing the decision rules of the CMFC SAM, the capability of the tool to detect a major depressive episode increased to 0.73 (sensitivity), whereas the specificity remained largely unchanged (0.92).

Generalized Anxiety Disorder Identification

A total of 12 studies (12/28, 43%) focused on the identification of GAD [21-23,25,31,32,34,35,38-40,43], comprising a total of 9 unique tools. The most popular assessments were the digitized version of the GAD–7, with sensitivity and specificity values ranging from 0.75 to 0.87 and 0.55 to 0.78, respectively [23,25,34,39]. Discrimination ability for digitized versions of the GAD–7 ranged from poor to good (AUC=0.65-0.86). Diagnostic validity for GAD identification was also assessed for the Computerized Adaptive Test–Anxiety (CAT–ANX), which comprises one of the modules of the CAT–MH. The sensitivity and specificity of the CAT–ANX varied depending on the sample type (entire sample: sensitivity=0.89, specificity=0.77; nonpsychiatric comparator: sensitivity=0.86, specificity=0.86 [31]). In addition, the study by Graham et al [32] demonstrated that the CAT–ANX was excellent at discriminating individuals with GAD from those without the condition (AUC=0.93).

Other tools included the digitized versions of the GAD–2, which was used by both Cano-Vindel et al [22] (sensitivity=0.77, specificity=0.80) and Donker et al [25] (sensitivity=0.83, specificity=0.61, AUC=0.76), as well as the GAD–SI (sensitivity=0.70, specificity=0.76 [25]), which showed fair discrimination ability (AUC=0.78). The GAD survey of the WMH–ICS demonstrated good to excellent accuracy (current: sensitivity=1.00, specificity=0.86, AUC=0.93; lifetime: sensitivity=0.97, specificity=0.79, AUC=0.88 [21]). In addition, the GAD item of the WSQ was used across 2 studies, with discrimination ability ranging from fair to good (Donker et al [23]: sensitivity=0.93, specificity=0.45, AUC=0.78; Meuldijk et al [38]: sensitivity=0.66, specificity=0.90, AUC=0.89).

GAD was assessed using the GAD module of the electronic psychological assessment screening system (e-PASS), which is based on the DSM–IV text revision criteria (sensitivity=0.78, specificity=0.68 [40]). The e-PASS assesses a total of 21

disorders; includes >540 questions; and is adaptive in nature, meaning that participants only answer questions based on their answers to previous items. It also includes a number of sociodemographic questions. The e-PASS is funded by the Australian Government Department of Health and Ageing and is available on the web for free. Upon completion, recommendations on what to do next (eg, referral to another service) are provided to individuals. If needed, the e-PASS provides e-therapist support via email, video, or chat. This is intended to help guide users and is not a replacement for face-to-face care.

Furthermore, GAD was also assessed using the Mental Health Screening Tool for Anxiety Disorders [35], which demonstrated excellent diagnostic accuracy (sensitivity=0.98, specificity=0.80, AUC=0.95). The tool comprises 11 questions based on existing questionnaires and diagnostic criteria, focus group interviews with patients with GAD, and an expert panel. Finally, the study by Rogers et al [43] used the CMFC. The initial screener had a sensitivity of 0.93 and a specificity of 0.63, whereas the SAM resulted in a sensitivity and specificity of 0.73 and 0.89, respectively. The sensitivity of the SAM increased to 0.90 when reviewing the module's decision rules, with the specificity remaining largely unchanged (0.86).

Panic Disorder Identification

Among the 7 studies (7/28, 25%) targeting the recognition of panic disorder [21,23,25,38,40,42,43], 8 unique digital mental health assessment tools were used. The most popular tool for panic disorder was the panic disorder item of the WSQ, which was used by Donker et al [23] (sensitivity=0.90, specificity=0.44, AUC=0.76), Meuldijk et al [38] (sensitivity=0.81, specificity=0.95, AUC=0.98), and Oromendia et al [42] (sensitivity=0.81, specificity=0.80, AUC=0.82). Other tools used included the digitized versions of the GAD–7 (sensitivity=0.88, specificity=0.37, AUC=0.62 [25]), GAD–2 (sensitivity=0.38, specificity=0.83, AUC=0.64 [25]), and GAD–SI (sensitivity=0.88, specificity=0.39, AUC=0.65 [25]) as well as the self-reported version of the Panic Disorder Severity Scale (AUC=0.70 [23]). In addition, the panic disorder questions of the e-PASS (sensitivity=0.71, specificity=0.91 [40]) and WMH–ICS (current: sensitivity=0.45, specificity=0.98, AUC=0.71; lifetime: sensitivity=0.71, specificity=0.83, AUC=0.77 [21]) were also used to assess the condition. Finally, the study by Rogers et al [43] used the CMFC. The initial screener had a sensitivity of 0.79 and a specificity of 0.52, whereas the SAM resulted in a sensitivity and specificity of 0.32 and 0.76, respectively.

Social Phobia Identification

A total of 5 studies (5/28, 18%) focused on the recognition of social phobia [23,25,38,40,43], comprising a total of 7 unique digital mental health assessment tools. The social phobia items of the WSQ were used across 2 studies (2/28, 7%; sensitivity=0.72, specificity=0.73, AUC=0.72 [23]; sensitivity=0.79, specificity=0.93, AUC=0.95 [38]). The accuracy of the GAD–7 (sensitivity=0.38, specificity=0.77 [25]) and GAD–2 (sensitivity=0.46, specificity=0.70 [25]) was also evaluated, and both presented AUCs <0.60, which is generally regarded as a fail. Other tools included the GAD–SI (sensitivity=0.69, specificity=0.39, AUC=0.76 [25]), the Fear Questionnaire (FQ; AUC=0.82 [23]), and the social phobia items of the e-PASS (sensitivity=0.60, specificity=0.90 [40]). In addition, the study by Rogers et al [43] used the CMFC. The initial screener had a sensitivity of 0.92 and a specificity of 0.53, whereas the SAM resulted in a sensitivity and specificity of 0.42 and 0.75, respectively.

PTSD Identification

A total of 5 studies (5/28, 18%) targeted PTSD [23,25,28,38,40], resulting in 7 unique digital mental health assessment tools with accuracies ranging from poor to good. The PTSD items of the WSQ were used by Donker et al [23] (sensitivity=0.83, specificity=0.47, AUC=0.65) and Meuldijk et al [38] (sensitivity=0.79, specificity=0.52, AUC=0.86). Other tools included the digitized versions of the GAD–7 (sensitivity=0.75, specificity=0.77, AUC=0.76 [25]), GAD–2 (sensitivity=0.88, specificity=0.71, AUC=0.74 [25]), GAD–SI (sensitivity=0.63, specificity=0.69, AUC=0.69 [25]), and Impact of Event Scale (AUC=0.82 [23]), which includes a total of 15 items. In addition, the PTSD items of the e-PASS (sensitivity=0.75, specificity=0.92 [40]) and M-3 (sensitivity=0.88, specificity=0.70 [28]) were used to assess for the presence of the disorder.

OCD Identification

OCD was assessed using 3 unique digital mental health assessments across 3 separate studies (3/28, 11%) [23,38,40]. The OCD item of the WSQ was used in 2 studies (2/28, 7%), with a sensitivity and specificity of 0.80 and 0.69 [23] and 0.67 and 0.91 [38], respectively, and a good discrimination ability in both studies (AUC=0.81 [23], AUC=0.82 [38]). The remaining 2 tools included the OCD items of the e-PASS (sensitivity=0.75, specificity=0.92 [40]) and the digitized version of the Yale–Brown Obsessive Compulsive Scale, which comprises a total of 10 questions and showed good accuracy (AUC=0.86 [23]).

Agoraphobia Identification

A total of 2 studies (2/28, 7%) targeted the identification of agoraphobia [23,38] with good accuracy. In both studies, the authors used the agoraphobia item of the WSQ (sensitivity=1.00, specificity=0.63, AUC=0.81 [23]; sensitivity=0.81, specificity=0.95, AUC=0.80 [38]). Donker et al [23] also used the digitized version of the FQ, which includes 5 questions to assess the condition (AUC=0.81).

MDD Identification

Among the 8 studies (8/28, 29%) focusing on MDD [20,22,26,29,30,32,40,41], a total of 6 digital mental health assessment tools were used. The most widely used tool was the Computerized Adaptive Diagnosis for MDD (CAD–MDD), which comprises one of the modules of the CAT–MH and consists of a total of 389 questions. The accuracy of the CAD–MDD varied across studies and sample types (sensitivity=0.77-0.96, specificity=0.64-1.00 [20,30,32]). The CAT–DI was used by Gibbons et al [29], with a sensitivity of 0.82 and a specificity of 0.85. The MDD module of the e-PASS was used by Nguyen et al [40] (sensitivity=0.86, specificity=0.79), whereas 2 studies (2/28, 7%) used the

digitized versions of the PHQ–9 with good accuracy (sensitivity=0.89, specificity=0.79, AUC=0.90 [26]) and the 2-item Patient Health Questionnaire (sensitivity=0.78, specificity=0.73 [22]). Finally, the study by Nielsen et al [41] used the Major Depression Inventory, which is a digital version of an existing questionnaire and includes 13 questions, resulting in poor accuracy (sensitivity=0.62, specificity=0.63, AUC=0.66).

BD or Bipolar Spectrum Disorder Identification

In total, 1 study (1/28, 4%) targeted lifetime bipolar spectrum disorder [28] using the 4 BD items of the M-3, which had a sensitivity of 0.88 and a specificity of 0.70. In addition, the study by Rogers et al [43] used the CMFC to detect BD in individuals who met the criteria for a major depressive episode. The initial screener had a sensitivity of 0.63 and a specificity of 0.79, whereas the SAM resulted in a sensitivity and specificity of 0.50 and 0.97, respectively.

ADHD Identification

A total of 1 study (1/28, 4%) assessed for ADHD [43] using the CMFC. The initial screener resulted in a sensitivity and specificity of 0.94 and 0.61, respectively, whereas the SAM had a sensitivity of 0.69 and a specificity of 0.86.

AUD and SUD Identification

A total of 5 studies (5/28, 18%) targeted the identification of AUD [23,37,38,40,44] using a total of 5 distinct digital mental health assessment tools with fair to good accuracy. The alcohol items of the WSQ were used by both Donker et al [23] (sensitivity=0.83, specificity=0.72, AUC=0.77) and Meuldijk et al [38] (sensitivity=0.56, specificity=0.92, AUC=0.82). Other tools included the alcohol module of the e-PASS (sensitivity=0.42, specificity=1.00 [40]) as well as the digitized versions of the single-item screening question (SISQ) for AUD (SISQ–alcohol; sensitivity=0.87, specificity=0.74, AUC=0.80 [37]); Tobacco, Alcohol, Prescription Medication, and Other Substance Use tool (sensitivity=0.97, specificity=0.99 [44]); and Alcohol Use Disorders Identification Test (AUC=0.75 [28]).

A total of 2 studies (2/28, 7%) focused on SUD. The study by McNeely et al [37] used the SISQ–drugs, which had a sensitivity of 0.85 and a specificity of 0.89 (AUC=0.87). The study by Rogers et al [43] used the CMFC. The initial screener had a sensitivity of 0.80 and a specificity of 0.92, whereas the SAM resulted in a sensitivity and specificity of 0.67 and 0.96, respectively.

Eating Disorders Identification

Regarding eating disorders, 1 study (1/28, 4%) [46] focused on anorexia nervosa and bulimia nervosa (BN) as well as binge eating disorder and eating disorder otherwise not specified using the Eating Disorder Questionnaire–Online (EDQ–O), which is based on the Mini-International Neuropsychiatric Interview–Plus and DSM–IV text revision criteria and comprises a total of 26 questions. The accuracy of the EDQ–O for the recognition of these conditions ranged from fair to good (anorexia nervosa: sensitivity=0.44, specificity=1.00, AUC=0.72; BN: sensitivity=0.78, specificity=0.88, AUC=0.83; binge eating disorder: sensitivity=0.66, specificity=0.98, AUC=0.82; eating disorder otherwise not specified: sensitivity=0.87, specificity=0.72, AUC=0.79). An additional study (1/28, 4%) [40] targeted BN using the bulimia module of the e-PASS, which had a sensitivity and specificity of 0.50 and 0.97, respectively.

Emotionally Unstable Personality Disorder Identification

When considering personality disorders, 2 studies (2/28, 7%) targeted emotionally unstable personality disorder (EUPD) [27,36], also known as borderline personality disorder. Fowler et al [27] used digitized versions of the Five Factor Model, with a sensitivity of 0.70 and a specificity of 0.62 for the neuroticism and agreeableness composites and a sensitivity and specificity of 0.71 and 0.62, respectively, for the neuroticism, agreeableness, and conscientiousness composites. Both combinations of composites had fair accuracy (AUC=0.72 and 0.73, respectively). The authors also used the self-report Structured Clinical Interview for DSM Axis II Disorders Personality Questionnaire, which had a sensitivity and specificity of 0.78 and 0.80, respectively, and good discrimination ability (AUC=0.86), and the Personality Inventory for the DSM–5 (sensitivity=0.81, specificity=0.76), which also showed good accuracy (AUC=0.87). Lohanan et al [36] used the screening instrument for borderline personality disorder, which is based on the Structured Clinical Interview for DSM Axis II Disorders and includes a total of 5 items. The sensitivity of the screening instrument for borderline personality disorder was 0.56, whereas the specificity was 0.92 with good accuracy (AUC=0.83).

Psychosis Identification

In total, 1 study (1/28, 4%) targeted psychosis [33] using the Computerized Adaptive Test–Psychosis (CAT–Psychosis), which is one of the tests available in the CAT–MH. The accuracy of the CAT–Psychosis was good (entire sample: AUC=0.85; including only those who had received the Structured Clinical Interview for DSM Axis I Disorders: AUC=0.80).

Suicidality Identification

A total of 2 studies (2/28, 7%) examined suicidality. The first study [43] used the CMFC, with the accuracy of the initial screener varying depending on the criteria examined (thoughts of own death: sensitivity=0.75, specificity=0.89; suicidal ideation: sensitivity=0.75, specificity=0.84; specific plan: sensitivity=1.00, specificity=0.80). The second study [47] used the Ultra Brief Checklist for Suicidality, which had a sensitivity of 0.91 and a specificity of 0.85 for the cutoff score with the highest Youden index.

Risk of Bias and Applicability Assessment

The evaluation of risk of bias and applicability for all 28 studies was conducted using the amended QUADAS–2 tool [17]. The results are summarized in Table 3, with scores for each signaling question available upon request. This assessment revealed a high risk of bias in most of the considered studies. For instance, with regard to patient selection, 12 studies (12/28, 43%) [20,24,29-33,38,40,42,45,47] had high risk of bias, primarily because of issues with enrollment and a failure to avoid a case–control sample, which may not fully reflect real-world

patient populations. A total of 9 studies (9/28, 32%) [21-23,25,34,36,39,44,46] did not provide enough information regarding their sample and sampling procedures. Similarly, risk of bias was an issue when considering index test administration, with 10 studies (10/28, 36%) [21,24-28,35,36,40,47] showing high risk of bias, which was primarily due to the studies not using a prespecified threshold. A total of 13 studies (13/28, 46%) [20,22,29-31,33,34,38,42-46] failed to provide enough information regarding the index test administration. This was particularly with regard to whether the results were interpreted without knowledge of the reference standard. In total, 1 study (1/28, 4%) [29] showed high risk of bias when considering the reference standard, with the results interpreted with knowledge of the results of the index test, whereas 14 studies (14/28, 50%) [20,22,24,27,30,31,33,34,36-38,42,44,45] did not provide sufficient information regarding the interpretation of the reference standard. Finally, flow and timing were also a consideration, with 4 studies (4/28, 14%) showing high risk of bias. In this regard, Guinart et al [33] did not re-administer the reference standard to patients who had received a diagnostic interview within the 12 months before taking part in the study, and the studies by Gibbons et al [29-31] included nonpsychiatric controls in the analyses who appeared not to have received the reference standard. A total of 11 studies (11/28, 39%) [20,22-25,34-36,38,39,45] did not provide enough information regarding the timing between the index test and reference standard.

In terms of applicability, given our review question and strict inclusion and exclusion criteria, all the included studies were judged to have low applicability concerns.

Table 3. Results of the amended quality assessment of the included studies.

Study	Risk of bias				Applicability concerns		
	Patient selection	Index test	Reference standard	Flow and timing	Patient selection	Index test	Reference standard
Achtyes et al [20]	☒ a	?[b]	?	?	☒ c	☒	☒
Ballester et al [21]	?	☒	☒	☒	☒	☒	☒
Cano-Vindel et al [22]	?	?	?	?	☒	☒	☒
Donker et al [23]	?	☒	☒	?	☒	☒	☒
Donker et al [24]	☒	☒	?	?	☒	☒	☒
Donker et al [25]	?	☒	☒	?	☒	☒	☒
Du et al [26]	☒	☒	☒	☒	☒	☒	☒
Fowler et al [27]	☒	☒	?	☒	☒	☒	☒
Gaynes et al [28]	☒	☒	☒	☒	☒	☒	☒
Gibbons et al [29]	☒	?	☒	☒	☒	☒	☒
Gibbons et al [30]	☒	?	?	☒	☒	☒	☒
Gibbons et al [31]	☒	?	?	☒	☒	☒	☒
Graham et al [32]	☒	☒	☒	☒	☒	☒	☒
Guinart et al [33]	☒	?	?	☒	☒	☒	☒
Kertz et al [34]	?	?	?	?	☒	☒	☒
Kim et al [35]	☒	☒	☒	?	☒	☒	☒
Lohanan et al [36]	?	☒	?	?	☒	☒	☒
McNeely et al [37]	☒	☒	?	☒	☒	☒	☒
Meuldijk et al [38]	☒	?	?	?	☒	☒	☒
Munoz-Navarro et al [39]	?	☒	☒	?	☒	☒	☒
Nguyen et al [40]	☒	☒	☒	☒	☒	☒	☒
Nielsen et al [41]	☒	☒	☒	☒	☒	☒	☒
Oromendia et al [42]	☒	?	?	☒	☒	☒	☒
Rogers et al [43]	☒	?	☒	☒	☒	☒	☒
Sanchez et al [44]	?	?	?	☒	☒	☒	☒
Schulte-van Maaren et al [45]	☒	?	?	?	☒	☒	☒
Ter Huurne et al [46]	?	?	☒	☒	☒	☒	☒
Yoon et al [47]	☒	☒	☒	☒	☒	☒	☒

[a]High risk.

[b]Unclear risk.

[c]Low risk.

Discussion

Overview

This systematic review set out to explore the current state and validity of question-and-answer–based digital mental health assessment tools targeting a wide range of mental health conditions. We believe that the findings of this review will provide health care professionals and researchers with a deeper understanding of the use of digital technologies for the screening and diagnosing of mental health conditions in adulthood, as well as of the challenges that remain and opportunities for the development of innovative digital mental health assessment tools moving forward.

Implications for Health Care Professionals

The digitization of existing pen-and-paper questionnaires and scales routinely used for mental health screening and assessment can offer various benefits, such as minimal delivery costs, efficient data collection, and increased convenience. For health care providers looking to digitize the use of existing pen-and-paper questionnaires in their clinical practice, the included studies report on 26 unique tools. Critically, most of these tools were designed to target a single condition rather than being comprehensive assessments of psychopathology, with most including <45 questions. Thus, a combination of these tools should be considered if a comprehensive mental health assessment is preferred.

Alternatively, tools targeting several conditions, such as the M-3 [28], WHM–ICS surveys [21], WSQ [23,38,42], e-PASS [40], and CMFC [43], may represent more attractive options for mental health screening in primary care settings and the first stages of triage. Notably, only the e-PASS includes sociodemographic questions, providing valuable information on factors that are known to be correlated with mental health concerns [48]. In addition, the e-PASS is adaptive in nature, meaning that participants only answer questions based on their answers to previous items, which can ensure that assessment completion is more time-efficient and only relevant symptom data are collected. Adaptive testing was also offered by the CMFC, which is eligible for reimbursement for primary care practices in the United States, as well as by the CAD–MDD, CAT–DI, CAT–ANX, and CAT–Psychosis, which are commercially available.

Overall, the intended settings of use should be carefully considered by health care professionals interested in implementing digital mental health assessment tools in their clinics. Similarly, the importance of accuracy measures in choosing relevant digital tools cannot be overstated. This systematic review revealed mixed findings regarding the validity of the included digital technologies, with accuracy values varying significantly between and within conditions and instruments as well as across different samples. Sensitivity and specificity values ranged from 0.32 to 1.00 and 0.37 to 1.00, respectively, and AUCs ranged from poor (0.57) to excellent (0.98).

Specifically, the GAD–7 and its more succinct versions, which represent the most frequently used instruments, generally demonstrated poor to fair discriminatory performance across a range of anxiety disorders [23,25,34]. An exception was the study by Munoz-Navarro et al [39], where the GAD–7 showed good accuracy in identifying GAD. The digitized versions of existing pen-and-paper questionnaires used by Schulte-van Maaren et al [45] with the aim of identifying any anxiety disorder had excellent accuracy, whereas digitized versions of the FQ, Impact of Event Scale–Revised, and Yale–Brown Obsessive Compulsive Scale demonstrated good discriminatory performance for a variety of anxiety disorders [23]. Regarding digitized versions of existing pen-and-paper questionnaires targeting conditions other than anxiety, the PHQ–9 demonstrated excellent accuracy for MDD [26], whereas the 2-item Patient Health Questionnaire was only fair [22], and the Major Depression Inventory demonstrated poor performance in identifying the condition [41]. SISQs for both AUD and SUD had good accuracy [37], whereas tools assessing for EUPD demonstrated fair to good discriminatory performance [27]. Importantly, although the screening or diagnostic accuracy of these digitized versions of existing pen-and-paper questionnaires appeared to vary significantly across studies, previous systematic reviews have generally revealed good interformat reliability between digital and paper versions, suggesting that these are comparable [49,50]. Therefore, differences in screening or diagnostic accuracy are likely to be due to study effects or methodological issues rather than the tools used being unreliable. Moving forward, there is a need for carefully designed, high-quality studies to further validate and assess the clinical utility of digitized versions of pen-and-paper questionnaires. This will help guide clinicians toward meaningful technologies.

Regarding tools that were not a digitized version of existing pen-and-paper questionnaires and instead gathered questions designed ex novo by mental health experts based on existing diagnostic tools and criteria, the WMH–ICS surveys demonstrated good to excellent accuracy for the identification of any anxiety and depressive disorder as well as GAD [21]. However, the accuracy of the WMH–ICS surveys was fair for any mood disorder and panic disorder [21]. In contrast, the Mental Health Screening Tool for Anxiety Disorders [35] and Tobacco, Alcohol, Prescription Medication, and Other Substance Use scale [44] were excellent at identifying GAD and AUD, respectively. Similarly, the SI-Bord demonstrated good accuracy for EUPD [36], whereas the Ultra Brief Checklist for Suicidality had a sensitivity and specificity of 0.91 and 0.85, respectively, for suicidality [47]. Regarding eating disorders, the EDQ–O presented fair to good discriminatory performance [46].

In addition, the accuracy of the WSQ varied from poor to excellent depending on the condition of interest and study [23,38,42]. Similarly, the clinical utility of the e-PASS varied considerably across conditions, with sensitivity and specificity values ranging from 0.42 to 0.86 and 0.68 to 1.00, respectively [40]. The accuracy of the CMFC also varied across conditions, with sensitivity and specificity values ranging from 0.63 to 1.00 and 0.61 to 0.92 (initial screener) and from 0.32 to 0.75 and

0.90 to 0.97 (SAMs), respectively [43]. Furthermore, the accuracy of the CAD–MDD, CAT–DI, CAT–ANX, and CAT–Psychosis varied across studies and depending on the comparison group (eg, nonpsychiatric comparator vs psychiatric comparator) [20,29-33]. Of these, the CAD–MDD was conceptualized and developed as a screening tool for depression in primary care, whereas the CAT–DI and CAT–ANX are better suited for assessing depression and anxiety severity, respectively [30,32]. Taken together in the form of the CAT–MH, these adaptive assessments could provide a valuable screening and assessment tool for depression and anxiety [32]. The CAT–Psychosis served as a discriminating tool for the presence of psychosis and as an assessment tool for symptom severity, thereby being well-placed in secondary care for psychosis screening and follow-up assessments. Finally, the accuracy of the M-3 varied across conditions, with sensitivity and specificity values ranging from 0.82 to 0.88 and 0.70 to 0.80, respectively [28].

Overall, the utility of the tools included in this review will strongly depend on clinical needs. For screening purposes, tools that have high sensitivity and that can be easily completed by patients are to be prioritized. In contrast, tools with high specificity perform well for diagnostic purposes in symptomatic patient populations. The implementation of digital mental health assessments in common practice workflows will likely require pilot-testing to tailor the tool to case-specific needs.

Recommendations for Research

In addition to reporting on digital mental health assessments' features and accuracy, this systematic review highlights tool development and study design considerations that may inform future research aims. Although the diagnosis of GAD, any depressive disorder, and MDD was investigated in several studies, fewer eligible studies were found for specific anxiety disorders, such as panic disorder and social phobia, as well as AUD. Notably, very few studies targeted the identification of BD, ADHD, SUD, psychosis, and suicidality. Thus, there remain opportunities for the development of more comprehensive digital diagnostic tools. Indeed, digital technologies have the capacity to collect a vast range of key sociodemographic and symptom data. Undeniably, by moving away from brief symptom count checklists such as the GAD–7 and PHQ–9, digital technologies can offer avenues toward a dimensional view of psychopathology, providing valuable information on the co-occurrence of symptoms and diagnoses. Indeed, digital technologies, including adaptive or nonlinear questionnaires where patients are required to answer questions based on previous answers, have the capacity to further streamline and personalize the collection of cross-disorder symptom data. Although outside the scope of this systematic review, combining clinical information with biomarker profiling strategies may allow clinicians and researchers to further shift the focus from categorical constructs to a dimensional approach to psychopathology. For instance, the combination of symptom data and serum analytes has been shown to predict the development of future depressive episodes in individuals presenting with social anxiety [51] and panic disorder [52]. In addition, combining digital symptom-based data with dried blood spot samples shows some promise as a noninvasive and cost-effective diagnostic test for both MDD [53] and BD [54], but research in this area remains largely unexplored.

In addition to suggesting opportunities for future research, this systematic review raises considerations of methodology and research reporting practices. Indeed, researchers and digital mental health innovators should pursue carefully designed, high-quality studies to validate and assess the clinical utility of their diagnostic tools. Of note, the study by Nielsen et al [41] stood out for their comprehensively written methods and well-designed study. For the remaining studies, risk of bias was a concern despite our amended and less stringent QUADAS–2 measures. This was often due to missing information regarding participant sampling procedures, the administration and interpretation of the index test and reference standard, and timing. Inevitably, the nondisclosure of methodological information can hinder the assessment of bias in current and future systematic review exercises aimed at determining the clinical utility of digital mental health assessments. In addition, missing information can prevent replicability studies from validating the findings. Moving forward, the QUADAS–2 measures could be used by researchers and peer reviewers as a checklist for study procedures that should be clearly reported in study methods in addition to complying with relevant guidelines such as the Standards for Reporting of Diagnostic Accuracy Studies [55]. In particular, careful consideration should be given to patient selection, the index test, the reference standard, and flow and timing. For instance, moving away from a case–control study design, digital mental health care researchers should consider evaluating digital mental health assessment tools within the intended context. This would allow for the appraisal of diagnostic technologies in real-world patient populations, thereby facilitating interoperability and guiding health care professionals toward clinically meaningful technologies.

Strengths and Limitations

To our knowledge, this is the first systematic review to assess the validity of question-and-answer–based digital mental health assessment tools targeting a wide range of mental health conditions. However, despite our comprehensive and carefully designed search strategies as well as the inclusion of any study design and language, it is possible that some relevant studies may have been missed. Furthermore, given the focus of this review where only digital tools that were exclusively question-and-answer–based were included, diagnostic technologies that collect passive data (eg, activity rhythms, sleep quality, sentiment, and language patterns) or a combination of active and passive data were not evaluated, with further research in this area being required.

Conclusions

The findings of this systematic review revealed that the field of digital mental health assessment tools is still in its early stages. Indeed, most of the included studies used digitized versions of existing pen-and-paper questionnaires as opposed to more sophisticated and comprehensive digital diagnostic technologies that can be easily integrated into routine clinical care. Furthermore, our review revealed mixed findings regarding the accuracy of the included digital technologies, which varied

significantly between and within conditions as well as across different samples. In addition, risk of bias was a concern with the included studies. This comprehensive systematic review has important implications for the development and implementation of digital mental health assessments. Namely, there exist opportunities for further innovation in the field of digital diagnostic technologies for mental health. Importantly, carefully designed, high-quality studies are essential to validate the clinical utility of these technologies. Finally, evaluating these tools within the intended context is likely to facilitate interoperability and help guide clinicians toward meaningful technologies.

Acknowledgments

This research was funded by the Stanley Medical Research Institute (grant number 07R-1888).

Conflicts of Interest

SB is a director of Psynova Neurotech Ltd and Psyomics Ltd and has financial interests in Psyomics Ltd. TSS had financial interests in Psyomics Ltd at the time of submission. The remaining authors have no conflicts of interest to declare.

Multimedia Appendix 1
Search strategies.
[DOCX File , 23 KB - mental_v9i3e32824_app1.docx]

Multimedia Appendix 2
Checklist summary of the mental health disorders investigated in the included studies.
[DOCX File , 27 KB - mental_v9i3e32824_app2.docx]

Multimedia Appendix 3
Diagnostic accuracy per index test separated by condition of interest.
[DOCX File , 111 KB - mental_v9i3e32824_app3.docx]

References

1. Steel Z, Marnane C, Iranpour C, Chey T, Jackson JW, Patel V, et al. The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013. Int J Epidemiol 2014 Apr;43(2):476-493. [doi: 10.1093/ije/dyu038] [Medline: 24648481]
2. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2095-2128. [doi: 10.1016/S0140-6736(12)61728-0] [Medline: 23245604]
3. Whiteford HA, Ferrari AJ, Degenhardt L, Feigin V, Vos T. The global burden of mental, neurological and substance use disorders: an analysis from the Global Burden of Disease Study 2010. PLoS One 2015;10(2):e0116820 [FREE Full text] [doi: 10.1371/journal.pone.0116820] [Medline: 25658103]
4. GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 2020 Oct 17;396(10258):1204-1222 [FREE Full text] [doi: 10.1016/S0140-6736(20)30925-9] [Medline: 33069326]
5. Kazdin AE, Blase SL. Rebooting psychotherapy research and practice to reduce the burden of mental illness. Perspect Psychol Sci 2011 Jan;6(1):21-37. [doi: 10.1177/1745691610393527] [Medline: 26162113]
6. Mojtabai R, Olfson M, Sampson NA, Jin R, Druss B, Wang PS, et al. Barriers to mental health treatment: results from the National Comorbidity Survey Replication. Psychol Med 2011 Aug;41(8):1751-1761 [FREE Full text] [doi: 10.1017/S0033291710002291] [Medline: 21134315]
7. Andersson G, Cuijpers P, Carlbring P, Riper H, Hedman E. Guided Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: a systematic review and meta-analysis. World Psychiatry 2014 Oct;13(3):288-295 [FREE Full text] [doi: 10.1002/wps.20151] [Medline: 25273302]
8. Cheng P, Luik AI, Fellman-Couture C, Peterson E, Joseph CL, Tallent G, et al. Efficacy of digital CBT for insomnia to reduce depression across demographic groups: a randomized trial. Psychol Med 2019 Feb;49(3):491-500. [doi: 10.1017/S0033291718001113] [Medline: 29792241]
9. van Ballegooijen W, Riper H, Cuijpers P, van Oppen P, Smit JH. Validation of online psychometric instruments for common mental health disorders: a systematic review. BMC Psychiatry 2016;16:45 [FREE Full text] [doi: 10.1186/s12888-016-0735-7] [Medline: 26915661]
10. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]

11. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, 5th Edition. Washington, DC: American Psychiatric Association; 2013.
12. International statistical classification of diseases and related health problems (11th ed). World Health Organization. 2019. URL: https://www.who.int/standards/classifications/classification-of-diseases [accessed 2022-02-09]
13. Torous J, Staples P, Shanahan M, Lin C, Peck P, Keshavan M, et al. Utilizing a personal smartphone custom app to assess the patient health questionnaire-9 (PHQ-9) depressive symptoms in patients with major depressive disorder. JMIR Ment Health 2015;2(1):e8 [FREE Full text] [doi: 10.2196/mental.3889] [Medline: 26543914]
14. Martin-Key NA, Schei TS, Barker EJ, Spadaro B, Funnell E, Benacek J, et al. The current state and diagnostic accuracy of digital mental health assessment tools for psychiatric disorders: protocol for a systematic review and meta-analysis. JMIR Res Protoc 2021 Jan 08;10(1):e25382 [FREE Full text] [doi: 10.2196/25382] [Medline: 33416508]
15. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
16. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983 Sep;148(3):839-843. [doi: 10.1148/radiology.148.3.6878708] [Medline: 6878708]
17. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011 Oct 18;155(8):529-536. [doi: 10.7326/0003-4819-155-8-201110180-00009] [Medline: 22007046]
18. Reitsma J, Rutjes A, Whiting P, Vlassov V, Leeflang M, Deeks J. Chapter 9: Assessing methodological quality. In: Deeks JJ, Bossuyt PM, Gatsonis C, editors. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0. London, United Kingdom: The Cochrane Collaboration; 2009.
19. Naaktgeboren CA, de Groot JA, Rutjes AW, Bossuyt PM, Reitsma JB, Moons KG. Anticipating missing reference standard data when planning diagnostic accuracy studies. Br Med J 2016 Feb 09;352:i402 [FREE Full text] [doi: 10.1136/bmj.i402] [Medline: 26861453]
20. Achtyes ED, Halstead S, Smart L, Moore T, Frank E, Kupfer DJ, et al. Validation of computerized adaptive testing in an outpatient nonacademic setting: the VOCATIONS trial. Psychiatr Serv 2015 Oct;66(10):1091-1096 [FREE Full text] [doi: 10.1176/appi.ps.201400390] [Medline: 26030317]
21. Ballester L, Alayo I, Vilagut G, Almenara J, Cebrià AI, Echeburúa E, UNIVERSAL Study Group. Accuracy of online survey assessment of mental disorders and suicidal thoughts and behaviors in Spanish university students. Results of the WHO World Mental Health - International College Student initiative. PLoS One 2019;14(9):e0221529 [FREE Full text] [doi: 10.1371/journal.pone.0221529] [Medline: 31487306]
22. Cano-Vindel A, Muñoz-Navarro R, Medrano LA, Ruiz-Rodríguez P, González-Blanch C, Gómez-Castillo MD, PsicAP Research Group. A computerized version of the Patient Health Questionnaire-4 as an ultra-brief screening tool to detect emotional disorders in primary care. J Affect Disord 2018 Jul;234:247-255. [doi: 10.1016/j.jad.2018.01.030] [Medline: 29549826]
23. Donker T, van Straten A, Marks I, Cuijpers P. A brief web-based screening questionnaire for common mental disorders: development and validation. J Med Internet Res 2009 Jul;11(3):e19 [FREE Full text] [doi: 10.2196/jmir.1134] [Medline: 19632977]
24. Donker T, van Straten SA, Marks I, Cuijpers P. Brief self-rated screening for depression on the internet. J Affect Disord 2010 May;122(3):253-259. [doi: 10.1016/j.jad.2009.07.013] [Medline: 19679358]
25. Donker T, van Straten A, Marks I, Cuijpers P. Quick and easy self-rating of generalized anxiety disorder: validity of the Dutch web-based GAD-7, GAD-2 and GAD-SI. Psychiatry Res 2011 Jun 30;188(1):58-64. [doi: 10.1016/j.psychres.2011.01.016] [Medline: 21339006]
26. Du N, Yu K, Ye Y, Chen S. Validity study of Patient Health Questionnaire-9 items for Internet screening in depression among Chinese university students. Asia Pac Psychiatry 2017 Sep;9(3) [FREE Full text] [doi: 10.1111/appy.12266] [Medline: 28856843]
27. Fowler JC, Madan A, Allen JG, Patriquin M, Sharp C, Oldham JM, et al. Clinical utility of the DSM-5 alternative model for borderline personality disorder: differential diagnostic accuracy of the BFI, SCID-II-PQ, and PID-5. Compr Psychiatry 2018 Jan;80:97-103. [doi: 10.1016/j.comppsych.2017.09.003] [Medline: 29069625]
28. Gaynes BN, DeVeaugh-Geiss J, Weir S, Gu H, MacPherson C, Schulberg HC, et al. Feasibility and diagnostic validity of the M-3 checklist: a brief, self-rated screen for depressive, bipolar, anxiety, and post-traumatic stress disorders in primary care. Ann Fam Med 2010 Apr;8(2):160-169 [FREE Full text] [doi: 10.1370/afm.1092] [Medline: 20212303]
29. Gibbons RD, Weiss DJ, Pilkonis PA, Frank E, Moore T, Kim JB, et al. Development of a computerized adaptive test for depression. Arch Gen Psychiatry 2012 Nov;69(11):1104-1112 [FREE Full text] [doi: 10.1001/archgenpsychiatry.2012.14] [Medline: 23117634]
30. Gibbons RD, Hooker G, Finkelman MD, Weiss DJ, Pilkonis PA, Frank E, et al. The computerized adaptive diagnostic test for major depressive disorder (CAD-MDD): a screening tool for depression. J Clin Psychiatry 2013 Jul;74(7):669-674 [FREE Full text] [doi: 10.4088/JCP.12m08338] [Medline: 23945443]

31. Gibbons RD, Weiss DJ, Pilkonis PA, Frank E, Moore T, Kim JB, et al. Development of the CAT-ANX: a computerized adaptive test for anxiety. Am J Psychiatry 2014 Feb;171(2):187-194 [FREE Full text] [doi: 10.1176/appi.ajp.2013.13020178] [Medline: 23929270]
32. Graham AK, Minc A, Staab E, Beiser DG, Gibbons RD, Laiteerapong N. Validation of the computerized adaptive test for mental health in primary care. Ann Fam Med 2019 Jan;17(1):23-30 [FREE Full text] [doi: 10.1370/afm.2316] [Medline: 30670391]
33. Guinart D, de Filippis R, Rosson S, Patil B, Prizgint L, Talasazan N, et al. Development and validation of a computerized adaptive assessment tool for discrimination and measurement of psychotic symptoms. Schizophr Bull 2021 Apr 29;47(3):644-652. [doi: 10.1093/schbul/sbaa168] [Medline: 33164091]
34. Kertz S, Bigda-Peyton J, Bjorgvinsson T. Validity of the Generalized Anxiety Disorder-7 scale in an acute psychiatric sample. Clin Psychol Psychother 2013;20(5):456-464. [doi: 10.1002/cpp.1802] [Medline: 22593009]
35. Kim S, Park K, Yoon S, Choi Y, Lee S, Choi K. A brief online and offline (Paper-and-Pencil) screening tool for generalized anxiety disorder: the final phase in the development and validation of the mental health screening tool for anxiety disorders (MHS: A). Front Psychol 2021;12:639366 [FREE Full text] [doi: 10.3389/fpsyg.2021.639366] [Medline: 33692730]
36. Lohanan T, Leesawat T, Wongpakaran T, Wongpakaran N, Karawekpanyawong N, Oon-Arom A, et al. Development and validation of a screening instrument for borderline personality disorder (SI-Bord) for use among university students. BMC Psychiatry 2020 Aug 17;20(1):479 [FREE Full text] [doi: 10.1186/s12888-020-02807-6] [Medline: 32998759]
37. McNeely J, Cleland CM, Strauss SM, Palamar JJ, Rotrosen J, Saitz R. Validation of self-administered single-item screening questions (SISQs) for unhealthy alcohol and drug use in primary care patients. J Gen Intern Med 2015 Dec;30(12):1757-1764 [FREE Full text] [doi: 10.1007/s11606-015-3391-6] [Medline: 25986138]
38. Meuldijk D, Giltay EJ, Carlier IV, van Vliet IM, van Hemert AM, Zitman FG. A validation study of the Web Screening Questionnaire (WSQ) compared with the Mini-International Neuropsychiatric Interview-Plus (MINI-Plus). JMIR Ment Health 2017 Aug 29;4(3):e35 [FREE Full text] [doi: 10.2196/mental.5453] [Medline: 28851674]
39. Muñoz-Navarro R, Cano-Vindel A, Moriana JA, Medrano LA, Ruiz-Rodríguez P, Agüero-Gento L, et al. Screening for generalized anxiety disorder in Spanish primary care centers with the GAD-7. Psychiatry Res 2017 Oct;256:312-317. [doi: 10.1016/j.psychres.2017.06.023] [Medline: 28666201]
40. Nguyen DP, Klein B, Meyer D, Austin DW, Abbott JM. The diagnostic validity and reliability of an internet-based clinical assessment program for mental disorders. J Med Internet Res 2015 Sep 21;17(9):e218 [FREE Full text] [doi: 10.2196/jmir.4195] [Medline: 26392066]
41. Nielsen MG, Ørnbøl E, Bech P, Vestergaard M, Christensen KS. The criterion validity of the web-based Major Depression Inventory when used on clinical suspicion of depression in primary care. Clin Epidemiol 2017;9:355-365 [FREE Full text] [doi: 10.2147/CLEP.S132913] [Medline: 28740432]
42. Oromendia P, Bonillo A, Molinuevo B. Web-based screening for panic disorder: validity of a single-item instrument. J Affect Disord 2015 Jul 15;180:138-141. [doi: 10.1016/j.jad.2015.03.061] [Medline: 25898334]
43. Rogers R, Hartigan SE, Sanders CE. Identifying mental disorders in primary care: diagnostic accuracy of the connected mind fast check (CMFC) electronic screen. J Clin Psychol Med Settings 2021 Oct 05:1-15 [FREE Full text] [doi: 10.1007/s10880-021-09820-1] [Medline: 34609692]
44. Sanchez K, Gryczynski J, Carswell SB, Schwartz RP. Development and feasibility of a Spanish language version of the tobacco, alcohol, prescription drug, and illicit substance use (TAPs) tool. J Addict Med 2021;15(1):61-67. [doi: 10.1097/ADM.0000000000000699] [Medline: 32657958]
45. Schulte-van Maaren YW, Giltay EJ, van Hemert AM, Zitman FG, de Waal MW, Carlier IV. Reference values for anxiety questionnaires: the Leiden Routine Outcome Monitoring study. J Affect Disord 2013 Sep 25;150(3):1008-1018. [doi: 10.1016/j.jad.2013.05.031] [Medline: 23810480]
46. ter Huurne ED, de Haan HA, ten Napel-Schutz MC, Postel MG, Menting J, van der Palen J, et al. Is the Eating Disorder Questionnaire-Online (EDQ-O) a valid diagnostic instrument for the DSM-IV-TR classification of eating disorders? Compr Psychiatry 2015 Feb;57:167-176. [doi: 10.1016/j.comppsych.2014.10.019] [Medline: 25464837]
47. Yoon S, Park K, Choi K. The ultra brief checklist for suicidality. J Affect Disord 2020 Nov 01;276:279-286 [FREE Full text] [doi: 10.1016/j.jad.2020.07.037] [Medline: 32697710]
48. Hudson CG. Socioeconomic status and mental illness: tests of the social causation and selection hypotheses. Am J Orthopsychiatry 2005 Jan;75(1):3-18. [doi: 10.1037/0002-9432.75.1.3] [Medline: 15709846]
49. Alfonsson S, Maathz P, Hursti T. Interformat reliability of digital psychiatric self-report questionnaires: a systematic review. J Med Internet Res 2014;16(12):e268 [FREE Full text] [doi: 10.2196/jmir.3395] [Medline: 25472463]
50. Gwaltney CJ, Shields AL, Shiffman S. Equivalence of electronic and paper-and-pencil administration of patient-reported outcome measures: a meta-analytic review. Value Health 2008;11(2):322-333. [doi: 10.1111/j.1524-4733.2007.00231.x] [Medline: 18380645]
51. Gottschalk MG, Cooper JD, Chan MK, Bot M, Penninx BW, Bahn S. Discovery of serum biomarkers predicting development of a subsequent depressive episode in social anxiety disorder. Brain Behav Immun 2015 Aug;48:123-131. [doi: 10.1016/j.bbi.2015.04.011] [Medline: 25929723]

52. Gottschalk MG, Cooper JD, Chan MK, Bot M, Penninx BW, Bahn S. Serum biomarkers predictive of depressive episodes in panic disorder. J Psychiatr Res 2016 Feb;73:53-62. [doi: 10.1016/j.jpsychires.2015.11.012] [Medline: 26687614]
53. Han SY, Tomasik J, Rustogi N, Lago SG, Barton-Owen G, Eljasz P, et al. Diagnostic prediction model development using data from dried blood spot proteomics and a digital mental health assessment to identify major depressive disorder among individuals presenting with low mood. Brain Behav Immun 2020 Nov;90:184-195 [FREE Full text] [doi: 10.1016/j.bbi.2020.08.011] [Medline: 32861718]
54. Tomasik J, Han SY, Barton-Owen G, Mirea D, Martin-Key NA, Rustogi N, et al. A machine learning algorithm to differentiate bipolar disorder from major depressive disorder using an online mental health questionnaire and blood biomarker data. Transl Psychiatry 2021 Jan 12;11(1):41 [FREE Full text] [doi: 10.1038/s41398-020-01181-x] [Medline: 33436544]
55. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, STARD Group. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. Clin Chem 2015 Dec;61(12):1446-1452. [doi: 10.1373/clinchem.2015.246280] [Medline: 26510957]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
AUC: area under the receiver operating characteristic curve
AUD: alcohol use disorder
BD: bipolar disorder
BN: bulimia nervosa
CAD–MDD: Computerized Adaptive Diagnosis for Major Depressive Disorder
CAT–ANX: Computerized Adaptive Test–Anxiety
CAT–DI: Computerized Adaptive Test–Depression Inventory
CAT–MH: Computerized Adaptive Test–Mental Health
CAT–Psychosis: Computerized Adaptive Test–Psychosis
CMFC: Connected Mind Fast Check
DSM: Diagnostic and Statistical Manual of Mental Disorders
EDQ–O: Eating Disorder Questionnaire–Online
e-PASS: electronic psychological assessment screening system
EUPD: emotionally unstable personality disorder
FQ: Fear Questionnaire
GAD: generalized anxiety disorder
GAD–2: 2-item Generalized Anxiety Disorder Scale
GAD–7: 7-item Generalized Anxiety Disorder Scale
GAD–SI: single-item Generalized Anxiety Disorder Scale
ICD: International Statistical Classification of Diseases and Related Health Problems
M-3: My Mood Monitor
MDD: major depressive disorder
OCD: obsessive-compulsive disorder
PHQ–9: 9-item Patient Health Questionnaire
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PTSD: posttraumatic stress disorder
QUADAS–2: Quality Assessment of Diagnostic Accuracy Studies 2
SAM: standardized assessment module
SISQ: single-item screening question
SUD: substance use disorder
WMH–ICS: World Health Organization World Mental Health International College Student
WSQ: Web-Based Screening Questionnaire

Edited by J Torous; submitted 17.08.21; peer-reviewed by M Gottschalk, D Guinart; comments to author 04.10.21; revised version received 28.10.21; accepted 11.11.21; published 30.03.22.

Please cite as:
Martin-Key NA, Spadaro B, Funnell E, Barker EJ, Schei TS, Tomasik J, Bahn S
The Current State and Validity of Digital Assessment Tools for Psychiatry: Systematic Review
JMIR Ment Health 2022;9(3):e32824
URL: https://mental.jmir.org/2022/3/e32824
doi:10.2196/32824
PMID:35353053

Original Paper

Informing the Future of Integrated Digital and Clinical Mental Health Care: Synthesis of the Outcomes From Project Synergy

Haley M LaMonica[1], BSc, MA, PhD; Frank Iorfino[1], BSc, MCMSc, PhD; Grace Yeeun Lee[1], MPH, MHM; Sarah Piper[1], GradDip, BSc; Jo-An Occhipinti[1], PhD; Tracey A Davenport[1], BA, eMBA; Shane Cross[1], BPsych, MPsych, PhD; Alyssa Milton[1], BSc, MAppSc, PhD; Laura Ospina-Pinillos[1], MD, PhD; Lisa Whittle[1], BSc; Shelley C Rowe[1], PhD; Mitchell Dowling[1], BPsych, PhD; Elizabeth Stewart[1], BA, DocClinPsych; Antonia Ottavio[1,2], DipHthSci, BN; Samuel Hockey[1], BA; Vanessa Wan Sze Cheng[1], PhD; Jane Burns[2], PhD; Elizabeth M Scott[1], MBBS; Ian B Hickie[1], AM, MD

[1]Brain and Mind Centre, The University of Sydney, Camperdown, Australia

[2]InnoWell Pty Ltd, Sydney, Australia

Corresponding Author:
Haley M LaMonica, BSc, MA, PhD
Brain and Mind Centre
The University of Sydney
94 Mallett Street
Camperdown, 2050
Australia
Phone: 61 426955658
Email: haley.lamonica@sydney.edu.au

Abstract

Background: Globally, there are fundamental shortcomings in mental health care systems, including restricted access, siloed services, interventions that are poorly matched to service users' needs, underuse of personal outcome monitoring to track progress, exclusion of family and carers, and suboptimal experiences of care. Health information technologies (HITs) hold great potential to improve these aspects that underpin the enhanced quality of mental health care.

Objective: Project Synergy aimed to co-design, implement, and evaluate novel HITs, as exemplified by the InnoWell Platform, to work with standard health care organizations. The goals were to deliver improved outcomes for specific populations under focus and support organizations to enact significant system-level reforms.

Methods: Participating health care organizations included the following: Open Arms–Veterans & Families Counselling (in Sydney and Lismore, New South Wales [NSW]); NSW North Coast headspace centers for youth (Port Macquarie, Coffs Harbour, Grafton, Lismore, and Tweed Heads); the Butterfly Foundation's National Helpline for eating disorders; Kildare Road Medical Centre for enhanced primary care; and Connect to Wellbeing North Coast NSW (administered by Neami National), for population-based intake and assessment. Service users, families and carers, health professionals, and administrators of services across Australia were actively engaged in the configuration of the InnoWell Platform to meet service needs, identify barriers to and facilitators of quality mental health care, and highlight potentially the best points in the service pathway to integrate the InnoWell Platform. The locally configured InnoWell Platform was then implemented within the respective services. A mixed methods approach, including surveys, semistructured interviews, and workshops, was used to evaluate the impact of the InnoWell Platform. A participatory systems modeling approach involving co-design with local stakeholders was also undertaken to simulate the likely impact of the platform in combination with other services being considered for implementation within the North Coast Primary Health Network to explore resulting impacts on mental health outcomes, including suicide prevention.

Results: Despite overwhelming support for integrating digital health solutions into mental health service settings and promising impacts of the platform simulated under idealized implementation conditions, our results emphasized that successful implementation is dependent on health professional and service readiness for change, leadership at the local service level, the appropriateness and responsiveness of the technology for the target end users, and, critically, funding models being available to support implementation. The key places of interoperability of digital solutions and a willingness to use technology to coordinate health care system use were also highlighted.

Conclusions: Although the COVID-19 pandemic has resulted in the widespread acceptance of very basic digital health solutions, Project Synergy highlights the critical need to support equity of access to HITs, provide funding for digital infrastructure and digital mental health care, and actively promote the use of technology-enabled, coordinated systems of care.

(JMIR Ment Health 2022;9(3):e33060) doi:10.2196/33060

KEYWORDS

mental health; technology; co-design; participatory research; health care reform; stakeholder participation; mobile phone

Introduction

Mental Health Service Reform

In 2014, the National Mental Health Commission invited the Young and Well Cooperative Research Centre (CRC) to provide advice about digital mental health solutions for young people [1]. Following interviews with mental health and industry leaders, the Young and Well CRC promoted the view that there was an urgent need to broaden research and development (R&D) focus from standalone mental health apps, digital health records, and other simple technologies (Digital Health 1.0) to highly coordinated integration of health information technologies (HITs) within more effective health care delivery systems (HIT Digital Health 2.0). That is, new systems of care would integrate both conventional and novel in-clinic and web-based mental health services, thereby broadening the rapid provision of scalable forms of high-quality care.

As a result, the National Mental Health Commission recommended that HITs be considered a cornerstone of mental health service reform in Australia. This would require that the mental health system be overhauled to embed digital health within conventional in-clinic services [2]. More recently, this recommendation was echoed by the Australian Productivity Commission review of mental health. It calls for increased use of technology through a national digital platform, emphasizing initial assessment and referral processes, as well as increasing service user access to a broader range of interventions and supports [3]. However, there is still a considerable need for increased practical knowledge as to how to best incorporate HITs into clinical services, as well as how to capitalize on the aggregate information available from HITs that could be used to drive more effective coordination of care, enhanced allocation of clinical resources, faster identification of risk, and greater emphasis on achieving a higher proportion of improved health, social, and economic outcomes.

Over the past 20 years, our research group has actively advocated for the use of technology to drive enhanced mental health care, and broader system reform, in Australia [4-8]. Although there is a growing evidence base supporting the use of digital technologies, the integration of digital solutions in mental health services remains relatively rare. As an example, Titov et al [9] have reported success in integrating internet-delivered cognitive behavioral therapy as part of clinical care with a mental health professional in 5 clinics internationally. Unfortunately, far more studies highlight barriers to successful implementations of digital health solutions [10-13], and further research is required to identify the most effective ways in which to design and implement evidence-based HITs as part of clinical care at scale.

Project Synergy

The Australian Government Department of Health commissioned Project Synergy in 2014 to be conducted by the Young and Well CRC in partnership with the University of Sydney's Brain and Mind Centre. However, following the completion of the Young and Well CRC's work in 2016, InnoWell, a joint venture enterprise between the University of Sydney and PwC (Australia), was established to deliver the remainder of the work. The overarching objective of Project Synergy is to explore how best to use such new HITs to drive mental health service reform toward delivering better outcomes for people experiencing mental ill-health [4].

Because of its scale, Project Synergy had 2 phases. Phase I (2014-2016; Aus $5.5 million [US $3.97 million]) investigated the use of digital (including smartphone) technologies to transform the provision of mental health services for Australian young people through co-designing, building, implementing, and evaluating prototypes with representative end user populations. These included young people along with their supportive others (eg, family, carers, and friends), health professionals, and service providers [4]. Four research studies were conducted on prototypical software and concepts that included the following: (1) a university-based health and well-being prototype [14]; (2) a synergized web-based system including a triage e-tool, health and well-being e-tool, and a *Mental Health eClinic* with a multidimensional web-based assessment, shared mental health dashboard, and *video visit* capabilities [15]; (3) wireframes for a tiered suicide risk management protocol [16]; (4) functionality for trusted supportive others [16-18]; (5) and a proof of concept study for implementing technologies into primary youth mental health services [19]. The details of these preliminary studies have been published previously [4].

The remainder of this paper focuses on phase II (2017-2021; Aus $30 million [US $21.63 million]), which concluded in June 2021. In phase II, prototypes from phase I were iterated into a beta version of a single web-based platform named the InnoWell Platform and implemented into both conventional in-clinic and web-based primary mental health services. The fundamental aim was articulated as delivering the right level of care to individuals upon their first presentation to service (*Right care, First time*) as underpinned by clinical stage, level of clinical need, and personal choice, with the aim of preventing progression to more serious complex and severe forms of illness.

InnoWell Platform

As detailed by Davenport et al [20], the InnoWell Platform collects both multidimensional self-report and informant (eg, supportive others and health professionals) information via

questionnaires to capture a holistic understanding of a service user's clinical presentation, level of need, and personal preferences, to then monitor their progress over time. Specifically, survey data were collected across a range of biopsychosocial health domains, such as psychological distress, depressed mood, sleep-wake cycle, physical health, and social and occupational functioning. Further objective behavioral data can also be collected via third-party integrations (eg, Fitbit fitness trackers [Google LLC]). As shown in Figure 1, the multidimensional assessment results are presented back to service users in a dashboard color coded with a traffic light system to ease interpretability (ie, green, yellow, orange, and red are used to denote increasing degrees of symptom severity or level of impairment for each health domain). Service users are also provided with a description of each health domain and a brief explanation of the result they received (Figure 2).

The assessment results can then be used collaboratively by the service user and their health professional to ensure both parties are informed and actively engaged in treatment decisions in a responsive way. This is designed to occur not only at the outset of the care journey but also on an ongoing basis, as service users and clinicians jointly track progress over time. Importantly, as shown in Figure 3, service users can also access recommended nonclinical care options (eg, fact sheets, apps, e-tools, and other web-based systems) to support their mental health and well-being proactively (eg, before receiving and in conjunction with the clinical care received through the service).

Figure 1. Example dashboard from the InnoWell Platform.

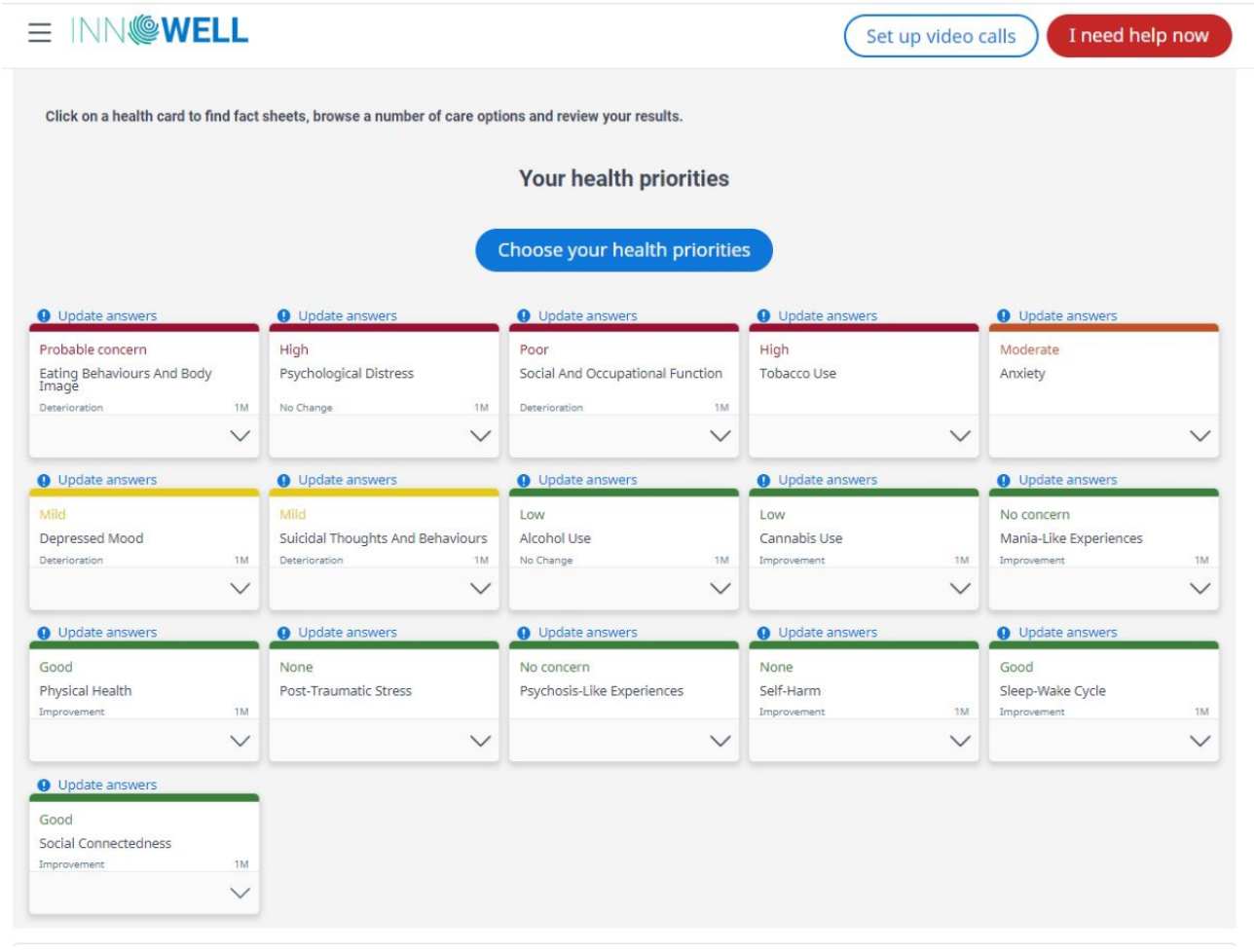


Figure 2. Example explanation of a health domain and assessment result from the InnoWell Platform.

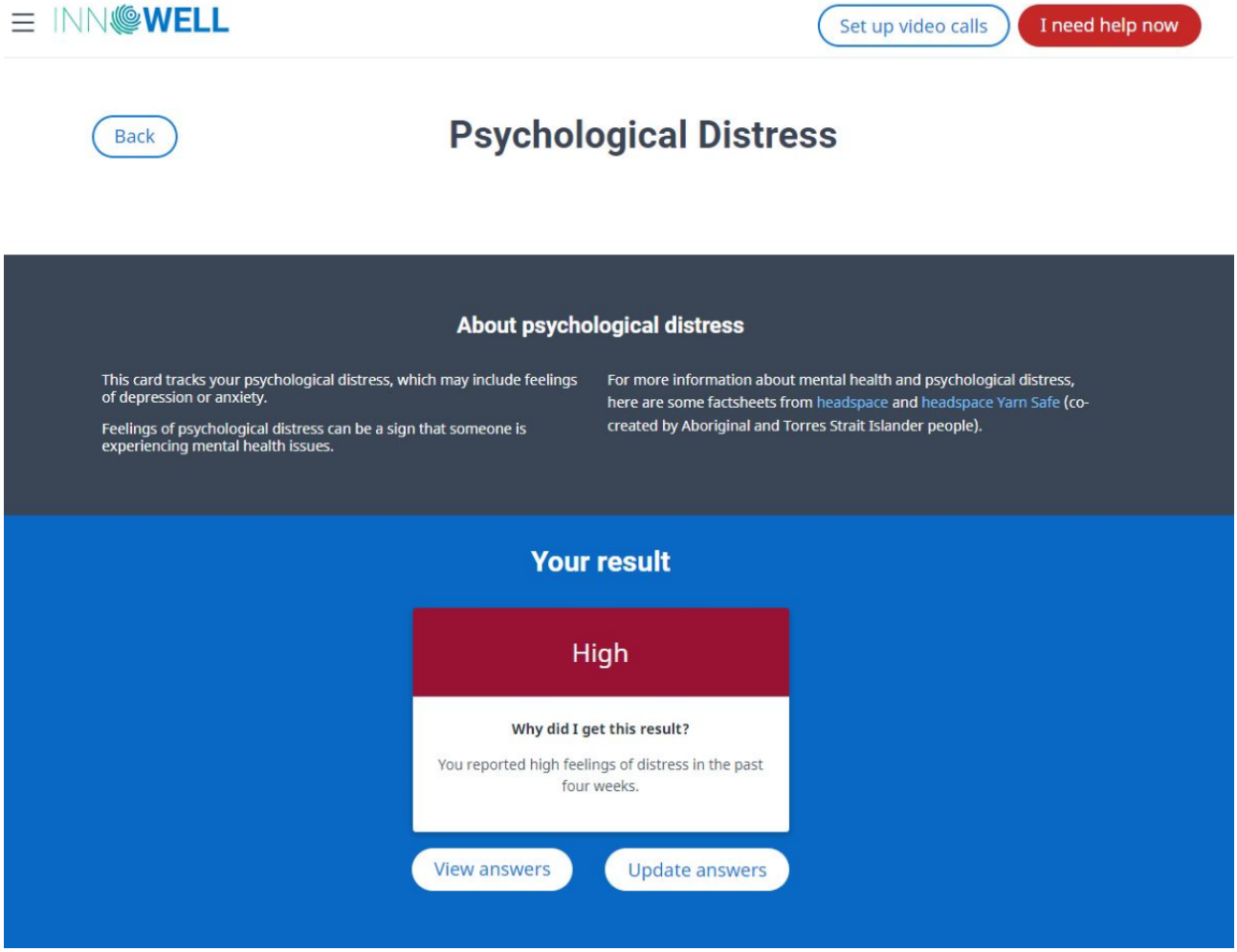
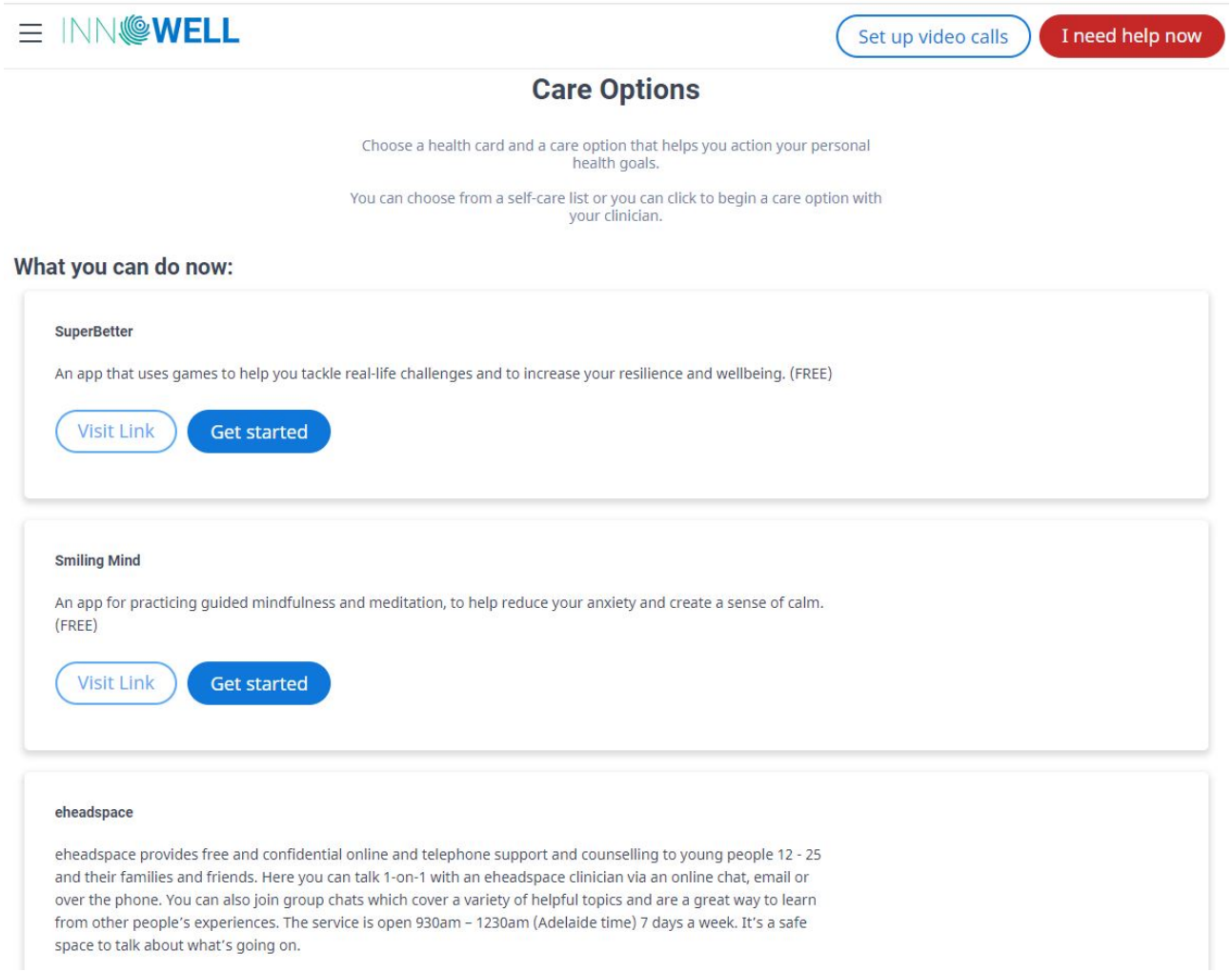


Figure 3. Example nonclinical care options from the InnoWell Platform.

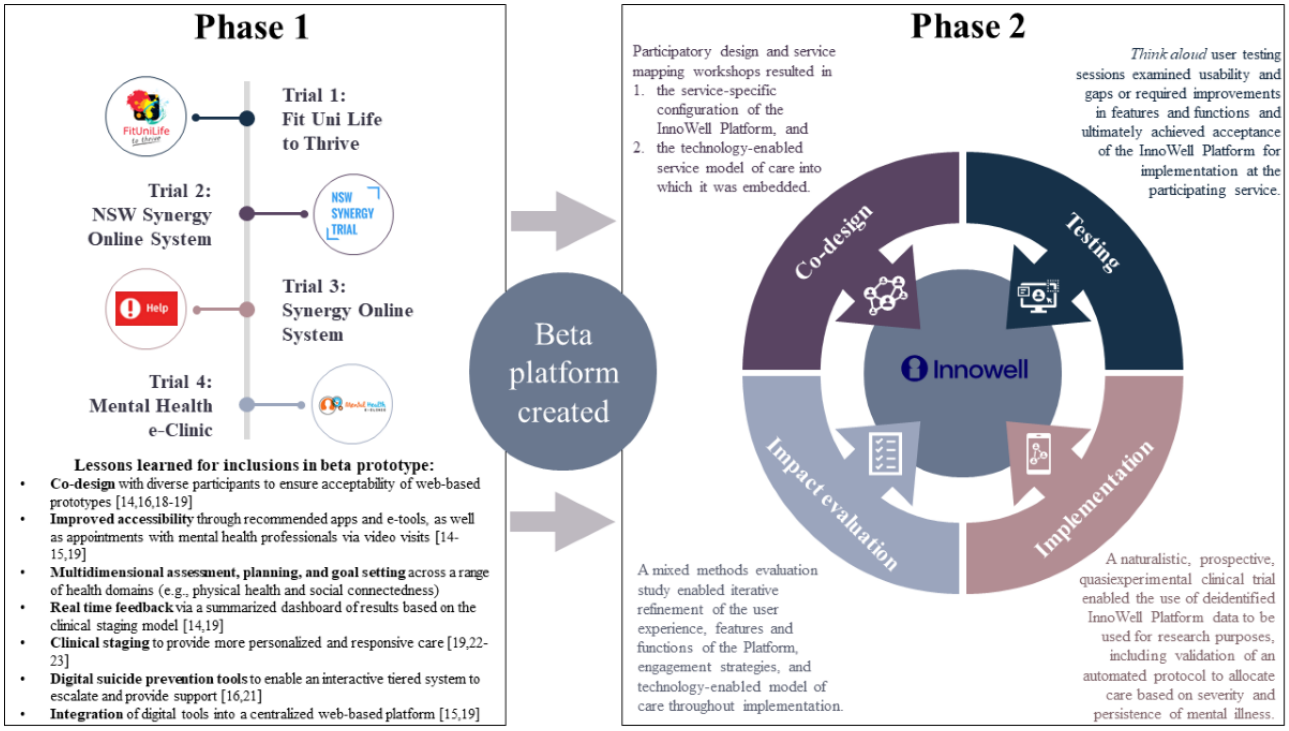


Objective

The InnoWell Platform was co-designed and implemented in diverse service settings across different age groups and populations. The Project Synergy effectiveness trials aimed to deliver critical insights about how HITs, exemplified by the InnoWell Platform, can be optimally configured and implemented as part of standard care to meet the needs of the populations under focus, their immediate care team (health professionals and service providers), and supportive others. Furthermore, a mixed methods approach was established to evaluate the web-based platform and the impact of its implementation (where available) in the individual services to iteratively refine the InnoWell Platform and the technology-enabled service model into which it was embedded.

As shown in Figure 4 [14-16,18-19,21-22,23], this paper seeks to summarize the methodologies used in phase II of Project Synergy and highlight their associated research outcomes. Specifically, we describe the process by which the original beta version of the InnoWell Platform was (1) iteratively co-designed, developed, and refined through participatory design workshops with individuals with lived experience of mental ill-health, health professionals, and service administrators to create service-specific configurations of the digital platform; (2) tested by end users to determine acceptance for each participating service or service user group; (3) implemented within a technology-enabled service model that had been co-designed with service staff through service mapping; and (4) evaluated to determine its impact at the level of the service user, health professional, and service. We provide an overview of the key outcomes from this research, highlight tools that can be used to drive the uptake of HITs by both service users and health professionals, and provide examples of how smart, digital technologies can drive innovation in models of care at the service and system level to better address demand management, early intervention, and rapid assessment and triage of at-risk users to specialized care.

Figure 4. A summary of the methodologies and key outcomes from phase II of Project Synergy [14-16,18-19,21-22,23]. NSW: New South Wales.



Methods

Overview of the Project Synergy Research Streams

Phase II of Project Synergy aimed to conduct 4 rolling research streams to evaluate the engagement with and efficiency and effectiveness of the InnoWell Platform. The research streams focused on four populations across the life span, namely, veterans and their families, young children and their families, young people, and older adults. Importantly, through the internal R&D process, these rolling research streams enabled the iterative development and redevelopment, evaluation, and refinement of essential functionality within the InnoWell Platform (eg, assessment, dashboard, algorithms, real time data tracking, aggregate service indicator dashboard, immediate *staged care* recommendations, and service pathways) over time. One or more research studies were planned in relation to each research stream, in collaboration with relevant mental health services when feasible or with specific participant groups (ie, young children and families). In relation to the former, potential participating services were identified based on the alignment between their service user group or groups and the research streams. After agreeing to participate in Project Synergy, a research agreement was formalized between the University of Sydney and the participating service. As described in this *Methods* section, the methodology that underpins Project Synergy's research trials includes co-design activities (ie, participatory design workshops, user acceptance testing, and service mapping sessions), implementation science, and a mixed methods evaluation of the impact of the InnoWell Platform and the associated technology-enabled model of care at the level of the individual, health professional, and service based on key performance indicators.

Participating Services and Service User Groups

The InnoWell Platform was co-designed with and tested by the target end users, including individuals seeking or engaged in mental health care. Digital health solutions are widely recognized as a potential means by which to address issues of demand management in mental health globally, and, to a degree, the COVID-19 pandemic fast-tracked the uptake of technology as part of mental health care. However, it is critical to recognize that factors such as age, technical skills, literacy, and health literacy affect the ways in which individuals engage with technology [24]. Despite the proliferation of digital health solutions, such as apps, in recent years, little attention has been paid to how these tools will be used by individuals from diverse populations [25]. In addition, only 30% of apps developed by established digital health companies have tested their apps with individuals with clinical conditions [26]. To ensure that the InnoWell Platform and technology-enabled model of care met the needs of the target end users, it was crucial to ensure that the research studies were conducted with individuals engaged in clinical care. Participating services and the service users to whom they provide care are presented in Textbox 1.

Textbox 1. Participating mental health services and their respective service user groups.

- Open Arms–Veterans & Families Counselling Sydney and Lismore, New South Wales (NSW), providing care to current and ex-serving military personnel and their families
- NSW North Coast *headspace* centers (Port Macquarie, Coffs Harbour, Grafton, Lismore, and Tweed Heads) providing mental health care to young people in regional NSW aged 12-25 years
- The Butterfly Foundation's National Helpline supporting those with eating disorders and negative body image via telephone, email, and web-based chat
- Kildare Road Medical Centre—a large general practice in Blacktown, NSW, with a socioeconomically and culturally diverse service user population, including a large number of Aboriginal and Torres Strait Islander service users
- Connect to Wellbeing North Coast NSW (administered by Neami National)—a community-based primary intake and referral service located within the footprint of the North Coast NSW Primary Health Network. Within the Australian context, Primary Health Networks are operated by not-for-profit companies and are responsible for commissioning and coordinating health services to meet the needs of the people in a designated region

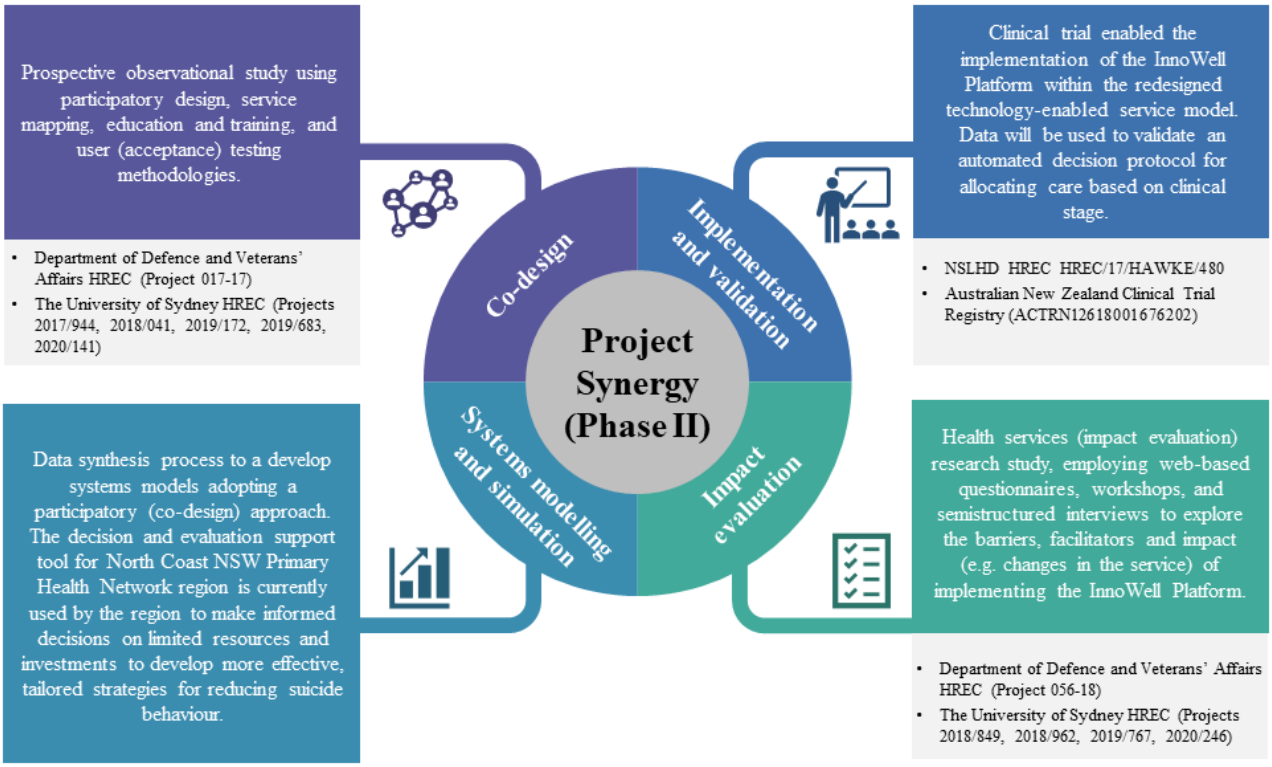
Additional co-design and preliminary evaluation research studies were conducted with specific service user groups not affiliated with a mental health service including older adults (aged ≥50 years) and children (aged ≥8 years) and families to inform future iterations of the InnoWell Platform. These participants were recruited through active research clinics at the University of Sydney's Brain and Mind Centre as well as via nongovernmental and private organizations associated with the Brain and Mind Centre. The co-design studies were advertised via posters and postcards at the advertising site as well as via their websites and social media pages. Importantly, although a diverse grouping of services and service users participated in Project Synergy, further investigation is still required to determine the usability, acceptability, and relevance of the InnoWell Platform for individuals who are from a culturally and linguistically diverse background, identify as Aboriginal and Torres Strait Islanders, or identify as having a disability.

Finally, approximately 50 local stakeholders from the North Coast NSW Primary Health Network, including representatives from health and social policy agencies, nongovernment organizations, primary care providers, emergency services, research institutions, community groups, and, importantly, people with lived experience of suicide, contributed to the development of a system dynamics model. Using a broad systems perspective, stakeholders identified key outcomes of interest and subsequently provided feedback to iteratively refine the model.

Ethics Approval

This research required multiple ethics approvals by various human research ethics committees (HRECs) owing to the diverse organizational structures governing each of the participating services. The clinical trial allowing users of the InnoWell Platform to share their deidentified data for research purposes was approved by the Northern Sydney Local Health District HREC (reference HREC/17/HAWKE/480) and was registered with the Australian New Zealand Clinical Trial Registry (ACTRN12618001676202). Separate ethics were sought for the co-design and impact evaluation research for each participating service. The governing bodies of some mental health services required applications to be submitted through their own internal HRECs (ie, the Department of Defence and Veterans' Affairs HREC Projects 056-18 and 017-17), whereas others preferred that required applications were submitted through the University of Sydney HREC (Projects 2017/944, 2018/849, 2018/041, 2018/962, 2019/172, 2019/683, 2019/767, 2020/141, and 2020/246). The development of the system dynamics model did not require ethics approval as it involved secondary analysis of a broad range of aggregate data sets and synthesis of existing research. Stakeholders were invited to contribute their expertise to this model but were not the subject of the research activity. Figure 5 outlines the diverse research methodologies that were used as part of Project Synergy, underpinning all aspects of the project from co-design and development of the InnoWell Platform through to impact evaluation and dissemination of outcomes.

Figure 5. Research methodologies and ethics approvals underpinning Project Synergy. HREC: human research ethics committee; NSW: New South Wales; NSLHD: Northern Sydney Local Health District.



Co-design

An R&D cycle, described by Davenport et al [17], was established and used to explicitly position end users as empowered participants in all stages of the research—from planning and design through to implementation and evaluation. Central to the R&D cycle are participatory design workshops to discover, prototype, and evaluate potential design solutions through the use of prompted discussion, review of working prototypes (wireframes) and existing InnoWell Platform functionality, personas (ie, a fictitious potential user for whom the product is being designed), and user journeys. Direct engagement with potential end users as well as other key stakeholders (ie, representatives from funding agencies) through targeted participatory workshops is essential to ensure the development and build of acceptable, usable, and scalable digital tools. The primary objective of the workshops was to co-design and configure the InnoWell Platform for each participating service or user group to ensure that it met the needs of the target end users, including people with lived experience accessing care and their supportive others as well as health professionals, service managers, and administrators.

Workshop agendas were tailored to each service by the research management team, comprising researchers, health professionals, and service administrators, to determine how best to discover how the InnoWell Platform might enhance standard care. Areas of focus included increased accessibility to care, waitlist management, risk identification, coordination of care with other services, and the potential for digital tools to support self-management and well-being. Participants worked together to create mockups of the technology, including the landing page, assessments, and dashboard, highlighting key features and functions that they thought would address identified *pain points* for the service and service users. Only after having captured participants' unbiased ideas were they then provided with paper wireframes or screenshots of the InnoWell Platform on which to provide specific feedback on prototyped or existing features and functionality. The final outcome of the workshops was a service-specific configuration of the InnoWell Platform.

Importantly, the service-specific version of the InnoWell Platform was then tested by end users in one-to-one, face-to-face, 90-minute user testing sessions. Each session was facilitated by a member of the research team, and a scribe was present to take notes to capture all feedback accurately. The aim of the user testing sessions was to (1) learn if participants were able to navigate the InnoWell Platform intuitively, (2) identify any features of the platform that the user would want to see changed before implementation of the platform into the service (ie, showstoppers), (3) gain insight into participants' experience of using the platform, (4) identify possibilities for improving the platform to better meet user needs and wants, and (5) learn how best to implement the platform into the participating service. Testers were observed interacting with the InnoWell Platform and asked to *think aloud* when exploring its features, providing both positive and negative feedback about the user experience. In addition to freely exploring the features and functionality of the InnoWell Platform, testers were also asked to complete specific scenarios relevant to their role (eg, service user, health professional, and service manager). For example, a health professional might be asked to try entering data into the InnoWell Platform about a service user and then to comment on their experience (ie, ease of completing the action and gaps in the functionality). The feedback was integrated back into the iterative R&D cycle to inform the

ongoing refinement of the product by our collaborators at InnoWell. Furthermore, user testing was also used to achieve *acceptance* of the technology, signally that it was ready for implementation within the service.

Finally, in parallel with the participatory design workshops and user testing sessions, our team also used service mapping, a process mapping methodology, to understand how best to integrate the InnoWell Platform within participating mental health services. As published in LaMonica et al [27], we have used the term *service mapping* to refer to the structured approach to understanding the needs, existing processes, gaps in care, and performance of a mental health service mapped against an individual's journey through the service. More specifically, participants worked collaboratively with the researchers to first map the current service user journey through the participating service on a whiteboard or butcher's paper. The gaps or limitations of the current care pathway were reviewed relative to the key features of high-quality mental health care to identify areas for improvement or reform. The existing map was then revised to include technical element supported by the InnoWell Platform (eg, web-based initial assessment and clinical and nonclinical care options). The outcome was a co-designed technology-enabled model of care highlighting the intersection and impact of the InnoWell Platform at key points in care (eg, referral, assessment, treatment planning, and routine outcome monitoring).

Table 1 summarizes the research aims and co-design activities for each of our participating services and service user groups.

Table 1. Summary of Project Synergy co-design activities.

Population	Service partner	Research aim	Co-design activities
Current and ex-serving Australian Defence Force personnel and their families	Open Arms–Veterans & Families Counselling Sydney and Lismore, NSW[a]	Co-design, develop, implement, and evaluate the InnoWell Platform to achieve improved outcomes	• 4 participatory design workshops (n=21)[b] • 2 service mapping workshops (n=21) • 18 user testing sessions
Young people aged 12-25 years residing in the NSW North Coast	NSW North Coast *headspace* centers (Port Macquarie, Coffs Harbour, Grafton, Lismore, and Tweed Heads)	Enhance quality and timeliness of mental health services for NSW North Coast youth	• 11 participatory design workshops (n=48) • 9 service mapping workshops (n=46) • 30 user testing sessions
Australians with eating disorders and body image issues	The Butterfly Foundation's National Helpline	Evaluate, prototype, and configure a digital system for a nonconventional (web-based, phone, and email) service providing support for eating disorders and body image issues through the use of co-design processes	• 6 participatory design workshops (n=47) • 2 service mapping workshops (n=12) • 3 user testing sessions
NSW North Coast residents	Connect to Wellbeing North Coast NSW (administered by Neami National)	Configure and implement the InnoWell Platform to support community-based primary intake and referrals across the North Coast NSW Primary Health Network	• 1 user testing sessionc
Service users in a community general practice setting	Kildare Road Medical Centre, Blacktown, NSW	Configure and implement the InnoWell Platform to support Kildare Road Medical Centre's delivery of mental health care	• 2 service mapping workshops (n=6)c
Older adults (aged ≥50 years)	N/A[d]	Co-design and configure the InnoWell Platform to evaluate the accessibility, engagement, and appropriateness of the technology for an older age group	• 4 participatory design workshops (n=21) • 19 older adults consented to provide user feedback on the InnoWell Platform after using it for up to 90 days
Younger children and families	N/A	Co-design a configuration of the InnoWell Platform that is appropriate for younger children and their families	• 3 participatory design workshops (n=15)

[a]NSW: New South Wales.

[b]A total of 9 cofacilitated participatory design workshops, including 2 service mapping workshops, were conducted in 2017 with current and former military personnel and their families as well as Open Arms' health professionals, service managers, and administrators from Sydney, Canberra, Maitland, Singleton, and Port Stephens [28]. Importantly, the outcomes from this collaboration served as the basis for the configuration of the InnoWell Platform with supplementary participatory workshops being conducted in 2018 specifically as part of Project Synergy.

[c]Further information informing the service-specific configurations and user acceptance of the InnoWell Platform was captured during routine project management meetings with each of the participating services.

[d]N/A: not applicable.

Implementation and the Clinical Trial

The effectiveness of the InnoWell Platform is being assessed in a naturalistic, prospective, quasiexperimental (ie, uncontrolled, nonrandomized) clinical trial. It is important to note that this methodology is rapidly gaining recognition as a practical and efficient method of assessing change over time in users under ecologically valid conditions (ie, real-world conditions of constantly changing apps and technologies) [29,30]. Importantly, by conducting group comparisons based on clinical stage, demographics, and locality, it is possible to infer strong evidence for the usefulness (or lack thereof) of the technology-enabled blended model of care.

Regarding implementation and conduct of the clinical trial, the co-designed and configured InnoWell Platform was implemented at each of the participating services within the redesigned, technology-enabled service model as part of standard practice in accordance with our established implementation science strategy [31]. Importantly, standard practice was defined by the service; in other words, participating services had the opportunity to implement the InnoWell Platform in varying ways depending on what best suited their service and their service user base. All service users aged ≥14 years who presented to a participating service were eligible to use the InnoWell Platform as part of their care. Upon creating an account in the InnoWell Platform, users were presented with a *Terms of Use* agreement informing them that their deidentified data collected by the InnoWell Platform would be used for research purposes unless they *opt out* [32]. In 2007, the Australian National Statement of Ethical Conduct in Human Research was updated to enable HRECs to approve an *opt-out* consent for low-risk clinical research [33]. *Opt-out* approaches require researchers to ensure that participants are aware of their potential inclusion in research and given the opportunity to indicate that they do not want their data used in this way (ie, thus opting out). Importantly, service users were able to update their data sharing permissions in the InnoWell Platform at any time, enabling them to, for example, withdraw from the clinical trial should they wish to do so. As the InnoWell Platform is an exempt medical device, it is not required to be included in the Australian Register of Therapeutic Goods and can be used by service users regardless of whether they opt out of the clinical trial.

In addition to the sharing of data for research purposes, service users are also informed that their data will be shared with the participating service at which they are receiving care. This information is made explicit to the user at the point at which the InnoWell Platform is originally explained to them by a staff member from the participating service using a simple script delivered via telephone, email, chat, or in person, and in the subsequent email invitation they receive inviting them to use the InnoWell Platform. In addition, this information is documented in the InnoWell *Privacy Policy*, which is available on the landing page, before creating an account and providing any personal information.

The primary objective of the above-mentioned health services trial is to validate an automated decision protocol for allocating care based on severity and persistence of mental health symptoms (ie, based on the clinical stage) [34] relative to the clinical stage as allocated by health professionals, multidisciplinary teams via consensus, and an expert clinical reference group [20]. The protocol aims to automate critical decision points to triage service users based on clinical stage. Specifically, the first decision determines whether there is clear evidence of at least one full-threshold, major discrete disorder, or a persistent and recurrent syndrome or syndromes, whereas the second decision aims to determine whether the syndrome is nonspecific or attenuated. By identifying the clinical stage and level of need, service users can then be automatically triaged to self-management strategies including self-directed apps and e-tools, ambulatory care services, or acute or specialized services at the point of service entry. As detailed by Sawrikar et al [35], clinical staging avoids the *fail first* approach typified by stepped care in favor of allocating the *Right care, First time*. The clinical trial allows for the collection of deidentified data for research purposes, and cross-sectional findings from the participating services are provided in Multimedia Appendix 1 [36-40]; however, the findings from the clinical trial are beyond the scope of this paper and will be published separately.

An automated suicide escalation protocol is embedded within the InnoWell Platform, designed to identify service users reporting suicidal thoughts or behaviors and, importantly, suggest pathways to appropriate and timely care [21]. Specifically, when a service user expresses suicidal thoughts or behaviors on the InnoWell Platform's multidimensional assessment, they are immediately provided with a pop-up message containing contact information for relevant 24-hour crisis support services (ie, Lifeline and Kids Helpline). In addition, to enable a rapid clinical response, a notification is also sent to the clinical service when the degree of risk meets a predetermined severity threshold established by the service for their service user base.

Impact Evaluation

Traditional clinical science approaches to the development and implementation of interventions rely on a lengthy linear process, from basic science through to effectiveness trials and dissemination [41], which can result in delays of up to 17 years for research translation into clinical practice [42]. Given the rapid advances that occur in the technology sector, this is not a practical approach to the evaluation of HITs or technology-enabled models of care. To that end, Mohr et al [29] highlight the benefits of evaluating the challenges and successes associated with tools such as the InnoWell Platform in targeted clinical settings, such as through the Project Synergy participating services, to iteratively refine and optimize usability, acceptability, and effectiveness to ensure that it meets identified clinical objectives. Therefore, as detailed previously by LaMonica et al [10], we used a mixed methods impact evaluation study, including web-based surveys, semistructured interviews, and workshops, to identify potential barriers to and facilitators of implementation and to evaluate the impact of the InnoWell Platform on the participating mental health services over time, including (1) digital literacy and competence of the service staff in relation to the implementation of the HIT in the service; (2) changes in the service in association with the implementation of the HIT-enabled solution; and (3) the quality,

usability, and acceptability of the solution. All participating staff involved in the implementation of the InnoWell Platform at each of the respective services, including health professionals, service managers, and administrators, were invited to participate in the impact evaluation study. Where staff from the services' funding or governing bodies (ie, service providers or primary health networks) were associated with implementation, these staff were also invited to participate and were required to provide written informed consent. The impact evaluation data were collected at baseline and every 3 months thereafter for the duration of the implementation, the length of which varied based on agreements made with participating services. Importantly, the use of mixed methods allowed the service staff to provide feedback in a manner that suited their work schedule (ie, web-based helpline operating on a 16-hour schedule vs traditional face-to-face counseling services operating during normal business hours), thus ensuring both breadth (via surveys) and depth (via the semistructured interviews and workshop) of evaluation from a broad range of participants.

Participatory Systems Modeling

As referenced previously, a system dynamics model was developed using a participatory approach that involved approximately 50 local stakeholders, including representatives from health and social policy agencies, nongovernment organizations, primary care providers, emergency services, research institutions, community groups, and, importantly, people with lived experience of suicide [43]. Over a series of 3 workshops, stakeholders were asked to draw on the deep tacit knowledge and diverse perspectives of these system actors to map the local health system and the role that social determinants play in psychological distress and suicidal behaviors. These workshops focused on prioritization of the key outcomes of interest for the model, the mapping of pathways and drivers of those outcomes, and the prioritization of interventions to be included in the model. A full description of this process, the model, and its results can be found in the study by Occhipinti et al [43].

Results

Co-design

The results of most of our team's co-design work have been published previously and will therefore not be presented in detail [27,28,36-38,44]. However, a summary of the key insights derived from the participatory design workshops is provided in Table 2, including findings regarding the (1) support for HITs as part of clinical care, (2) critical factors to consider in relation to the design and content of HITs, and (3) potential for incongruity between health professional and service manager attitudes and beliefs toward HITs during co-design relative to implementation.

In addition to the participatory design outcomes, service mapping also proved an effective way to engage key stakeholders in reflecting on existing service and system pathways and, in turn, identifying current barriers to and facilitators of quality mental health care. As highlighted in LaMonica et al [27], by working collaboratively with participating services, service mapping highlighted key gaps in care that might be improved through the implementation of the InnoWell Platform, while also identifying aspects of care that could be improved through changes to specific service-level processes (ie, intake). In this way, the service mapping process prevented a *one-size-fits-all* approach to service reform, ensuring that specific contextual factors impacting the quality of care at each service were considered, thus optimizing the service delivery. Each participating service was able to determine how best to integrate the InnoWell Platform as part of standard care as it pertained to the respective services. For example, at Open Arms and most *headspace* services, service users were invited to complete the InnoWell Platform's multidimensional assessment before their first appointment, whereas at Kildare Road Medical Centre (KRMC), a general practice, service users were only invited to the InnoWell Platform after having an appointment with a mental health nurse who determined whether the digital tool was appropriate for their individual care needs. Importantly, our research shows that the process of mapping technology-enabled service models can lead to improvements in service delivery pathways, both in relation to and independent of the technology [28].

Table 2. Key insights learned through participatory design workshops.

Key insights	Specific findings
Overwhelming support for integrating HITs[a] into mental health care across populations and service settings [4,27,28,36-38,44].	• The results of a lived-experience-led national community consultation program showed considerable service user interest in and experience with digital health solutions. On the basis of data collected through 5 digital engagements, 8 face-to-face consultations, and 2 community engagement events, 81% of respondents indicated they were comfortable sharing mental health experiences on the web and 94% had already done so [45].
Service users and carers further emphasized the importance of being provided reputable (academic, government, or nonprofit mental health organization) care options and information [37].	• The results of 4 participatory design workshops conducted with 21 community dwelling older adults (aged ≥50 years), including carers, highlighted the need for information delivered via HITs to come from a credible source to be perceived as trustworthy and reliable [37].
Concerns with data privacy and security of personal and health information were prevalent across all service user groups [28,37,46]; however, health professionals questioned whether all users would be wary of security risks.	• Young people emphasized the need for privacy information to be readily available to allow a user to be completely comfortable when entering sensitive information into a HIT ("Always ask, could this site be more secure with my information." [Member of the headspace Youth Reference Group, participatory design workshop in Wollongong, August 30, 2018]). • Given the ubiquity of technology use by young people, some health professionals questioned whether they would be suitably concerned about sharing personal data ("Young people may be so used to this as they have grown up with it that they would not see sharing personal data as a major barrier to accessing a health and wellbeing e-tool" [Health Professional, participatory design workshop in Coffs Harbour, December 4, 2018]). • Members of the military community emphasized the need to be fully transparent in relation to limitations to confidentiality and data sharing ("Data security needs to be highlighted, particularly that the information is not shared with DVA [Department of Veterans' Affairs]." [Veteran, participatory design workshop, August 24, 2017]) [28]. • Older adults indicated that data privacy and security risks are a primary barrier to the use of HITs ("Anything on the Internet I just don't really trust, I don't want to put my information of any kind out there." [Older Adult, participatory design workshop in Sydney October 9, 2019]) [37].
There is a gap in what is expressed during the co-design process and actual implementation; often, clinicians are very active and willing supporters during the co-design process, however not in practice [47].	• The 48 participants, including young people, supportive others, health professionals, service managers, and administrators, from 10 participatory design workshops with headspace services in the North Coast PHNb, recognized the potential for HITs to improve service quality and efficiency; however, a qualitative review of 70 fortnightly logs completed by on-the-ground implementation officers working across 5 headspace centers revealed persistent resistance to change [47].

[a]HIT: health information technology.

[b]PHN: Primary Health Network.

Implementation and the Clinical Trial

Overview

To address known barriers, as well as those that might be specific to individual services, the R&D team developed a protocol for the implementation of digital solutions into mental health services, including digital tools and technology-enabled models of care [31]. The protocol supports active collaboration between the researchers, health services, and service providers and users and is intended to promote mitigation strategies that might enable the successful translation of new innovations into standard clinical practice. This not only includes a strong emphasis on co-design but also acknowledges that a well-designed technology product is not sufficient to drive improved outcomes or mental health reform. Therefore, the protocol highlights the need to (1) upskill the mental health workforce regarding the evidence-based digital, clinical, service, and safety elements essential to quality mental health care, as well as how digital solutions can support the delivery of such care; (2) provide on-the-ground support to both service users and staff in the early stages of implementation to troubleshoot any issues with the technology and to help interpret and report clinical data in a meaningful way; and (3) refine iteratively the digital solution, as well as the technology-enabled model of care, in response to service-level changes in outcomes and processes or user feedback on the quality, usability, and acceptability of the solution. In relation to the latter, it is crucial to continue to refine digital solutions to ensure they become an integral part of standard care, enabling improved outcomes on key service performance indicators. Concurrently, minimizing increased burden on service staff associated with the implementation is challenging. However, striking this balance may afford the best opportunity for the sustainability and scalability of the solution.

The naturalistic, quasiexperimental clinical trial to validate the InnoWell Platform is in progress. At the time of this publication,

1644 service users aged ≥14 years consented to share their data for research purposes. Cross-sectional data from the initial multidimensional assessment for each participating service are provided in Multimedia Appendix 1. Further analysis of the data collected by the InnoWell Platform will be conducted at the service-specific level and is beyond the scope of this paper. Further, the data to validate the automated decision protocol for allocating care based on clinical stage will be analyzed and published separately. Importantly, however, the automated suicide escalation protocol embedded within the InnoWell Platform has been shown to effectively identify service users requiring urgent risk assessment. For example, 22.6% (24/106) of the young people using the InnoWell Platform as part of their care at a *headspace* service in the regional North Coast NSW Primary Health Network footprint reported high levels of suicidal thoughts and behaviors (STBs), thus triggering an immediate notification to the service. A similar proportion (18/62, 29%) of users from Open Arms underwent a rapid risk assessment after the service received a notification of moderate to high levels of STB as self-reported via the assessments in the InnoWell Platform. Although the management of such notifications can be challenging, services recognized the need for this functionality, as it fast-tracks care.

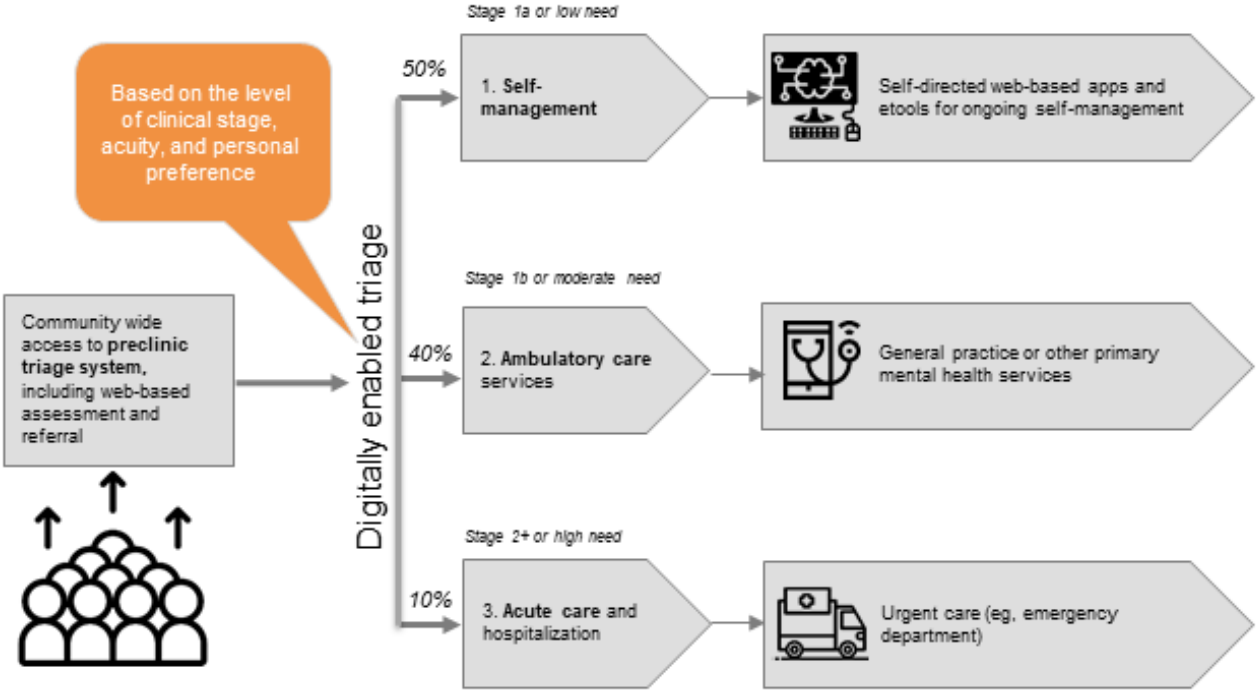
Innovative Models of Enhanced Primary Care

The implementation science strategies resulted in an innovative model of enhanced primary care, best suited to local implementation. Specifically, KRMC is a large Western Sydney family practice with 26 general practitioners (GPs) seeing an average of 25 to 30 service users per day. As part of their engagement in Project Synergy, KRMC used a mental health nurse to support their GPs with assessment and triage, treatment planning and intervention, and case review of those who presented with symptoms suggesting mental ill-health. The mental health nurse was a critical facilitator of the co-designed technology-enabled service delivery model at KRMC. For example, the application of the InnoWell Platform's suicide escalation protocol [21] identified 13% (17/131) of those service users as having STBs, thus enabling the mental health nurse to triage these individuals rapidly to appropriate care. The mental health nurse also reviewed the initial assessment results in collaboration with each service user to guide a discussion about available care options, considering service user preferences. Importantly for KRMC, the technology-enabled model of care, as led by the mental health nurse, resulted in expanded treatment options, including the availability of brief psychological therapies conducted by the mental health nurse as well as the apps and e-tools embedded within the InnoWell Platform. Furthermore, the mental health nurse was also able to use the multidimensional assessment to draft a mental health care plan for later review by the GP. Consequently, this reduced the assessment time required by the GP, allowing them to focus instead on delivering *value-added* care to more service users. In light of the success of this pilot study, KRMC sought funding to support the ongoing implementation of this innovative technology-enabled, mental health nurse-led service delivery model.

An Innovative Model of Enhanced Community-Based Assessment and Triage

As previously reported [39], the R&D team partnered with Neami National (who deliver the central community-based intake service of the NSW North Coast Primary Health Network—*Connect to Wellbeing North Coast*) to co-design and implement a new technology-enabled assessment and triage system for the range of services offered [39]. Specifically, the InnoWell Platform's multidimensional assessments were used to (1) identify service users with risk factors and symptoms or impairment (stage 1a) who were immediately triaged to self-management including self-directed and clinician-supported apps and e-tools, (2) assist those with attenuated or subthreshold major disorders (stage 1b+) to clinical care within ambulatory care services including general practice or another primary care service, and (3) provide immediate clinical assessment for those with a risk of self-harm and determine the need for acute care or hospitalization. This model could serve as a prototype for a direct-to-service user preclinic triage system (Figure 6). Such a system is scalable and directly addresses the Productivity Commission's call to develop a sustainable national digital platform to facilitate assessment and referral processes that match levels of mental health care more directly with service users' actual needs [3].

Figure 6. Preclinic triage system enabled by the InnoWell Platform (as originally published by Davenport et al [45]).



Impact Evaluation

As shown in Table 3, data from the impact evaluation results highlight consistent agreement regarding the potential for HITs to improve outcomes for service users and mental health services alike. However, participants, including health professionals, service managers, and administrators, indicated that technologies are generally not well integrated into current service delivery models. Service readiness for change (eg, existing technology infrastructure and the digital literacy of staff and service users) was noted to be a potential barrier to successful implementation, with less than half of respondents indicating that their service was ready to implement new technologies to enhance mental health care. Furthermore, there was considerable variability among clinical staff as to whether it was their responsibility to recommend technology as part of standard care.

The impact evaluation findings were used to iteratively refine the technology-enabled model of care throughout the implementation of the InnoWell Platform at each participating service. For example, based on findings from the evaluation activities, services modified (1) the point in the care journey at which the InnoWell Platform was offered to participants, (2) the required qualifications for the person presenting the platform to service users for the first time (eg, mental health nurse, counselor, or clinical psychologist rather than administrative staff), (3) the technology available to health professionals at the service to be able to best use the InnoWell Platform actively during sessions, and (4) processes by which the data collected by the InnoWell Platform were used to inform care. Furthermore, the findings were fed back to service leadership to help find ways to better support health professionals and administrative staff to use the InnoWell Platform effectively and efficiently as part of standard care. Finally, feedback regarding the usability and acceptability of the InnoWell Platform as well as identified gaps or limitations in its features and functionality were fed back to InnoWell as part of the internal R&D cycle to inform the development or redevelopment and refinement of the technology.

Table 3. Impact evaluation outcomes.

Key outcome	Specific findings
Staff across multiple service settings consistently support the use of technology as part of their work; however, they also list digital literacy of both service users and health professionals as well as service readiness for change as potential barriers to widespread adoption [10].	• A total of 81% (38/47) of health professionals and service administrators reported benefits of using technology as part of their work. • Most staff (26/45, 57%) questioned whether their service users' digital literacy was sufficient to use technology as part of their mental health care; however, of potential users, young people, who were considered digital natives, were expected to be most likely to access and adopt technologies as part of care. • Although approximately two-thirds (27/45, 60%) of staff indicated that their service's policies support the belief that technologies can improve service user outcomes by providing more efficient and effective care, only 44% (20/45) of service staff indicated that their service was ready to implement new technologies to enhance mental health care. • Furthermore, only 53% (24/45) of staff reported that their service actively encourages the integration of technologies as part of standard care.
There must be organizational leadership (PHN[a]-level decision-making) as well as a local champion at the service level to support a successful implementation [4].	• Qualitative feedback was collected from 40 staff from across 5 headspace centers in the Central Eastern PHN who were involved in implementing a prototype of a web-based mental health clinic, known as the Mental Health eClinic. Their feedback highlighted the significant benefit of both (1) high-level endorsement and coordination at the PHN level and 2) the presence and engagement of on-the-ground leadership to assist in solving day-to-day implementation challenges [4]. • Qualitative analysis of 70 logs completed by on-the-ground implementation officers at 5 headspace services implementing the InnoWell Platform showed a consistent interest among service staff in implementing the InnoWell Platform as a demand management tool and to provide better care; however, a lack of strong and deliberate leadership was highlighted as a barrier to change. Notably, this finding did not change appreciably in response to COVID-19 [47]. • To facilitate a successful implementation, service leadership need to establish clear strategies to mitigate potential barriers to implementation identified by on-the-ground staff responsible for driving the implementation [10].
Health professionals are often confident about the effectiveness of their current service models (ie, business as usual) and express reluctance to change their usual practices [10].	• A primary implementation barrier identified by health professionals relates to concerns that digital tools could replace clinical expertise; however, all participants denied this as a personal concern.

[a]PHN: Primary Health Network.

Participatory Systems Modeling of the North Coast Primary Health Network Mental Health Services

The R&D team was engaged to work with the North Coast Primary Health Network to model the mental health services available across the region and test the relative impacts of a range of interventions on mental health and prevention of suicide. This work was critical for Project Synergy, as one of the primary interventions of interest was technology-enabled care coordination. The capacity to compare the impact of this system-level intervention with other clinical and health system alternatives was a major advance. The model was validated against historic time series data (2011-2017), and then projected population-level trajectories of suicidal behavior in the region using a 20-year time horizon (2021-2041) were derived. Against the baseline trajectory, implementation of technology-enabled care coordination was projected to decrease suicide deaths by 5.6% (95% CI 4.8%-6.5%) [43]. The model also found that the most effective combination of interventions to reduce suicidal behavior was supporting social connectedness, technology-enabled coordinated care, postsuicide attempt assertive aftercare, reductions in childhood adversity, and increasing youth employment [43]. This combination of interventions was projected to reduce self-harm hospitalizations (indicative of suicide attempts) by 28.5% (95% CI 26.3%-30.8%) and suicide deaths by 29.3% (95% CI 27.1%-31.5%) among a youth population (aged 15-24 years). Introducing additional interventions beyond this best performing suite of interventions produced only marginal improvement. That is, "more is not necessarily better" [43]. Further analysis demonstrated that technology-enabled care coordination led to greater reductions in suicide deaths, the total number of self-harm hospitalizations or mental health–related ED hospitalizations, and the prevalence of high psychological distress in the population, when accompanied by increases in service capacity growth by 20% (including standard telehealth practices [48]). Findings from these simulation models indicate the significant promise offered by HITs when effective implementation is achieved.

Tools to Support the Adoption of Digital Solutions in Clinical Practice

As noted previously, service and health professional readiness for change are 2 key factors that impact the implementation of digital solutions. To help build confidence in the use of apps and e-tools in practice, the R&D team adapted the original Mobile App Rating Scale (MARS; a reliable and internationally recognized app rating system [49]) to ensure it is appropriate for health-related apps as well as e-tools (eg, websites and web-based courses). It is now known as the A-MARS [50]. A

complimentary quality assurance protocol was also developed to guide health professionals in the evaluation of the quality and safety of health-related apps and e-tools to determine their appropriateness for use in clinical practice. This approach includes (1) a broad exploration for available apps and e-tools; (2) shortlisting of those health-related apps and e-tools deemed to match the needs of the service users, health professionals, or service; (3) evaluation using the newly developed A-MARS; and (4) review of the ratings compared with service-specific criteria to determine appropriateness for recommendation.

Given the consistent concerns regarding data security and privacy highlighted in our co-design work, the R&D team also developed a privacy risk assessment tool to assess whether current apps and e-tools are meeting privacy standards [46]. Our review revealed consistently poor readability, resulting in marked limitations in the transparency of the information presented. This, in turn, can undermine a service user's trust in the privacy of their personal and health information. To assist health professionals and service providers in understanding potential privacy risks, we developed easy-to-use guidelines for their consideration before promoting individual apps and e-tools as part of care [46]. It is our recommendation that these guidelines be adopted to ensure that HITs are used to their full potential to maximize service user health outcomes while minimizing risk and users are informed of privacy and security considerations to be able to make educated decisions as to how they would like to share their personal and health information [46].

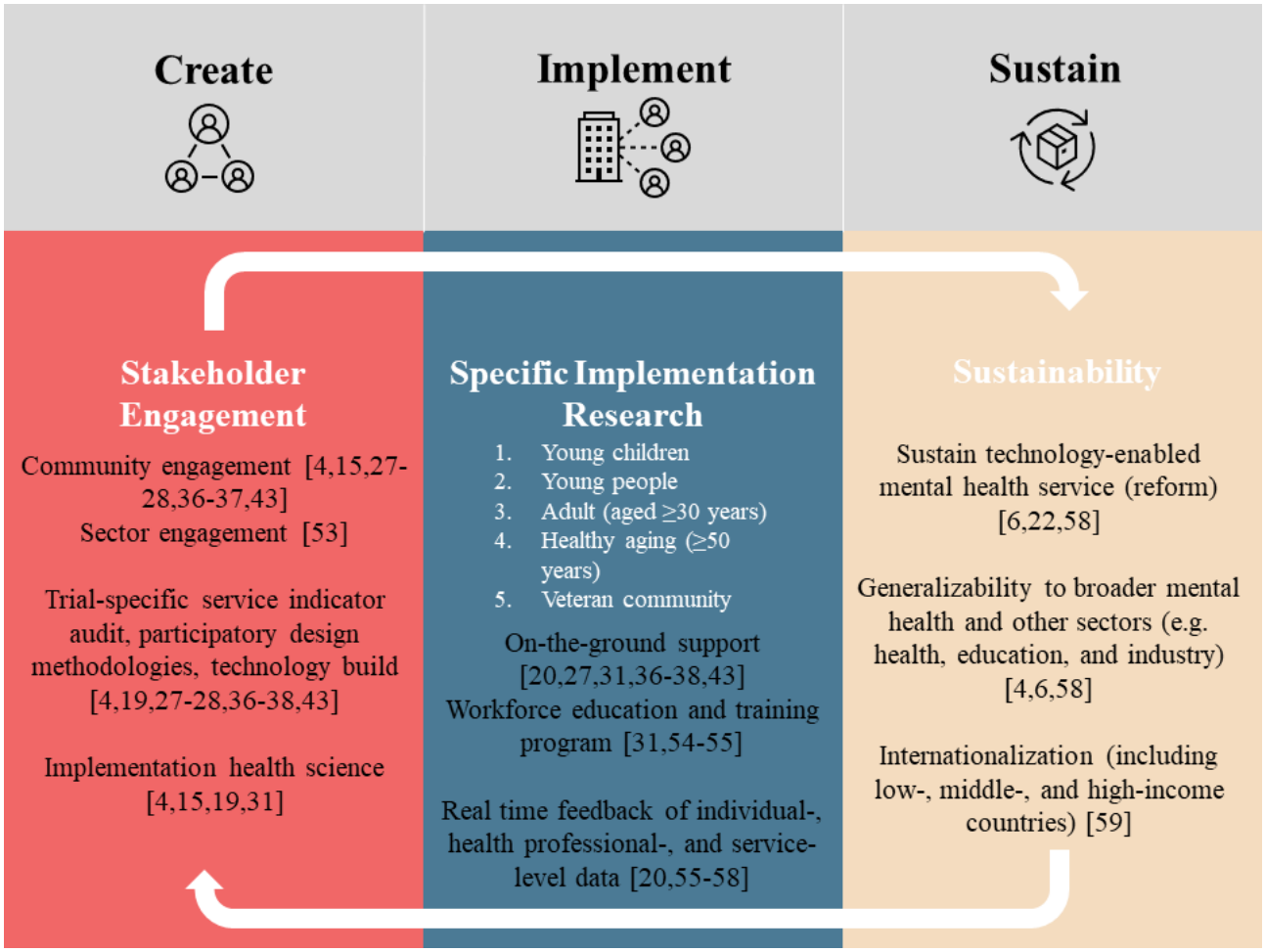
Discussion

Principal Findings

Recent enhanced funding for mental health care in Australia has been focused largely on increasing access to traditional clinic-based psychological therapy [51], with little to no emphasis on delivering quality care matched to the level of need, coordinating systems of care, or accounting for the unequal distribution of resources across urban, regional, and rural communities. Although the increased investment in mental health included in the Australian Government Federal Budget of May 2021 signals recognition of the need for system reform, there is a relative lack of funding dedicated to service innovation, health systems research, or evaluation [52]. These omissions are particularly concerning given evidence highlighting that technology-enabled care coordination has the potential to differentially improve health outcomes (ie, reductions in deaths by suicide, mental health–related hospitalizations, and rates of high psychological distress in the community) and assist with demand management over and above any impact owing to a 20% increase in service capacity and standard telehealth services [48]. Furthermore, technology-enabled coordinated care in conjunction with programs targeting the social determinants of health would appear to offer a good return on investment, particularly in regions with limited health resources [43].

With the stage set by the population-level modeling outcomes, Project Synergy has now delivered critical outcomes regarding the practicalities of designing, implementing, and evaluating digital mental health solutions (Figure 7 [4,6,15,19-20,22,27-28,31,36-38,43,53-59]). These outputs are critical pieces needed to inform delivery of effective technology-enabled and coordinated mental health care, as called for by the Productivity Commission [3] and the World Economic Forum [60]. We now have the methodologies, tools, and essential insights to build robust, dynamic, sustainable, scalable, and coordinated systems of mental health care. This has the capacity to firmly establish Australia as a leader in the delivery of digital mental health care.

Figure 7. Critical learnings from Project Synergy [Figure 7 [4,6,15,19-20,22,27-28,31,36-38,43,53-59]].



On the basis of the findings from Project Synergy, it is critical to ensure that all digital solutions and innovative models of technology-enhanced systems of care be developed in collaboration with the intended target audiences, including those with lived experience of mental ill-health. However, it is important to recognize that the co-design process must also consider the needs and time demands on health professionals. Historically, health professionals have been reluctant to drive service reforms, frequently arguing instead simply for more of the same (ie, *endless therapy*) [61]. Similarly, our results highlight the potential for health professionals to be reluctant to implement digital tools as part of standard practice. To that end, research indicates that health professionals are most likely to accept changes that they are prepared for and that they value, especially those that are likely to result in direct benefits to service users [62]. Furthermore, it is essential to provide health professionals with the necessary education and training to support their understanding and use of digital health solutions. It needs to be very clear that the primary goal is to deliver high-quality and person-centered care, not simply operational or financial efficiency. Until digital health care is included as part of the core curriculum for standard health professional education, services are also likely to benefit from including a digital navigator on their care team. These new personnel can evaluate the suitability of available digital solutions for the service, troubleshoot technical difficulties, and assist in interpreting and reporting data collected via digital tools in a meaningful way to inform care and self-management strategies [63].

As existing digital infrastructure is so poorly integrated, health professionals are frequently required to enter the same data across multiple systems, including client management software, electronic medical records, minimum data sets, and personal health records. Only by reducing this administrative burden are health professionals likely to engage with new digital solutions; therefore, there is an urgent need to ensure open application program interfaces are built into all digital solutions to enable interoperability or the transfer of relevant personal and health information between various systems. Such integration can inform shared decision-making, track progress over time in coordination with and between care teams and supportive services, and empower the service user, along with families and carers, to be informed and active partners in care.

Systematically collecting multidimensional assessment data through self-report and clinical digital tools, as well as wearables and smartphone sensors, will provide the opportunity to examine service user outcomes, health professional practices, and service-level performance in relation to key indicators, including safety, acceptability, efficiency, accessibility, effectiveness, coordination, and workforce capability [64]. More specifically, assessment data enables services to understand the clinical complexity of those service users presenting for care. This can ensure that the best care options are offered immediately, as

well as promoting coordination of care with other relevant specialist or support services. This type of data may also help services provide necessary training opportunities for their workforce. Furthermore, routine outcome monitoring enables data-informed decisions regarding the effectiveness or lack thereof of treatment plans to facilitate necessary changes or referrals to specialist care as needed. Finally, data linkage at the state and national levels, using data from health records (ie, Medicare Benefit Scheme and Pharmaceutical Benefits Scheme), provides the necessary information for analysis of the costs and effectiveness of various models of care. Taken together, these methods of evaluation offer an opportunity for accountability in service delivery.

Having successfully delivered the necessary building blocks to overhaul the mental health care system through Project Synergy, it is now time to translate these insights into action. With the ever-increasing demand for more personalized and higher-quality services [65], it is clear that simply increasing investment into various clinical services delivered in isolation will not be enough to improve the outcomes for the whole system; instead, new models of integrated care, and the digital infrastructure to support them, is needed to accompany these new clinical investments. Project Synergy supports calls for increased use of digital technologies in mental health care; however, it also highlights that the way these technologies are deployed really matters. Specifically, dynamic systems modeling shows that when digital technologies are limited to standard telehealth practices (ie, by extending existing services on the web), without changing the underlying model of care, the impact is low [48]. Although standard telehealth will now be implemented more widely, following the COVID-19 pandemic, to date, little effort has been made to use these technologies in ways that promote multidimensional team-based care or maximize the benefits that these technologies can really provide. By capitalizing on digital tools and investing in digital infrastructure, we now have the opportunity to recalibrate the whole mental health system, resulting in a greater impact on outcomes as opposed to those expected by simply improving the capacity across individual components of the existing mental health system.

Conclusions

Although the Australian Government continues to invest in more clinical services, the well-documented failings of the mental health system (notably restricted and delayed access to quality care, siloed services, inadequate use of routine outcome monitoring, and care plans that do not match a service user's level of need), may be left unaddressed. Project Synergy highlights the power of more innovative, digitally enhanced systems to increase the efficiencies of health care systems by addressing demand management, reducing delays in access to appropriate care, enabling deployment of early intervention strategies, and rapidly assisting at-risk service users to access acute and specialist care. System-level innovation can be achieved via the engagement of the appropriate service leadership and active promotion of the interoperability of the new technologies.

Acknowledgments

This research was conducted on behalf of the Australian Government Department of Health as part of Project Synergy (2017-2021). InnoWell was formed by the University of Sydney and PricewaterhouseCoopers (PwC) Australia to deliver the Aus $30 million (US $21.63 million) Australian Government–funded Project Synergy.

Project Synergy was made possible by many researchers over the lifetime of the project. Sincerest thanks go to Amelia English, William Capon, Dominic FitzGerald, Tanya Jackson, Cristina Ricci, Anna E Roberts, Vilas Sawrikar, Joan Torony, and Hannah Yee for their contributions. The authors would also like to thank participants from Open Arms–Veterans & Families Counselling (Sydney and Lismore), the North Coast Primary Health Network, the *headspace* centers (Coffs Harbour, Grafton, Lismore, Port Macquarie, and Tweed Heads), Neami National, Connect to Wellbeing, Butterfly Foundation's National Helpline 1800 ED HOPE, and Kildare Road Medical Centre as well as older adults, young children, and family members for their involvement in this research. In addition, the authors acknowledge and thank the Project Synergy Lived Experience Advisory Function led by Ms Sue Muller for their contribution to this research. Finally, the authors thank their collaborators at InnoWell, including former CEOs Ayse Burke and Will Fellowes, as well as the Board of Directors chaired by Michael Ronaldson with Deputy Alan Griffin, Kristin Stubbins from PwC (Australia), and Professor Laurent Rivory from the University of Sydney, for supporting the research objectives of Project Synergy.

Authors' Contributions

Authors IBH, TAD, SC, and JB were integral in securing funding to support the study. The studies encompassed by Project Synergy were designed by HML, FI, JO, TAD, SC, AM, LOP, LW, SCR, MD, ES, and TO with subsequent contributions by GYL, SP, SH, and EMS. Data analyses were conducted by HML, FI, AM, SCR, and MD, with contributions from VWSC and oversight and advice from IBH. The system dynamics modeling was undertaken by Dr Adam Skinner (Research Fellow, Brain and Mind Centre, University of Sydney) with contributions from authors JO and FI. All authors contributed to and have approved the final manuscript.

Conflicts of Interest

IBH is codirector of Health and Policy at the Brain and Mind Centre (BMC) at the University of Sydney. The BMC operates an early-intervention youth services at Camperdown under contract to headspace. He is the Chief Scientific Advisor to, and a 5% equity shareholder in, InnoWell Pty Ltd. InnoWell was formed by the University of Sydney (45% equity) and PwC (Australia; 45% equity) to deliver the Aus $30 million (US $21.63 million) Australian Government–funded Project Synergy (2017-2020; a 3-year program for the transformation of mental health services) and to lead the transformation of mental health services internationally through the use of innovative technologies. JO is both Head of Systems Modelling, Simulation & Data Science at the University of Sydney's BMC and Managing Director of Computer Simulation & Advanced Research Technologies. TAD is now Director (Research and Evaluation) of the Design and Strategy Division, Australian Digital Health Agency. EMS is Principal Research Fellow at the BMC, University of Sydney. She is Discipline Leader of Adult Mental Health, School of Medicine, University of Notre Dame, and a consultant psychiatrist. She was the Medical Director at the Young Adult Mental Health Unit, St Vincent's Hospital Darlinghurst, until January 2021. She has received honoraria for educational seminars related to the clinical management of depressive disorders supported by Servier and Eli-Lilly pharmaceuticals. She has participated in a national advisory board for the antidepressant compound Pristiq, manufactured by Pfizer. She was the National Coordinator of an antidepressant trial sponsored by Servier. The source of funding does not entail any potential conflict of interest for the other members of the Project Synergy Research and Development Team.

Multimedia Appendix 1
Data from the InnoWell Platform for Participating Services.
[DOCX File , 469 KB - mental_v9i3e33060_app1.docx]

References

1. Burns J. NMHC Advice-on-innovative-technologies-in-eMental-health. National Mental Health Commission Review. 2015. URL: https://www.researchgate.net/publication/283010660_NMHC_Advice-on-Innovative-Technologies-in-eMental-Health [accessed 2022-02-20]
2. National Review of Mental Health Programmes and Services – Preliminary analysis. National Mental Health Commission. URL: http://mhaustralia.org/fact-sheets/national-review-mental-health-programmes-and-services-preliminary-analysis [accessed 2022-02-20]
3. Productivity Commission Report no.95. Canberra, ACT, Australia: Productivity Commission; 2020.
4. Hickie IB, Davenport TA, Burns JM, Milton AC, Ospina-Pinillos L, Whittle L, et al. Project Synergy: co-designing technology-enabled solutions for Australian mental health services reform. Med J Aust 2019 Oct;211 Suppl 7:S3-39. [doi: 10.5694/mja2.50349] [Medline: 31587276]
5. Groom G, Hickie I, Davenport T. 'Out of Hospital, Out of Mind!' : A Report Detailing Mental Health Services in Australia in 2002 and Community Priorities for National Mental Health Policy for 2003-2008. Canberra: Mental Health Council of Australia; 2003.
6. Hickie IB. Moving beyond stepped care to staged care using a novel, technology-enabled care model for youth mental health. Med J Aust 2019 Nov 03;211(9):404-405. [doi: 10.5694/mja2.50379] [Medline: 31679168]
7. Burns JM, Birrell E, Bismark M, Pirkis J, Davenport TA, Hickie IB, et al. The role of technology in Australian youth mental health reform. Aust Health Rev 2016 Nov;40(5):584-590. [doi: 10.1071/AH15115] [Medline: 26934382]
8. Rosenberg S, Hickie IB, Mendoza J. National mental health reform: less talk, more action. Med J Aust 2009 Feb 16;190(4):193-195. [doi: 10.5694/j.1326-5377.2009.tb02346.x] [Medline: 19220184]
9. Titov N, Dear B, Nielssen O, Staples L, Hadjistavropoulos H, Nugent M, et al. ICBT in routine care: a descriptive analysis of successful clinics in five countries. Internet Interv 2018 Sep;13:108-115 [FREE Full text] [doi: 10.1016/j.invent.2018.07.006] [Medline: 30206525]
10. LaMonica HM, Milton A, Braunstein K, Rowe SC, Ottavio A, Jackson T, et al. Health information technology-enabled solutions for Australian mental health services reform: impact evaluation. JMIR Form Res 2020 Mar (forthcoming). [doi: 10.2196/preprints.18759]
11. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of eHealth interventions: systematic review of the literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
12. Vis C, Mol M, Kleiboer A, Bührmann L, Finch T, Smit J, et al. Improving implementation of eMental health for mood disorders in routine practice: systematic review of barriers and facilitating factors. JMIR Ment Health 2018 Mar 16;5(1):e20 [FREE Full text] [doi: 10.2196/mental.9769] [Medline: 29549072]
13. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
14. Davenport T, Ospina-Pinillos L, Ricci C, Milton A, Burns J, Hickie I. Study 1: Fit Uni Life to thrive: an online health and wellbeing prototype for young people. Med J Aust 2019;211(7 Suppl).

15. Ospina-Pinillos L, Davenport TA, Ricci CS, Milton AC, Scott EM, Hickie IB. Developing a mental health eClinic to improve access to and quality of mental health care for young people: using participatory design as research methodologies. J Med Internet Res 2018 May 28;20(5):e188 [FREE Full text] [doi: 10.2196/jmir.9716] [Medline: 29807878]
16. Davenport T, Milton A, Whittle L, McLoughlin L, Ospina-Pinillos L, Mendoza J, et al. Study 3: co-designing, user testing and evaluating digital suicide prevention functionality. Med J Aust 2019;7:S24-S29.
17. Davenport TA, Milton AC, Ospina-Pinillos L, Whittle L, Ricci CS, Burns JM, et al. Project Synergy research and development cycle: iterative processes of participatory design, user testing, implementation and feasibility testing. Med J Aust 2019 Oct;211 Suppl 7.
18. Davenport T, Whittle L, Ospina-Pinillos L, Milton A, Burns J, Hickie I. Evaluating engagement with a prototypic online platform to improve the mental health and wellbeing of young people living in disadvantaged communities. Med J Aust 2019(Suppl (7)):S17-S23.
19. Cross S, Piper S, Davenport T, Milton A, Iorfino F, Rici C, et al. Implementation study of a prototypic e-clinic being integrated into youth mental health services: staff experiences and reported service quality improvements. Med J Aust 2019;7:S30-S36.
20. Davenport TA, LaMonica HM, Whittle L, English A, Iorfino F, Cross S, et al. Validation of the InnoWell platform: protocol for a clinical trial. JMIR Res Protoc 2019 May 31;8(5):e13955 [FREE Full text] [doi: 10.2196/13955] [Medline: 31152524]
21. Iorfino F, Davenport TA, Ospina-Pinillos L, Hermens DF, Cross S, Burns J, et al. Using new and emerging technologies to identify and respond to suicidality among help-seeking young people: a cross-sectional study. J Med Internet Res 2017 Jul 12;19(7):e247 [FREE Full text] [doi: 10.2196/jmir.7897]
22. Cross S, Davenport T, Scott E. A service delivery model to support highly personalised and measurement-based care in youth mental health. Med J Aust 2019;211(9):S42-S46.
23. Hickie IB. The role of new technologies in monitoring the evolution of psychopathology and providing measurement-based care in young people. World Psychiatry 2020 Feb;19(1):38-39 [FREE Full text] [doi: 10.1002/wps.20697] [Medline: 31922686]
24. Hall AK, Bernhardt JM, Dodd V, Vollrath MW. The digital health divide: evaluating online health information access and use among older adults. Health Educ Behav 2015 Apr 25;42(2):202-209 [FREE Full text] [doi: 10.1177/1090198114547815] [Medline: 25156311]
25. Sarkar U, Gourley GI, Lyles CR, Tieu L, Clarity C, Newmark L, et al. Usability of commercially available mobile applications for diverse patients. J Gen Intern Med 2016 Dec;31(12):1417-1426 [FREE Full text] [doi: 10.1007/s11606-016-3771-6] [Medline: 27418347]
26. Winkle BV, Carpenter N, Moscucci M. Why aren't our digital solutions working for everyone? AMA J Ethics 2017 Nov 01;19(11):1116-1124 [FREE Full text] [doi: 10.1001/journalofethics.2017.19.11.stas2-1711] [Medline: 29168683]
27. LaMonica HM, Davenport TA, Ottavio A, Rowe SC, Cross SP, Iorfino F, et al. Optimising the integration of technology-enabled solutions to enhance primary mental health care: a service mapping study. BMC Health Serv Res 2021 Jan 15;21(1):68 [FREE Full text] [doi: 10.1186/s12913-021-06069-0] [Medline: 33451328]
28. LaMonica HM, Davenport TA, Burns J, Cross S, Hodson S, Veitch J, et al. Technology-enabled mental health service reform for open arms - veterans and families counselling: participatory design study. JMIR Form Res 2019 Sep 19;3(3):e13662 [FREE Full text] [doi: 10.2196/13662] [Medline: 31538937]
29. Mohr DC, Schueller SM, Riley WT, Brown CH, Cuijpers P, Duan N, et al. Trials of intervention principles: evaluation methods for evolving behavioral intervention technologies. J Med Internet Res 2015 Jul 08;17(7):e166 [FREE Full text] [doi: 10.2196/jmir.4391] [Medline: 26155878]
30. Fleming TM, de Beurs D, Khazaal Y, Gaggioli A, Riva G, Botella C, et al. Maximizing the impact of e-therapy and serious gaming: time for a paradigm shift. Front Psychiatry 2016;7:65 [FREE Full text] [doi: 10.3389/fpsyt.2016.00065] [Medline: 27148094]
31. LaMonica HM, Davenport TA, Braunstein K, Ottavio A, Piper S, Martin C, et al. Technology-enabled person-centered mental health services reform: strategy for implementation science. JMIR Ment Health 2019 Sep 19;6(9):e14719 [FREE Full text] [doi: 10.2196/14719] [Medline: 31538938]
32. Innowell website terms of use. InnoWell. URL: http://www.innowell.org/website-terms/ [accessed 2022-02-20]
33. The Australian Research Council and Universities Australia. National Statement on Ethical Conduct in Human Research 2007 (Updated 2018). Canberra: National Health and Medical Research Council; 2007.
34. Iorfino F, Cheng VW, Cross SP, Yee HF, Davenport TA, Scott EM, et al. Right care, first time: developing a theory-based automated protocol to help clinically stage young people based on severity and persistence of mental illness. Front Public Health 2021;9:621862 [FREE Full text] [doi: 10.3389/fpubh.2021.621862] [Medline: 34513775]
35. Sawrikar V, Stewart E, LaMonica HM, Iorfino F, Davenport TA, Cross S, et al. Using staged care to provide "right care first time" to people with common affective disorders. Psychiatr Serv 2021 Jun;72(6):691-703. [doi: 10.1176/appi.ps.202000145] [Medline: 33765863]
36. Milton AC, Hambleton A, Dowling M, Roberts AE, Davenport T, Hickie I. Technology-enabled reform in a nontraditional mental health service for eating disorders: participatory design study. J Med Internet Res 2021 Feb 16;23(2):e19532 [FREE Full text] [doi: 10.2196/19532] [Medline: 33591283]

37. LaMonica HM, Davenport TA, Roberts AE, Hickie IB. Understanding technology preferences and requirements for health information technologies designed to improve and maintain the mental health and well-being of older adults: participatory design study. JMIR Aging 2021 Jan 06;4(1):e21461 [FREE Full text] [doi: 10.2196/21461] [Medline: 33404509]
38. LaMonica HM, Roberts AE, Davenport TA, Hickie IB. Evaluation of the usability and acceptability of the InnoWell platform as rated by older adults: survey study. JMIR Aging 2021 Apr 21;4(2):e25928 [FREE Full text] [doi: 10.2196/25928] [Medline: 33881410]
39. Davenport TA, LaMonica HM, Rowe S, Sturgess J, Scott EM, Lee G, et al. Innovative preclinic triage system to guide Australians to the right mental health care first time. Aust Health Rev 2021;45(5):578. [doi: 10.1071/ah20233]
40. Weathers FW, Litz BT, Keane TM, Palmieri PA, Marx BP, Schnurr PP. The PTSD Checklist for DSM-5 (PCL-5). US Department of Veterans Affairs. 2013. URL: https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp [accessed 2022-02-11]
41. Onken LS, Carroll KM, Shoham V, Cuthbert BN, Riddle M. Reenvisioning clinical science: unifying the discipline to improve the public health. Clin Psychol Sci 2014 Jan 01;2(1):22-34 [FREE Full text] [doi: 10.1177/2167702613497932] [Medline: 25821658]
42. Balas E, Boren S. Managing clinical knowledge for health care improvement. Yearb Med Inform 2018 Mar 07;09(01):65-70. [doi: 10.1055/s-0038-1637943]
43. Occhipinti J, Skinner A, Iorfino F, Lawson K, Sturgess J, Burgess W, et al. Reducing youth suicide: systems modelling and simulation to guide targeted investments across the determinants. BMC Med 2021 Mar 12;19(1):61 [FREE Full text] [doi: 10.1186/s12916-021-01935-4] [Medline: 33706764]
44. Rowe SC, Davenport TA, Easton MA, Jackson TA, Melsness J, Ottavio A, et al. Co-designing the InnoWell Platform to deliver the right mental health care first time to regional youth. Aust J Rural Health 2020 Apr;28(2):190-194. [doi: 10.1111/ajr.12617] [Medline: 32281183]
45. National community consultation program final report. InnoWell Lived Experience Advisory Function Working Group. 2020. URL: https://www.innowell.org/reports/project-synergy/nccp-june2020-final.pdf [accessed 2022-02-20]
46. LaMonica HM, Roberts AE, Lee GY, Davenport TA, Hickie IB. Privacy practices of health information technologies: privacy policy risk assessment study and proposed guidelines. J Med Internet Res 2021 Sep 16;23(9):e26317 [FREE Full text] [doi: 10.2196/26317] [Medline: 34528895]
47. LaMonica HM, Roberts AE, Jackson TA, Rowe S, Hickie IB. The InnoWell Platform: Challenges in integrating and implementing health information technologies into youth regional mental health care. 2021 Presented at: Digital Health & Informatics Network; Feb 10, 2021; Virtual conference.
48. Iorfino F, Occhipinti J, Skinner A, Davenport T, Rowe S, Prodan A, et al. The impact of technology-enabled care coordination in a complex mental health system: a local system dynamics model. J Med Internet Res 2021 Jun 30;23(6):e25331 [FREE Full text] [doi: 10.2196/25331] [Medline: 34077384]
49. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
50. Roberts A, Davenport T, Wong T, Moon C, Hickie I, LaMonica H. Evaluating the quality and safety of health-related apps and e-tools: adapting the Mobile App Rating Scale and developing a quality assurance protocol. Internet Interv 2021 Apr;24:100379 [FREE Full text] [doi: 10.1016/j.invent.2021.100379] [Medline: 33777705]
51. Budget 2020-21: prioritising mental health – doubling better access initiative sessions. Australian Government Department of Health. 2020. URL: https://www.health.gov.au/resources/publications/budget-2020-21-prioritising-mental-health-doubling-better-access-initiative-sessions [accessed 2022-02-20]
52. Budget 2021 - 2022: budget strategy and outlook. Commonwealth of Australia. URL: https://parlinfo.aph.gov.au/parlInfo/search/display/display.w3p;query=Id%3A%22library%2Fbudget%2F2021_01%22 [accessed 2022-02-20]
53. Burns JM, Davenport TA, Durkin LA, Luscombe GM, Hickie IB. The internet as a setting for mental health service utilisation by young people. Med J Aust 2010 Jun 07;192(S11):S22-S26. [doi: 10.5694/j.1326-5377.2010.tb03688.x] [Medline: 20528703]
54. Ospina-Pinillos L, Davenport T, Iorfino F, Tickell A, Cross S, Scott EM, et al. Using new and innovative technologies to assess clinical stage in early intervention youth mental health services: evaluation study. J Med Internet Res 2018 Sep 10;20(9):e259 [FREE Full text] [doi: 10.2196/jmir.9966] [Medline: 30201602]
55. Davenport TA, Cheng VW, Iorfino F, Hamilton B, Castaldi E, Burton A, et al. Flip the clinic: a digital health approach to youth mental health service delivery during the COVID-19 pandemic and beyond. JMIR Ment Health 2020 Dec 15;7(12):e24578 [FREE Full text] [doi: 10.2196/24578] [Medline: 33206051]
56. Ellis LA, McCabe KL, Rahilly KA, Nicholas MA, Davenport TA, Burns JM, et al. Encouraging young men's participation in mental health research and treatment: perspectives in our technological age. Clin Investigation 2014 Oct;4(10):881-888. [doi: 10.4155/cli.14.61]
57. Ellis LA, Collin P, Hurley PJ, Davenport TA, Burns JM, Hickie IB. Young men's attitudes and behaviour in relation to mental health and technology: implications for the development of online mental health services. BMC Psychiatry 2013 Apr 20;13:119 [FREE Full text] [doi: 10.1186/1471-244X-13-119] [Medline: 23601273]

58. Cross S, Davenport T, LaMonica H, Hickie I. Potential of real-time and integrated clinical data to drive continuous quality improvement in youth mental health services. Med J Aust 2019;211(7):S37-S39.
59. Ospina-Pinillos L, Davenport T, Mendoza Diaz A, Navarro-Mancilla A, Scott EM, Hickie IB. Using participatory design methodologies to co-design and culturally adapt the Spanish version of the mental health eClinic: qualitative study. J Med Internet Res 2019 Aug 02;21(8):e14127 [FREE Full text] [doi: 10.2196/14127] [Medline: 31376271]
60. Empowering 8 billion minds: enabling better mental health for all via the ethical adoption of technologies. World Economic Forum. URL: https://www.weforum.org/whitepapers/empowering-8-billion-minds-enabling-better-mental-health-for-all-via-the-ethical-adoption-of-technologies [accessed 2022-02-20]
61. Rosenberg S, Hickie I, Rock D. Rethinking mental health in Australia. The University of Sydney. URL: https://www.sydney.edu.au/content/dam/corporate/documents/brain-and-mind-centre/youthe/rethinking-the-mental-health-of-australia.pdf [accessed 2022-02-20]
62. Nilsen P, Seing I, Ericsson C, Birken SA, Schildmeijer K. Characteristics of successful changes in health care organizations: an interview study with physicians, registered nurses and assistant nurses. BMC Health Serv Res 2020 Feb 27;20(1):147 [FREE Full text] [doi: 10.1186/s12913-020-4999-8] [Medline: 32106847]
63. Wisniewski H, Torous J. Digital navigators to implement smartphone and digital tools in care. Acta Psychiatr Scand 2020 Apr;141(4):350-355 [FREE Full text] [doi: 10.1111/acps.13149] [Medline: 31930477]
64. National Mental Health Performance Subcommittee. Key performance indicators for Australian public mental health services. National Mental Health Strategy. 2013. URL: https://www.aihw.gov.au/getmedia/f9bb1a07-a43b-458a-9b73-64ef19d8aedd/Key-Performance-Indicators-for-Australian-Public-Mental-Health-Services-Third-Edition.pdf.aspx [accessed 2022-02-20]
65. Morris P. The Australian mental health crisis: a system failure in need of treatment. Australas Psychiatry 2006 Sep;14(3):332-333. [doi: 10.1080/j.1440-1665.2006.02300_1.x] [Medline: 16923051]

Abbreviations

CRC: Cooperative Research Centre
GP: general practitioner
HIT: health information technology
HREC: human research ethics committee
KRMC: Kildare Road Medical Centre
MARS: Mobile App Rating Scale
NSW: New South Wales
R&D: research and development
STB: suicidal thoughts and behavior

Edited by J Torous; submitted 26.08.21; peer-reviewed by V Mylonopoulou, AM Stupinski; comments to author 31.10.21; revised version received 14.12.21; accepted 30.12.21; published 09.03.22.

Please cite as:
LaMonica HM, Iorfino F, Lee GY, Piper S, Occhipinti JA, Davenport TA, Cross S, Milton A, Ospina-Pinillos L, Whittle L, Rowe SC, Dowling M, Stewart E, Ottavio A, Hockey S, Cheng VWS, Burns J, Scott EM, Hickie IB
Informing the Future of Integrated Digital and Clinical Mental Health Care: Synthesis of the Outcomes From Project Synergy
JMIR Ment Health 2022;9(3):e33060
URL: https://mental.jmir.org/2022/3/e33060
doi:10.2196/33060
PMID:34974414

Original Paper

PTSD Coach Version 3.1: A Closer Look at the Reach, Use, and Potential Impact of This Updated Mobile Health App in the General Public

Haijing Wu Hallenbeck[1,2], PhD; Beth K Jaworski[1], PhD; Joseph Wielgosz[1,2,3], PhD; Eric Kuhn[1,2], PhD; Kelly M Ramsey[1], BA; Katherine Taylor[1], MPH, PsyD; Katherine Juhasz[1], MS; Pearl McGee-Vincent[1], PsyD; Margaret-Anne Mackintosh[1], PhD; Jason E Owen[1], MPH, PhD

[1]Dissemination and Training Division, National Center for Posttraumatic Stress Disorder, Veterans Affairs Palo Alto Health Care System, Menlo Park, CA, United States

[2]Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA, United States

[3]Sierra Pacific Mental Illness Research Education and Clinical Center, Veterans Affairs Palo Alto Healthcare System, Palo Alto, CA, United States

Corresponding Author:
Haijing Wu Hallenbeck, PhD
Dissemination and Training Division
National Center for Posttraumatic Stress Disorder
Veterans Affairs Palo Alto Health Care System
795 Willow Rd
Menlo Park, CA, 94025
United States
Phone: 1 650 493 5000
Email: haijing.hallenbeck@stanford.edu

Abstract

Background: With widespread smartphone ownership, mobile health apps (mHealth) can expand access to evidence-based interventions for mental health conditions, including posttraumatic stress disorder (PTSD). Research to evaluate new features and capabilities in these apps is critical but lags behind app development. The initial release of PTSD Coach, a free self-management app developed by the US Departments of Veterans Affairs and Defense, was found to have a positive public health impact. However, major stakeholder-driven updates to the app have yet to be evaluated.

Objective: We aimed to characterize the reach, use, and potential impact of PTSD Coach Version 3.1 in the general public. As part of characterizing use, we investigated the use of specific app features, which extended previous work on PTSD Coach.

Methods: We examined the naturalistic use of PTSD Coach during a 1-year observation period between April 20, 2020, and April 19, 2021, using anonymous in-app event data to generate summary metrics for users.

Results: During the observation period, PTSD Coach was broadly disseminated to the public, reaching approximately 150,000 total users and 20,000 users per month. On average, users used the app 3 times across 3 separate days for 18 minutes in total, with steep drop-offs in use over time; a subset of users, however, demonstrated high or sustained engagement. More than half of users (79,099/128,691, 61.46%) accessed one or more main content areas of the app (ie, Manage Symptoms, Track Progress, Learn, or Get Support). Among content areas, features under Manage Symptoms (including coping tools) were accessed most frequently, by over 40% of users (53,314/128,691, 41.43% to 56,971/128,691, 44.27%, depending on the feature). Users who provided initial distress ratings (56,971/128,691, 44.27%) reported relatively high momentary distress (mean 6.03, SD 2.52, on a scale of 0-10), and the use of a coping tool modestly improved momentary distress (mean −1.38, SD 1.70). Among users who completed at least one PTSD Checklist for DSM-5 (PCL-5) assessment (17,589/128,691, 13.67%), PTSD symptoms were largely above the clinical threshold (mean 49.80, SD 16.36). Among users who completed at least two PCL-5 assessments (4989/128,691, 3.88%), PTSD symptoms decreased from the first to last assessment (mean −4.35, SD 15.29), with approximately one-third (1585/4989, 31.77%) of these users experiencing clinically significant improvements.

Conclusions: PTSD Coach continues to fulfill its mission as a public health resource. Version 3.1 compares favorably with version 1 on most metrics related to reach, use, and potential impact. Although benefits appear modest on an individual basis, the app provides these benefits to a large population. For mHealth apps to reach their full potential in supporting trauma recovery,

future research should aim to understand the utility of individual app features and identify strategies to maximize overall effectiveness and engagement.

(JMIR Ment Health 2022;9(3):e34744) doi:10.2196/34744

KEYWORDS

posttraumatic stress disorder; trauma; mental health; mHealth; mobile app; public health; self-management; mobile phone

Introduction

Background on PTSD Coach

With 85% of US adults now owning a smartphone [1], mobile health (mHealth) apps remain one of the most promising avenues for disseminating evidence-based interventions for mental health [2]. Increased dissemination of mental health interventions is sorely needed, as only a minority of individuals with mental health concerns receive services. For example, only one-third of US adults with a moderately severe mental health condition received treatment in a given year [3]. There are a variety of barriers to traditional, in-person services, including mental health stigma and limited access to care, both of which are more pronounced for racial and ethnic minorities [4,5]. mHealth apps are well-positioned to mitigate these barriers, with their discreet nature and a similar rate of smartphone ownership across White, Black, and Hispanic groups in the United States [1]. Responding to this apparent potential, the development of mHealth apps has exploded in popularity; however, systematic research on them has lagged far behind [6-8]. To understand their value from a public health perspective, it is important to characterize the reach, use, and potential benefits of these apps. Furthermore, evaluating mHealth apps on an ongoing basis, especially as apps are modified or updated with new features, can provide an important feedback loop to inform researchers about what is working well and what could be improved within apps [9]. This study focuses on assessing the public health impact of an mHealth app for posttraumatic stress disorder (PTSD), with substantial updates to both the features in the app and the analytic capabilities for understanding its use.

PTSD is a significant mental health condition due to its often debilitating effects on psychosocial functioning and quality of life [10,11] and its high prevalence, especially among military veterans [12,13]. As part of a portfolio of mHealth apps [14], the PTSD Coach app was developed by the US Department of Veterans Affairs (VA) National Center for PTSD (NCPTSD) and the US Department of Defense Center for Telehealth and Technology. PTSD Coach was designed with veterans and service members in mind and was also intended as a public health resource to help any individual impacted by trauma. As such, the app has been available to the public since 2011 on both iOS and Android platforms. Drawing from evidence-based treatments (eg, cognitive behavioral therapy [15]), PTSD Coach provides psychoeducation, self-assessments, coping tools, and resources for support and professional care; the self-management app is not meant to replace treatment with a mental health professional. Importantly, PTSD Coach is provided free of charge and protects users' privacy by collecting data anonymously (ie, no identifying information). The app is offered in both English and Spanish, is accessible to people with visual and hearing impairments, and can be used without internet connectivity. PTSD Coach is available worldwide, through its US version (described in this paper) as well as separate versions developed in 6 other countries (a result of sharing source code with international partners [16]).

In 2015, Owen et al [17] sought to provide an initial characterization of the reach, use, and potential impact of PTSD Coach in the general public. To do so, they examined *in the wild* data (ie, data from people who are using the publicly available version of the app in their everyday lives), thus enabling the assessment of naturalistic patterns of use. Evaluating PTSD Coach, version 1, between March 2011 and February 2014, the authors found that the app had been broadly disseminated with over 150,000 downloads and over 10,000 active users per month, had reached its target audience (eg, veterans and civilians with PTSD symptoms, their family members, and mental health providers), and was reviewed positively by users. The authors provided descriptive statistics on patterns of app use; for example, showing that users used the app an average of 6 times and for a total duration of 5 minutes. Of note, although most users had steep drop-offs in use of the app over time, as is the case with self-management apps more broadly [18], there was a subset of high-engagement users who reported that they incorporated the app as part of a daily routine and continued to use the app even a year later. Lastly, on average, users who completed self-assessments endorsed PTSD symptoms above clinical threshold and rated high levels of momentary distress; after using a coping tool, momentary distress decreased by an average of 2 points (on a scale of 0-10), highlighting the benefits of the app during times of need. Findings from the study by Owen et al [17] are consistent with findings from controlled research studies (eg, randomized controlled trials) on PTSD Coach, in which the app was associated with positive user experiences [19,20] and benefits [21-23] in both veteran and civilian samples.

Updates to PTSD Coach

Since the study by Owen et al [17] was published, the PTSD Coach app has undergone several substantial updates to its design and features to address stakeholder feedback from users and health care professionals. The look and feel of PTSD Coach was revamped to have a clean, modern design (see Figures 1A and 1B as well as 2A and 2B for examples), and the app incorporated updated information about PTSD that was consistent with the Diagnostic and Statistical Manual of Mental Disorders, 5th Edition [24]. In addition, new evidence-based therapeutic features were added to the app, including mindfulness and relationship tools and the option to complete a safety plan for suicide prevention. For increased convenience, users can now access favorite tools and view an inspiring quote on the home screen of the app. Users can also more easily share

the app with family and friends and provide feedback about the app to the development team. Finally, the app's development team has improved its ability to detect, track, and resolve problems in the app on both the iOS and Android platforms. This may be especially critical for Android, which has substantial heterogeneity in its smartphones and had stability issues with its earlier operating system versions. In the previous evaluation of version 1, Android users reviewed the app less positively and were found to have lower rates of use and smaller benefits compared with iOS users [17].

Figure 1. Screenshots of the PTSD Coach home screen with 4 main content areas: (A) home screen from version 1; (B) home screen from version 3.1 (current version).

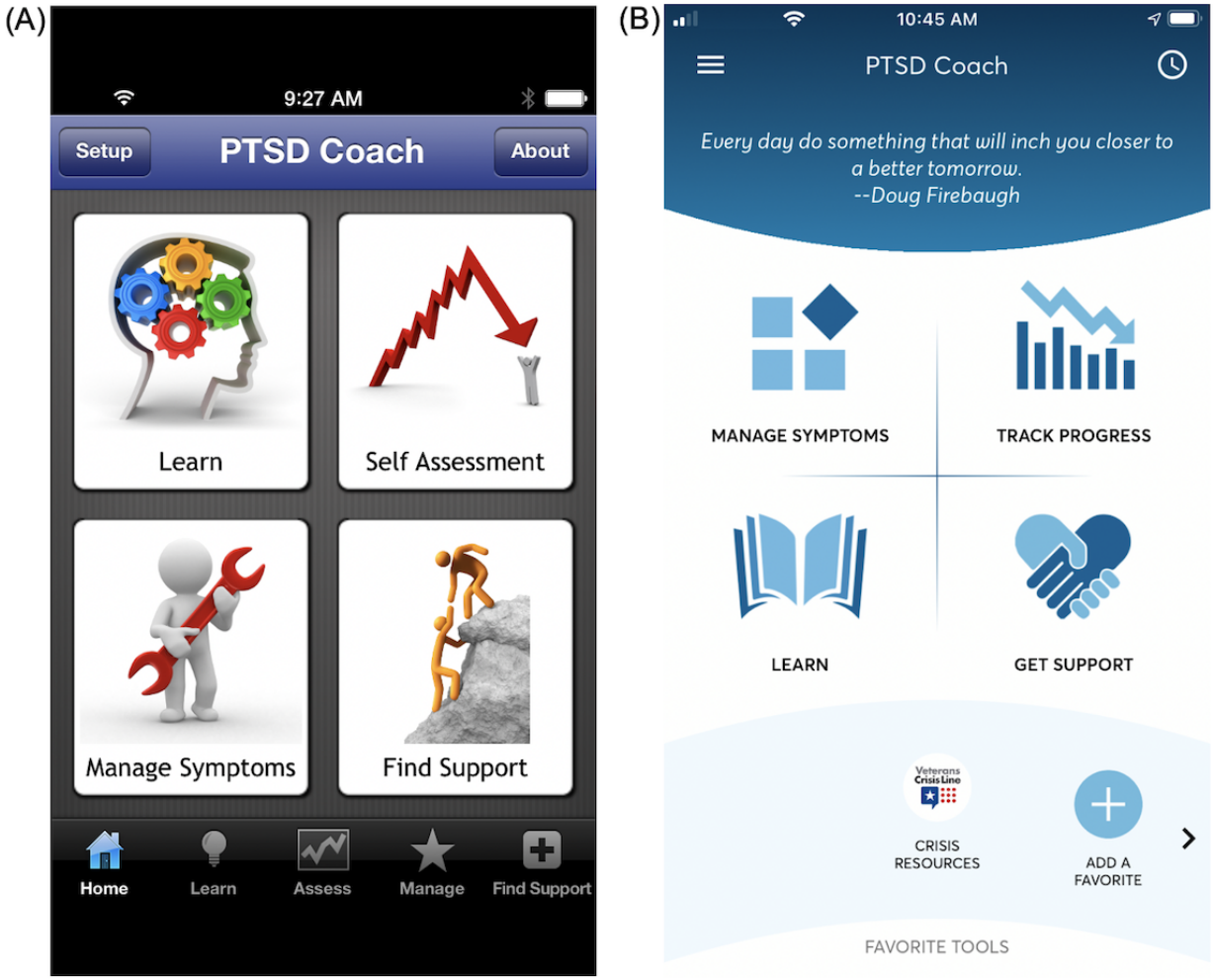
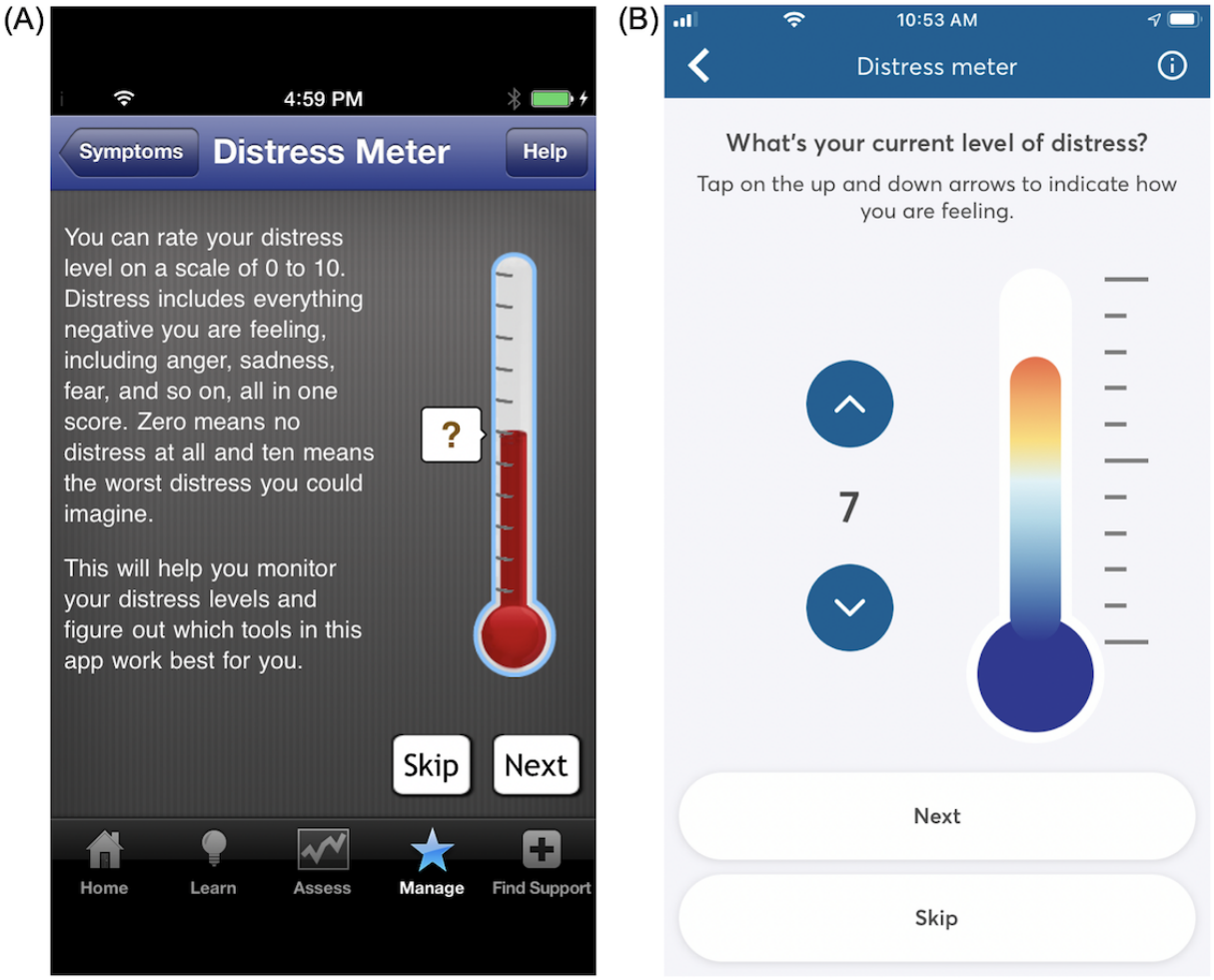


Figure 2. Screenshots of the PTSD Coach distress meter with which users can rate subjective units of distress on a scale of 0 to 10 rating: (A) distress meter from version 1; (B) distress meter from version 3.1 (current version).



Another significant update is that PTSD Coach can now accommodate user-level analyses on all features of the app. To do so, anonymous install codes (generated upon the first launch of the app) are used as a proxy for individual users. Previously, to establish a baseline on the reach, use, and potential impact of PTSD Coach in the general public, Owen et al [17] analyzed data which were available in aggregate form only (ie, at the group level). The data were primarily available in aggregate form for two groups: first-time users and returning users. It was not possible to link which first-time users went on to become returning users; consequently, they could not analyze any individual user's data (eg, change in PTSD symptoms over time). Furthermore, some metrics (eg, levels of momentary distress) were not available to be analyzed for users as a whole (ie, first-time and returning users combined), and other metrics (eg, use of specific coping tools) were not available at all. Thus, this limited obtaining detailed information about how users interacted with the app. Having this information could inform researchers of what users' needs are (eg, most frequently reported PTSD symptoms) or what users finding helpful in the app (eg, most frequently used coping tools).

Objectives of This Study

In this study, we sought to characterize the reach, use, and potential impact of the current version of PTSD Coach, version 3.1, which was released on April 20, 2020. In particular, we aimed to build upon the study by Owen et al [17] by characterizing the use of specific features of the app, within the main content areas, to obtain a fine-grained picture of how users are interacting with the app. This aim is consistent with the growing recognition in the literature that evaluating the *quality* of engagement with an app is critical in addition to the *quantity* of engagement [25]. Given the differences observed between iOS and Android users with version 1 of the app, we also compared reach, use, and impact by platform. Notably, we did not seek to directly replicate Owen et al [17] because of changes in the instrumentation of the app (eg, differences in the availability of data), mismatch in the length of data observation periods (ie, 3 years vs 1 year), and relevant environmental factors that we could not control (eg, increasing proliferation of smartphone devices). However, given the updates to the PTSD Coach app, we expected that the findings for version 3.1 would generally demonstrate maintained or improved metrics related to reach, use, and potential impact, compared with version 1.

Regarding relevant environmental factors, we note that our data observation period (April 2020 to April 2021) fell squarely within the timeframe of the COVID-19 pandemic. Because the pandemic contributed to rising levels of mental health symptoms [26], it was possible that we would observe evidence of wider reach and increased use of PTSD Coach as a function of the stressors associated with the pandemic. Because we did not have version 3.1 data before April 2020, we could not rule out this possibility, and we certainly hope that the app has been helpful to those experiencing heightened PTSD symptoms during this challenging time. In this evaluation, we consider the potential role of the pandemic by contextualizing our findings with this lens (eg, seeing whether users are endorsing higher levels of PTSD symptoms now compared with in the study by Owen et al [17] and seeing whether use of the app seemed to fluctuate alongside peaks of COVID-19 cases in the country).

Our data observation period also overlapped with 2 large-scale initiatives within the VA health care system. First, the VA Office of Connected Care (OCC) started a program in 2016 that distributes tablets to veterans engaging in telehealth services; this program is ongoing and, more recently, has focused on iOS tablets specifically [27,28]. Between April 2020 and April 2021, the OCC downloaded PTSD Coach along with other health care–related apps onto 95,000 iOS tablets provided to veterans. Second, between January 2020 and December 2020, the NCPTSD trained 1100 VA staff members at 19 sites across the country on the use of VA mHealth apps in the care of veterans [29]. As part of these training sessions, VA staff downloaded and explored the PTSD Coach app, including a safety plan for suicide prevention. Because of the anonymous nature of the app data, we were unable to determine which PTSD Coach users were veterans using VA-provided iOS tablets or VA staff members undergoing mHealth app training. However, we discuss aspects of our findings that were likely shaped by these initiatives.

Methods

Data Sources

We utilized download and use data from the public version of the PTSD Coach mobile app between April 20, 2020, and April 19, 2021. The download data came from the Apple App Store (for iOS devices) or the Google Play Store (for Android devices) and included the country code associated with a user's Apple ID or Google Play account. The use data corresponded to in-app events, that is, actions taken by the user in the app (eg, screens selected and buttons pressed), and were collected for quality improvement purposes. There were use data for each install code, which was a unique, random sequence of numbers and letters generated upon the first launch of the app. These install codes were used as proxies for individual app users. Although we do not think this was the case for most users, a user could have had more than one install code as a result of deleting and installing the app multiple times or installing the app on multiple devices. Users can also opt out of sharing their use data within the app settings. All use data were anonymous and encrypted and stored on a secure server. No identifying or device information (other than whether the platform was iOS or Android) was collected or stored.

Ethics Approval

These data were collected by the NCPTSD mobile mental health program as part of ongoing quality improvement, which was approved by the Palo Alto VA Research and Development Committee (RDIS No. ROS0021). The Institutional Review Board at Stanford University School of Medicine reviewed the project and determined that it was non-research.

The PTSD Coach, Version 3.1, Mobile App

Onboarding and Home Screen

After users accepted the End User License Agreement (EULA), they were shown a brief tutorial about the four main content areas of the app: Manage Symptoms, Track Progress, Learn, and Get Support. Then, they could opt to personalize the app (eg, by adding pictures, music, support contacts, or switching to using the app in Spanish) or they could proceed directly to the app content (which took them to the home screen). From the home screen, users could access the 4 main content areas (Figure 1B). In addition, from the home screen, users could open a lateral menu with the option of completing a safety plan for suicide prevention. This lateral menu also had options for users to learn more about and personalize the app, manage their data, and share and give feedback about the app.

Manage Symptoms

Users could indicate a current PTSD-related symptom that they were experiencing: Reminded of Trauma, Avoiding Triggers, Disconnected From People, Disconnected From Reality, Sad/Hopeless, Worried/Anxious, Angry, and Unable to Sleep. Users could access a coping tool either through a recommendation based on a selected symptom or by viewing the complete list of tools; they could also access a list of tools previously marked as favorites. There were a total of 23 coping tools (see Multimedia Appendix 1 for the list). Before and after using a tool, users were asked to rate their current level of distress (ie, momentary distress) using a visual thermometer analog corresponding to a scale from 0 to 10 (Figure 2B); users had the option to skip or turn off this rating feature. We refer to these ratings as pretool and posttool subjective units of distress (SUDs).

Track Progress

Users could complete and receive feedback on a self-assessment of their PTSD symptoms, view a graph and details of their past self-assessments, and set a reminder to take future self-assessments. The self-assessment used was the PTSD Checklist for DSM-5 (PCL-5 [30]), which has 20 items that are answered on a 5-point scale (0=*not at all* to 4=*extremely*) about how much a person was bothered by individual symptoms in the past month. The PCL-5 was found to have good reliability and validity in both civilian [31] and veteran [32] samples. Scores of 31 to 33 or higher correspond with a likely PTSD diagnosis [32], and decreases of approximately ≥5 points and ≥10 points indicate reliable and clinically significant change, respectively [30].

Learn

Users could read psychoeducational information organized under 3 categories: About PTSD, Getting Professional Help, and PTSD and the Family. There are 21 learn topics under About PTSD, 22 learn topics under Getting Professional Help, and 12 learn topics under PTSD and the Family (see Multimedia Appendix 1 for the list).

Get Support

Users could access resources for additional support organized under 3 categories: Crisis Resources, Find Professional Care, and Grow Your Support. Under Crisis Resources, information for suicide prevention and crisis hotlines were included, as well as the option to add a personal support contact. Under Find Professional Care, a broad range of mental health treatment resources were listed, including information for military-specific treatment options (eg, VAs and Vet Centers) and options open to the general public (eg, Psychology Today). Under Grow Your Support, there were ideas about ways to reach out to and connect with others, including information for joining both military-specific (eg, Team Red, White, and Blue) and general groups (eg, Meetup groups). Resources under all 3 categories contained direct links to phone numbers and websites.

Safety Plan for Suicide Prevention

Located in the lateral menu, the safety plan for suicide prevention was based on the Safety Planning Intervention [33]. In a tutorial, users were first oriented to the purpose of the safety plan, encouraged to discuss the safety plan with a provider, and given crisis resources and options for finding professional care. The safety plan was divided into 6 sequential steps, which comprised a predetermined and individualized set of strategies designed to help individuals manage mental health crises (eg, suicidal urges) instead of acting on impulse. These steps correspond to identifying (1) warning signs, (2) self-coping strategies, (3) social contacts and settings for purposes of distraction, (4) family and friends for purposes of crisis management, (5) mental health professionals and agencies, and (6) ways to restrict access to lethal means. Steps 3 and 6 have 2 parts, and steps 3 through 6 involve adding a contact. For analysis, a safety plan was considered complete if a user filled out complete information for all 6 steps.

Analysis Plan

There were 9,415,339 in-app events in the data observation period between April 20, 2020, and April 19, 2021. Each *event* contained an event name, date and time stamp, and the associated user's install code and platform (ie, iOS or Android), as well as any nontext data entered by the user (eg, SUDs ratings). Following Kozlov et al [34], we considered events with time stamps within 30 minutes of one another as part of a single *visit* to the app; we took this approach to avoid inconsistencies in the default app instrumentation for marking the end of a visit. Thus, visit duration reflected the time that passed between the first and last events that belonged to the same visit. If a user had only one event in total (the result of opening the app without responding to the EULA), visit duration was encoded as 0 minutes; this was also the case for total app use duration and time between the first and last app use.

Data preprocessing was conducted using Python (version 3.7.7; Python Software Foundation) with the pandas (version 1.05) [35] and pyodbc (version 4.0.0) [36] libraries. Event data were extracted from the server and labeled with their corresponding visit numbers via Python scripts with embedded SQL queries. Next, the data were run through a second Python script to generate per-user summaries. This script first cleaned the data and removed duplicate events, then extracted user-level metrics in a table format and combined the user-level metrics into a unified data set. We used the final data set of user-level metrics to run descriptive statistics and difference tests using SPSS Statistics (version 26; IBM) to evaluate the reach, use, and potential impact of PTSD Coach, for all users and then separately for iOS and Android users. For difference tests, the *t* test (2-tailed) effect size was Cohen *d* (small: 0.2, medium: 0.5, and large: 0.8), and the chi-square test effect size was Cramer *V* (small: 0.1 to <0.3, medium: 0.3 to <0.5, and large: ≥0.5). Variables that were not normally distributed were Box-Cox transformed ($\lambda=-0.3$, with an additional 0.001 constant added to handle values of 0) before performing difference tests and calculating effect sizes.

For *reach*, we examined numbers of downloads and users, including number of active users per month (ie, users who used the app at least once during a given month) to see if reach was sustained over the data observation period. For *use*, we examined overall use of the app (eg, total number of visits), use of the app over time (eg, retention), and use of specific features of the app (eg, whether a coping tool was accessed). Regarding use of specific app features, we determined the most frequently used features at the level of the user, in which the number of users who accessed a specific app feature at least once (collapsing across all visits) was divided by the total number of users. For *impact*, we examined first SUDs ratings and first PCL-5 scores as well as changes in these metrics, by subtracting posttool SUDs ratings from pretool SUDs ratings and by subtracting the last PCL-5 score from the first PCL-5 score. To accurately characterize *use* and *impact*, we limited analyses to install codes whose first event fell inside the observation period; we considered these to be new users who started using the app after version 3.1 was released and for whom we had maximum use data.

Lastly, regarding use of the app over time, for calculating retention, we divided the 12-month period into the following bins: days 1 to 7, weeks 1 to 4, and months 1 to 12. The starting point was the user's first event. Subsequent events after the first event could fall into different bins—for example, ≥0 and <24 hours after the first event (day 1), ≥0 and <7 days after the first event (week 1), and ≥0 and <30 days after the first event (month 1). For the user percentage calculation for each bin, a user was included in the numerator if they had a qualifying subsequent event relative to their first event. The denominator was adjusted for each bin to reflect the number of users with potential observable data, which was based on when users started using the app. For example, a user whose first event occurred on the first day of the observation period (April 20, 2020) would be included in the denominator for all bins. In contrast, a user whose first event occurred on the second-to-last day of the observation period (April 18, 2021) would be included in the

denominator for the day 1 bin only, as it was not possible to further observe their data with the observation period cutoff (April 19, 2021). We evaluated both classic retention (ie, app use on a specific day relative to first use) and rolling retention (ie, app use on or after a specific day, relative to first use).

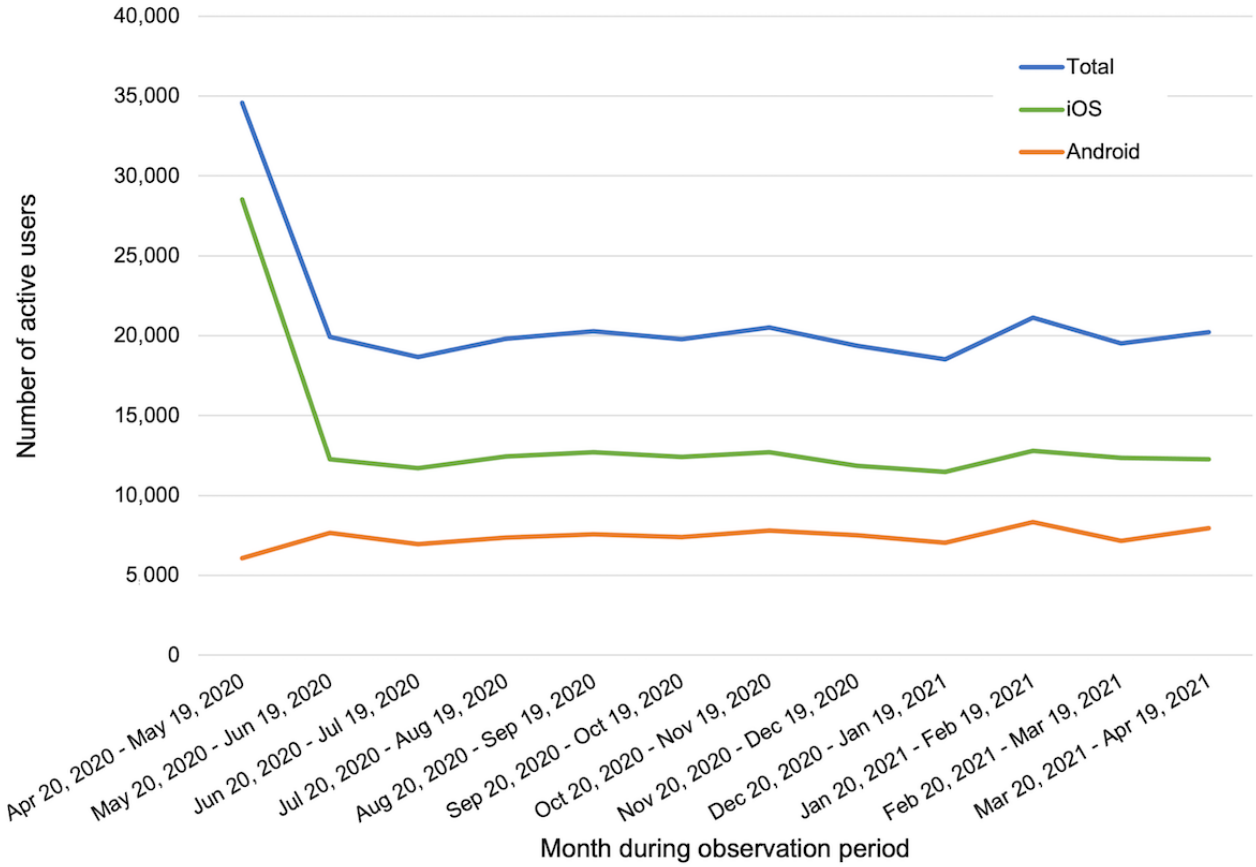
Results

All Users

Reach

During the observation period between April 20, 2020, and April 19, 2021, there were 207,001 downloads of the PTSD Coach app. Most of these (188,203/207,001, 90.92%) were downloads from user accounts based in the United States. There were 148,354 app users (ie, users who used PTSD Coach at any point during the observation period) and an average of 21,032 active monthly users (ie, users who used PTSD Coach at least once during a given month). The number of active monthly users per month during the observation period is shown in Figure 3. This number was highest in the first month, fell in the second month, and remained relatively steady for the remainder of the year—a pattern that was driven by the number of iOS users (see the section *Reach* under the subheading *Comparison of iOS and Android Users* below for an explanation). Among the 148,354 total users, 128,691 (86.74%) were new users (ie, they started using the app during the observation period).

Figure 3. Reach of PTSD Coach, version 3.1, between April 20, 2020, and April 19, 2021.



Use

Overall Use

The metrics of overall use for new users of PTSD Coach are presented in Table 1. On average, users of PTSD Coach visited the app approximately 3 times, with each visit having an average duration of 5 minutes and involving 18 events. In total, users spent approximately 18 minutes using the app across 3 unique days. Compared with these means, the medians and modes were lower, indicating that the means were positively skewed by extreme values; indeed, maximum values were much larger than the means, whereas IQRs remained relatively small. These findings illustrate that there was a subset of users with much higher levels of engagement than the average user. For example, 2.02% (2601/128,691) of new users visited the app on average ≥18 times (ie, ≥2 SDs above the mean for all new users), corresponding to a total of approximately 230 minutes spent using the app across 26 unique days.

Table 1. Overall use of PTSD Coach, version 3.1, among new users (n=128,691) between April 20, 2020, and April 19, 2021.

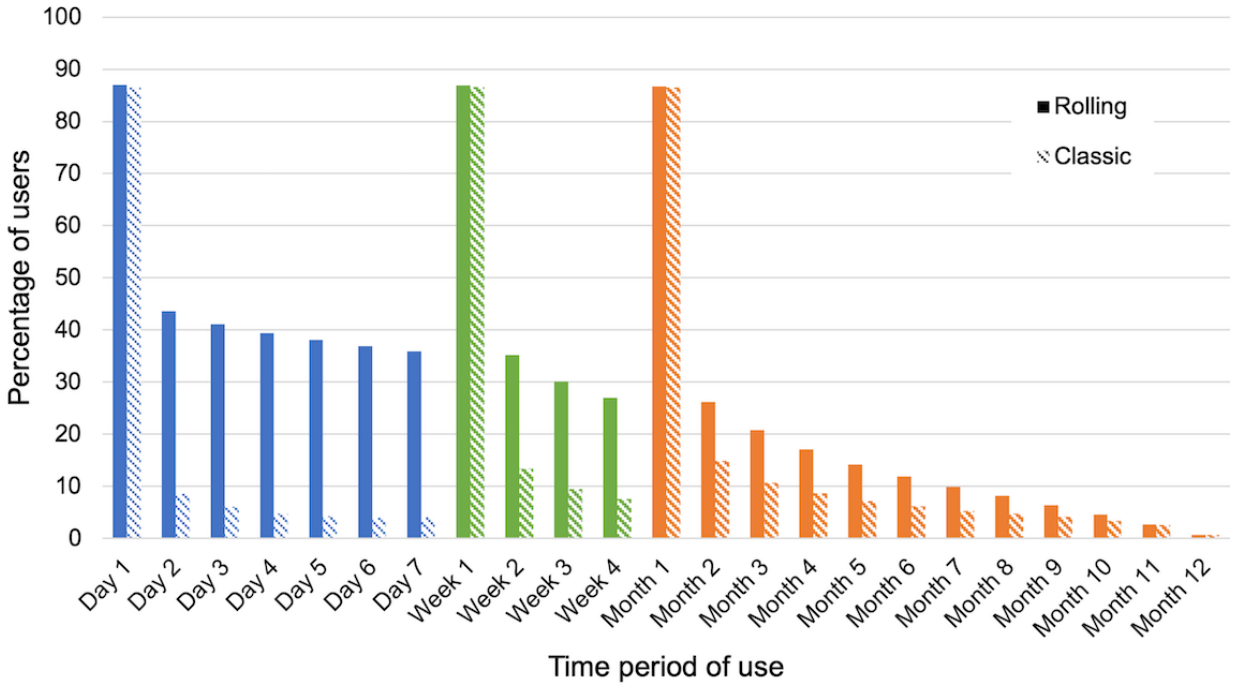
Category	Mean (SD)	Median[a]	Mode[a]	Range	IQR[a]
Number of visits	3.26 (7.41)	2	1	1.00-501.00	2
Number of events per visit	17.89 (25.70)	9	1	1.00-1052.00	21
Visit duration (minutes)	4.60 (7.14)	2	0	0.00-203.20	6
Total duration (minutes)	17.55 (58.62)	4	0	0.00-6472.37	16
Number of unique days	2.70 (4.88)	1	1	1.00-222.00	2

[a]This was calculated after rounding values to the nearest whole integer for the number of events per visit, visit duration, and total duration.

Use Over Time

The rates of classic retention (app use on a specific day, relative to first use) and rolling retention (app use on or after a specific day, relative to first use) for new users during the 12-month observation period are displayed in Figure 4. As expected, rolling retention rates were higher than classic retention rates. For both classic and rolling retention, approximately 87% of users (110,971/128,305, 86.49% and 111,660/128,305, 87.03%, respectively) used the app during the first day of opening it (day 1). Because week 1 and month 1 are inclusive of day 1, retention rates across these 3 periods are very similar. Beyond day 1, 8.53% (10,923/128,032) of users used the app during day 2 (classic retention), and 43.54% (55,744/128,032) of users used the app during day 2 or later (rolling retention). At the other extreme, for month 12, classic and rolling retention were identical because of the observation period cutoff, with 0.69% (62/8953) of users using the app in the 12th month after the initial opening of the app. Lastly, relevant to rolling retention specifically, new users had on average 31.54 (SD 64.01) days that spanned their first and last use of the app. Upon restriction to the subset of high-engagement users identified above (2601/128,691, 2.02% of new users), the average time between the first and last use of the app was 173.76 (SD 96.91) days.

Figure 4. Rolling and classic retention of PTSD Coach, version 3.1, new users.



Use of Specific Features

After accepting the EULA and completing onboarding, 65.86% (84,763/128,691) of new users arrived at the home screen. A total of 61.46% (79,099/128,691) of users visited a content area. Specifically, 25.69% (33,057/128,691) of the users visited 1 content area, 13% (16,702/128,691) visited 2 content areas, 9.58% (12,325/128,691) visited 3 content areas, and 13.22% (17,015/128,691) visited all 4 content areas. The mean number of content areas visited was 1.33 (SD 1.41).

First, regarding specific content areas visited, 55.68% (71,649/128,691) of users visited the Manage Symptoms. Within this content area, 43.78% (56,347/128,691) of users selected at least one of the eight PTSD-related symptoms. The three most frequently selected symptoms were Reminded of Trauma (29,960/128,691, 23.23%), Avoiding Triggers (19,387/128,691, 15.06%), and Unable to Sleep (18,127/128,691, 14.08%). Also within this content area, 41.43% (53,314/128,691) of users accessed at least one coping tool. The three most frequently accessed tools were Change Your Perspective (19,988/128,691, 15.53%), Inspiring Quotes (19,668/128,691, 15.28%), and

Ambient Sounds (17,753/128,691, 13.8%). Providing ratings of momentary distress, 44.27% (56,971/128,691) of users rated their SUDs before using a coping tool, and 13.86% (17,841/128,691) of users rated their SUDs before and after using a tool.

Second, 27.57% (35,480/128,691) of users visited the Track Progress content area. Within this content area, 13.67% (17,589/128,691) of users completed at least one PCL-5 assessment, and 3.88% (4989/128,691) of users completed 2 or more assessments. In addition, 9.05% (11,647/128,691) of users set a reminder to complete future PCL-5 assessments.

Third, 28.49% (36,666/128,691) of users visited the Learn content area. Within this content area, 16.29% (20,964/128,691) of users viewed at least one learn topic under any category. Under the About PTSD category, 12.55% (16,152/128,691) of users viewed at least one learning topic, with the most frequent topic being PTSD Facts (8904/128,691, 6.92%). Under Getting Professional Help, 3.47% (4463/128,691) of users viewed at least one learn topic, with the most frequent topic being tools for PTSD (2141/128,691, 1.66%). Under PTSD and the Family, 5.12% (6587/128,691) of users viewed at least one learn topic, with the most frequent topic being Fighting Fair (3366/128,691, 2.62%).

Fourth, 21.52% (27,701/128,691) of users visited the Get Support content area. Within this content area, 9.7% (12,491/128,691) of users viewed Crisis Resources, 5.76% (7409/128,691) viewed Find Professional Care, and 6.68% (8600/128,691) viewed Grow Your Support. From 1 of these 3 categories, 3.52% (4535/128,691) of users accessed a website or phone resource. Also within this content area, 4.76% (6129/128,691) of users added personal support contact.

Fifth, regarding the safety plan for suicide prevention, 2.47% (3184/128,691) of users opened the plan, and 0.25% (320/128,691) of users also completed the plan. Of note, the 1100 VA staff who participated in the NCPTSD mHealth app training likely comprised about one-third of the users who opened the plan. It is not clear how many staff completed the plan, given that it was not required as part of the training.

Lastly, when restricting to the subset of high-engagement users identified above (2601/128,691, 2.02% of new users), a much higher proportion of these users visited key app features compared with the proportion among all users. For example, 87.43% (2274/2601) of these high-engagement users accessed at least one coping tool, 55.36% (1440/2601) completed at least one PCL-5 assessment, 52.67% (1370/2601) viewed at least one learn topic, 18.22% (474/2601) accessed a Get Support website or phone resource, and 2.04% (53/2601) completed a safety plan for suicide prevention. These percentages are between 2 and 10 times greater than the percentages for all users.

Impact

Among the users who had pretool SUDs ratings (n=56,971), first SUDs ratings reflected relatively high momentary distress (mean 6.03, SD 2.52). Among the users who had both pretool and posttool SUDs ratings (n=17,841), SUDs ratings decreased after using a tool, with a mean difference of −1.38 (SD 1.70; 95% CI −1.41 to −1.36), which was significantly different from 0 ($t_{17,840}$=−108.23; P<.001). The mean SUDs decrease was comparable among the subset of high-engagement users who had both pretool and posttool SUDs ratings (n=1447; mean −1.48, SD 1.42). Excluding these high-engagement users from analysis did not impact the mean for the remaining users (mean −1.37, SD 1.73), suggesting that the potential benefit of momentary distress relief was not driven by the high-engagement users.

Among the users who had at least one PCL-5 score (n=17,589), initial scores reflected high levels of PTSD symptoms (mean 49.80, SD 16.36), with 87.13% (15,326/17,589) of these users with a score of ≥31 points (ie, the threshold for likely PTSD diagnosis). Among users with at least two PCL-5 scores (n=4989), scores decreased from the first to last PCL-5, with a mean difference of −4.35 (SD 15.29; 95% CI −4.77 to −3.92) that was significantly different from zero, (t_{4988} =−20.07; P<.001). Specifically, 31.77% (1585/4989) of users with at least two PCL-5 scores had clinically significant decreases (ie, ≥10 points), and 44.34% (2212/4989) of users with at least two PCL-5 scores had reliable decreases (ie, ≥5 points) in their scores.

Compared with the mean for all users, the mean PCL-5 decrease was somewhat larger for the subset of high-engagement users who had at least two PCL-5 scores (n=1051; mean −5.95, SD 16.44). Excluding these high-engagement users from analysis slightly attenuated the mean for the remaining users (mean −3.92, SD 14.94). These findings suggest that users as a whole are still experiencing a potential benefit in PTSD symptom reduction and that this benefit may be more pronounced for the subset of high-engagement users.

Comparison of iOS and Android Users

Reach

Among the 207,001 downloads of the PTSD Coach app during the observation period, 73.65% (152,461/207,001) were iOS downloads and 26.35% (54,540/207,001) were Android downloads. Of note, among the 152,461 iOS downloads, 95,000 (62.31%) originated from the VA OCC initiative to provide iOS tablets to veterans engaging in telehealth services. Among the 148,354 total app users, 96,143 (64.81%) were iOS users, and 52,211 (35.19%) were Android users. There was an average of 13,628 active monthly iOS users and 7404 active monthly Android users. Figure 3 displays active monthly users from both platforms for each month during the observation period. In contrast to the number of Android users staying relatively consistent throughout the months, the number of iOS users peaked in the first month before it fell and remained steady, which was likely related to VA OCC striving to meet telehealth demands by distributing iOS tablets during the early months of the COVID-19 pandemic. Lastly, among the 128,691 new users, 80,006 (62.17%) were iOS users, and 48,685 (37.83%) were Android users.

Use

In Table 2, the metrics for overall use, use over time, and use of specific features are displayed separately for iOS and Android

new users. Across all metrics, iOS users showed lower levels of use compared with Android users. These differences were statistically significant (P<.001) and were associated with generally small effect sizes, with the exception of medium effect sizes for number of events per visit, visit duration, and total duration.

Table 2. Use of PTSD Coach, version 3.1, among new users between April 20, 2020, and April 19, 2021, separated by platform.

Use category	iOS users (n=80,006)	Android users (n=48,685)	Difference test[a]		Effect size	
			t test (*df*)	Chi-square (*df*)	Cohen *d*	Cramer *V*
Overall use						
Number of visits, mean (SD)	3.18 (7.17)	3.39 (7.78)	−3.66 (128,689)	N/A[b]	0.02	N/A
Number events per visit, mean (SD)	15.39 (23.57)	22.00 (28.40)	−102.16 (128,689)	N/A	0.61	N/A
Visit duration (in minutes), mean (SD)	4.12 (6.67)	5.40 (7.78)	−101.95 (128,689)	N/A	0.64	N/A
Total duration (in minutes), mean (SD)	16.03 (53.95)	20.06 (65.50)	−99.52 (128,689)	N/A	0.62	N/A
Number of unique days, mean (SD)	2.62 (4.67)	2.83 (5.16)	−5.80 (128,689)	N/A	0.03	N/A
Use over time						
Days between first and last use, mean (SD)	30.21 (64.33)	33.72 (66.04)	−5.44 (128,689)	N/A	0.03	N/A
Use of specific features						
Home, n (%)	47,718 (59.64)	37,045 (76.09)	N/A	3643.14 (1)	N/A	0.17
Number of content areas, mean (SD)	1.24 (1.40)	1.49 (1.40)	−30.52 (128,689)	N/A	0.18	N/A
Manage Symptoms content area, n (%)	41,730 (52.16)	29,919 (61.45)	N/A	1060.20 (1)	N/A	0.09
Selected symptom	34,396 (42.99)	21,951 (45.09)	N/A	54.02 (1)	N/A	0.02
Rated pretool SUDs[c]	33,087 (41.36)	23,884 (49.06)	N/A	727.86 (1)	N/A	0.08
Rated pretool and posttool SUDs[c]	7828 (9.78)	10,013 (20.57)	N/A	2946.87 (1)	N/A	0.15
Accessed coping tool	31,056 (38.82)	22,258 (45.72)	N/A	594.30 (1)	N/A	0.07
Track Progress content area, n (%)	20,275 (25.34)	15,205 (31.23)	N/A	525.90 (1)	N/A	0.06
Completed 1 PCL-5[d]	10,137 (12.67)	7452 (15.31)	N/A	178.27 (1)	N/A	0.04
Completed ≥2 PCL-5s	2785 (3.48)	2402 (4.93)	N/A	232.07 (1)	N/A	0.04
Set assessment reminder	6507 (8.13)	5140 (10.56)	N/A	216.20 (1)	N/A	0.04
Learn content area, n (%)	20,766 (25.96)	15,900 (32.66)	N/A	667.72 (1)	N/A	0.07
Accessed learn topic	11,863 (14.83)	9101 (18.69)	N/A	331.82 (1)	N/A	0.05
Get Support content area, n (%)	16,407 (20.51)	11,294 (23.2)	N/A	129.80 (1)	N/A	0.03
Accessed web or phone resource	2346 (2.93)	2189 (4.5)	N/A	217.76 (1)	N/A	0.04
Added support contact	2707 (3.38)	3422 (7.03)	N/A	886.82 (1)	N/A	0.08
Completed safety plan for suicide prevention, n (%)	162 (0.2)	158 (0.32)	N/A	18.18 (1)	N/A	0.01

[a]All differences were statistically significant at P<.001.

[b]N/A: not applicable.

[c]SUDs: subjective units of distress.

[d]PCL-5: Posttraumatic Stress Disorder Checklist for DSM-5.

Impact

Among the iOS (n=33,087) and Android (n=23,884) users who had pretool SUDs ratings, ratings from iOS users (mean 5.85, SD 2.56) reflected lower momentary distress (P<.001; Cohen d=0.16) than those from Android users (mean 6.26, SD 2.46). Among the iOS (n=7828) and Android (n=10,013) users who had both pretool and posttool SUDs ratings, iOS users (mean −1.70, SD 1.80) had larger reductions (P<.001; Cohen d=0.33) than Android users (mean −1.14, SD 1.59).

Among the iOS (n=10,137) and Android (n=7452) users who had at least one PCL-5 score, initial scores reflected lower levels of PTSD symptoms (P<.001; Cohen d=0.06) for iOS users (mean 49.39, SD 16.32) than for Android users (mean 50.36, SD 16.40). Among iOS (n=2590) and Android (n=2399) users with at least two PCL-5 scores, there was a similar decrease in scores (P=.26; Cohen d=0.03) from the first to last PCL-5 (iOS: mean −4.58, SD 15.45; Android: mean −4.10, SD 15.12).

Discussion

Overview

Developed by VA NCPTSD and Department of Defense Center for Telehealth and Technology, PTSD Coach is an evidence-based, secure app that is available for free to the general public for the self-management of PTSD symptoms. In a previous study, version 1 of PTSD Coach was found to have been positively received and had a wide reach among members of the public, who used the app to varying extents and found it helpful in reducing momentary distress [17]. Because PTSD Coach has been updated in the last several years with new features and analysis capabilities, in this study, we examined the reach, use, and potential impact of the current version of PTSD Coach, version 3.1, through utilizing public use data between April 2020 and April 2021. In addition, as part of evaluating use, we were able to extend prior work by characterizing the frequency of use of specific app features in PTSD Coach, thereby establishing a baseline on *how* the app is being used as part of examining engagement [25].

Principal Findings and Comparison With Previous Work

First, we found that the PTSD Coach app continued to achieve broad dissemination to the general public, with approximately 210,000 downloads, 150,000 total users, 130,000 new users, and 20,000 active users per month during the 1-year data observation period. Among the 210,000 downloads, 95,000 downloads originated from VA OCC's initiative to distribute iOS tablets that were preloaded with a range of health care apps to veterans. Even after accounting for these institutional downloads, the reach of PTSD Coach appears to be considerably expanded for version 3.1 compared with version 1 (which had approximately 150,000 downloads and 10,000 active users per month during a 3-year observation period [17]).

Second, PTSD Coach, version 3.1, was used, on average, 3 times across 3 separate days for a total duration of 18 minutes of use. Outside of the mean, other values (eg, median, maximum, and IQR) for our use metrics revealed that there was a subset of users with much higher levels of engagement than the average user; we used an example cutoff (ie, ≥2 SDs above the mean for the number of visits) to illustrate the use patterns for this subgroup. In terms of use over time, we observed sharp attrition rates; however, there were also some users who were still using the app 12 months later, indicating potential long-term use. Overall, these metrics demonstrate a similar pattern of use that was previously found with version 1 [17]. There is preliminary evidence, however, that version 3.1 was being used for a longer total duration than version 1 (18 vs 5 minutes, respectively), but we interpret this cautiously, given the different approaches used to define the end of a visit.

For the use of specific features, which we were able to examine for the first time with version 3.1, we found that most users (79,099/128,691, 61.46%) arrived at the home screen and proceeded to a main content area. Among all the content areas, Manage Symptoms was accessed most frequently. Within this content area, over 40% (53,314/128,691, 41.43% to 56,971/128,691, 44.27%; depending on the feature) of users selected a current symptom they wished to address, rated their SUDs, and accessed a coping tool. Users indicated that they most frequently wanted help with PTSD re-experiencing symptoms (ie, Reminded of Trauma), and they frequently accessed tools involving a cognitive restructuring component (ie, Change Your Perspective and Inspiring Quotes); 1 caveat was that these tools were also the most frequently recommended across the different symptoms. In contrast to these Manage Symptoms features, the use of specific features in other content areas and parts of the app was lower (eg, the next highest was 20,964/128,691, 16.29% of users accessing a Learn topic) and was lowest for the safety plan for suicide prevention (with only 320/128,691, 0.25% of users completing the plan). For the safety plan, the actual frequency of use was likely lower, as our numbers were influenced by the NCPTSD training for VA staff on the use of VA mHealth apps. To better increase access to this feature, the NCPTSD is working on building a stand-alone safety plan app.

Third, users of PTSD Coach who provided SUDs ratings or completed self-assessments generally endorsed high levels of momentary distress and high levels of PTSD symptoms with most above clinical threshold. Thus, the reach of the app includes members of the general population who are experiencing difficulties in coping with trauma. In terms of potential impact, the average decrease in momentary distress after coping tool use (approximately 1 point on a scale of 0-10) and the average decrease in PTSD symptoms (approximately 4 points on the PCL-5) were both modest. It is important to note that these averages were calculated from a small proportion of users (eg, 4989/128,691, 3.88%, with at least two PCL-5 assessments). However, even a minority of users in this study still represents a large number of people who experienced potential benefits (eg, approximately one-third, 1585/4989, 31.77% of users with at least two PCL-5 assessments had scores reflecting clinically significant improvement in PTSD symptoms). Comparing across the 2 app versions, users appeared to be similarly distressed. However, when compared with version 1, version 3.1 appears to be associated with slightly attenuated effects for SUDs change after tool use (ie, a 1-point vs 2-point average decrease); it is possible that, as the number

of tools increased within the app, or as the number of mHealth apps increased, a user might have experienced less satisfaction even for the same tool (a general phenomenon known as the paradox of choice [37]). Changes in PCL scores were not available in version 1 for comparison with version 3.1. Although the impact on momentary distress may be small, there was a positive trajectory in the reduction of PTSD symptoms over time with version 3.1.

Notably, the positive trajectory in PTSD symptom reduction appears to reflect a potential overall benefit from the app. Average PTSD symptom reduction (approximately 4 points on the PCL-5) for all users with at least two PCL-5 assessments (n=4989) was not primarily driven by a subset of high-engagement users with at least two PCL-5 assessments (n=1051). This subgroup, however, had a slightly greater PTSD symptom reduction (approximately 6 points). Considering the use patterns described above (eg, a mean of 3 visits for 18 minutes of use), we think that most users may be experiencing benefits from the app by practicing coping tools or learning information about PTSD during times of distress. It is our hope that, after discontinuing use of the app, these users can continue to use the coping tools that they learned or are more empowered to make decisions about how to manage PTSD. In contrast to average users, the subset of high-engagement users visited key features of the app more frequently and over an extended period. These users may be incorporating app features as part of a self-care routine; for example, regularly tracking PTSD symptoms by completing self-assessments.

Lastly, we compared the reach, use, and potential impact of PTSD Coach, version 3.1, for iOS and Android users. The app continues to reach more iOS users than Android users, which makes sense given that iOS users make up most smartphone users in the United States [38]. However, iOS users used the app to a lesser extent. iOS users also showed lower levels of momentary distress and PTSD symptoms. There were mixed findings on whether the 2 groups benefited similarly from the app, with larger decreases in pre- to posttool SUDs for iOS users than for Android users but similar decreases in PTSD symptoms. Of note, most of these statistically significant differences (which were not surprising given our large sample size) were associated with small effect sizes, suggesting that any difference in experience for an individual user across the 2 platforms was relatively small. The greater reach, lower use, and lower distress among iOS users could have been shaped by VA OCC's broad distribution of iOS tablets to veterans (even though veterans themselves may be more likely to own Android smartphones [39]). Some veterans may have opened the app on the tablet but were not motivated to continue to use it because they did not have a particular need for it, in contrast to other veterans or users who searched, found, and installed the app on their own. In fact, approximately only 40% of the veterans receiving a tablet during the observation period had a diagnosis of PTSD (Cindie Slightam, MPH, email communication, October 13, 2021). Taking into account VA OCC's potential influence on version 3.1 iOS metrics, we may effectively be seeing that differences between the 2 platforms have leveled out over time (because with version 1, there was an opposite pattern with lower rates of use for Android users than for iOS users [17]).

Limitations and Future Directions

Although our study had several strengths (eg, a large sample size, a naturalistic approach, and the examination of specific app features), it also had the following limitations. Our PTSD Coach data were collected exclusively during the COVID-19 pandemic, which could have limited the generalizability of our findings. We note that version 3.1 and version 1 users endorsed similar levels of PTSD symptoms and that use of version 3.1 did not seem to fluctuate alongside peaks of COVID-19 cases in the country. This increased our confidence that use of PTSD Coach during this time was still linked to self-management of PTSD symptoms, rather than self-management of more general distress.

We were able to shed light on how different features of the app were being used, but these frequency findings could have been influenced by order effects. For example, among the more frequently accessed parts of the app, the Manage Symptoms content area is located in the upper left quadrant of the home screen, and Reminded of Trauma is at the top of the list of symptoms. To draw stronger conclusions about which app features users are attracted to, future research could use A/B testing designs (eg, switching the order of app features and examining the resulting impact).

Owing to the anonymous nature of the data, we did not have information about our users, beyond their completed ratings and self-assessments. We note that, compared with the percentage of users who endorsed clinically significant PTSD symptoms, a lower percentage of users accessed information about getting professional care within the app. It may be helpful to find ways to highlight these resources in the app. However, it is possible that many of these users are already under the care of a mental health professional. In addition to future research investigating the treatment status of PTSD Coach users, it would be valuable to know the characteristics of users (eg, demographics, veteran status, and trauma history) who experienced clinically significant symptom improvements. For example, we would hope to see that users in this group include both veterans with histories of combat trauma or military sexual trauma as well as nonveterans experiencing other types of trauma (eg, motor vehicle accidents). Gaining traction in the literature, the precision medicine endeavor to answer, "What works well for whom?" [40] should include testing self-management mHealth apps as an intervention format that may be a particularly good fit for certain individuals. Matching people appropriately to using a self-management app could potentially reduce the strain on the mental health system and allow providers to maximize their time (eg, in this case, possibly allowing for the reallocation of 476 direct clinical care hours, if multiplying the n=1585 with clinically significant PTSD symptom improvement by 18 minutes of app use).

Finally, although the raw use data contained timestamps for individual events, the summary of metrics extracted for each user was not in a longitudinal format. Thus, we could not examine the order in which users engaged with different features. For example, we extracted first and last PCL-5

assessment scores, but we did not know when these self-assessments occurred relative to other events in the app. Having a longitudinal data set in which the use of key features is logged in chronological order would enable researchers to better investigate questions of how to optimize engagement with the app. Given that there was a subset of high-engagement users in PTSD Coach, which was consistent with naturalistic studies of other VA self-management apps [34,41], future research could investigate factors that are associated with increased engagement [42] to try to underscore these factors in the app. One such factor could be the completion of a self-assessment upon first using an app, as this was recently linked to using an app on more days as well as using more coping tools, within the COVID Coach app [41].

Conclusions

In summary, we found evidence that PTSD Coach, version 3.1, is serving its intended purpose as a public health resource. The app reached a large number of people, including those who were experiencing significant levels of PTSD symptoms (ie, the target population), and likely expanded access to evidence-based interventions and resources. Most users visited the app only a few times but engaged with key app content. Some app features (eg, coping tools) were accessed more frequently than others (eg, self-assessments), giving researchers a sense of what was appealing to users and what could potentially be improved within the app. Although benefits in momentary distress and PTSD symptoms were generally small on a per-individual basis, the app made these benefits available to the population on a large scale, which could have resulted in a cumulative, positive impact on public health (ie, with impact defined as the product of reach and efficacy [43]). Future research should aim to more flexibly examine the utility of different app features (eg, through A/B testing), as well as to investigate questions on understanding effectiveness (eg, to better match the intervention format to the person) and optimizing engagement (eg, to enhance the likelihood of a meaningful impact) with the app. Pursuing research through these avenues will help to ensure that mHealth apps can reach their full potential to alleviate symptoms and to enhance well-being and functioning for individuals with PTSD.

Acknowledgments

This research was supported by a Department of Veterans Affairs Office of Academic Affiliations Advanced Fellowship in Mental Illness Research and Treatment awarded to HWH. The views expressed in this paper are those of the authors only and do not reflect the official policy or position of the institutions with which the authors are affiliated, the Department of Veterans Affairs, or the US Government.

Authors' Contributions

BKJ, EK, KMR, KT, KJ, PMV, MAM, and JEO contributed to the development of the app. HWH, JW, and JEO contributed to data extraction and formatting. HWH contributed to the data analysis and writing of the manuscript. All authors have contributed to the editing of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of coping tools and learn topics in the PTSD Coach Version 3.1 mobile app.
[DOCX File , 16 KB - mental_v9i3e34744_app1.docx]

References

1. Mobile fact sheet. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2022-02-20]
2. Price M, Yuen EK, Goetter EM, Herbert JD, Forman EM, Acierno R, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. Clin Psychol Psychother 2014;21(5):427-436 [FREE Full text] [doi: 10.1002/cpp.1855] [Medline: 23918764]
3. Demyttenaere K, Bruffaerts R, Posada-Villa J, Gasquet I, Kovess V, Lepine JP, WHO World Mental Health Survey Consortium. Prevalence, severity, and unmet need for treatment of mental disorders in the World Health Organization World Mental Health Surveys. JAMA 2004 Jun 02;291(21):2581-2590. [doi: 10.1001/jama.291.21.2581] [Medline: 15173149]
4. Eylem O, de Wit L, van Straten A, Steubl L, Melissourgaki Z, Danışman GT, et al. Stigma for common mental disorders in racial minorities and majorities a systematic review and meta-analysis. BMC Public Health 2020 Jun 08;20(1):879 [FREE Full text] [doi: 10.1186/s12889-020-08964-3] [Medline: 32513215]
5. Mental Health Culture, Race, and Ethnicity : A Supplement to Mental Health : A Report of the Surgeon General · Volume 2. Rockville, MD: Department of Health and Human Services, U. S. Public Health Service; 2001.
6. Anthes E. Pocket psychiatry: mobile mental-health apps have exploded onto the market, but few have been thoroughly tested. Nature 2016;532(7597):20 [FREE Full text]

7. Torous J, Firth J, Huckvale K, Larsen ME, Cosco TD, Carney R, et al. The emerging imperative for a consensus approach toward the rating and clinical recommendation of mental health apps. J Nerv Ment Dis 2018 Aug;206(8):662-666. [doi: 10.1097/NMD.0000000000000864] [Medline: 30020203]
8. Wang K, Varma DS, Prosperi M. A systematic review of the effectiveness of mobile apps for monitoring and management of mental health symptoms or disorders. J Psychiatr Res 2018 Dec;107:73-78. [doi: 10.1016/j.jpsychires.2018.10.006] [Medline: 30347316]
9. Mohr DC, Cheung K, Schueller SM, Hendricks Brown C, Duan N. Continuous evaluation of evolving behavioral intervention technologies. Am J Prev Med 2013 Oct;45(4):517-523 [FREE Full text] [doi: 10.1016/j.amepre.2013.06.006] [Medline: 24050429]
10. Holowka D, Marx B. Assessing PTSD-related functional impairment and quality of life. In: The Oxford Handbook of Traumatic Stress Disorders. New York, NY: Oxford University Press; 2012.
11. Schnurr PP, Lunney CA, Bovin MJ, Marx BP. Posttraumatic stress disorder and quality of life: extension of findings to veterans of the wars in Iraq and Afghanistan. Clin Psychol Rev 2009 Dec;29(8):727-735. [doi: 10.1016/j.cpr.2009.08.006] [Medline: 19744758]
12. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
13. Wisco BE, Marx BP, Wolf EJ, Miller MW, Southwick SM, Pietrzak RH. Posttraumatic stress disorder in the US veteran population. J Clin Psychiatry 2014 Dec 24;75(12):1338-1346. [doi: 10.4088/jcp.14m09328]
14. Owen JE, Kuhn E, Jaworski BK, McGee-Vincent P, Juhasz K, Hoffman JE, et al. VA mobile apps for PTSD and related problems: public health resources for veterans and those who care for them. Mhealth 2018;4:28 [FREE Full text] [doi: 10.21037/mhealth.2018.05.07] [Medline: 30148141]
15. Beck A, Rush A, Shaw B, Emery G. Cognitive Therapy of Depression. New York, NY: The Guilford Press; 1979.
16. Kuhn E, van der Meer C, Owen JE, Hoffman JE, Cash R, Carrese P, et al. PTSD Coach around the world. Mhealth 2018;4:15 [FREE Full text] [doi: 10.21037/mhealth.2018.05.01] [Medline: 29963560]
17. Owen JE, Jaworski BK, Kuhn E, Makin-Byrd KN, Ramsey KM, Hoffman JE. mHealth in the wild: using novel data to examine the reach, use, and impact of PTSD coach. JMIR Ment Health 2015;2(1):e7 [FREE Full text] [doi: 10.2196/mental.3935] [Medline: 26543913]
18. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
19. Kuhn E, Greene C, Hoffman J, Nguyen T, Wald L, Schmidt J, et al. Preliminary evaluation of PTSD Coach, a smartphone app for post-traumatic stress symptoms. Mil Med 2014 Jan;179(1):12-18. [doi: 10.7205/MILMED-D-13-00271] [Medline: 24402979]
20. Miner A, Kuhn E, Hoffman JE, Owen JE, Ruzek JI, Taylor CB. Feasibility, acceptability, and potential efficacy of the PTSD Coach app: a pilot randomized controlled trial with community trauma survivors. Psychol Trauma 2016 May;8(3):384-392. [doi: 10.1037/tra0000092] [Medline: 27046668]
21. Keen SM, Roberts N. Preliminary evidence for the use and efficacy of mobile health applications in managing posttraumatic stress disorder symptoms. Health Systems 2017 Dec 19;6(2):122-129. [doi: 10.1057/hs.2016.2]
22. Kuhn E, Kanuri N, Hoffman JE, Garvert DW, Ruzek JI, Taylor CB. A randomized controlled trial of a smartphone app for posttraumatic stress disorder symptoms. J Consult Clin Psychol 2017 Mar;85(3):267-273. [doi: 10.1037/ccp0000163] [Medline: 28221061]
23. Possemato K, Kuhn E, Johnson E, Hoffman JE, Owen JE, Kanuri N, et al. Using PTSD Coach in primary care with and without clinician support: a pilot randomized controlled trial. Gen Hosp Psychiatry 2016;38:94-98. [doi: 10.1016/j.genhosppsych.2015.09.005] [Medline: 26589765]
24. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders: Dsm-5™. Washington, DC: American Psychiatric Association; 2013.
25. Zhang R, Nicholas J, Knapp AA, Graham AK, Gray E, Kwasny MJ, et al. Clinically meaningful use of mental health apps and its effects on depression: mixed methods study. J Med Internet Res 2019 Dec 20;21(12):e15644 [FREE Full text] [doi: 10.2196/15644] [Medline: 31859682]
26. Vindegaard N, Benros ME. COVID-19 pandemic and mental health consequences: systematic review of the current evidence. Brain Behav Immun 2020 Oct;89:531-542 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.048] [Medline: 32485289]
27. Zulman D, Wong E, Slightam C, Gregory A, Jacobs JC, Kimerling R, et al. Making connections: nationwide implementation of video telehealth tablets to address access barriers in veterans. JAMIA Open 2019 Oct;2(3):323-329 [FREE Full text] [doi: 10.1093/jamiaopen/ooz024] [Medline: 32766533]
28. Miliard M. VA working with Apple to broaden telehealth access for veterans. Healthcare IT News. URL: https://www.healthcareitnews.com/news/va-working-apple-broaden-telehealth-access-veterans [accessed 2022-02-20]
29. PTSD: National Center for PTSD. U.S. Department of Veterans Affairs. URL: https://www.ptsd.va.gov/professional/tech-care/tech_lectures.asp [accessed 2022-02-20]

30. Weathers F, Litz BT, Keane TM, Palmieri PA PA, Marx B, Schnurr PP. The PTSD Checklist for DSM-5 (PCL-5). National Center for PTSD. URL: www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp [accessed 2021-09-04]
31. Blevins CA, Weathers FW, Davis MT, Witte TK, Domino JL. The posttraumatic stress disorder checklist for DSM-5 (PCL-5): development and initial psychometric evaluation. J Trauma Stress 2015 Dec;28(6):489-498. [doi: 10.1002/jts.22059] [Medline: 26606250]
32. Bovin MJ, Marx BP, Weathers FW, Gallagher MW, Rodriguez P, Schnurr PP, et al. Psychometric properties of the PTSD Checklist for Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition (PCL-5) in veterans. Psychol Assess 2016 Nov;28(11):1379-1391. [doi: 10.1037/pas0000254] [Medline: 26653052]
33. Stanley B, Brown GK. Safety planning intervention: a brief intervention to mitigate suicide risk. Cognit Behav Practice 2012 May;19(2):256-264. [doi: 10.1016/j.cbpra.2011.01.001]
34. Kozlov E, Bantum E, Pagano I, Walser R, Ramsey K, Taylor K, et al. The reach, use, and impact of a free mHealth mindfulness app in the general population: mobile data analysis. JMIR Ment Health 2020 Nov 27;7(11):e23377 [FREE Full text] [doi: 10.2196/23377] [Medline: 33245289]
35. McKinney W. Data structures for statistical computing in Python. In: Proceedings of the 9th Python in Science Conference. 2010 Presented at: Proceedings of the 9th Python in Science Conference; Jun 28 - Jul 3, 2010; Austin, Texas. [doi: 10.25080/majora-92bf1922-00a]
36. Project description for pyodbc. pyodbc. URL: https://pypi.org/project/pyodbc/ [accessed 2022-02-20]
37. Morita P. Design of mobile health technology. In: Design for Health Applications of Human Factors. Cambridge, Massachusetts, United States: Academic Press; 2020.
38. Wallen J. Why is Android more popular globally, while iOS rules the US? Tech Republic. URL: https://www.techrepublic.com/article/why-is-android-more-popular-globally-while-ios-rules-the-us/ [accessed 2022-02-20]
39. Edwards-Stewart A, Smolenski DJ, Reger GM, Bush N, Workman DE. An analysis of personal technology use by service members and military behavioral health providers. Military Med 2016 Jul;181(7):701-709. [doi: 10.7205/milmed-d-15-00041]
40. Bickman L, Lyon AR, Wolpert M. Achieving precision mental health through effective assessment, monitoring, and feedback processes : introduction to the special issue. Adm Policy Ment Health 2016 May;43(3):271-276 [FREE Full text] [doi: 10.1007/s10488-016-0718-5] [Medline: 26887937]
41. Jaworski BK, Taylor K, Ramsey KM, Heinz A, Steinmetz S, Pagano I, et al. Exploring usage of COVID coach, a public mental health app designed for the COVID-19 pandemic: evaluation of analytics data. J Med Internet Res 2021 Mar 01;23(3):e26559 [FREE Full text] [doi: 10.2196/26559] [Medline: 33606656]
42. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
43. Abrams D, Orleans C, Niaura R, Goldstein M, Prochaska J, Velicer W. Integrating individual and public health perspectives for treatment of tobacco dependence under managed health care: a combined stepped-care and matching model. Ann Behav Med 1996 Dec;18(4):290-304 [FREE Full text] [doi: 10.1007/bf02895291]

Abbreviations

EULA: End User License Agreement
mHealth: mobile health
NCPTSD: National Center for PTSD
OCC: Office of Connected Care
PCL-5: Posttraumatic Stress Disorder Checklist for DSM-5
PTSD: posttraumatic stress disorder
SUDs: subjective units of distress
VA: Department of Veterans Affairs

Edited by J Torous; submitted 05.11.21; peer-reviewed by M Price, P Yellowlees; comments to author 20.12.21; revised version received 24.01.22; accepted 24.01.22; published 29.03.22.

Please cite as:
Hallenbeck HW, Jaworski BK, Wielgosz J, Kuhn E, Ramsey KM, Taylor K, Juhasz K, McGee-Vincent P, Mackintosh MA, Owen JE
PTSD Coach Version 3.1: A Closer Look at the Reach, Use, and Potential Impact of This Updated Mobile Health App in the General Public
JMIR Ment Health 2022;9(3):e34744
URL: https://mental.jmir.org/2022/3/e34744
doi:10.2196/34744
PMID:35348458

Original Paper

Measuring Adherence Within a Self-Guided Online Intervention for Depression and Anxiety: Secondary Analyses of a Randomized Controlled Trial

Maria Hanano[1], BA; Leslie Rith-Najarian[2,3], PhD; Meredith Boyd[1], MA; Denise Chavira[1], PhD

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

[2]Department of Psychology, Harvard University, Cambridge, MA, United States

[3]Dean of Student's Office, Harvard University, Cambridge, MA, United States

Corresponding Author:
Maria Hanano, BA
Department of Psychology
University of California, Los Angeles
1285 Franz Hall, Box 951563
Los Angeles, CA, 90095
United States
Phone: 1 9493501523
Email: mariahanano@g.ucla.edu

Abstract

Background: Self-guided online interventions offer users the ability to participate in an intervention at their own pace and address some traditional service barriers (eg, attending in-person appointments, cost). However, these interventions suffer from high dropout rates, and current literature provides little guidance for defining and measuring online intervention adherence as it relates to clinical outcomes.

Objective: This study aims to develop and test multiple measures of adherence to a specific self-guided online intervention, as guided by best practices from the literature.

Methods: We conducted secondary analyses on data from a randomized controlled trial of an 8-week online cognitive behavioral program that targets depression and anxiety in college students. We defined multiple behavioral and attitudinal adherence measures at varying levels of effort (ie, low, moderate, and high). Linear regressions were run with adherence terms predicting improvement in the primary outcome measure, the 21-item Depression, Anxiety, and Stress Scale (DASS-21).

Results: Of the 947 participants, 747 initiated any activity and 449 provided posttest data. Results from the intent-to-treat sample indicated that high level of effort for behavioral adherence significantly predicted symptom change ($F_{4,746}=17.18$, $P<.001$; and $\beta=-.26$, $P=.04$). Moderate level of effort for attitudinal adherence also significantly predicted symptom change ($F_{4,746}=17.25$, $P<.001$; and $\beta=-.36$, $P=.03$). Results differed in the initiators-only sample, such that none of the adherence measures significantly predicted symptom change ($P=.09$-$.27$).

Conclusions: Our findings highlight the differential results of dose-response models testing adherence measures in predicting clinical outcomes. We summarize recommendations that might provide helpful guidance to future researchers and intervention developers aiming to investigate online intervention adherence.

Trial Registration: ClinicalTrials.gov NCT04361045; https://clinicaltrials.gov/ct2/show/NCT04361045

(JMIR Ment Health 2022;9(3):e30754) doi:10.2196/30754

KEYWORDS

self-guided; adherence; depression; anxiety; online intervention

Introduction

There has been a proliferation of online interventions aimed at preventing and treating mental health disorders (eg, [1]). Online interventions have the potential to reach a wide audience while bypassing barriers that are more common to traditional face-to-face interventions, such as financial cost, inaccessibility, and stigma [2]. The flexible nature of online interventions

provides its users autonomy to interact with content according to their unique schedule and preferences [3]. Self-guided online interventions, in particular, maintain anonymity, address some concerns related to stigma, are less costly, and require little time from mental health professionals [4,5]. Recent meta-analyses have shown that both guided and self-guided online interventions can be effective for treating a range of problems such as depression and anxiety [1,3,6]. Moreover, research has shown that online interventions attract a large number of individuals (eg, 38,000 registrants to MoodGYM) who experience significant mental health symptoms [2].

Despite the promising nature of self-guided online interventions, multiple reviews report high dropout and poor adherence rates [1,7]. A meta-analysis found that while 72% of adults adhered to guided online interventions (ie, guided by a mental health professional), only 26% adhered to self-guided online interventions [8]. As one example, an online self-guided and publicly available cognitive behavioral therapy program aimed at preventing depression and anxiety attracted 38,000 registrants, but only 3.9% adhered to the intervention (adherence was defined as completing 3 of the 5 modules; [2]). Eysenbach [9] described this "law of attrition" as a fundamental challenge for online intervention trials—relative to drug or psychosocial therapeutic trials—because participants are less closely supervised and thus, they receive more sporadic doses of an intervention, or even none at all. Because of the variability in adherence rates, it is difficult to measure and make conclusions about the effectiveness of such interventions. Presently, there is limited understanding about how adherence within self-guided online interventions affects clinical outcomes, which in turn limits our ability to identify those intervention components that might be the necessary mechanisms of change.

There are currently many challenges to understanding adherence and its relation to outcomes within online interventions. Such challenges include varying ways of operationalizing and measuring adherence, which reduce our ability to compare adherence rates across various trials [10,11]. The term "adherence" is often used interchangeably with terms such as engagement, user retention, or dropout (eg, [12]). Focusing on definitions for "adherence," Sieverink and colleagues [11] analyzed how 62 studies operationalized adherence to online interventions (both guided and self-guided), and found that operationalizations fell into 3 categories: (1) "the more usage, the better"; (2) researcher-defined "intended use" but without justification (eg, a user is adherent when logging in at least once a week for 3 weeks); and (3) researcher-defined "intended use" justified using theory, evidence, or rationale (eg, We know from previous research that users benefit the most from the technology when finishing module 4, so a user is adherent once module 4 is completed). Beintner and colleagues [10] found that an array of usage measures can define adherence, such as percentage of participants completing all modules, percentage of participants completing each module, percentage of participants who visited the website, average number of log-ins, and average duration of visit. This variability in measurement has prevented convergence of evidence on which adherence measures are valid. Consequently, the lack of standardized adherence measures perpetuates a cycle where researchers use a wide variety of adherence measures for self-guided online interventions. The result is less clarity on a conceptual framework of adherence as it applies to online interventions.

Various reviews provide actionable recommendations for improving the standardization of how we define, measure, and report adherence to online interventions (eg, [10,11]). We have distilled various recommendations into 2 broad topics we believe to be particularly useful and feasible: (1) creating and reporting multiple measures, and (2) relating these measures to outcomes.

The first recommendation—to create and report multiple measures of adherence—facilitates our understanding of the multiple ways by which participants may adhere to online interventions [10]. According to Beintner and colleagues [10], such measures should be both universal (eg, measures of average number of completed sessions) and intervention specific (eg, completing diaries or discussion boards) [10]. Universal measures are most frequently used, allowing for the comparison of such metrics across studies. Measures of adherence specific to the intervention should also be created and reported, in order for study designers to understand whether study-specific components are beneficial. It is also recommended that measures reflect the interventions' intended use [11]. This recommendation is in line with the World Health Organization's definition of adherence as "the extent to which a person's behavior – taking medication, following a diet, and/or executing lifestyle changes – corresponds with agreed recommendations from a health care provider" [13]. Sieverink and colleagues [11] propose that it is most useful to understand the threshold required, or how much adherence is necessary, in order for it to predict improved outcomes, rather than assuming that more adherence is always better. In summary, it may be advantageous to have some universal adherence measures and some intervention-specific adherence measures based on the intervention's intended/recommended minimum use.

The second recommendation is to test the relationship between measures of adherence and clinical outcomes to inform the validity and utility of adherence measures. This process is critical to make correct interpretations about the impact of adherence on target outcomes and the effectiveness of the online intervention [10,14]. A recent review found that treatment adherence (defined as the total number of sessions completed by the participant divided by the total number of treatment sessions) was significantly related to outcomes within self-guided online interventions [4]. Another study, which examined multiple types of adherence measures (activities completed per login, total time spent online, total time spent online per log-in, combined modules, and activities measure), found that only the number of activities completed per login was significantly associated with better outcomes for those who received the online intervention [15]. They also divided patterns of usage into 3 levels (low, medium, and high) and found that medium-level users did not differentially benefit from the intervention compared with low-level users [15]. In turn, measures of adherence that are predictive of symptom improvement should inform an online intervention's engagement strategies.

In this study, we follow these recommendations to test questions about *how much* adherence and *which* measures of adherence matter in predicting symptom improvement within a self-guided online intervention. The first aim of the study is to demonstrate an example of the process of testing multiple measures as applied to a specific self-guided online intervention. The second aim is to translate the results of this process into recommendations for future researchers and interventionists to consider when making decisions about creating and testing measures of adherence in online interventions.

Methods

Design

We conducted a secondary analysis using data from a randomized controlled trial (RCT) that tested a self-guided web-based mental health skills program for universal prevention of anxiety and depression in university students [16]. Primary results of this trial showed small intervention effects overall [16]. Participants were randomly assigned to an immediate intervention condition, (ie, they could access the intervention upon signing up), or a delayed access condition (ie, they were on a waitlist initially and only granted access to the online platform after the immediate intervention condition was over). The timing of both conditions was staggered such that the start and finish week of the intervention access for both conditions corresponded to equivalent weeks within the respective academic quarter. For the purposes of this study, data collected from both conditions were collapsed such that pre-/postscores reflect each participant's status immediately before receiving the intervention and immediately after receiving the intervention.

Participants

Participants were at least 18 years old, undergraduate and graduate students at the University of California, Los Angeles. Recruitment efforts included, but were not limited to, department-wide emails, flyers posted around the university, social media, and announcements in psychology courses. Compensation for research survey participation included entry into US $10-US $100 gift card drawings or course credit. Exclusion criteria were being enrolled in a similar anxiety and depression treatment study, invalid data reporting (eg, straight-lining or high inconsistency in responses), and not verifying one's online intervention account or not completing the account setup process. Out of a total of 947 participants, 747 initiated any activity, and 449 had posttest outcome data.

Intervention

A more detailed description of the intervention and screenshots of the platform can be found in the primary intervention [16]. In this section, we describe the information most relevant for understanding adherence within the context of the tested intervention. The program consisted of 8 weeks, each of which focused on an evidence-based skills theme. Participants were allowed to choose what they would like to practice from a list of activities relevant to the respective weeks' theme. For example, the "Change Your Thinking" week provided instructions for cognitive restructuring strategies, and activities that participants could practice and log included "identify any unhelpful thinking patterns," "identify evidence for and against the unhelpful thought," "shift your attention," etc. As another example, the "Pause" week focused on strategies to foster mindfulness practice in daily life, and provided instructions for activities such as "eat mindfully," "listen mindfully," "meditate mindfully." With regard to how many skills participants would practice each week, there were no such requirements in the instructions of the intervention, but rather the online platform used a virtual medal system to incentivize more practice. Participants were awarded medals depending on the amount of activities logged each week: "Bronze" is awarded when a participant completes at least one log for that week; "Silver" is awarded when a participant completes three logs for that week; "Gold" is awarded when a participant completes at least five logs for that week. Finally, in addition to logging any skills they practiced, participants were prompted to submit an end-of-week check-in comprising 2 reflective questions about skills practiced that week. Examples of reflective questions that participants could answer during weekly check-ins were as follows: Week 3: "Which technique was most helpful for you?"; "Did this week move you closer or not to your goals?". Week 7: "Did being more mindful make your more aware of anything in your life or daily experiences?"; "Did this week move you closer or not to your Life 2.0?" Submitting a check-in was completely optional, however, they are required to log any activities they practice on the activity log tab on the platform.

Measures

Adherence

Following the first recommendation, multiple measures of adherence were created that were (1) both universal and study specific and (2) designed to capture the intervention's intended use.

Staudt's model of adherence (referred to as *engagement* in Staudt 2007 [17]) informed the selection of our universal measures. Specifically, Staudt proposes that adherence can be thought of universally as involving behavioral and attitudinal aspects. This model defines behavioral adherence as, "client performance of the tasks that are necessary to implement treatment and to ultimately achieve outcomes." Examples of behavioral adherence in face-to-face interventions include maintenance of appointments, homework completion, and responsiveness to the practitioner. The model defines attitudinal adherence as, "the emotional investment in and commitment to treatment that follow from believing that it is worthwhile and beneficial" [17]. Examples of attitudinal adherence in face-to-face interventions include positive attitude toward the intervention, perceiving the intervention as worthy of time and energy, perceiving the benefits outweigh the costs of the treatment. The author explains that a person's attitude toward the intervention represents the "heart" of adherence and is necessary for participants to make meaningful changes during an intervention. In the context of our study, to ensure that study-specific behavioral and attitudinal adherence measures were defined based on the interventions' intended use, we referred to the intervention's weekly instructions provided to participants. The resulting adherence measures are described in more detail below and summarized in Table 1.

Table 1. Definition of 6 measures of adherence based on recommendations.

Measure	Behavioral adherence[a,b]	Attitudinal adherence[c,d]
Minimal effort	Number of weeks with at least one skill practice log	Number of weeks with check-in word count of any length
Moderate effort	Number of weeks with at least three skill practice logs	Number of weeks with check-in word count ≥ respective average
High effort	Number of weeks with at least five logs	Number of weeks with check-in wordcount ≥ respective average +1 SD

[a]Universal definition: performance of intervention-related tasks.

[b]Intervention-specific definition: practice of skills, per amount of weekly logged activity within user account.

[c]Universal definition: emotional investment.

[d]Intervention-specific definition: elaborateness of written responses to reflection prompts, per word count of weekly check-ins within user account.

Behavioral Adherence

Regarding behavioral adherence, as a reminder, participants practiced skills each week and were awarded 1 of 3 medals depending on the number of skills practiced (see the "Intervention" section; [16]). As such, behavioral adherence was operationalized by the skills practice logs within each user's account. We created 3 behavioral adherence measures, which reflected the number of weeks with behavioral adherence at 3 effort levels (ie, minimal, moderate, and high), categorized according to the medal system: (1) Minimal behavioral adherence: number of weeks with at least one skills practice log; (2) Moderate behavioral adherence: number of weeks with at least three logs; and (3) High behavioral adherence: number of weeks with at least five logs.

Attitudinal Adherence

In order to have somewhat parallel operationalizations of behavioral and attitudinal measures, we also applied the 3 levels of effort to attitudinal adherence. For attitudinal adherence, participants were prompted to respond to 2 weekly check-in questions, though they were not instructed on how much to write. As such, attitudinal adherence was operationalized as the extensiveness of user's open-ended reflective responses on weekly check-ins. Because these check-ins are optional and encourage the participants to reflect on their experience and growth, we believed that the act of electing to complete a check-in would reflect the participant's emotional investment in the intervention. Moreover, writing a longer response to an optional check-in question, rather than briefly answering the question, reflects varied levels of adherence effort by participants. To create minimal, moderate, and high levels of attitudinal adherence, we used word count on weekly check-ins. One previous study found that diary entry word count in an online intervention was correlated with the number of activities that the individual logged [18]. First, we cleaned weekly check-ins to remove (1) duplicate responses, and (2) random or nonalphanumeric characters. Second, we obtained means of word count on end-of-week check-ins for each week and used them as cut-offs determining each level of effort. Third, we created the attitudinal adherence measures at each effort level, defined as

- Minimal: number of weeks with at least one word.
- Moderate: number of weeks with word mean at mean or above for each respective week.
- High level: number of weeks with word mean at 1 SD from mean and above for each respective week.

For example, the mean word count on the third week of the intervention was 32.87; participants at or above this mean for the third week were considered to have moderate attitudinal adherence to the intervention for that week.

Primary Outcome: Depression, Anxiety, and Stress

The primary outcome measure in this study was the 21-item Depression, Anxiety, and Stress Scale (DASS-21), which assesses self-reported symptoms of depression, anxiety, and stress [19]. In previous studies the measure had demonstrated adequate internal consistency (Cronbach α=0.83-0.90) and construct validity [19]. In this RCT, internal consistency of the DASS-21 was adequate (total: Cronbach α=.92; depression: Cronbach α=.89; anxiety: Cronbach α=.79; stress: Cronbach α=.82) [16]. DASS-21 total symptom change scores were calculated (post-pre scores) and used as the primary outcome. We expected use of DASS-21 total scores (as opposed to subscale scores, for example) to maximize the power of our analyses, given that it showed the largest effect size per our primary intervention main effect analyses [16] (Exploratory linear regression analyses served to support that there were indeed decreased strength and significance in the relationships between adherence variables and DASS-21 subscales.).

Covariates

Covariates selected for the study included (1) condition (intervention upon signing up vs delayed access condition); (2) suicidal ideation at baseline measured through the use of question 9 on the 9-item Patient Health Questionnaire (PHQ-9) [20]; and (3) gender. The PHQ-9 is a 9-question measure widely used to assess self-reported symptoms of depression. We operationalized suicidal ideation through question 9, which directly assesses for suicidal ideation [20]. Gender was defined as binary "female" or "male," based on the demographics data linked from participants official student records. The rationale for examining suicidal ideation and gender as covariates is outlined in the primary RCT [16]. Inclusion of these 3 covariates was thus consistent with those tested in the primary online intervention analyses. These 3 covariate variables were also included as auxiliary variables informing our multiple imputation model.

Other Measures

Remaining variables were included only as auxiliary variables for the multiple imputation model in predicting missing data in our primary outcome variable (see the "Data Analysis" section for more details). The Grit Scale (GRIT) is a 12-item measure

aimed at measuring traits of perseverance, maintaining focus, and interest in long-term goals [21]. The Treatment Motivation Questionnaire (TMQ) [22] assesses reasons for initiating and remaining in treatment; we used an adapted version of this measure to apply to the tested online intervention [16]. The Subjective Happiness Scale (SHS) [23] is a 4-tem scale that measures global subjective happiness.

Statistical Analysis

Preliminary Analyses

First, given the amount of missing outcome data at posttest, we conducted a series of independent unpaired *t* tests and chi-square tests to determine whether any variables were significantly related to posttest missing data status, which would suggest if data were not missing completely at random (MCAR). Relatedly, to support our selection of the multiple imputation model to deal with missingness, we aimed to verify that our data set was not MCAR using the Little MCAR test. Second, to identify necessary covariates for the regression analyses, we conducted the same series of analyses but this time to assess all variables as predictors for significant differences in DASS-21 change scores. Third, we conducted linear regression assumption checks to determine if (1) the necessary conditions were met, and (2) conducting multiple linear regressions with more than 1 adherence predictor simultaneously was appropriate [24].

Multiple Imputation

Given the degree of missing DASS-21 outcome data at posttest (370/947, 39.1%) and results of preliminary analyses (reported below), we determined that analyzing data only from complete cases would likely produce biased findings with decreased power. As such, we implemented multiple imputation procedures for missing posttest DASS-21 values (ie, outcome), predicted by auxiliary variables: the 6 measures of adherence, condition, gender, suicidal ideation, GRIT, TMQ, PHQ-9, SHS, and total baseline DASS-21. We were liberal in our selection of auxiliary variables, and included all measures collected at baseline in the imputation model to preserve any complex associations that may exist among the variables, especially considering that adding too many variables is unlikely to produce bias [25]. We also elected to run 50 imputations to decrease standard errors and produce stable estimates, which require between 50 and 100 imputations [25].

Linear Regressions

Multiple linear regressions were run including all identified covariates in step 1, and then adding the respective adherence measure as a variable in step 2, predicting the main outcome variable of DASS-21 change scores (posttest – baseline scores). Results from the 50 multiple imputations were then pooled and analyzed. Given that Type II error rates increase as more models are tested, we calculated Benjamini-Hochberg critical values [26], an alternative to simply applying a *P* value of .05 across all models.

Ethics Approval

This study was approved by the appropriate University of California Los Angeles Institutional Review board (IRB# 17-000761).

Results

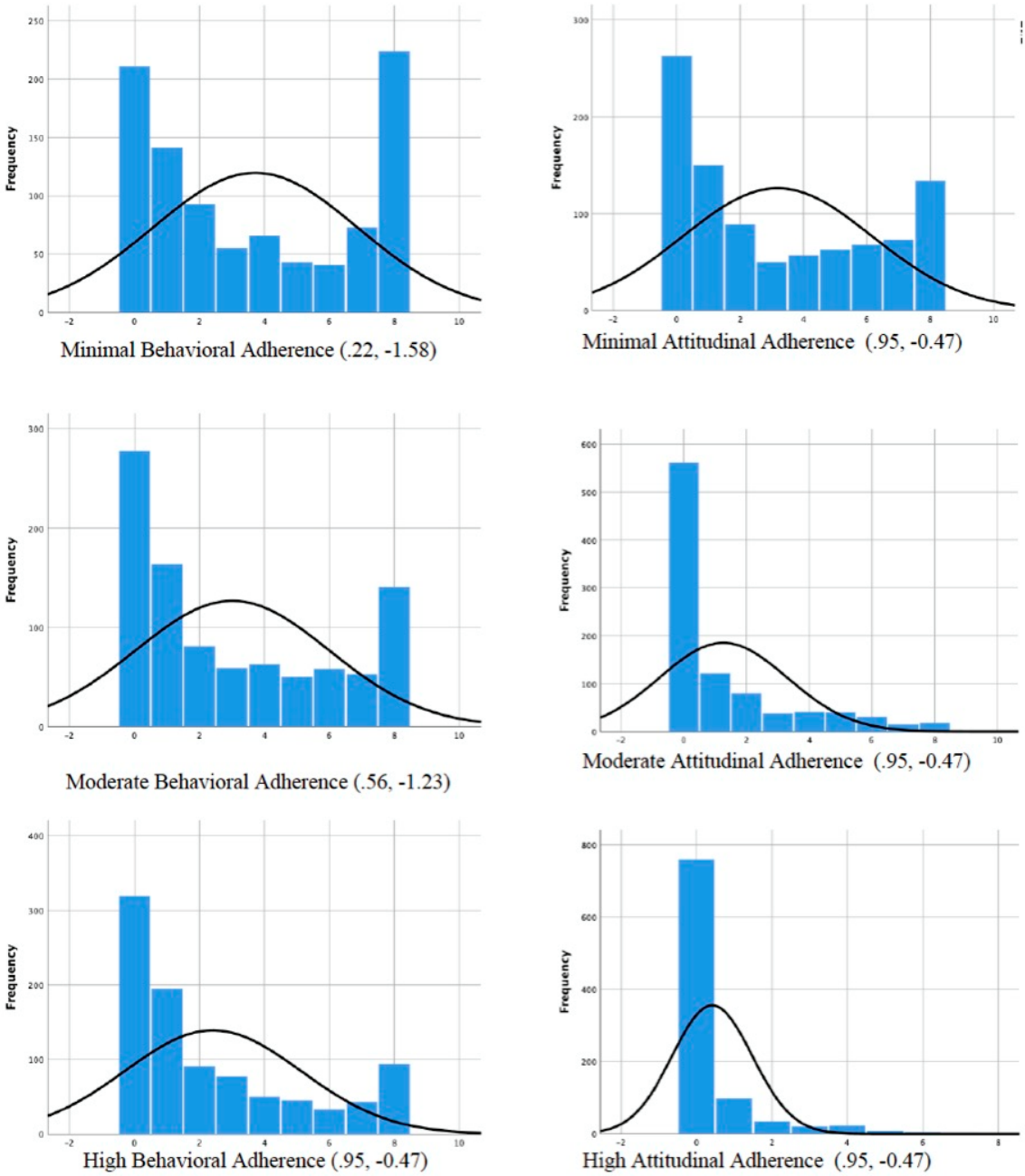
Participants

The sample consisted of 947 students, with a mean age of 23.01 (SD 5.56), with 729 (77.0%) identified as female, 299 (31.6%) as White, and 121 (12.8%) as international. The groups that received the intervention immediately consisted of 587 participants, and the delayed intervention group consisted of 360 participants. The number of program initiators in this sample, that is, those with any activity stored within their user account, was 747 students. Baseline DASS-21 total scores had a mean of 17.30 (SD 10.47), and the mean DASS-21 change was –2.67 (SD 9.38). The number of participants with posttest outcome data (DASS-21) was 449/947 (47.4% of sample).

Adherence

For behavioral adherence, the mean number of skills practice logs per module overall was 2.14 (n=947). The percentage of 0 skills practice logs on each module ranged from 47.7% (452/947) to 63% (597/947). When participants with 0 logs on each separate module were excluded, the mean number of skills practice logs was 4.51. The module with the highest average skills practice logs was the Welcome module (mean 5.06 [SD] 2.25), and the one with the lowest was the Physical Exercise module (mean 3.58 [SD] 1.92). For attitudinal adherence, the mean weekly check-in word count for modules overall was 57.70. The module with the highest check-in word count was the Wrap-up week (mean 53.03 [SD] 37.91), and the one with the lowest was the Welcome week (mean 25.13 [SD] 22.34). Figure 1 presents the full distribution of participants meeting respective adherence criteria across 1-8 modules.

Figure 1. Distribution plots of each adherence measure (Adherence measure, Skewness, Kutorsis).



Preliminary Analyses

Assessment of Missing Data

Results of independent *t* tests and chi-square tests treating posttest data missing status as the independent variable in relation to other variables revealed that only condition and program initiation were significantly related to posttest data missing status (*P*<.001). By contrast, there were no significant differences between posttest data missing status and GRIT, TMQ, and PHQ-9 (*P*=.81-.08). The Little MCAR test confirmed that our data set was not MCAR and therefore utilizing the listwise deletion would result in biased coefficients (χ^2_{38}=399.95, *P*<.001). For this reason, multiple imputation was used instead to handle missing data in our subsequent linear regression models.

Covariate Identification Analyses

Results of independent *t* tests and chi-square tests revealed that suicidal ideation was significantly related to DASS-21 change scores (*P*<.001). Results also revealed that GRIT, TMQ, PHQ-9, SHS, program initiation, gender, and condition were not significantly related to DASS-21 change scores (*P*=.07-.9). However, gender had been a significant covariate in the main RCT analyses [16], and the condition variable is conceptually meaningful, given that our data set collapsed pre-post outcomes from each condition at different respective assessment periods. Therefore, suicidal ideation, gender, and condition were entered as covariates in our regression models.

Assumption Checks

We conducted assumption checks to confirm utilizing multiple linear regressions was appropriate [24]. The following

assumptions were met: (1) 1 continuous dependent variable; (2) multiple independent variables; (3) independence of observations (Durbin-Watson=1.99); (4) linearity between the dependent and independent variables (confirmed via separate scatter plots); and (5) homoscedasticity (confirmed via residual scatter plots based on a model including all adherence variables and covariates). However, assumption 6 requiring that data must not show multicollinearity between all variables was not met (variance inflation factor=1.00-12.77). Therefore, moving forward we tested each independent variable in its own model (including covariates), and there was no multicollinearity (variance inflation factor=1.00-1.01). We also identified outliers and potential influential cases through examination of models' scatter plots, leverage, Cook *D*, and DFBETA values. Although some participants exceeded respective diagnostic test values, DFBETAs revealed little influence specifically from our adherence variables of interest. Specifically, only 2 participants were influential on 1 adherence variable, with no participants identified as influential for the remaining 5 adherence variables. Therefore, no cases were removed from subsequent analyses. Assumption 8 requiring normally distributed residuals was confirmed by graphs of each model's standardized residuals: histograms looked approximately normal and each Q-Q plot showed a linear line.

Main Findings (Intent to Treat)

See Table 2 for results of all 6 regressions. Linear regressions revealed that high behavioral adherence significantly predicted symptom improvement (unstandardized *B*=–0.26, *P*=.04). Additionally, moderate attitudinal adherence also significantly predicted symptom improvement (unstandardized *B*=–0.36, *P*=.03). All other measures of behavioral and attitudinal adherence did not significantly predict symptom improvement (*P*=.05-.10). None of the observed *P* values from these 6 regressions (*P*=.03-.10) fell below their respective Benjamini-Hochberg critical value (0.008-0.50), indicating that significant results may be due to false discovery rate.

Table 2. Results from linear regressions examining 6 measures of adherence.

Adherence term in model and variables	Intent-to-treat sample (n=947)		
	Pooled R^2	Pooled *B*	*P* value
Behavioral minimal effort	0.068		
Gender		–7.21	.77
Condition		–0.26	.17
SI[a]		–1.03	<.001
Adherence		–0.24	.05
Behavioral moderate effort	0.067		
Gender		–7.18	.73
Condition		–0.31	.16
SI		–0.99	<.001
Adherence		–0.22	.07
Behavioral high effort	0.067		
Gender		–0.33	.70
Condition		–0.99	.16
SI		–7.16	<.001
Adherence		–0.26	.04
Attitudinal minimal effort	0.067		
Gender		–0.28	.74
Condition		–0.94	.19
SI		–7.20	<.001
Adherence		–0.23	.07
Attitudinal moderate effort	0.069		
Gender		–0.32	.71
Condition		–1.00	.16
SI		–7.11	<.001
Adherence		–0.36	.03
Attitudinal high effort	0.065		
Gender		–0.38	.66
Condition		–0.96	.18
SI		–7.04	<.001
Adherence		–0.51	.10

[a]SI: suicidal ideation.

Post Hoc Findings (Initiators-Only Sample)

As a post hoc check, we reran the main analyses excluding those that did not initiate any activity, for quantitative and conceptual reasons. Quantitatively, the distribution plots of the adherence variables reveal that data are largely skewed to the left (Figure 1), which is largely due to 21.1% (200/947) of participants not initiating any type of activity on the platform. Additionally, preliminary *t* tests revealed that program initiation was significantly related to postdata missingness (t_{945}=–9.29, *P*<.001). Conceptually, previous research has shown that despite a large number of individuals enrolling in self-guided online interventions, very few actually initiate the program [9]. For example, a systematic review of self-guided online interventions for depression and anxiety found that 33%-88% of users who downloaded an app actually used it at least once [2]. Thus, we expected that an initiator sample might reveal different findings about the relationship between adherence and outcomes for the self-guided online program.

The initiator sample consisted of 747 students, with a mean age of 23.01 (SD 5.56), who were 78.8% (n=589) female, 31.3% (n=234) White, and 11.0% (n=82) international. At baseline, the mean DASS-21 total score was 17.55 (SD 10.59). In this sample, the number of people with post-DASS-21 data was 410

(54.9%), indicating this subsample had more complete pre-post data than the full sample. The *t* and chi-square tests examining the differences between program initiators and noninitiators revealed that gender was significantly related to program initiation, such that female students were more likely to initiate than male students (*P*=.01). All other variables (suicidal ideation, *P*=.21; TMQ, *P*=.18; GRIT, *P*=.46; PHQ-9, *P*=.07; and SHS, *P*=.98) were not significantly related to program initiation. Linear regressions revealed that none of the adherence terms in all 6 linear regressions models significantly predicted symptom improvement (*P*=.28-.08).

Discussion

Summary of Findings

Given that online interventions offer users ease of accessibility, autonomy, and flexibility, many researchers are greatly interested in identifying indicators of adherence to online interventions that are predictive of symptom improvement. Because of the lack of consensus in the literature regarding the operationalization and measurement of adherence, our first aim was to demonstrate a process for intentionally defining multiple measures of adherence to test their utility in predicting symptom improvement. To achieve this aim, we created multiple measures that fit into universal categories (eg, behavioral, attitudinal) but that were still intervention specific (eg, number of skill practice logs, word count on weekly check-ins). Each adherence measure was specified based on the interventions' intended use (eg, 3 levels of effort based on the platform's virtual medal system). We believe that "behavioral" and "attitudinal" dimensions of engagement are universal enough to be widely applicable across intervention designs, allowing for comparison between interventions, though the intervention-specific adherence metrics may still vary. For example, these categories could be applied to Headspace [27], a popular meditation online application. In this case, behavioral adherence could be defined as the number of meditation modules used or the number of minutes listed; attitudinal adherence could be defined as completing the check-in questions in the "Journey" section or customizing the notifications in the "Settings" section.

Results from the intent-to-treat sample demonstrated that behavioral and attitudinal measures of adherence were predictive of symptom improvement at differing levels of effort. Specifically, high-effort behavioral adherence (ie, number of modules with at least five logs; *P*=.04) and moderate-effort attitudinal adherence (ie, number of modules with check-in word count at or above respective average; *P*=.03) predicted significantly more decrease in DASS-21 total scores. By contrast, results from the initiator sample revealed none of the adherence measures as predictive of symptom improvement. In other words, when our adherence measures were tested with a more conservative definition of the intervention user sample, the previously detected adherence effects disappeared. In summary, whether or not adherence effects surmounted conventional levels of statistical significance (ie, 5% probability of being observed due to random chance) depended on definitions of *how* the intervention is used (ie, type of adherence), *how much* it is used (ie, effort of adherence), and also *who* is a user (ie, intervention sample criteria).

Implications

Although the aggregate of our findings could be interpreted as weak evidence for adherence effects within self-guided online mental health interventions, we instead interpret them as supporting just how challenging it is to measure such dose-response effects. If such adherence were truly unimportant, then we would have expected larger *P* values for most or all adherence terms (Table 2). By contrast, our 6 main models converged such that the negative relationship between intervention adherence and symptoms had only a 3%-10% probability of being observed due to random chance. Regarding the relatively low amount of variance explained, it is likely due at least in part to statistical constraints. On the one hand, we are measuring an extremely diverse independent variable: the seemingly infinite ways and degrees that individuals can adhere to self-guided online interventions [10]. One the other hand, we are simultaneously trying to detect changes in a constrained dependent variable: self-guided online mental health interventions generally produce smaller effects with a restricted range of improvement [4,28]. Indeed, the main trial results for the currently tested intervention found robust but small effects [16]. This dilemma can be characterized as testing high-variance dose (ie, adherence) in the prediction of small-effects response. Unfortunately, we are still left with the unresolved questions of *which* and *how much* adherence should be prioritized for users of self-guided online interventions. In service of resolving these questions in the future, we have translated some lessons learned into recommendations that future researchers and developers could use when investigating the role of adherence within a specific online intervention (Table 3). We hope that these recommendations will assist others in understanding and measuring adherence in a more thorough and standardized manner. We elaborate below on some potential additional benefits that following our outlined recommendations could provide.

Table 3. Recommendations for research on adherence-outcome effects of online interventions.

Item	Recommendation	Section cross-reference
Operationalize adherence measures	• Literature review: Identify relevant research definitions of adherence and recommendations for developing adherence definitions/measures. Given that measures of adherence on guided or self-guided online interventions are variably defined in the current literature, it is important to use past knowledge to move toward standardization. • Define multiple measures: Based on that review, create multiple measures of adherence to the given online intervention. Because of the highly variable designs and features of online interventions, adherence to them can be measured in many ways. Defining multiple measures a priori and reporting on all of them allow for testing of differential effects of adherence (ie, to which features? How much?) on improvement in outcomes.	• See the "Introduction," "Adherence," "Behavioral Adherence," and "Attitudinal Adherence" sections for relevant literature. • See the "Adherence" section for results supporting variability in respective adherence rates. • See the "Main Findings (Intent to Treat)" section for main results demonstrating significance tests varying for each adherence term.
Select primary outcome measure	• Carefully select your outcome measure with attention to maximizing detection of adherence-outcome effects. Given that online interventions are prone to small main intervention effects, the sensitivity of an outcome measure is especially crucial for adherence-outcome effects. • Note: If you have multiple outcome measures or multiple subscales, you might maximize your power by selecting that which showed largest effect size per your primary intervention main effect analyses.	• See the "Primary Outcome" section for methodological support of our choice of primary outcome, based on previously reported results. See Table 2 for demonstration of small effects.
Select an appropriate data analytic plan	• Model selection: Design an analysis plan that is appropriate for the specific goal of testing adherence-outcome effects. For example, because users can be simultaneously adhering to multiple aspects of an online intervention, it is all the more important to check for collinearity. If such assumptions would be violated, adherence measures must be separately tested as predictors. • Covariates selection: Select covariates with primary intervention analyses in mind. Because adherence-outcome effects are presumably tested after primary intervention effects, covariation selection should be consistent across both.	• See the "Preliminary Analyses" and "Linear Regressions" sections for justification of model selection. See the "Assumption Checks" section for model assumption check results. • See the "Covariates" section for justification of covariate selection. • See the "Covariate Identification Analyses" section for results supporting our covariate selection.
Identify method to deal with missing data	• First, identify the rate of missingness in your primary outcome measure. Once a rate is identified, select an appropriate method for dealing with missing data (ie, last observation carried forward, raw data, completer only). Because online interventions often experience high dropout, a larger proportion of data may be missing, and thus results could drastically change by imputation decision.	• See the "Multiple Imputation" section for methodological rationale. • See the "Assessment of Missing Data" section for results supporting choice of imputation method.
Define your sample	• Defining your intent-to-treat and initiator sample can be less clear-cut for online intervention research. Study enrollment does not guarantee a user has completed intervention enrollment (eg, created an account, downloaded the app). Furthermore, enrollment in the online intervention does not guarantee intervention initiation (eg, signing in to read content at least once, completing at least one practice activity). Thus, for studies investigating adherence to online interventions, it is important to consider which sample had true measurement of adherence (eg, versus failure to complete intervention enrollment).	• See the "Participants" section for methodological justification. • See the "Assessment of Missing Data" section for results demonstrating different adherence dose-response results for intent-to-treat versus initiator-only samples.
Report all your results	• Ensure that all your findings regarding each adherence measure are clearly reported in your paper. Because there is such much variability in research findings about online intervention adherence-outcome effects, knowing null results will help future researchers and intervention developers better disentangle where adherence matters most.	• See the "Results" section and Table 2 for thorough reporting of preliminary, main, and post hoc analyses.

Testing multiple measures of adherence is critical for understanding dose-response effects across diverse users. According to Sieverink and colleagues [11], there appears to be an assumption in the literature that a user must interact with all aspects of the intervention to benefit. As such, researchers often define and operationalize adherence based on this

assumption. However, online interventions are highly variable in their designs, components, instructions, and goals. Indeed, our results support the idea that users can adhere strongly to some aspects of an intervention, but not adhere to others. If we had assumed otherwise in this study, we would not have been able to see that multiple types of adherence at varying levels of effort could be associated with symptom improvement in our intent-to-treat sample. Therefore, perhaps there will never be a single answer about *which* type of adherence matters; rather, researchers should continue testing multiple adherence measures to better understand dose-response effects for the wide audience of self-guided online intervention users.

Our study also demonstrated that discrepant adherence results can arise depending on how the "user" sample is determined. When all individuals who created an account (ie, intent-to-treat sample) were included in the sample, results revealed that relatively higher levels of effort (ie, high behavioral adherence and moderate attitudinal adherence) predicted symptom improvement. When only those individuals with account activity within the intervention (ie, program initiators) were included, none of the adherence measures were predictive of symptom improvement. Unfortunately, the differentiation between study enrollment, intervention enrollment, and intervention initiation is rarely reported in online intervention trials. To use 2 specific trials as examples, Donkin et al [15] included participants based on study enrollment (ie, those who completed a 3-month follow-up assessment), whereas Christensen et al [29] included participants based on intervention initiation (ie, those who completed at least one activity in the intervention). Yet the difference in study enrollment rates versus intervention initiation rates may be especially pronounced in the context of open-source online interventions, given the ease of access to signing up (compared with face-to-face interventions). Indeed, as previously mentioned, a systematic review of self-guided online interventions found that 33%-88% of users who downloaded an app actually used it at least once [2]. Knowing that attrition and sporadic use are to be expected in the context of online interventions [9], we strongly recommend that researchers both (1) clearly report and justify how they defined their "user" sample, and (2) report results based on different user sample types.

Limitations

The main limitation in this study, which is expected in the context of online interventions, is the large proportion of missing postintervention data. To ameliorate this, we used multiple imputations, and generated a large number of data sets while including any potential auxiliary variables to support model generation of estimates. As compared with other methods of dealing with missing data such as listwise and pairwise deletion, multiple imputation is considered a "state-of-the-art" technique and is a recommended procedure in the methodological literature [30]. However, the inability to confirm that the data set was missing at random could have affected the interpretation of our results. Next, detection of any significant adherence-outcome effects was difficult given the overall small intervention effects, per our small R^2 values. Such small effects are unfortunately inherent to research on an online prevention program (see [16]), and there is also a prohibitive "floor effect" when a nonclinical sample can only improve so much. Another main limitation to the study is the generalizability of our results. There may be more efficient and accurate ways of measuring adherence to interventions for substance use disorders, for example, or for younger or older populations. Replication of findings will be important. An additional limitation in our study was our use of word count to operationalize attitudinal adherence, which has not been a previously established way of defining attitudinal adherence. Word count was selected as an indicator in this study given the high feasibility and ease of collection. However, future research is needed to validate word count as a measure of attitudinal adherence. Specifically, researchers might consider qualitative coding to identify themes associated with attitudinal adherence and examine correlation with word count.

Conclusion

First and foremost, researchers are encouraged to use the checklist in Table 3 as a resource for recommendations when planning studies, and to report any further decisions pertinent to examining adherence to online interventions. Researchers are also encouraged to reproduce and expand the behavioral and attitudinal categories that we have outlined. Although results of this study provide some preliminary evidence for the predictive validity, additional research is needed to determine the generalizability of these candidates. Finally, researchers are also encouraged to examine individual characteristics (eg, personal traits, baseline symptom severity, technology preference) that may moderate the relationship between adherence and outcome. Such findings could be used to inform customization of interventions to maximize benefit based on relevant personal characteristics.

In conclusion, online interventions offer its users the autonomy to interact with the platform, resulting in a variety of ways that individuals could adhere to the intervention. This poses a challenge for researchers who aim to understand the role of adherence in improving these interventions. However, through the accumulation of high-quality and transparent research into the numerous forms of adherence that result in symptom change, it will be possible to prioritize features and make design decisions to maximize effectiveness of online interventions. With this end goal in mind, this paper does not claim to have identified the best way to measure adherence, rather we add to the ongoing discussion by summarizing our lessons learned to facilitate this discussion and the process of defining and measuring adherence. By creating a more efficient and standardized process for future researchers, we hope to facilitate the creation of high-quality transparent research to understand the role of adherence in self-guided online interventions.

Acknowledgments

We thank Shreya Chadda and Hayley Hammons for their contributions as research assistants during the original trial study and their assistance with database setup for this study.

Conflicts of Interest

LR-N is the developer of the online platform used in this study and owns the copyright as registered with the United States Copyright Office. The authors have no other conflicts of interest to disclose.

References

1. Beatty L, Binnion C. A Systematic review of predictors of, and reasons for, adherence to online psychological interventions. Int J Behav Med 2016 Dec 8;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]
2. Fleming TM, de Beurs D, Khazaal Y, Gaggioli A, Riva G, Botella C, et al. Maximizing the impact of e-Therapy and serious gaming: time for a paradigm shift. Front Psychiatry 2016;7:65 [FREE Full text] [doi: 10.3389/fpsyt.2016.00065] [Medline: 27148094]
3. Hollis C, Falconer CJ, Martin JL, Whittington C, Stockton S, Glazebrook C, et al. Annual Research Review: Digital health interventions for children and young people with mental health problems - a systematic and meta-review. J Child Psychol Psychiatry 2017 Apr 10;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
4. Karyotaki E, Kleiboer A, Smit F, Turner DT, Pastor AM, Andersson G, et al. Predictors of treatment dropout in self-guided web-based interventions for depression: an 'individual patient data' meta-analysis. Psychol. Med 2015 Apr 17;45(13):2717-2726. [doi: 10.1017/s0033291715000665]
5. Cuijpers P, Marks IM, van Straten A, Cavanagh K, Gega L, Andersson G. Computer-aided psychotherapy for anxiety disorders: a meta-analytic review. Cogn Behav Ther 2009 Jun;38(2):66-82. [doi: 10.1080/16506070802694776] [Medline: 20183688]
6. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009 Dec 15;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
7. Waller R, Gilbody S. Barriers to the uptake of computerized cognitive behavioural therapy: a systematic review of the quantitative and qualitative evidence. Psychol. Med 2008 Sep 24;39(5):705-712. [doi: 10.1017/s0033291708004224]
8. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
9. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
10. Beintner I, Vollert B, Zarski A, Bolinski F, Musiat P, Görlich D, et al. Adherence reporting in randomized controlled trials examining manualized multisession online interventions: systematic review of practices and proposal for reporting standards. J Med Internet Res 2019 Aug 15;21(8):e14181 [FREE Full text] [doi: 10.2196/14181] [Medline: 31414664]
11. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
12. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
13. De Geest S, Sabaté E. Adherence to long-term therapies: evidence for action. European Journal of Cardiovascular Nursing 2016 Jun 22;2(4):323-323. [doi: 10.1016/s1474-5151(03)00091-4]
14. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
15. Donkin L, Hickie IB, Christensen H, Naismith SL, Neal B, Cockayne NL, et al. Rethinking the dose-response relationship between usage and outcome in an online intervention for depression: randomized controlled trial. J Med Internet Res 2013 Oct 17;15(10):e231 [FREE Full text] [doi: 10.2196/jmir.2771] [Medline: 24135213]
16. Rith-Najarian L. A dissemination and implementation approach to preventing anxiety and depression in young people. Doctoral dissertation, University of California Los Angeles. 2019. URL: https://escholarship.org/content/qt6786366q/qt6786366q.pdf [accessed 2022-01-26]
17. Staudt M. Treatment engagement with caregivers of at-risk children: gaps in research and conceptualization. J Child Fam Stud 2006 Aug 23;16(2):183-196. [doi: 10.1007/s10826-006-9077-2]
18. Van der Zanden R, Curie K, Van Londen M, Kramer J, Steen G, Cuijpers P. Web-based depression treatment: associations of clients' word use with adherence and outcome. J Affect Disord 2014 May;160:10-13 [FREE Full text] [doi: 10.1016/j.jad.2014.01.005] [Medline: 24709016]
19. Norton PJ. Depression Anxiety and Stress Scales (DASS-21): psychometric analysis across four racial groups. Anxiety Stress Coping 2007 Sep;20(3):253-265. [doi: 10.1080/10615800701309279] [Medline: 17999228]

20. Beard C, Hsu K, Rifkin L, Busch A, Björgvinsson T. Validation of the PHQ-9 in a psychiatric sample. J Affect Disord 2016 Mar 15;193:267-273. [doi: 10.1016/j.jad.2015.12.075] [Medline: 26774513]
21. Duckworth AL, Quinn PD. Development and validation of the short grit scale (grit-s). J Pers Assess 2009 Mar 17;91(2):166-174. [doi: 10.1080/00223890802634290] [Medline: 19205937]
22. Ryan RM, Plant RW, O'Malley S. Initial motivations for alcohol treatment: Relations with patient characteristics, treatment involvement, and dropout. Addictive Behaviors 1995 May;20(3):279-297. [doi: 10.1016/0306-4603(94)00072-7]
23. Lyubomirsky S, Lepper HS. Measure of subjective happiness: preliminary reliability and construct validation. Social Indicators Research 1999 Feb;46(2):137-155. [doi: 10.1023/A:1006824100041]
24. Ghani IMM, Ahmad S. Stepwise multiple regression method to forecast fish landing. Procedia - Social and Behavioral Sciences 2010;8:549-554. [doi: 10.1016/j.sbspro.2010.12.076]
25. Enders CK. Applied Missing Data Analysis. New York, NY: Guilford Press; 2010.
26. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society: Series B (Methodological) 2018 Dec 05;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
27. Headspace. URL: http://headspace.com/ [accessed 2022-01-26]
28. Harrer M, Adam SH, Baumeister H, Cuijpers P, Karyotaki E, Auerbach RP, et al. Internet interventions for mental health in university students: A systematic review and meta-analysis. Int J Methods Psychiatr Res 2019 Jun 26;28(2):e1759 [FREE Full text] [doi: 10.1002/mpr.1759] [Medline: 30585363]
29. Christensen H, Griffiths KM, Korten AE, Brittliffe K, Groves C. A comparison of changes in anxiety and depression symptoms of spontaneous users and trial participants of a cognitive behavior therapy website. J Med Internet Res 2004 Dec 22;6(4):e46. [doi: 10.2196/jmir.6.4.e46] [Medline: 15631970]
30. Peugh JL, Enders CK. Missing data in educational research: a review of reporting practices and suggestions for improvement. Review of Educational Research 2016 Jun 30;74(4):525-556. [doi: 10.3102/00346543074004525]

Abbreviations

DASS-21: 21-item Depression, Anxiety, and Stress Scale
GRIT: Grit Scale
MCAR: missing completely at random
PHQ-9: 9-item Patient Health Questionnaire
RCT: randomized controlled trial
SHS: Subjective Happiness Scale
TMQ: Treatment Motivation Questionnaire

Edited by J Torous; submitted 12.06.21; peer-reviewed by K Patel, M Salimi; comments to author 08.07.21; revised version received 25.08.21; accepted 16.09.21; published 28.03.22.

Please cite as:
Hanano M, Rith-Najarian L, Boyd M, Chavira D
Measuring Adherence Within a Self-Guided Online Intervention for Depression and Anxiety: Secondary Analyses of a Randomized Controlled Trial
JMIR Ment Health 2022;9(3):e30754
URL: https://mental.jmir.org/2022/3/e30754
doi:10.2196/30754
PMID:35343901

Original Paper

The Effectiveness of a Nonguided Mindfulness App on Perceived Stress in a Nonclinical Dutch Population: Randomized Controlled Trial

Leonieke W Kranenburg[1], PhD; Jamie Gillis[1], MSc; Birgit Mayer[2], PhD; Witte J G Hoogendijk[3], MD, PhD

[1]Section of Medical Psychology, Department of Psychiatry, Erasmus University Medical Center, Rotterdam, Netherlands

[2]Department of Psychology, Education and Child studies, Erasmus University Rotterdam, Rotterdam, Netherlands

[3]Department of Psychiatry, Erasmus University Medical Center, Rotterdam, Netherlands

Corresponding Author:
Leonieke W Kranenburg, PhD
Section of Medical Psychology
Department of Psychiatry
Erasmus University Medical Center
PO Box 2040
Rotterdam, 3000 CA
Netherlands
Phone: 31 10 7040139
Email: l.kranenburg@erasmusmc.nl

Abstract

Background: Mindfulness has become increasingly popular, and positive outcomes have been reported for mindfulness-based interventions (MBIs) in reducing stress. These findings make room for innovative perspectives on how MBIs could be applied, for instance through mobile health (mHealth).

Objective: The aim of this study is to investigate whether a nonguided mindfulness mobile app can decrease perceived stress in a nonclinical Dutch population over the course of 8 weeks, with follow-up at 6 months.

Methods: A randomized controlled trial was performed to compare an experimental group that made use of a structured 8-week mHealth mindfulness program and a control group after 8 weeks, with follow-up after 6 months. Participants were recruited via a national television program. The primary outcome measure was perceived stress as measured by the Perceived Stress Scale, secondary outcomes were symptoms of burnout (measured using the visual analog scale [VAS]) and psychological symptoms (measured using the Four-Dimensional Symptom Questionnaire [4DSQ] at follow-up). Outcomes were analyzed using a multilevel regression model.

Results: At baseline, 587 respondents were included. Results showed no postintervention differences between groups for the level of perceived stress. With regard to the secondary outcome measures, the VAS for *emotional exhaustion* and *physical exhaustion* showed significantly lower scores for the experimental group after 8 weeks (P=.04 and P=.01, respectively), but not at follow-up. There were no differences between groups for psychological symptoms measured using the 4DSQ.

Conclusions: These findings do not support our hypothesis that using the mindfulness app would reduce stress levels. However, our findings related to diminished exhaustion at 8 weeks are encouraging and require further investigation.

Trial Registration: ClinicalTrials.gov NCT05246800; https://clinicaltrials.gov/show/NCT05246800

(JMIR Ment Health 2022;9(3):e32123) doi:10.2196/32123

KEYWORDS

mHealth; mindfulness; stress; burnout; non-clinical population; nonclinical

Introduction

Mindfulness practice has become increasingly popular, both in the general public [1,2] and in health care [3]. Mindfulness derives from Buddhism and can be defined as an ability to observe one's bodily sensations, feelings, and thoughts with an open, nonjudgmental, and accepting mind toward one's experiences [4]. Mindfulness has been found to alleviate intense

emotional states [5,6] and enhance emotional coping mechanisms in response to stress [7], and it is believed to induce shifts in processing of negative emotions under stress [8]. The core principles of mindfulness are incorporated in a variety of psychological treatments that are referred to as mindfulness-based interventions (MBIs).

MBIs are traditionally delivered face to face. New types of applications include web-based programs and mobile health (mHealth) [9]. Advantages of digital applications of MBIs include their availability and accessibility, avoiding waiting lists, saving travelling time, reduced costs, being able to work in one's own environment, and nonrequirement of a therapist [10,11]. Especially during the current COVID-19 outbreak, these are favorable assets. Reviews of studies on the efficacy of these web-based MBIs found up to moderate effects on stress and depression [7,12-16]. A recent meta-analysis on the efficacy of mindfulness meditation apps on users' well-being and mental health–related outcomes concluded that mindfulness apps seemed promising in improving well-being and mental health, but those results should be interpreted with caution [17]. The strongest effects were observed on stress, depression, and burnout. However, regarding burnout, only 3 studies could be included, indicating that this may be a relatively new outcome variable in this research field. The findings of Gál et al [17] are interesting in this respect as chronic stress, burnout, and depression can be viewed as a continuum. Chronic stress can lead to burnout and a great overlap exists between burnout and depression, with shared features including motivational problems and exhaustion [18,19]. In total, 17% of all employees in the Netherlands report burnout symptoms [20], and 5% of all Dutch adults are annually affected by depression [21]. During the COVID-19 pandemic, these numbers may even be higher, as reports worldwide point to increased mental health complaints for various population subgroups, such as health care professionals, teachers, and those working from home [22-25]. The strain on mental health care budgets and practices, characterized by long waiting lists and shortness of qualified personnel, make it even more important to invest in new technologies to reduce mental health problems in the general population [22]. Therefore, the aim of this study is to investigate whether a nonguided mindfulness mobile app can decrease perceived stress levels and burnout symptoms in a nonclinical Dutch population.

Methods

Design and Recruitment

A randomized controlled trial with follow-up measures at 6 months was performed. Participants were recruited through the television (TV) program *Kassa*, in the context of 4 broadcasts on the topic "stress" in March 2018. The study was announced during this TV show, and in that particular moment, viewers were invited to visit the program's website and click the weblink with more information about the study and the possibility to apply for it. There were no eligibility criteria, other than being an adult. After having read and accepted the terms and conditions (informed consent), participants were referred to the web-based questionnaires. Randomization took place after they had filled out the baseline (T0) questionnaires and was performed with a built-in randomization algorithm. The experimental group was provided access to the mindfulness mobile app, which contained an 8-week nonguided mindfulness program developed to reduce stress symptoms. The control group was suggested to read information about stress and burnout on the *Kassa* website. Participants were not blinded to their condition, as they knew whether or not they received access to the mindfulness app. There were three time points of measurement: T0 (baseline, before randomization), T1 (at the end of the program, 8 weeks after randomization), and T2 (6 months after randomization).

Intervention

The mindfulness application was developed by Minddistrict, an eHealth company in the Netherlands. The content of the app was developed by professionals in the field of mental health care and in accordance with the principles of mindfulness-based stress reduction and mindfulness-based cognitive therapy [4,26]. This mindfulness mobile app was the first app version derived from the one already existing web-based mindfulness program at the time used in mental health care settings. The app consisted of a structured program, with chapters on psycho-education on mindfulness and the importance of practicing; acting on auto-pilot, conscious attention; nonjudgmental attention, awareness; doing versus being mode; attention for breath and body, conscious response; acceptance; a mindful attitude toward thoughts; and applying mindfulness in daily life and staying mindful. Each chapter started with a short explanation of a specific mindfulness principle and was followed by relevant exercises, such as the body scan, Raisin Exercise, breath exercises, and sitting meditation. After completion of the exercises, participants were asked about their experiences, and the participant received an encouraging standard feedback message to keep practicing the exercises for optimal mindfulness training. There was no real-life contact (either in person or on the internet) with a mindfulness trainer. There was the possibility to create a personal program with favorite exercises.

Measures

Demographics

Participants age, sex, level of education, and occupation were assessed.

Perceived Stress Scale

The PSS measures perceived stress levels [27]. The 14-item Dutch version was used in this study. All items are rated on a 4-point Likert scale, with higher scores indicating more perceived stress. Cronbach α ranges between .84 and .86 [27] and overall psychometric properties are evaluated as acceptable [28]. The Cronbach α for this sample was >.89 for all measurement time points.

Visual Analogue Scale

Burnout symptoms were assessed using 8 visual analogue scales (VASs). The symptoms measured were as follows: *control over emotions*, *memory and concentration*, *sleep*, *work interest*, *work performance*, *interest in others*, *emotional exhaustion*, *and*

physical exhaustion. Each symptom was rated on a 0-100 scale, with higher scores indicating higher difficulty.

Four-Dimensional Symptom Questionnaire

The Four-Dimensional Symptom Questionnaire (4DSQ) consists of 50 items rated on a 4-point Likert scale [29]. The 50 items can be grouped into four dimensions: Distress (n=16), Depression (n=6), Anxiety (n=12), and Somatization (n=16). Sum scores are calculated for each dimension. The reliability of these dimensions was good, with Cronbach α>.79 for all subscales [30]. The Cronbach α for this sample was >.88 for all subscales. This measure was applied at T2 only.

Statistical Analyses

The experimental group and the control group were compared using a multilevel regression analysis, with participants as the upper level and their repeated measures as the lower level. Time (repeated measures at T0, T1, and T2), treatment group (experimental or control), and the time–group interaction were postulated as fixed effects. The difference in change in the PSS and VAS between the groups at follow-up is considered the primary contrast. At T2, differences between groups for the 4DSQ were analyzed with an independent samples *t* test (2-tailed). Two-sided *P* values of <.05 were considered significant. Data were analyzed using descriptive statistics in SPSS (version 25; IBM Corp).

Ethical Considerations

The study was approved by the Erasmus University Medical Ethical Committee and evaluated as not subject to the Dutch act on medical scientific research involving human subjects (METC 2017-1117).

Results

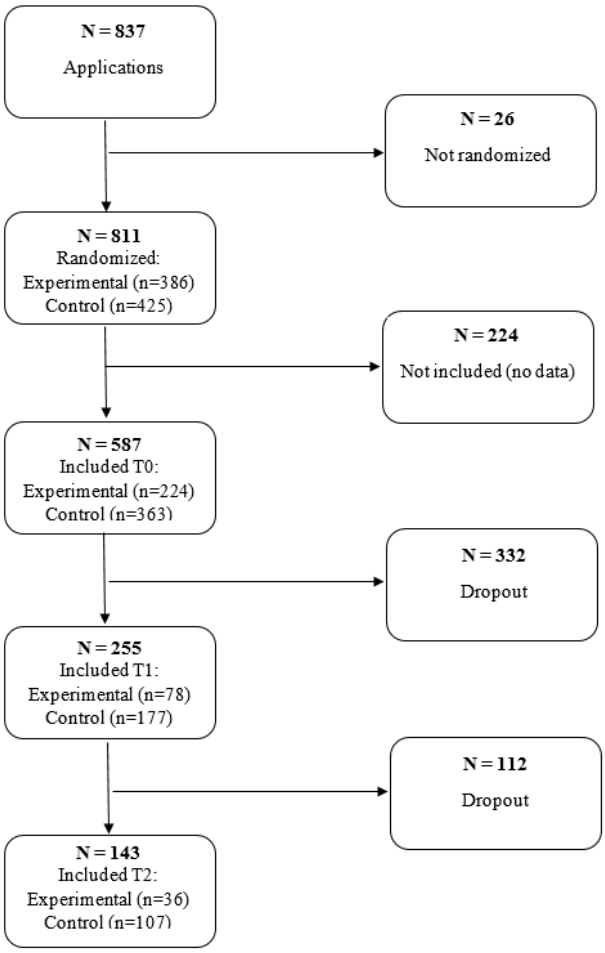
Participants

The final sample at T0 included 587 participants. This sample consisted mostly of highly educated (64.5%), employed (74.7%), and female (74.6%) individuals with a mean age of 46.05 (SD 13.64) years. More detailed information about the initial and final sample is provided in Table 1. Figure 1 provides the participant randomization flowchart.

Table 1. Demographic variables.

Variables	Time points	
	T0	T2
Age (years), mean (SD)	45.86 (13.69)	47.70 (12.80)
Sex, n (%)		
Female	435 (74.1)	137 (74.8)
Male	152 (25.9)	36 (25.2)
Education, n (%)		
Elementary school	2 (0.3)	1 (0.7)
Middle and high school	72 (12.3)	14 (9.8)
Secondary education	129 (22.0)	34 (23.8)
Higher education	384 (65.4)	94 (65.7)
Employment status, n (%)		
Social welfare	7 (1.1)	3 (2.1)
Informal care	9 (1.5)	0 (0)
Unemployed	14 (2.4)	6 (4.2)
Volunteers	21 (3.6)	5 (3.5)
Domestic household	21 (3.6)	4 (2.8)
Retired	34 (5.8)	11 (7.7)
Sick leave	38 (6.5)	13 (9.1)
Part-time	206 (35.0)	47 (32.9)
Full-time	237 (40.0)	54 (37.8)

Figure 1. Flowchart of inclusion and dropout.



Primary Outcomes

Perceived stress was measured with the PSS at T0, T1, and T2. Table 2 provides descriptive statistics of the PSS over time for both groups. Changes between groups over time were analyzed with a multilevel regression model. The interaction between group and time was not a significant predictor of perceived stress measured using the PSS at T1 ($F_{1,501.26}$=0.70, P=.40) and T2 ($F_{1,490.35}$=1.36, P=.24; Table 3).

Table 2. Descriptive statistics of the perceived stress scale (PSS).

Group	PSS score, mean (SD)	Participants, n
T0		
Control	29.33 (8.32)	363
Experimental	30.10 (7.67)	224
T1		
Control	27.51 (9.21)	174
Experimental	26.59 (8.83)	73
T2		
Control	26.30 (9.98)	107
Experimental	26.00 (9.72)	36

Table 3. Estimates of the fixed effects measures of the perceived stress scale (PSS; dependent variable: PSS score).

Interaction	Estimate (95% CI); SE	*P* value
Group×T1	–0.71 (–2.38 to 0.96); 0.85	.40
Group×T2	–1.29 (–3.47 to 0.89); 1.11	.24

Secondary Outcomes

There were 8 VASs measuring burnout symptoms at T0, T1, and T2. Changes between groups over time were measured with a multilevel regression model. The VAS scales of *emotional exhaustion* and *physical exhaustion* were both significant at T1 ($F_{1,520.41}$=4.16, P=.04 and $F_{1,528.76}$=6.29, P=.01, respectively), for the time–group interaction, with the experimental group presenting lower exhaustion scores. Upon 6-month follow-up (T2), this effect was not maintained with the outcomes ($F_{1,510.18}$=0.03, P=.87 and $F_{1,520.06}$=0.04, P=.84, respectively). Changes between the two groups over time for the other 6 VAS scales did not differ at T1 and T2. Outcomes for the 4DSQ subscales at T2 showed no significant differences between both groups.

Discussion

Principal Findings

Our primary research question served to investigate whether an 8-week nonguided mindfulness mobile app can decrease perceived stress levels in a nonclinical Dutch population. Our findings do not support our hypothesis that using a nonguided mindfulness app reduces perceived stress levels. Not observing an effect on stress in this study might be explained by the complete lack of personal contact in this study; that is, there was no mindfulness trainer reachable through the app, nor was there the possibility to engage in social contact with other participants. This potential hypothesis is supported by the recent review of Borghouts [31], which points toward a generally higher engagement for guided (vs unguided) interventions and to the importance of social connectedness as a facilitator of user engagement. In addition, a recent meta-analysis [32] focusing specifically on web-based mindfulness found that web-based MBIs resulted in higher effect sizes for stress when offered guidance. As there was also no personal contact between the research team and the respondents, one could say that our study results may be similar to what one might expect to find in a real-world user situation, where despite high levels of app download, only a small portion of users actually use the apps for a longer period [33,34].

Another finding of our study was that the experimental group reported lower levels of both emotional and physical exhaustion after 8 weeks of using the app. This finding is particularly interesting, as different definitions of burnout all share exhaustion as a central component [35-38]. For instance, according to Schaufeli et al [38], burnout is conceptualized as a state of mental exhaustion, leading to both an inability and an unwillingness to act. Furthermore, in the process model of burnout, emotional exhaustion is one of the first symptoms to develop [39]. Dealing with stressors in everyday life can result in the depletion of cognitive and emotional resources, and these can cause exhaustion [40-42]. It is striking that the possible gain of the app may lie in reducing feelings of exhaustion in the broad sense. This might be related to the role mindfulness plays in autonomous self-regulation [43], which preserves vitality and energy [44]. Hence, a better spending and preservation of cognitive and emotional resources through increased self-regulation might have resulted in a reduction of physical and emotional exhaustion after 8 weeks. This finding is in line with other studies that reported that mindfulness interventions are related to a reduction of emotional exhaustion, both in health care professionals [45,46] and other employees [47]. However, when it comes to mobile mindfulness apps, to our best knowledge, only one previous study specifically reported on exhaustion outcomes. In this study, significant effects were found for reduced emotional exhaustion [48]. These findings might indicate that a mindfulness app has the potential to be used as preventive intervention for burnout in a nonclinical population.

Clinical Implications and Directions for Future Research

Future research on the effects of mobile mindfulness apps on burnout is warranted. Given the long waiting lists for mental health care, an ideal setting for further testing this or other mobile mindfulness apps would be general practitioner clinics and mental health care institutions. That is, driven by these long waiting lists, it has become more common to offer patients bridge interventions with low personnel costs, including eHealth modules. This provides a great opportunity to implement a study design with conditions that vary in the amount and type of personal contact. Such a design could even be advanced by using randomization schemes that allow for including patient preferences with regard to these aspects. Next, given our findings that both mental and physical exhaustion decreased in the app user group, it would be of great interest to conduct frequent measures of individual stress and burnout symptoms with an experience sampling method [49]. This can help build general networks of how burnout symptoms develop and worsen over time [50]. By drawing on such data, future app-based interventions could be personalized to increase effectiveness. Building a user-friendly experience sampling method incorporated in the mindfulness app would then be the next challenge for app developers. In addition, build-in measures for actual time spent on practicing mindfulness exercises would also contribute to the field, as until now such data are limited [1]. Of course, such new features should first be subjected to acceptability and feasibility studies.

Limitations

The fact that participants were unequally distributed between the experimental and the control group was a limitation of this study. This was owing to a previously unnoticed error in the automated randomization algorithm, which could unfortunately not be repaired afterward. Another limitation is the lack of an adherence measure, which makes it impossible to look at a possible dose-response relationship [1]. Furthermore, the open kind of recruitment via a TV program may have biased our sample. For instance, it could be that extravert persons were more likely to spontaneously act upon the invitation to visit the website and register for the study. As extraversion is a barrier for user engagement of digital mental health interventions [31], the way participants were recruited could have biased our sample. This of course is only a hypothesis, as we have no data on personality traits of our sample, or on other relevant user characteristics, such as mental health status and previous

experiences with mindfulness or meditation, which may have influenced both user engagement and the outcomes.

Conclusions

Our study did not find an effect of using a mindfulness app on perceived stress levels in a nonclinical Dutch population. However, this may be owing to the type of respondent recruitment and a lack of control on adherence levels. Our results show diminished emotional and physical exhaustion in the app user group after 8 weeks. These findings are encouraging as they suggest that a mindfulness app has potential to be used as a preventive intervention for burnout.

Acknowledgments

The authors wish to thank Reinier Timman for his support in performing the statistical analyses.

Authors' Contributions

LWK designed the study, collected and analyzed the data, and drafted the manuscript. JG analyzed the data and drafted the manuscript. BM supervised coauthor JG and drafted, reviewed, and edited the manuscript. WJGH designed the study and drafted, reviewed, and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 16993 KB - mental_v9i3e32123_app1.pdf]

References

1. Taylor H, Strauss C, Cavanagh K. Can a little bit of mindfulness do you good? A systematic review and meta-analyses of unguided mindfulness-based self-help interventions. Clin Psychol Rev 2021 Nov;89:102078. [doi: 10.1016/j.cpr.2021.102078] [Medline: 34537665]
2. Mrazek AJ, Mrazek MD, Cherolini CM, Cloughesy JN, Cynman DJ, Gougis LJ, et al. The future of mindfulness training is digital, and the future is now. Curr Opin Psychol 2019 Aug;28:81-86. [doi: 10.1016/j.copsyc.2018.11.012] [Medline: 30529975]
3. Gotink RA, Chu P, Busschbach JJV, Benson H, Fricchione GL, Hunink MGM. Standardised mindfulness-based interventions in healthcare: an overview of systematic reviews and meta-analyses of RCTs. PLoS One 2015;10(4):e0124344 [FREE Full text] [doi: 10.1371/journal.pone.0124344] [Medline: 25881019]
4. Kabat-Zinn J. Full Catastrophe Living, Revised Edition: How to cope with stress, pain and illness using mindfulness meditation. Boston, MA: Little, Brown Book Group; 2013.
5. Baer RA. Mindfulness training as a clinical intervention: A conceptual and empirical review. Clin Psychol 2003;10(2):125-143. [doi: 10.1093/clipsy.bpg015]
6. Keng S, Smoski MJ, Robins CJ. Effects of mindfulness on psychological health: a review of empirical studies. Clin Psychol Rev 2011 Aug;31(6):1041-1056 [FREE Full text] [doi: 10.1016/j.cpr.2011.04.006] [Medline: 21802619]
7. Lyzwinski LN, Caffery L, Bambling M, Edirippulige S. A systematic review of llectronic mindfulness-based therapeutic interventions for weight, weight-related behaviors, and psychological stress. Telemed J E Health 2018 Mar;24(3):173-184. [doi: 10.1089/tmj.2017.0117] [Medline: 28885896]
8. Davidson RJ, Kabat-Zinn J, Schumacher J, Rosenkranz M, Muller D, Santorelli SF, et al. Alterations in brain and immune function produced by mindfulness meditation. Psychosom Med 2003;65(4):564-570. [doi: 10.1097/01.psy.0000077505.67574.e3] [Medline: 12883106]
9. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
10. Andersson G, Titov N. Advantages and limitations of Internet-based interventions for common mental disorders. World Psychiatry 2014 Feb;13(1):4-11 [FREE Full text] [doi: 10.1002/wps.20083] [Medline: 24497236]
11. Cuijpers P, Marks IM, van Straten A, Cavanagh K, Gega L, Andersson G. Computer-aided psychotherapy for anxiety disorders: a meta-analytic review. Cogn Behav Ther 2009;38(2):66-82. [doi: 10.1080/16506070802694776] [Medline: 20183688]
12. Spijkerman MPJ, Pots WTM, Bohlmeijer ET. Effectiveness of online mindfulness-based interventions in improving mental health: A review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016 Apr;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]

13. Jayewardene WP, Lohrmann DK, Erbe RG, Torabi MR. Effects of preventive online mindfulness interventions on stress and mindfulness: A meta-analysis of randomized controlled trials. Prev Med Rep 2017 Mar;5:150-159 [FREE Full text] [doi: 10.1016/j.pmedr.2016.11.013] [Medline: 28050336]
14. Sevilla-Llewellyn-Jones J, Santesteban-Echarri O, Pryor I, McGorry P, Alvarez-Jimenez M. Web-based mindfulness interventions for mental health treatment: systematic review and meta-analysis. JMIR Ment Health 2018 Sep 25;5(3):e10278 [FREE Full text] [doi: 10.2196/10278] [Medline: 30274963]
15. Toivonen KI, Zernicke K, Carlson LE. Web-based mindfulness interventions for people with physical health conditions: systematic review. J Med Internet Res 2017 Aug 31;19(8):e303 [FREE Full text] [doi: 10.2196/jmir.7487] [Medline: 28860106]
16. Victorson DE, Sauer CM, Wolters L, Maletich C, Lukoff K, Sufrin N. Meta-analysis of technology-enabled mindfulness-based programs for negative affect and mindful awareness. Mindfulness 2020 Apr 27;11(8):1884-1899. [doi: 10.1007/s12671-020-01373-y]
17. Gál É, tefan S, Cristea IA. The efficacy of mindfulness meditation apps in enhancing users' well-being and mental health related outcomes: a meta-analysis of randomized controlled trials. J Affect Disord 2021 Jan 15;279:131-142. [doi: 10.1016/j.jad.2020.09.134] [Medline: 33049431]
18. Bianchi R, Schonfeld IS, Laurent E. Burnout-depression overlap: a review. Clin Psychol Rev 2015 Mar;36:28-41. [doi: 10.1016/j.cpr.2015.01.004] [Medline: 25638755]
19. Wurm W, Vogel K, Holl A, Ebner C, Bayer D, Mörkl S, et al. Depression-burnout overlap in physicians. PLoS One 2016;11(3):e0149913 [FREE Full text] [doi: 10.1371/journal.pone.0149913] [Medline: 26930395]
20. Nationale Enquête Arbeidsomstandigheden. Methodologie en globale resultaten. TNO. 2019. URL: https://www.monitorarbeid.tno.nl/nl-nl/publicaties/nea-2019/ [accessed 2022-03-01]
21. Zicht op depressie: de aandoening, preventie en zorg. Kerngegevens uitgelicht. Depressie in Nederland: feiten en cijfers. Trimbos-instituut; 2018. URL: https://www.trimbos.nl/kennis/cijfers/depressie [accessed 2022-11-23]
22. Tement S, Ketiš ZK, Miroševič Š, Selič-Zupančič P. The impact of Psychological Interventions with Elements of Mindfulness (PIM) on empathy, well-being, and reduction of burnout in physicians: a systematic review. Int J Environ Res Public Health 2021 Oct 25;18(21):11181 [FREE Full text] [doi: 10.3390/ijerph182111181] [Medline: 34769700]
23. Leo CG, Sabina S, Tumolo MR, Bodini A, Ponzini G, Sabato E, et al. Burnout among healthcare workers in the COVID 19 era: a review of the existing literature. Front Public Health 2021;9:750529 [FREE Full text] [doi: 10.3389/fpubh.2021.750529] [Medline: 34778184]
24. Vargas Rubilar N, Oros LB. Stress and burnout in teachers during times of pandemic. Front Psychol 2021;12:756007 [FREE Full text] [doi: 10.3389/fpsyg.2021.756007] [Medline: 34899498]
25. Johnson MS, Skjerdingstad N, Ebrahimi OV, Hoffart A, Johnson SU. Parenting in a pandemic: parental stress, anxiety and depression among parents during the government-initiated physical distancing measures following the first wave of COVID-19. Stress Health 2021 Dec 13. [doi: 10.1002/smi.3120] [Medline: 34902219]
26. Segal Z, Williams M, Teasdale J. Mindfulness-Based Cognitive Therapy for Depression (2nd edition). New York, NY: Guilford Publications; 2018.
27. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
28. Lee E. Review of the psychometric evidence of the perceived stress scale. Asian Nurs Res (Korean Soc Nurs Sci) 2012 Dec;6(4):121-127 [FREE Full text] [doi: 10.1016/j.anr.2012.08.004] [Medline: 25031113]
29. Terluin B, Rijmen F, van Marwijk H, Stalman W. Waarde van de Vierdimensionale Klachtenlijst (4DKL) voor het detecteren van depressieve stoornissen. HUWE 2007 Jul;50(7):494-501. [doi: 10.1007/bf03085226]
30. Terluin B, Rhenen WV, Schaufeli WB, De Haan M. The four-dimensional symptom questionnaire (4DSQ): measuring distress and other mental health problems in a working population. Work & Stress 2004 Jul;18(3):187-207. [doi: 10.1080/02678370420002975351]
31. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
32. Sommers-Spijkerman M, Austin J, Bohlmeijer E, Pots W. New evidence in the booming field of online mindfulness: an updated meta-analysis of randomized controlled trials. JMIR Ment Health 2021 Jul 19;8(7):e28168 [FREE Full text] [doi: 10.2196/28168] [Medline: 34279240]
33. Baumel A, Edan S, Kane JM. Is there a trial bias impacting user engagement with unguided e-mental health interventions? A systematic comparison of published reports and real-world usage of the same programs. Transl Behav Med 2019 Nov 25;9(6):1020-1033. [doi: 10.1093/tbm/ibz147] [Medline: 31689344]
34. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
35. Bakker AB, Demerouti E, Schaufeli WB. Validation of the Maslach Burnout Inventory - General Survey: An internet Study. Anxiety, Stress & Coping 2002 Jan;15(3):245-260. [doi: 10.1080/1061580021000020716]

36. Maslach C, Jackson SE. The measurement of experienced burnout. J Organiz Behav 1981 Apr;2(2):99-113. [doi: 10.1002/job.4030020205]
37. Van AM, Oeij S, Seeleman J, Starmans R, Terluin B, Wewerinke A. Overspanning en burn-out. Nederlands Huisartsen Genootschap Richtlijnen (Guidelines of the Dutch academy for General Practicioners). 2018. URL: https://richtlijnen.nhg.org/standaarden/overspanning-en-burn-out [accessed 2022-12-15]
38. Schaufeli WB, Desart S, De Witte H. Burnout Assessment Tool (BAT)-development, validity, and reliability. Int J Environ Res Public Health 2020 Dec 18;17(24):9495 [FREE Full text] [doi: 10.3390/ijerph17249495] [Medline: 33352940]
39. Leiter MP, Maslach C. The impact of interpersonal environment on burnout and organizational commitment. J Organiz Behav 1988 Oct;9(4):297-308. [doi: 10.1002/job.4030090402]
40. Baumeister RF, Bratslavsky E, Muraven M, Tice DM. Ego depletion: is the active self a limited resource? J Pers Soc Psychol 1998 May;74(5):1252-1265. [doi: 10.1037//0022-3514.74.5.1252] [Medline: 9599441]
41. Jonsdottir IH, Nordlund A, Ellbin S, Ljung T, Glise K, Währborg P, et al. Cognitive impairment in patients with stress-related exhaustion. Stress 2013 Mar;16(2):181-190. [doi: 10.3109/10253890.2012.708950] [Medline: 22746338]
42. Weisberg J, Sagie A. Teachers' physical, mental, and emotional burnout: impact on intention to quit. J Psychol 2010 Apr 02;133(3):333-339. [doi: 10.1080/00223989909599746]
43. Brown KW, Ryan RM. The benefits of being present: mindfulness and its role in psychological well-being. J Pers Soc Psychol 2003 Apr;84(4):822-848. [doi: 10.1037/0022-3514.84.4.822] [Medline: 12703651]
44. Ryan RM, Deci EL. From ego depletion to vitality: Theory and findings concerning the facilitation of energy available to the self. Soc Personal Psychol Compass 2008;2(2):702-717. [doi: 10.1111/j.1751-9004.2008.00098.x]
45. Jiménez-Picón N, Romero-Martín M, Ponce-Blandón JA, Ramirez-Baena L, Palomo-Lara JC, Gómez-Salgado J. The relationship between mindfulness and emotional intelligence as a protective factor for healthcare professionals: systematic review. Int J Environ Res Public Health 2021 May 20;18(10):5491 [FREE Full text] [doi: 10.3390/ijerph18105491] [Medline: 34065519]
46. Suleiman-Martos N, Gomez-Urquiza JL, Aguayo-Estremera R, Cañadas-De La Fuente GA, De La Fuente-Solana EI, Albendín-García L. The effect of mindfulness training on burnout syndrome in nursing: A systematic review and meta-analysis. J Adv Nurs 2020 May;76(5):1124-1140. [doi: 10.1111/jan.14318] [Medline: 32026484]
47. Janssen M, Heerkens Y, Kuijer W, van der Heijden B, Engels J. Effects of Mindfulness-Based Stress Reduction on employees' mental health: A systematic review. PLoS One 2018;13(1):e0191332 [FREE Full text] [doi: 10.1371/journal.pone.0191332] [Medline: 29364935]
48. Möltner H, Leve J, Esch T. [Burnout prevention and mobile mindfulness: evaluation of an app-based health training program for employees]. Gesundheitswesen 2018 Mar;80(3):295-300 [FREE Full text] [doi: 10.1055/s-0043-114004] [Medline: 29017191]
49. van Os J, Verhagen S, Marsman A, Peeters F, Bak M, Marcelis M, ESM-MERGE Investigators PhD, et al. The experience sampling method as an mHealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice. Depress Anxiety 2017 Jun;34(6):481-493. [doi: 10.1002/da.22647] [Medline: 28544391]
50. Borsboom D. A network theory of mental disorders. World Psychiatry 2017 Feb;16(1):5-13 [FREE Full text] [doi: 10.1002/wps.20375] [Medline: 28127906]

Abbreviations

4DSQ: Four-Dimensional Symptom Questionnaire
MBI: mindfulness-based interventions
mHealth: mobile health
PSS: perceived stress scale
TV: television
VAS: visual analogue scale

Edited by J Torous; submitted 15.07.21; peer-reviewed by M EIsenstadt, A Eisenstadt; comments to author 04.10.21; revised version received 26.11.21; accepted 20.12.21; published 18.03.22.

Please cite as:
Kranenburg LW, Gillis J, Mayer B, Hoogendijk WJG
The Effectiveness of a Nonguided Mindfulness App on Perceived Stress in a Nonclinical Dutch Population: Randomized Controlled Trial
JMIR Ment Health 2022;9(3):e32123
URL: https://mental.jmir.org/2022/3/e32123
doi:10.2196/32123
PMID:35302504

Original Paper

Treatment Interruptions and Telemedicine Utilization in Serious Mental Illness: Retrospective Longitudinal Claims Analysis

Marcy Ainslie[1], BA, BS, MS, EdD; Mary F Brunette[2], MD; Michelle Capozzoli[3], PhD

[1]Department of Nursing, University of New Hampshire, Durham, NH, United States

[2]Department of Psychiatry, Geisel School of Medicine at Dartmouth, Lebanon, NH, United States

[3]Department of Mathematics & Statistics, University of New Hampshire, Durham, NH, United States

Corresponding Author:
Marcy Ainslie, BA, BS, MS, EdD
Department of Nursing
University of New Hampshire
4 library way, hewitt hall
Durham, NH, 03824
United States
Phone: 1 603 562 8833
Email: marcy.ainslie@unh.edu

Abstract

Background: Avoiding interruptions and dropout in outpatient care can prevent mental illness symptom exacerbation and costly crisis services, such as emergency room visits and inpatient psychiatric hospitalization. During the COVID-19 pandemic, to attempt to maintain care continuity, telemedicine services were increasingly utilized, despite the lack of data on efficacy in patients with serious mental illness. Patients with serious mental illness are challenging to enroll and sustain in randomized controlled trials over time due to fluctuations in disease exacerbation. However, capturing and examining utilization and efficacy data in community mental health center (CMHC) patients with serious mental illness during the pandemic is a unique opportunity to inform future clinical and policy decision-making.

Objective: We aimed to identify and describe the characteristics of CMHC patients with serious mental illness who experienced treatment interruptions and who utilized telemedicine during the pandemic.

Methods: We conducted a retrospective observational study of treatment interruptions and telemedicine use during the period from December 2019 to June 2020 (compared to the period from December 2018 to June 2019) in New Hampshire CMHC patients. The study population included all Medicaid beneficiaries with serious mental illness engaged in treatment 3 months prior to the declaration of a state of emergency in response to the COVID-19 pandemic. We used chi-square tests of independence and logistic regression to explore associations between treatment interruptions and variables (gender, age, rurality, and diagnosis). Telemedicine utilization was categorized as low (<25%), medium (25%-75%), or high (>75%) use.

Results: A total of 16,030 patients were identified. New Hampshire CMHCs demonstrated only a 4.9% increase in treatment interruptions compared with the year prior. Patients who were male (odds ratio [OR] 1.27, 95% CI 1.17-1.38; *P*<.001), under the age of 18 years (ages 0-12 years: OR 1.37, 95% CI 0.62-0.86, *P*<.001; aged 13-17 years: OR 1.49, 95% CI 0.57-0.79, *P*<.001), or among milder diagnostic categories, such as anxiety disorders (OR 3.77, 95% CI 3.04-4.68; *P*<.001) and posttraumatic stress disorder (OR 3.69, 95% CI 2.96-4.61; *P*<.001), were most likely to experience treatment interruptions. Patients who were female (OR 0.89, CI 0.65-0.74), 18 to 34 years old (OR 0.74, CI 0.70-0.79), or among milder diagnostic categories, such as anxiety disorder (OR 0.69, CI 0.65-0.74) or posttraumatic stress disorder (OR 0.77, CI 0.72-0.83), and with major depressive disorder (OR 0.73, CI 0.68-0.78) were less likely to be in the low telemedicine utilization group.

Conclusions: The integration of telemedicine supported care continuity for most CMHC patients; yet, retention varied by subpopulation, as did telemedicine utilization. The development of policies and clinical practice guidelines requires empirical evidence on the effectiveness and limitations of telemedicine in patients with serious mental illness.

(JMIR Ment Health 2022;9(3):e33092) doi:10.2196/33092

KEYWORDS

telemedicine; mental health; serious mental illness; retention; mental illness; telehealth

Introduction

In response to the COVID-19 pandemic, the use of mental health telemedicine broadly expanded in US community mental health centers (CMHCs). CMHCs are designated by states to provide long-term outpatient behavioral, rehabilitation, and medication mental health services to people with serious mental illness, such as disabling schizophrenia, bipolar disorder, major depression, posttraumatic stress disorder, or other anxiety disorders [1]. More than 10 million Americans (approximately 5%) have serious mental illness, and such mental illnesses are a leading source of disability and treatment expenses [2,3], with schizophrenia alone costing approximately US $37.7 billion per year [4].

Prior to the pandemic, telemedicine was used to maintain mental health care continuity when patients and providers were separated by a distance, address transportation or childcare-related barriers, and address provider shortages [5,6]. Avoiding interruptions and dropout in outpatient care has been shown to prevent mental illness symptom exacerbation and the need for costly crisis services, such as emergency room visits and inpatient psychiatric hospitalization [7,8]. Medicaid is the most common payor for patients with serious mental illness due to related disability with resulting low income [9]. Prior research has demonstrated deficits in serious mental illness patient access, utilization, and efficacy of telemedicine as a modality for care delivery. Access concerns are related to limited digital bandwidth in rural areas and device ownership in low socioeconomic status households [10]. Some serious mental illness services require in-person contact that is not possible to deliver via telemedicine [11,12]; additionally, patients with high levels of symptoms and disorganization, which can occur with schizophrenia and bipolar disorder, may have difficulty utilizing this form of treatment.

Prior to the pandemic, CMHC telemedicine services were delivered via videoconference to a small, but growing, number of patients [13]. Typically, patients presented to the local CMHC office, where necessary electronic devices and connectivity were provided, in order to connect with a mental health provider located at a distance. In this prepandemic model, the role of the local CMHC was to mitigate telemedicine access and utilization concerns.

As the pandemic emerged in the United States, federal and state governments, followed by the Centers for Medicaid and Medicare Services, put a hold on regulatory requirements that had created barriers to telemedicine utilization in the delivery of health care services prior to the pandemic, specifically regarding Health Insurance Portability and Accountability Act–approved technology and the required location of the provider during the time of the patient visit [14]. State legislation and policy developments regarding telemedicine followed; these emergency changes broadened the scope of providers who may deliver services via telemedicine and permitted patients to receive these services from their own homes [15,16]. This transition occurred prior to addressing access and utilization concerns and despite little empiric evidence on the efficacy of such services for people with serious mental illness, in general, or for people with schizophrenia and bipolar disorders, in particular.

Research about telemedicine prior to the pandemic found that user perceptions influenced the success of its implementation, users required more technology support than was available, and reimbursement presented a barrier. Clinical efficacy trials [10,11,17-24] of mental health telemedicine utilization, albeit with small sample sizes, indicated that although addressing clinician concerns, logistical problems, technology, and staffing would be necessary [17,18], telephone-based cognitive behavioral therapy for psychosis showed high therapeutic alliance [19] and treatments by phone or video were effective for major depressive disorder [20-23], posttraumatic stress disorder [10,11], and general outpatients [24]. Telemedicine via telephone facilitated low-threshold support to 120 patients with serious mental illness to promote psychotropic mediation adherence for 6 months [25]; however, no large randomized trials broadly evaluating utilization and efficacy of telemedicine in patients with serious mental illness were identified.

Initial findings during the pandemic with respect to telemedicine for people with serious mental illness are mixed but indicate that many patients are willing and able to use video- or telephone-based telemedicine from their homes [26-29]. Another study [15] on service delivery for people with all types and severities of mental illness demonstrated a widening telemedicine utilization disparity between general and minority populations that occurred in the presence of an overall increase in mental health service utilization during the pandemic and suggested that there were increased barriers to telemedicine for minority populations. Additionally, a national survey about telemedicine utilization in patients with serious mental illness demonstrated a need for improved technical support and appointment availability, while at the same time suggesting telehealth visits can promote self-care strategies and resilience [30].

The dramatic transition to mental health telemedicine that occurred during the pandemic provides an important opportunity. The pivot to telemedicine in the serious mental illness population offers a vast, natural experiment to address the literature gap resulting from the challenges of enrolling and sustaining this population in randomized trials over time due to fluctuating symptom presentation and disease severity. Objective data on utilization of telemedicine and continuity of care in CMHC patients with serious mental illness during the pandemic will inform future clinical and policy making. It is critical to recognize the diversity that exists within the serious mental illness community, mitigate biases and assumptions regarding the prospects of telemedicine in this population, and identify characteristics of specific subgroups that may fare better or worse with treatment delivered by telemedicine. The purpose of this study was to (1) describe the characteristics of patients with serious mental illness associated with disruption in services despite the telemedicine expansion during the initial 3 months after the state of emergency declaration in response to COVID-19, (2) describe the characteristics of patients with serious mental illness who were most and least likely to use telemedicine, and (3) determine the extent to which various subpopulations utilized telemedicine to receive treatment.

Methods

Overview

We conducted an observational retrospective study using New Hampshire Medicaid service claims in CMHCs delivering serious mental illness services. The examination compared the 3-month period after the declaration state of emergency (study retention period) to the 3-month period prior to the declaration (study base period), encompassing December 1, 2019 through June 30, 2020. Additionally, in order to assess and account for baseline variability in treatment retention in this vulnerable population, claims were examined from 1 year prior (December 1, 2018 through June 30, 2019).

Ethics

The University of New Hampshire institutional review board reviewed the study protocol and, given that claims data did not contain identifiable protected health information, determined that this study did not require approval.

Study Population

Service claims for New Hampshire Medicaid beneficiaries were included in the analysis if the beneficiary (1) was active or eligible in Medicaid for at least one day within the study base or retention periods and (2) received at least one treatment service from a CMHC during the first 3 months of the study period (December 1, 2019 through February 29, 2020). Patients were excluded if they did not have a treatment service in the study base period.

Study Periods

Due to the COVID-19 pandemic, all 10 of the New Hampshire CMHCs rapidly transitioned most services to telemedicine with patients and providers in their home environment. Clinical providers, in collaboration with patients, determined the treatment delivery modality (ie, onsite versus telemedicine). All New Hampshire CMHCs transitioned to providing at least 50% of services by telemedicine on or before April 1, 2020. March 2020 was a transitional month, and thus, was eliminated from the data set. Therefore, the defined periods were (1) the study base period, from December 1, 2019 through February 29, 2020; (2) the study retention period, from April 1, 2020 through June 30, 2020; (3) the time-trends comparison to study base period, from December 1, 2018 through February 28, 2019; and (4) the time-trends comparison to study retention period, from April 1, 2019 through June 30, 2019.

Claims Acquisition and Preparation

New Hampshire Medicaid claims data, which included CMHC treatment service claims, patient diagnoses, and patient demographic information, were obtained from the New Hampshire Department of Health and Human Services Enterprise Business Intelligence data warehouse in November 2020. Claims for all services provided in CMHCs as defined by National Provider Identifier during the 4 study periods were included. Files were excluded if demographic data were incomplete.

Beneficiaries with CMHC treatment service claims were then selected. Treatment services were defined as codes for services that required therapeutic interaction between a mental health provider and patient (Multimedia Appendix 1). Files with case management and administrative codes that reflected activities independent of patient engagement were, therefore, excluded.

Measures

Outcomes

Treatment interruption was defined as instances in which those who presented with at least one treatment claim during a base period had no treatment claim during the corresponding retention period. Telemedicine use was identified by service claim codes and categorized, based on percentage of total treatment services during the retention period, into low (<25%), medium (25%-75%), or high (>75%). The study population was described by gender (male or female), age groups (0-12 years old, 13-17 years old, 18-34 years old, 35-54 years old, or 55 years and older), ZIP code (urban, representing an area with a population greater than 10,000 people, or rural, representing an area with a population of 10,000 or less), and diagnosis, which was categorized hierarchically (in the following order: schizophrenia, bipolar disorder, major depression, posttraumatic stress disorder, anxiety disorders, and all other conditions), with a designation for each beneficiary in the base and retention periods independently, because most beneficiaries had multiple diagnoses attached to their claims. Thus, if an individual had diagnoses of schizophrenia and major depression, they were included in the schizophrenia diagnosis group.

Statistical Analyses

The study periods and time-trends periods were compared by characteristics of gender, age group, ZIP code, and diagnosis. Summary statistics were used to calculate the change in percentage probability of a serious mental illness treatment interruption from the time trends period to the study period. Each categorical variable in the study and time-trends retention periods were analyzed with chi-square tests for independence. Primary logistic regression included all variables and was used to examine patients who were not retained in services, and again, to examine the patients who were retained in services. The misclassification rate was the number of observations that are classified incorrectly given a cut-off probability of 0.5.

Telemedicine use (low, medium, or high) was analyzed using chi-square tests for independence. The odds ratio (OR) was the proportional odds (ie, the exponent of the estimates) with the low category as the comparator—the odds of going from 25% service use to 25%-75% and >75% service use categories combined. All analyses were performed using JMP software (version 15; SAS Institute).

Results

CMHCs in New Hampshire experienced a 15.0% increase in the number of patients using treatment services from 2019 (n=13,456) to 2020 (n=15,471). In the study retention period, in the quarter after the state of emergency declaration, 18.3% (12,635/15,471) of serious mental illness beneficiaries were not

retained in community mental health treatment; in the analogous period in 2019, 13.4% (11,492/13,456) of serious mental illness beneficiaries were not retained.

There was a 3.0% higher probability of service interruption in male patients versus female patients from 2019 to 2020 (Table 1), and the probability of service disruption increased from 2019 to 2020 in each age group (0-12 years old: 6.9%; 13-17 years old: 5.8%; 18-34 years old: 4.2%; 35-54 years old: 4.7%; 55 years and older: 3.3%).

The probability of service disruption from 2019 to 2020 increased from 5.6% for rural ZIP codes and 4.9% for urban ZIP codes, and the probability of service interruption from 2019 to 2020 increased 2% for patients with schizophrenia, 1% for patients with bipolar disorder, 4.6% for patients with major depression, 5.1% for patients with posttraumatic stress disorder, and 6.8% for patients with anxiety and all other disorders.

A logistic regression model was used to examine the association of categorical variables with age group 55 years and older serving as dependent variable for age group comparison and schizophrenia serving as the dependent variable for diagnosis comparison (Table 2).

Most beneficiaries (11,672/12427, 93.9%) participated in at least one telemedicine visit during the period from April through June 2020; in contrast, during the analogous period in 2019, a very small percentage of beneficiaries (390/13456, 2.9%) (Table 3). All subpopulations within the CMHCs were able to access and utilize telemedicine for at least a part of the treatment plan. Low, medium, and high telehealth utilization in April through June 2020 are shown by gender (Figure 1), age group (Figure 2), ZIP code (Figure 3), and diagnosis (Figure 4) categories.

Female patients had lower odds (OR 0.87, 95% CI 0.86-0.92) than male patients of going from low utilization to either moderate or high utilization. Compared with patients 55 years and older, patients 0 to 12 years old (OR 1.18, 95% CI 1.09-1.27) and 13 to 17 years old (OR 1.16, 95% CI 1.09-1.25) had greater odds and patients 18 to 34 years old (OR 0.74, 95% CI 0.70-0.79) and 35 to 54 years old (OR 0.79, 95% CI 0.74-0.84) had lower odds of going from low utilization to either moderate or high utilization. Except for patients with bipolar disorder (OR 0.93, 95% CI 0.84-1.02), patients with diagnoses other than schizophrenia had lower odds of going from low utilization to either moderate or high utilization (major depression: OR 0.73, 95% CI 0.68-0.78; posttraumatic stress disorder: OR 0.77, 95% CI 0.72-0.83; anxiety or other disorders: OR 0.69, 95% CI 0.65-0.74).

Table 1. CMHC patients with serious mental illness with treatment interruptions.

Characteristic	2020 (n=15,471)		2019 (n=13,456)		Change from 2019 to 2020
	Treatment interruption, n	Probability (*P*<.001), %	Treatment interruption, n	Probability (*P*=.30), %	
Gender					
Female	1495	18.0	1027	14.3	3.7
Male	1548	21.6	937	14.9	6.7
Age group (years)					
0-12	689	22.3	470	15.4	6.9
13-17	565	22.7	377	16.9	5.8
18-34	873	23.6	551	19.4	4.2
35-54	623	16.6	373	11.9	4.7
≥55	293	12.1	193	8.8	3.3
ZIP code					
Rural	1168	21.2	754	15.6	5.6
Urban	1860	18.8	1191	13.9	4.9
Diagnosis					
Schizophrenia	105	6.8	72	4.8	2.0
Bipolar disorder	134	11.0	107	10.0	1.0
Major depression	762	20.1	493	15.5	4.6
Posttraumatic stress disorder	778	21.9	516	16.8	5.1
Anxiety or other	1265	23.5	776	16.7	6.8

Table 2. Logistic regression results for patients experiencing treatment interruption.

Variable	OR[a] (95% CI)	*P* value
Gender		
Female	0.78 (0.72-0.85)	<.001
Male	1.27 (1.17-1.38)	<.001
Age (years)		
0-12 years	1.37 (1.17-1.61)	<.001
13-17 years	1.49 (1.27-1.75)	<.001
18-34 years	1.83 (1.58-2.12)	<.001
35-54 years	1.29 (1.11-1.50)	.001
≥55	Reference	N/A[b]
ZIP code		
Rural	1.12 (1.03-1.22)	.006
Urban	0.89 (0.82-0.97)	.006
Diagnosis		
Schizophrenia	Reference	N/A
Bipolar	1.67 (1.28-2.17)	<.001
Major depression	3.32 (2.67-4.13)	<.001
Posttraumatic stress disorder	3.69 (2.96-4.61)	<.001
Anxiety or other	3.77 (3.04-4.68)	<.001

[a]OR: odds ratio.

[b]N/A: not applicable.

Table 3. Telemedicine utilization among patients with serious mental illness in the study retention period, after the pandemic state of emergency.

Variables	All	Low use		Medium use		High use	
	N	n	%	n	%	n	%
All	12,427	1845	14.84	6294	50.64	4288	34.50
Gender							
Female	6792	850	12.51	3394	49.97	2548	37.51
Male	5631	994	17.65	2897	51.45	1740	30.90
Age, in years							
0-12	2397	255	10.64	1436	59.91	706	29.45
13-17	1929	218	11.30	1096	56.82	615	31.88
18-34	2833	360	12.70	1321	46.63	1152	40.66
35-54	3132	491	15.68	1440	45.98	1201	38.34
55+	2132	520	24.39	998	46.81	614	28.80
Diagnosis							
Schizophrenia	1447	535	36.97	653	45.12	259	17.90
Bipolar	1080	175	16.20	502	46.48	403	37.31
Major depression	2949	366	12.41	1456	49.37	1127	38.21
Posttraumatic stress disorder	2769	291	10.51	1534	55.40	944	34.09
Other	4182	478	11.43	2149	51.38	1555	37.18
ZIP code							
Rural	4353	517	11.87	2221	51.02	1615	37.10
Urban	8037	1321	16.43	4056	50.46	2660	33.09

Figure 1. Telemedicine use by gender.

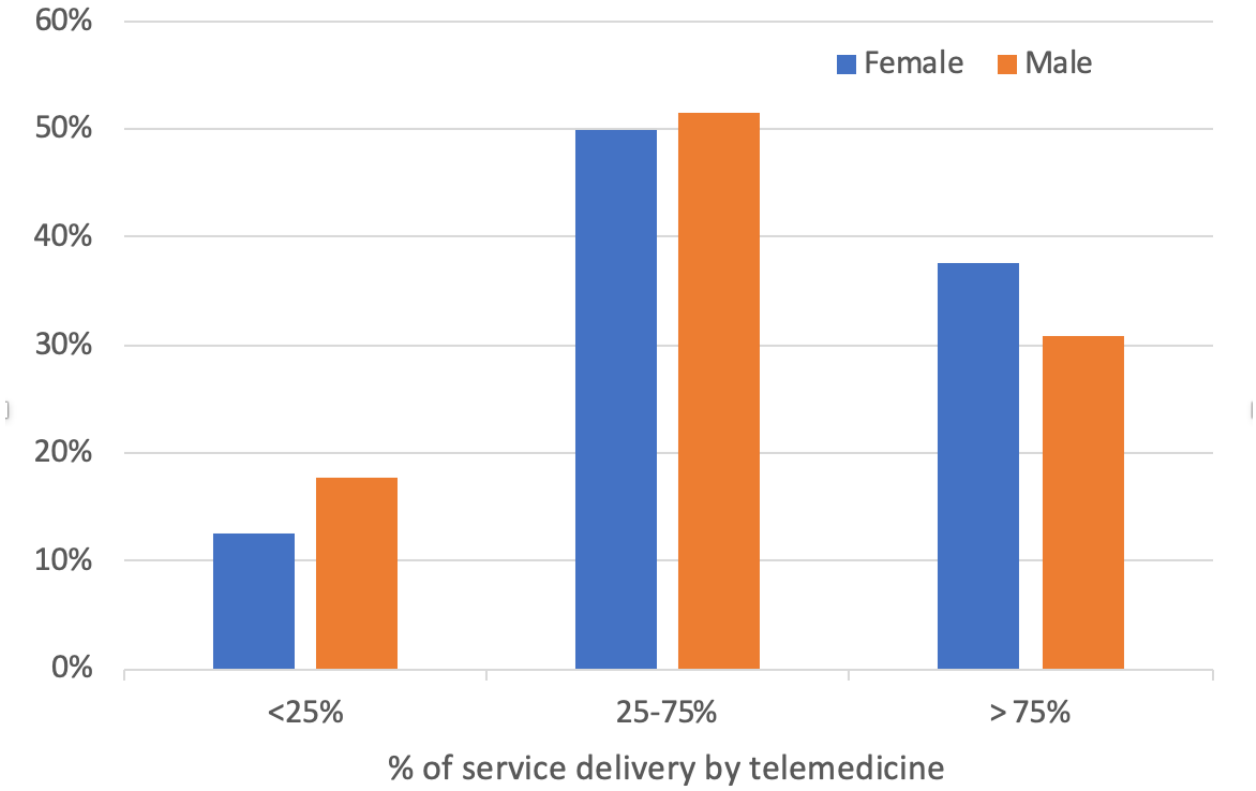


Figure 2. Telemedicine use by age group.

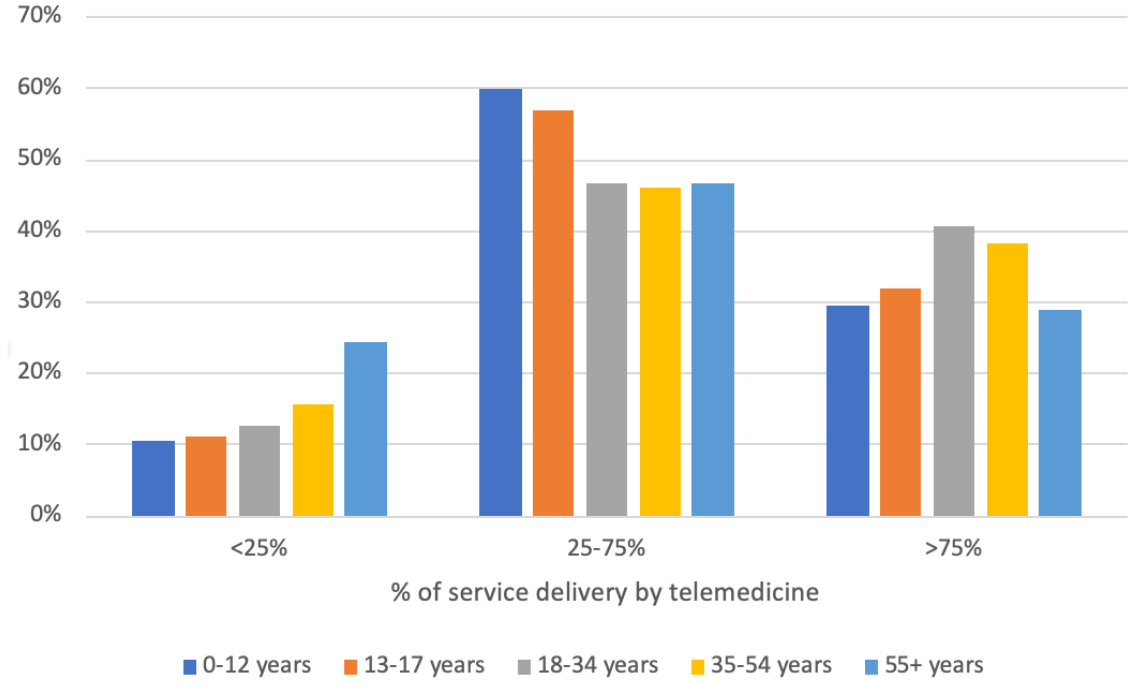


Figure 3. Telemedicine use by rurality.

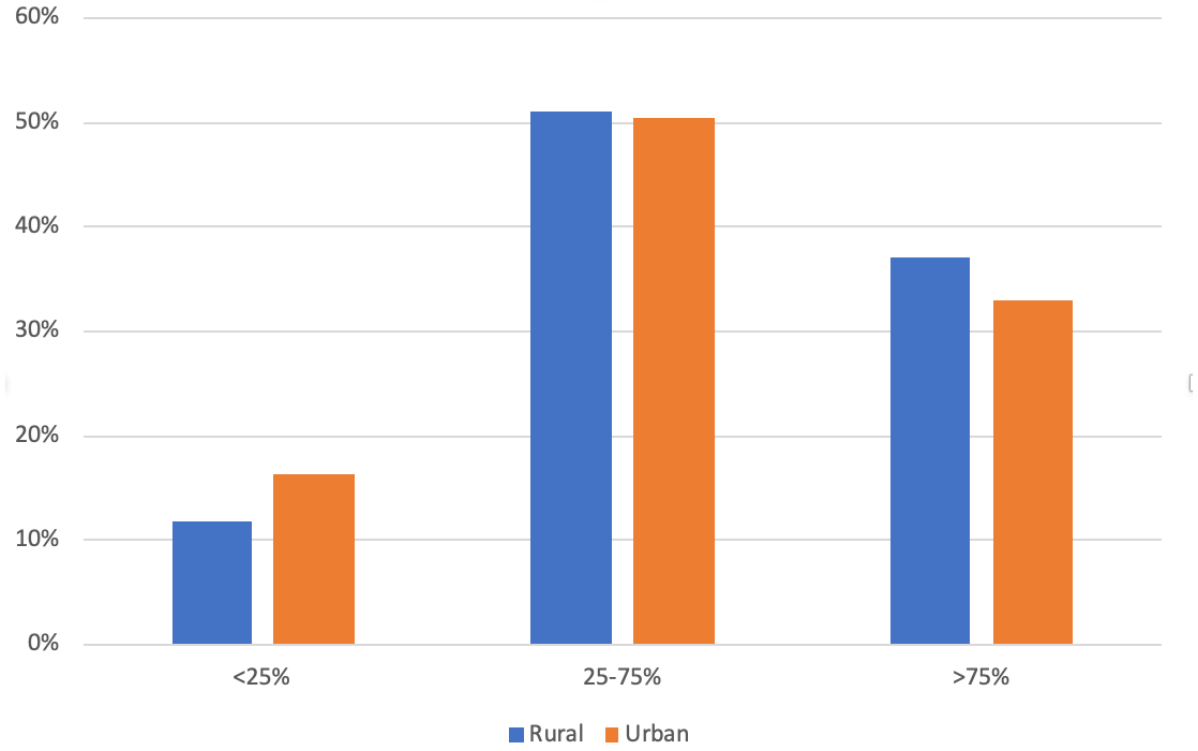
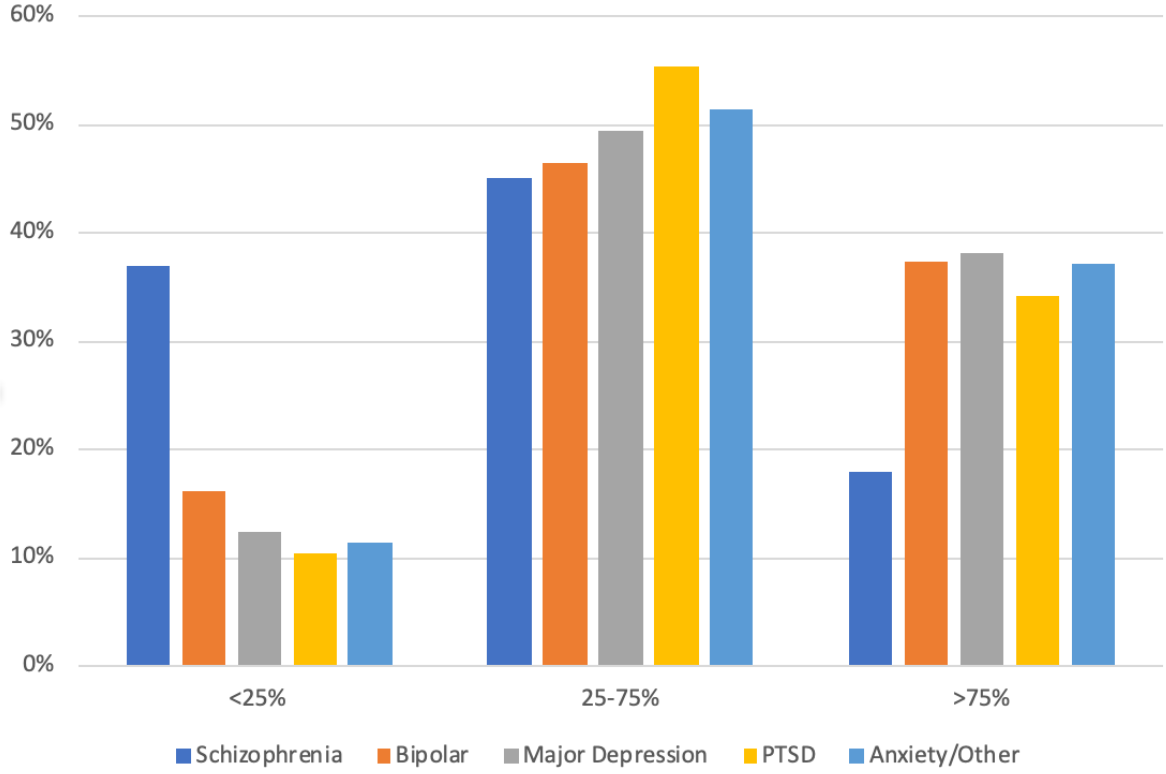


Figure 4. Telemedicine use by condition. PTSD: posttraumatic stress disorder.



Discussion

Telemedicine was utilized by the majority of CMHC patients in the months following the pandemic, likely supporting continuity of care for many vulnerable patients with serious mental illness. This is consistent with the results of national surveys on telemedicine utilization in serious mental illness [27,30,31]. Yet, even with the substantial rollout of telemedicine, retention in treatment was less than retention the prior year, and some subpopulations were more at risk for treatment interruptions than others. With limited data available on telemedicine care delivery in persons with schizophrenia and bipolar disorder in particular, this data may be used to inform the decision of how best to deliver care in this population.

Older patients and patients with more severe disorders (ie, schizophrenia) were more likely to be retained in treatment; however, they were also the least likely to utilize telemedicine. This finding shows that the decision for choosing modality of care can be trusted and empowered at the clinical level, as these vulnerable patients demonstrated higher than average retention rates. The findings of this study suggest that health care professional are able to individually identify for whom and when telemedicine was a viable option. The choice of treatment modality is nuanced and may affect whether a patient is retained in treatment. With treatment retention as the overarching goal, these findings support individualized decision-making about treatment delivery modalities through patient–provider collaboration.

Female patients were more likely to use more telemedicine and more likely to be retained than male patients. Although all age groups used a lot of telemedicine services, we found that patients 55 years and older had the lowest rates of interrupted service and were in the lowest telemedicine utilization category. While some older adults require assistance navigating digital platforms [32], a systematic review [33] of telemedicine feasibility and acceptability in older adults suggested patients demonstrated high levels of feasibility and acceptability, health care providers perceived patients of this age group to have physical, sensory, cognitive, and visual–spatial challenges to successful telemedicine use. These perceptions demonstrate a bias among telemedicine use in older adult patients that is impeding this method of care delivery [33]. Based on these findings, exploring decision-making around modality choice with older adults must include recognition of individual and systemic biases that may be limiting a meaningful means of service delivery. Furthermore, adequate technical support must be put in place to ensure an equitable health care delivery system.

Youth under 18 years old, have been found, prior to the pandemic, to have high rates of acceptability and satisfaction with telemedicine services [12,34,35]; however, nevertheless, consistent with previous findings [36], the findings of our study showed that youth and adolescents had the greatest increase in service interruption compared with all other age groups year over year. Thus, the pediatric service interruptions were not expected and the cause for this should be further explored. While the under 18-year-old patient population utilized more telemedicine than most other age groups, telemedicine use during pandemic lockdowns would have required internet access, device access, and parental support or supervision. Adequate internet access was a challenge during the pandemic as work and school demands transitioned to remote access [37]. Beyond internet access, Wi-Fi–enabled devices (ie, mobile phones, laptops, and tablets) were needed by adults and children to meet this new form of engagement with work and school. Additionally, parents and caregivers were strained balancing

home and work responsibilities, therefore, parents limited capabilities, to manage minor's mental health appointments, might be expected. Finally, with most schools pivoting to web-based learning, there were fewer adults witnessing the behaviors and mental health needs of the students to encourage outreach for mental health services [38]. However, before this service interruption rate of our youth is accepted as a result of pandemic specific circumstances and not representative of broad telemedicine mental health services, root causes must be explored. Because traumatic childhood events can have a long-term impact, facilitating youth engagement in mental health services overtime must be improved.

Living in urban or rural area did not significantly impact likelihood of retention in services (P=.006) or telemedicine utilization (P=.009). In contrast, Chu et al [31] found that urban-dwelling patients demonstrated a larger increase in telemedicine utilization residing in Ontario, Canada.

Social isolation and the spread of misinformation during the pandemic has been documented to have precipitated symptom exacerbation in preexisting mental illness [39] and some psychotic events in those struggling with schizophrenia. However, stable levels of psychotic symptoms and an increase in a sense of well-being are documented in early literature exploring this disease during the COVID-19 outbreak [40]. These positive outcomes are consistent with our findings that persons with schizophrenia were the most likely beneficiaries to be retained for services during the study period. Of all diagnoses, schizophrenia demonstrated the lowest use of telemedicine services.

Offering a mixed modality of service options enabled New Hampshire CMHCs to have a very high level of retention across demographic and diagnostic variables. It should be noted in this discussion that a large majority of serious mental illness beneficiaries were able to demonstrate access, utilization, and know-how to pivot to telemedicine services for continuity of care during this exceptional public health emergency. Given the scope of challenges facing all beneficiaries and health care facilities from April to June 2020, during the onset of the pandemic, retention in services continued at a rate of only 4.9% below that of the prior year during the same time.

There were 4 main limitations to this research. First, service modality specific coding and billing modifiers used to differentiate between televideo and telephone services were not available from the outset of the state of emergency. This was likely due to parity in billing and other pressing needs during the outset of the pandemic. A national survey of patients with serious mental illness during the same period as that of our study (April through June 2020) found that approximately 64% of telemedicine visits occurred via the telephone [27] versus 23% via televideo and 13% via a combination of telephone and televideo; however, these findings were based on self-reported information and could not be validated through claims data. Second, race and ethnicity reporting in Medicaid claims was incomplete, and thus, race/ethnicity could not be included as a variable. The majority of the New Hampshire Medicaid beneficiaries receive benefits through privately managed companies for plan administration; adherence to collecting the data on race and ethnicity was poor among privately managed companies. Third, qualitative data were not collected to understand the provider–patient decision-making about choice of service modality, which would have further identified variables impacting successful engagement with care. Fourth, we collected data under the conditions of a global pandemic. It is not clear if the provider or patient behaviors and actions exhibited during this time are representative of those when there is no pandemic.

It is well documented that people with serious mental illness have greater difficulties coping with disaster events, including higher avoidance, less resilience, and a stronger likelihood of an adverse exacerbation of symptoms to a distant event, let alone one happening in the present day. Treatment continuity is critical to prevent personally and financially costly exacerbations. This research demonstrated that individualized service modality decisions to promote engagement are effectively determined between patient and provider. Furthermore, telemedicine promoted continuity of care during the pandemic across all subpopulations of CMHC patients with serious mental illness, with some populations demonstrating more utilization than others.

This research examined the initial months after the state of emergency; a more comprehensive evaluation of the effectiveness of telemedicine for people with serious mental illness is needed. Stakeholders, including patients, providers, administrators, and policymakers, require data demonstrating how best to sustain engagement in care for CMHC patients in order to make decisions about when and for whom telemedicine is efficacious.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Community mental health center treatment codes.
[DOCX File , 13 KB - mental_v9i3e33092_app1.docx]

References

1. Dismuke CE, Egede LE. Association of serious psychological distress with health services expenditures and utilization in a national sample of US adults. Gen Hosp Psychiatry 2011;33(4):311-317. [doi: 10.1016/j.genhosppsych.2011.03.014] [Medline: 21762826]

2. Olfson M, Wang S, Wall M, Marcus SC, Blanco C. rends in serious psychological distress and outpatient mental health care of US adults. JAMA Psychiatry 2019 Feb 01;76(2):152-161 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.3550] [Medline: 30484838]
3. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):617-627 [FREE Full text] [doi: 10.1001/archpsyc.62.6.617] [Medline: 15939839]
4. Cloutier M, Aigbogun MS, Guerin A, Nitulescu R, Ramanakumar AV, Kamat SA, et al. The economic burden of schizophrenia in the United States in 2013. J Clin Psychiatry 2016 Jun;77(6):764-771. [doi: 10.4088/JCP.15m10278] [Medline: 27135986]
5. Kvedar J, Coye MJ, Everett W. Connected health: a review of technologies and strategies to improve patient care with telemedicine and telehealth. Health Aff (Millwood) 2014 Feb;33(2):194-199. [doi: 10.1377/hlthaff.2013.0992] [Medline: 24493760]
6. Choi KR, Easterlin MC. Intervention models for increasing access to behavioral health services among youth: a systematic review. J Dev Behav Pediatr 2018 Dec;39(9):754-762. [doi: 10.1097/DBP.0000000000000623] [Medline: 30334855]
7. Brown JD, Barrett A, Hourihan K, Caffery E, Ireys HT. State variation in the delivery of comprehensive services for Medicaid beneficiaries with schizophrenia and bipolar disorder. Community Ment Health J 2015 Jul;51(5):523-534. [doi: 10.1007/s10597-015-9857-5] [Medline: 25786723]
8. Mueser KT, Corrigan PW, Hilton DW, Tanzman B, Schaub A, Gingerich S, et al. Illness management and recovery: a review of the research. Psychiatr Serv 2002 Oct;53(10):1272-1284. [doi: 10.1176/appi.ps.53.10.1272] [Medline: 12364675]
9. Saloner B, Bandara S, Bachhuber M, Barry CL. Insurance coverage and treatment use under the Affordable Care Act among adults with mental and substance use disorders. Psychiatr Serv 2017 Jun 01;68(6):542-548. [doi: 10.1176/appi.ps.201600182] [Medline: 28093059]
10. McCarthy JF, Blow FC, Valenstein M, Fischer EP, Owen RR, Barry KL, et al. Veterans Affairs Health System and mental health treatment retention among patients with serious mental illness: evaluating accessibility and availability barriers. Health Serv Res 2007 Jun;42(3 Pt 1):1042-1060 [FREE Full text] [doi: 10.1111/j.1475-6773.2006.00642.x] [Medline: 17489903]
11. Douglas MD, Xu J, Heggs A, Wrenn G, Mack DH, Rust G. Assessing telemedicine utilization by using Medicaid claims data. Psychiatr Serv 2017 Feb 01;68(2):173-178 [FREE Full text] [doi: 10.1176/appi.ps.201500518] [Medline: 27691381]
12. Sauers-Ford HS, Hamline MY, Gosdin MM, Kair LR, Weinberg GM, Marcin JP, et al. Acceptability, usability, and effectiveness: a qualitative study evaluating a pediatric telemedicine program. Acad Emerg Med 2019 Sep;26(9):1022-1033 [FREE Full text] [doi: 10.1111/acem.13763] [Medline: 30974004]
13. Creedon TB, Schrader KE, O'Brien PL, Lin JR, Carroll CD, Mulvaney-Day N. Rural-nonrural differences in telemedicine use for mental and substance use disorders among Medicaid beneficiaries. Psychiatr Serv 2020 Aug 01;71(8):756-764. [doi: 10.1176/appi.ps.201900444] [Medline: 32290806]
14. Goldman ML, Druss BG, Horvitz-Lennon M, Norquist GS, Kroeger Ptakowski K, Brinkley A, et al. Mental health policy in the era of COVID-19. Psychiatr Serv 2020 Nov 01;71(11):1158-1162. [doi: 10.1176/appi.ps.202000219] [Medline: 32517639]
15. Yang J, Landrum MB, Zhou L, Busch AB. Disparities in outpatient visits for mental health and/or substance use disorders during the COVID surge and partial reopening in Massachusetts. Gen Hosp Psychiatry 2020;67:100-106 [FREE Full text] [doi: 10.1016/j.genhosppsych.2020.09.004] [Medline: 33091782]
16. Connolly SL, Stolzmann KL, Heyworth L, Weaver KR, Bauer MS, Miller CJ. Rapid increase in telemental health within the Department of Veterans Affairs during the COVID-19 pandemic. Telemed J E Health 2020 Oct 14:1. [doi: 10.1089/tmj.2020.0233] [Medline: 32926664]
17. Choi NG, Hegel MT, Marti N, Marinucci ML, Sirrianni L, Bruce ML. Telehealth problem-solving therapy for depressed low-income homebound older adults. Am J Geriatr Psychiatry 2014 Mar;22(3):263-271 [FREE Full text] [doi: 10.1097/JGP.0b013e318266b356] [Medline: 23567376]
18. Muir SD, de Boer K, Nedeljkovic M, Meyer D. Barriers and facilitators of videoconferencing psychotherapy implementation in veteran mental health care environments: a systematic review. BMC Health Serv Res 2020 Nov 01;20(1):999 [FREE Full text] [doi: 10.1186/s12913-020-05858-3] [Medline: 33131495]
19. Mulligan J, Haddock G, Hartley S, Davies J, Sharp T, Kelly J, et al. An exploration of the therapeutic alliance within a telephone-based cognitive behaviour therapy for individuals with experience of psychosis. Psychol Psychother 2014 Dec;87(4):393-410. [doi: 10.1111/papt.12018] [Medline: 24464969]
20. Shigekawa E, Fix M, Corbett G, Roby DH, Coffman J. The current state of telehealth evidence: a rapid review. Health Aff (Millwood) 2018 Dec;37(12):1975-1982. [doi: 10.1377/hlthaff.2018.05132] [Medline: 30633674]
21. Choi NG, Marti CN, Wilson NL, Chen GJ, Sirrianni L, Hegel MT, et al. Effect of telehealth treatment by lay counselors vs by clinicians on depressive symptoms among older adults who are homebound: a randomized clinical trial. JAMA Netw Open 2020 Aug 03;3(8):e2015648 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.15648] [Medline: 32865577]

22. Choi NG, Marti CN, Bruce ML, Hegel MT, Wilson NL, Kunik ME. Six-month postintervention depression and disability outcomes of in-home telehealth problem-solving therapy for depressed, low-income homebound older adults. Depress Anxiety 2014 Aug;31(8):653-661 [FREE Full text] [doi: 10.1002/da.22242] [Medline: 24501015]
23. Fortney JC, Pyne JM, Kimbrell TA, Hudson TJ, Robinson DE, Schneider R, et al. Telemedicine-based collaborative care for posttraumatic stress disorder: a randomized clinical trial. JAMA Psychiatry 2015 Jan;72(1):58-67. [doi: 10.1001/jamapsychiatry.2014.1575] [Medline: 25409287]
24. O'Reilly R, Bishop J, Maddox K, Hutchinson L, Fisman M, Takhar J. Is telepsychiatry equivalent to face-to-face psychiatry? results from a randomized controlled equivalence trial. Psychiatr Serv 2007 Jun;58(6):836-843. [doi: 10.1176/ps.2007.58.6.836] [Medline: 17535945]
25. Schulze LN, Stentzel U, Leipert J, Schulte J, Langosch J, Freyberger HJ, et al. Improving medication adherence with telemedicine for adults with severe mental illness. Psychiatr Serv 2019 Mar 01;70(3):225-228. [doi: 10.1176/appi.ps.201800286] [Medline: 30651059]
26. Talley RM, Brunette MF, Adler DA, Dixon LB, Berlant J, Erlich MD, et al. Telehealth and the community SMI population: reflections on the disrupter experience of COVID-19. J Nerv Ment Dis 2021 Jan;209(1):49-53. [doi: 10.1097/NMD.0000000000001254] [Medline: 33003053]
27. Guinart D, Marcy P, Hauser M, Dwyer M, Kane JM. Patient attitudes toward telepsychiatry during the COVID-19 pandemic: a nationwide, multisite survey. JMIR Ment Health 2020 Dec 22;7(12):e24761 [FREE Full text] [doi: 10.2196/24761] [Medline: 33302254]
28. Lynch DA, Medalia A, Saperstein A. The design, implementation, and acceptability of a telehealth comprehensive recovery service for people with complex psychosis living in NYC during the COVID-19 crisis. Front Psychiatry 2020;11:581149 [FREE Full text] [doi: 10.3389/fpsyt.2020.581149] [Medline: 33101093]
29. Chen JA, Chung W, Young SK, Tuttle MC, Collins MB, Darghouth SL, et al. COVID-19 and telepsychiatry: early outpatient experiences and implications for the future. Gen Hosp Psychiatry 2020;66:89-95 [FREE Full text] [doi: 10.1016/j.genhosppsych.2020.07.002] [Medline: 32750604]
30. Costa M, Reis G, Pavlo A, Bellamy C, Ponte K, Davidson L. Tele-mental health utilization among people with mental illness to access care during the CoViD-19 pandemic. Community Ment Health J 2021 May;57(4):720-726 [FREE Full text] [doi: 10.1007/s10597-021-00789-7] [Medline: 33566269]
31. Chu C, Cram P, Pang A, Stamenova V, Tadrous M, Bhatia RS. Rural telemedicine use before and during the COVID-19 pandemic: a repeated cross-sectional study. J Med Internet Res 2021 Mar 24:e26960 [FREE Full text] [doi: 10.2196/26960] [Medline: 33769942]
32. Lam K, Lu AD, Shi Y, Covinsky KE. Assessing telemedicine unreadiness among older adults in the United States during the COVID-19 pandemic. JAMA Intern Med 2020 Oct 01;180(10):1389-1391. [doi: 10.1001/jamainternmed.2020.2671] [Medline: 32744593]
33. Batsis JA, DiMilia PR, Seo LM, Fortuna KL, Kennedy MA, Blunt HB, et al. Effectiveness of ambulatory telemedicine care in older adults: a systematic review. J Am Geriatr Soc 2019 May 08:1737-1749. [doi: 10.1111/jgs.15959] [Medline: 31066916]
34. Vyas S, Murren-Boezem J, Solo-Josephson P. Analysis of a pediatric telemedicine program. Telemed J E Health 2018 Apr 24:993-997. [doi: 10.1089/tmj.2017.0281] [Medline: 29688823]
35. Boydell KM, Hodgins M, Pignatiello A, Teshima J, Edwards H, Willis D. Using technology to deliver mental health services to children and youth: a scoping review. J Can Acad Child Adolesc Psychiatry 2014 May;23(2):87-99 [FREE Full text] [Medline: 24872824]
36. Nicholas J, Bell IH, Thompson A, Valentine L, Simsir P, Sheppard H, et al. Implementation lessons from the transition to telehealth during COVID-19: a survey of clinicians and young people from youth mental health services. Psychiatry Res 2021 May;299:113848. [doi: 10.1016/j.psychres.2021.113848] [Medline: 33725578]
37. Mitchell S. Ethical concerns of school closures for low-income school aged children. Voice Bioethics 2020;6:1 [FREE Full text] [doi: 10.7916/vib.v6i.7057]
38. Hoffman JA, Miller EA. Addressing the consequences of school closure due to COVID-19 on children's physical and mental well-being. World Med Health Policy 2020 Aug 20;59(12):1309-1310 [FREE Full text] [doi: 10.1002/wmh3.365] [Medline: 32904951]
39. Usher K, Bhullar N, Jackson D. Life in the pandemic: social isolation and mental health. J Clin Nurs 2020 May 06;29(15-16):2756-2757 [FREE Full text] [doi: 10.1111/jocn.15290]
40. Barlati S, Nibbio G, Vita A. Schizophrenia during the COVID-19 pandemic. Curr Opin Psychiatry 2021 May 01;34(3):203-210. [doi: 10.1097/YCO.0000000000000702] [Medline: 33587494]

Abbreviations

CMHC: community mental health center
OR: odds ratio

Edited by J Torous; submitted 23.08.21; peer-reviewed by RA Karin, J Torous; accepted 24.01.22; published 21.03.22.

Please cite as:
Ainslie M, Brunette MF, Capozzoli M
Treatment Interruptions and Telemedicine Utilization in Serious Mental Illness: Retrospective Longitudinal Claims Analysis
JMIR Ment Health 2022;9(3):e33092
URL: https://mental.jmir.org/2022/3/e33092
doi:10.2196/33092
PMID:35311673

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org