JMIR Mental Health

Journal Impact Factor (JIF) (2022): 5.2
Volume 8 (2021), Issue 11 ISSN 2368-7959 Editor in Chief: John Torous, MD, MBI

Contents

Original Papers

Reviews

Viewpoint

Original Paper

Suicide Risk and Protective Factors in Online Support Forum Posts: Annotation Scheme Development and Validation Study

Stevie Chancellor[1], PhD; Steven A Sumner[2], MSc, MD; Corinne David-Ferdon[3], PhD; Tahirah Ahmad[4], BSc; Munmun De Choudhury[4], PhD

[1]Department of Computer Science & Engineering, University of Minnesota - Twin Cities, Minneapolis, MN, United States

[2]Office of Strategy and Innovation, National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, Atlanta, GA, United States

[3]Division of Violence Prevention, Centers for Disease Control and Prevention, Atlanta, GA, United States

[4]School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA, United States

Corresponding Author:
Stevie Chancellor, PhD
Department of Computer Science & Engineering
University of Minnesota - Twin Cities
200 Union Street SE
4-189 Keller Hall
Minneapolis, MN, 55455
United States
Phone: 1 612 625 4002
Email: steviec@umn.edu

Abstract

Background: Online communities provide support for individuals looking for help with suicidal ideation and crisis. As community data are increasingly used to devise machine learning models to infer who might be at risk, there have been limited efforts to identify both risk and protective factors in web-based posts. These annotations can enrich and augment computational assessment approaches to identify appropriate intervention points, which are useful to public health professionals and suicide prevention researchers.

Objective: This qualitative study aims to develop a valid and reliable annotation scheme for evaluating risk and protective factors for suicidal ideation in posts in suicide crisis forums.

Methods: We designed a valid, reliable, and clinically grounded process for identifying risk and protective markers in social media data. This scheme draws on prior work on construct validity and the social sciences of measurement. We then applied the scheme to annotate 200 posts from r/SuicideWatch—a Reddit community focused on suicide crisis.

Results: We documented our results on producing an annotation scheme that is consistent with leading public health information coding schemes for suicide and advances attention to protective factors. Our study showed high internal validity, and we have presented results that indicate that our approach is consistent with findings from prior work.

Conclusions: Our work formalizes a framework that incorporates construct validity into the development of annotation schemes for suicide risk on social media. This study furthers the understanding of risk and protective factors expressed in social media data. This may help public health programming to prevent suicide and computational social science research and investigations that rely on the quality of labels for downstream machine learning tasks.

(JMIR Ment Health 2021;8(11):e24471) doi:10.2196/24471

KEYWORDS

online communities; suicide crisis; construct validity; annotation scheme; Reddit; annotation

Introduction

Background

In the United States, suicide is a leading cause of death and a pressing public health concern [1,2]. Suicide rates have increased >30% over the past 20 years [2]. Suicide is preventable [1,3]—early identification and support of people at risk, such as those with suicidal ideation, is a proven strategy that can reduce suicide [3].

Digital communities and social networking platforms provide support to individuals who may be considering self-harm or suicide. Examples of such communities are r/SuicideWatch on Reddit [4-6], ReachOut [7-9], and TalkLife [10], which offer dynamic and organic support that assists those in need. For this aim, social media data have been harnessed as a naturalistic and unobtrusive source of information about how to improve suicide prevention [4,6,11-13].

A focus in recent research has been to conceptualize and quantify risk from web-based posts, thereby identifying who may be most in need of assistance. *Suicide risk* estimation assesses the likelihood that someone may attempt or die by suicide. For extracting measures of risk from these data, prior work has often conceptualized risk into categorical or ordinal groups—risky and not risky [4,6], a *stoplight* system of green, yellow, and red [7-9], or a 0 to 3 scale [13,14]. Categories are then mapped to training data for computational linguistic analysis and the development of machine learning models to quantify risk and potentially predict behavior [9,13,15-17].

Current quantifiable risk evaluations of suicidality map to a single perspective of evaluating risk, which focuses on aggregated notions of riskiness that may determine a response from a clinician. Instead of collapsing the notion of risk into a singular point, clinical and public health professionals instead often examine and track *risk factors* or attributes and characteristics that increase an individual's likelihood of attempting suicide in the future [18]. Such health professionals also explore *protective factors* or characteristics and behaviors that decrease the probability of suicidal ideation, planning, or attempts [18,19]. These include both psychological factors, such as access to mental health care and treatment, and social factors, such as supportive family members. These factors are important as they provide resilience and a *buffer* against suicide [19,20]. Assessing both risk and protective factors provides a more nuanced and holistic view of the risk for suicide.

Labeling social media data for complex behaviors such as suicidality is simultaneously pervasive within research and challenging. Social media data do not include clinically validated signals of distress or diagnosis, and labels must therefore be generated. Agreeing on and applying these labels to data sets is difficult in part as the evaluation of mental health (especially for suicide risk) is more subjective and requires complex labeling schemes [21,22]. However, there are no current schemas that study risk and protective factors for suicidal ideation in social media data. Moreover, there are no practical guidelines on how to construct and validate annotation systems and schema for complex mental health behaviors in social media [23]. This is of critical concern given that recent research on mental health and social media has identified numerous challenges in how clinical and health signals are constructed, annotated, and verified in data sets [23-25]. The reliability and validity of these signals are essential for ensuring studies on social media data accurately measure what they claim to measure [26,27].

Objective

To address this problem, we draw on the vocabulary and tools of construct validity measurement from the social sciences to formalize an annotation scheme. Measures of validity have a long and rich history in social sciences (under the name of measurement modeling) [27], computational linguistics [28,29], and psychometrics [30,31]. In this study, we focused on construct validity, or "making inferences from the sampling particulars of a study to the higher-order constructs they represent" [27,32]. In our case, this allows us to translate the higher-order clinical concept of *risk and protective factors* to those in social media. By using construct validity as an anchoring concept for our research, we aim to produce more accurate, representative, and reliable labels of risk and protective factors from digital text.

In this study, we provide the development process, a first validation, and results for a framework for operationalizing and testing clinical concepts via social media data. We do so by assessing the risk and protective factors of suicidal ideation in r/SuicideWatch, a Reddit community dedicated to social support during a suicide crisis event. A team of experts in social media, mental health, public health, and suicide worked collaboratively to develop this annotation scheme. We have provided detailed descriptions and procedures for iterative development and validation. Finally, we tested this approach on 200 posts from the community and discussed the initial results of our annotations and how they reflect on studying suicidal ideation in social media.

Our work provides a formalized approach for developing annotation data for suicide and social media data. We have discussed the implications of this research as they relate to the development of rigorous and validated frameworks for assessing mental health on the web. This work also considers downstream applications, such as expert annotation, training laypersons for generating training data in machine learning, closed coding for qualitative analysis or for grounded evaluation of machine learning model outcomes that assess suicide risk and buffers.

Methods

Our Approach to Labeling

Our research goals connect to the larger area of labeling data—a problem that applies across fields outside of computer science, such as linguistics [28] and psychometrics [30,31]. Given these considerations and our priorities for exploring construct validity through labeling [33], we designed a novel and iterative process for building an annotation scheme to evaluate suicide risk and protective factors in social media posts. We adopted the socioecological framework as the basis for labeling these factors. Initially focused on the sociological study of human development [34], socioecological models help conceptualize

the dynamic and interrelated factors that influence outcomes in psychological behaviors [34], and in our case, suicidal ideation [2,35]. In addition to personal and individual factors, this model accounts for circumstances such as relationships, community, and social pressures that affect well-being.

Creating labels that accurately capture what is of interest is tied to *construct validity*, the degree to which a practical measure or label captures what theoretical concept it claims to measure [32]. Guided by the approach of Simms and Watson [31] to construct validity and psychometric instruments, we formalized an approach to annotation scheme development that aligns around two core questions for construct validity:

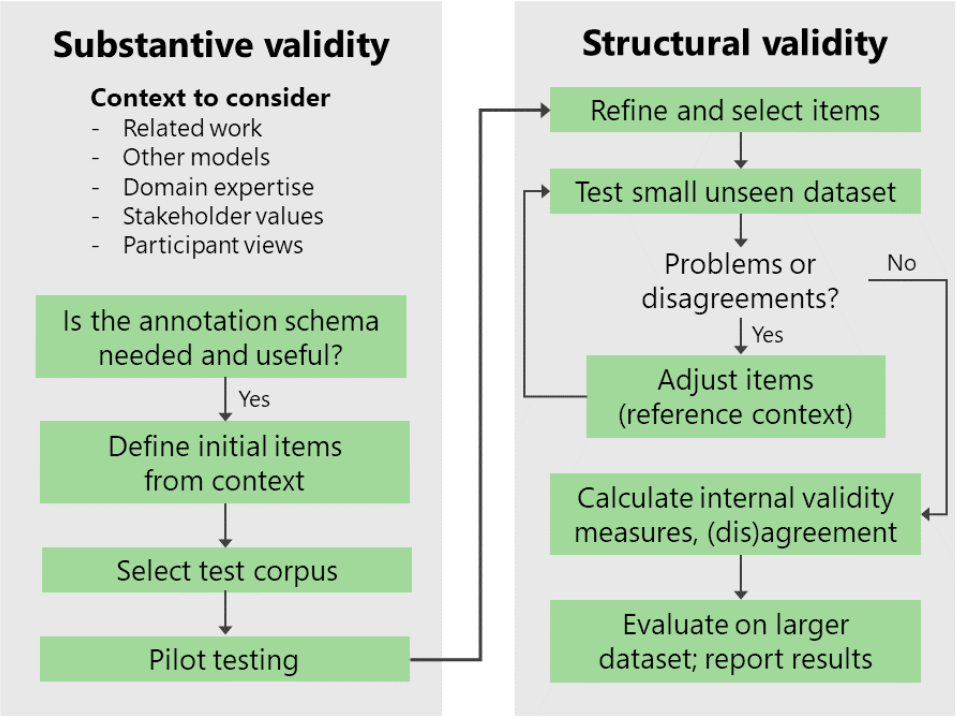
1. Is this annotation system needed, useful, and in alignment with prior work and expertise?
2. Is the annotation system reflective of the prior literature and able to be applied reliably across a research team?

We present an overview of this process in Figure 1. The first question approximates the process of *substantive validity*, which Simms and Watson [31] argue is "centered on the tasks of construct conceptualization and development of the initial item pool." We expand on the approach by Simms and Watson [31] for development to include crucial input from stakeholders and possibly participants for whom labels will be applied, adopting a stakeholder-driven and human-centered approach to social media data analysis [36]. The data set of interest and a pilot annotation scheme can then be developed.

Next comes the *structural validity* phase, where the scheme was tested against the construct in practical and measurable ways. We focused on two strategies for reaching consensus: small-scale testing and refinement of items and intergroup reliability testing. Raters apply the ratings to a random but small set of new examples from the social media corpus and engage in group discussions to adjust items and themes. Once consensus was reached, the raters independently annotated a larger batch of posts and recorded the metrics of interrater reliability to evaluate the consistency of the scheme.

In the following sections, we describe our application of this procedure to suicide risk and protective factors in social media data.

Figure 1. An overview of our annotation process.



Data Collection and Preparation

Source of Data

We used data from Reddit, a social media site organized into *subreddits*, individual communities organized around topic areas. We chose to study r/SuicideWatch, given its focus and interest from prior work [4,13,37,38] and ample text space for content (50,000 characters).

In June 2019, we gathered our data set from r/SuicideWatch (r/SW) from archived, public Reddit data through Google's BigQuery data storage platform, acquiring all data between January 2016 and February 2019. We then prefiltered the data set to remove content deleted by either moderators or users, as indicated by the *[deleted]* and *[removed]* tags. We also removed content posted by the subreddit's moderators and the user u/AutoModerator, a Reddit bot designed to automate moderation tasks.

Next, we selected 1000 posts to build an annotation data set, randomly sampled without replacement, for constructing all piloting data sets and the final annotation data set. We discarded posts that had short (>5 words, including the title) or long text content (>1500 words), as requested after a few rounds of piloting by 2 members of the research team. Short posts were removed because of the difficulty in providing meaningful annotations about the risk or protective factors; annotators found

it difficult to evaluate concrete factors with no details. Long posts often contained so much information that they overwhelmed the labeling schema. Together, these posts were very rare in our sample of 1000 posts—<20 posts or <2%. We also manually inspected each post to remove those that asked for help on behalf of someone else or that were about suicide bereavement (around 10 posts or 1%). We then gathered descriptive statistics for our data set, as shown in Table 1.

Table 1. Summary statistics for the 1000-candidate post data set.

Characteristics	Values
Total number of unique users	984
Post length (words), mean (SD)	221.81 (267.98)
Median length	141

Data Deidentification

Following best practices for detailed annotation of suicide content [15,33,39], we deidentified each post to remove personal details. We first removed any mention of usernames or links from the posts. Next, we tagged all person, organization, or location names using Stanford's Named Entity Recognizer through the nltk Python library. We replaced any tagged words with placeholder text (eg, named locations with the term LOCATION). To verify that these data were deidentified, the researcher responsible for gathering the data set manually checked and edited any posts to remove identifiable information, as necessary. After this step, the data were passed to the broader research team for coding.

Research Team and Positionality

The research team included 4 experts with complementary experiences across social media, mental health, public health, and suicide. This approach represents an interdisciplinary collaboration that considers public health, clinical, and social computing perspectives.

A total of 2 researchers are public health experts with additional backgrounds in psychology and clinical medicine. The other 2 researchers are computer scientists who are experts in social media and mental health. The team also included people with lived experiences of mental illness. Together, they have extensive experience working in high-risk mental health behaviors, such as suicide, expressed through social media.

Designing the Annotation Scheme

Phase 1: Evaluation of Context and Preliminary Item Development

To begin the initial development, we drew on several sources to understand the risk and protective factors. First, we reviewed the classification schemes used by the Centers for Disease Control and Prevention's National Violent Death Reporting System (NVDRS). NVDRS is a state-based system that collects data from multiple sources (eg, death certificates, coroner and medical examiner reports, law enforcement, and toxicology reports) to provide context for violence-related deaths, including suicide [40]. The NVDRS collects information about many risk factors for suicide, such as preceding health and mental health problems, as well as social and environmental factors associated with suicide. In addition, the research team drew on other work in suicide and social media [13,41], the relevant literature on risk and protective factors [40,42], and their experiences engaging with online mental health communities to create an initial version of the scheme.

This annotation scheme included questions related to suicidality and public health. Each item contained an overview defining and clarifying the item and excluding other categories. For example, the risk factor "crisis in past 2 weeks or upcoming 2 weeks" was paired with the following text for annotators: "Direct language that the event caused or contributed to the suicidal ideation or behavior is not required to code 'yes'. Use judgment to determine the time frame. Variable may overlap with other categories (eg, house foreclosure, court date for criminal offense)."

In addition to risk and protective factors, we also captured supplementary information useful in contextualizing risk and protective factors and complementing the use of this survey by stakeholders in suicide prevention (eg, national public health authorities, web-based moderators, and supportive others). The literature points to discussion or intentions with methods of harm as a key part of assessing intention and risk; therefore, we developed an item related to potential methods that an individual may discuss. We also included demographic information volunteered by the poster in r/SW, including self-stated gender and age, as well as whether the poster states that they are in the United States. See Multimedia Appendix 1 for the extended items, definitions, and clarifications of the annotators.

Initial Piloting and Adjustments

A total of 2 members of the team then independently piloted the scheme on 25 random posts drawn from the candidate data set. A total of 2 undergraduate research assistants also piloted the scheme for clarity and interpretability. The 2 members of the research team reported taking 30-45 minutes on the task, and the undergraduates took longer, between 45 minutes and an hour. All took detailed notes on their experiences; then, the team discussed their findings to come to a consensus and refined the scheme based on content:

- Assumptions around depressed mood: there was substantial conversation around annotating if the poster had "depressed mood and mental health problems." The nature of posting in a suicide crisis forum would be inferred to indicate the presence of suicidal thoughts or considerations and some common mental health conditions, such as depression. On the basis of the pilot and to increase precision and sensitivity to identifying mental health conditions that may be contributing to suicide risk, the initial risk factor of "depressed mood and mental health" was refocused to

include only explicit mentions of mental health diagnosis and symptoms other than suicidality.
- Write-in risk and protective factors: the annotation task presented situations where the annotators found risk or protective factors unaccounted for in the categories, though not prevalent enough to warrant a separate category (such as care for family members and dependents). For these, we added a write-in option to the risk and protective factor questions.

During this iteration phase, we also refined the scheme for practical concerns with labeling:

- *Unable to determine* signifier: for demographic questions, we added an option of *cannot tell/not indicated* to assist annotators in indicating their confidence that not enough information was provided to assess the poster's gender, age category, or possible location.
- Removed *any risk or protective factor present* category: this category was duplicated with other labels in the other categories, and the annotators did not feel it was useful to potential future efforts to connect factors to suicide interventions.

Phase 2: Formal Testing and Refinement, Initial Evaluations

After the initial version of the scheme was piloted, 2 team members annotated three rounds of posts randomly sampled without replacement from our candidate data set. Each time, they annotated 20-25 posts and then began checking for internal agreement. Between each round, all researchers met to clarify inconsistencies and better separate categories. Inconsistencies often involved discussing a single post's annotations or how to finesse the descriptions and definitions of items to strengthen consensus.

Interrater Agreement and Reliability Measurements

To evaluate internal validity, we selected 20 posts that were independently annotated by 2 raters on the team with experience in public health and mental health. For categories that were borderline on good-to-strong agreement, we supplemented those with an additional 20 posts for annotation. To quantitatively evaluate the agreement for each subitem or question, we used Gwet AC-1 over Cohen κ or raw percentage agreement. Although Cohen κ is frequently used for interrater reliability evaluations [43], Cohen κ does not adjust for rare category representations within data sets [44,45]. Gwet AC-1 manages rare or infrequent events better than Cohen κ and avoids the pitfalls of large class sizes when evaluating straightforward percentage agreements. In the final version, we saw strong agreement (Gwet AC-1>0.6) across all but one item (*explicit statement of mental health symptoms or diagnosis other than suicidality*).

Final Ratings for 200 Posts and Exploratory Factor Analysis

After establishing interrater agreement and consistency for evaluation, the 2 annotators rated 100 posts each. They rated these items independently, and we counted their annotations together for a total of 200 posts. The expert raters reported that this took between 1 and 3 minutes per post, depending on the post's length. The results from this analysis are presented in Table 2.

Table 2. Independent annotations of 200 posts (N=200).

Annotation category	Values, n (%)
Risk factors (included at least one)	164 (82)
Crisis in past 2 weeks or upcoming 2 weeks	11 (5.5)
Social or relationship problem	102 (51)
Finance or job problem	48 (24)
Physical health problem	18 (9)
Alcohol dependence	10 (5)
Other substance use problem	11 (5.5)
Legal problem	4 (2)
School- or academic-related problem	26 (13)
Death of a friend or family member	7 (3.5)
Explicit statement of mental health symptoms or diagnosis other than suicidality	98 (49)
History of abuse or witnessing violence in childhood	17 (8.5)
Protective factors (included at least one)	128 (64)
Positive social support presence in life	92 (46)
Desire to get better or feel better	59 (29.5)
Lack of means to harm self (perceived or actual)	8 (4)
Engagement in activities	15 (7.5)
Sense of purpose or hope	9 (4.5)
Access to health or mental health care	33 (16.5)
Gender	86 (43)
Male	48 (24)
Female	29 (14.5)
Transgender	9 (4.5)
Cannot tell or not indicated	114 (57)
Age (years)	101 (50.5)
High school or younger (<18)	24 (12)
College (18-22)	21 (10.5)
Postcollege (23-29)	14 (7)
Young adult unspecified (any age<30)	30 (15)
Adult (>30)	12 (6)
Cannot tell or not indicated	99 (49.5)
Mechanism	62 (31)
Firearm	12 (6)
Suffocation, hanging, or strangulation	11 (5.5)
Poisoning	23 (11.5)
Harm using sharp instruments or cutting	23 (11.5)
Fire or burns	0 (0)
Fall	9 (4.5)
Drowning	3 (1.5)
Motor vehicle or train accident	7 (3.5)
Post from inside the United States	189 (94.5)
Yes	175 (87.5)

Annotation category	Values, n (%)
No	14 (7)
Cannot tell or not indicated	11 (5.5)

Finally, we conducted a correlational analysis using tetrachoric correlations and exploratory factor analysis (EFA) to determine the relationships between individual protective and risk factors. EFA is a technique commonly used in scale development and psychometrics to evaluate whether any variables in a scheme, survey, or instrument are correlated such that they may be explained by unobservable or underlying variables called *factors* [46]. This allows us to inspect for potential overlap with correlations and how the schema may be reduced in future work. For EFA, we separated risk and protective factors, as these items were developed distinct from each other and had the most potential for common concepts and underlying factors. We conducted EFA using minimum residuals on the tetrachoric correlational matrix, and we reported the results using parallel analysis [46]. These are available in the psych package in R (R Core Team).

Results

Agreement Between 2 Raters

The final scheme items and their agreement scores are listed in Table 3. We saw strong results that indicated our annotation scheme was consistent between the 2 raters.

Table 3. Interrater reliability between raters in the final data set.

Question	Agreement, %	AC-1
Risk factors		
Crisis in past 2 weeks or upcoming 2 weeks	90	0.87
Social or relationship problem	81	0.64
Finance or job problem	90	0.85
Physical health problem	95	0.94
Alcohol dependence	100	1
Other substance use problem	98	0.97
Legal problem	100	1
School- or academic-related problem	90	0.86
Death of a friend or family member	100	1
Explicit statement of mental health symptoms or diagnosis other than suicidality[a]	78	0.57
History of abuse or witnessing violence in childhood	100	1
Protective factors		
Positive social support present in life[a]	90	0.81
Desire to get better or feel better[a]	80	0.68
Lack of means to harm self (perceived or actual)	95	0.95
Engagement in activities	93	0.92
Sense of purpose or hope	88	0.86
Access to physical or mental health care	90	0.86
Gender		
Male	93	0.89
Female	100	1
Transgender	100	1
Cannot tell or not indicated	90	0.81
Age (years)		
High school or younger (<18)	98	0.97
College (18-22)	98	0.97
Postcollege (23-29)	98	0.97
Young adult unspecified (any age<30)	88	0.83
Adult (>30)	100	1
Cannot tell or not indicated	85	0.72
Method		
Firearm	98	0.97
Suffocation, hanging, or strangulation	nan[b]	nan
Poisoning	98	0.97
Sharp instrument or cutting	100	1
Fire or burns	nan	nan
Fall	98	0.97
Drowning	100	1
Motor vehicle or train	nan	nan
Post in the United States[a]		

Question	Agreement, %	AC-1
Yes	80	0.71
No	95	0.94
Cannot tell or not indicated	85	0.81

[a]Indicates where we added 20 posts in the final rating for disambiguating challenging categories.

[b]Indicates that it was not present in the annotations.

Results of Annotation for 200 Posts

In Table 2, we present the independent annotations of 200 posts. We provide quotes for the context that has been edited to protect participants' identities.

Out of 200 posts, about 164 (82%) posts included at least one risk factor, and 128 (64%) posts included at least one protective factor. Related to this, 31% (62/200) described a possible mechanism for a current or past attempt at suicide. Of all 200 posts, 189 (94.5%) had either a protective or risk factor, leaving only 11 (5.5%) posts that did not. We manually inspected these 11 posts, and these tended to be very short posts with little information about the person's unique circumstances (eg, the entire post was "that calm when you finally decided–yea i'm gonna do it"). We noted that very short posts were difficult for annotators because of their length, and this limited the ability to apply this scheme to them. However, most posts were rich enough for annotation by the data set and indicated that community members were willing to disclose suicide risk and protective factors.

Analysis of Risk and Protective Factors

In Tables 4 and 5, we present histograms of the count of risk and protective factors by post. This shows that many posters have more than one risk or protective factor that they mention, indicating multiple avenues for support that may not have been captured through the evaluation of risk in a binary classification system.

Table 4. Risk factors per post.

Number of risk factors present in post	Count in data set
0	36
1	56
2	49
3	35
4	14
5	8
6	2
7	0

Table 5. Protective factors per post.

Number of protective factors present in post	Count in data set
0	72
1	65
2	41
3	16
4	6
5	0

The most prevalent risk factors were social or relationship problems (102/200, 51%), mental health symptoms (98/200, 49%), financial or job problems (48/200, 24%), school or academic problems (26/200, 13%), and physical health problems (18/200, 9%). We noted that over half of all posts mentioned social and relationship problems in their posts. These included trouble with family members ("I can't stay with my family for another 10 months"), breakups ("I miss my ex so much, but he doesn't care about me and has forgotten me with his new girlfriend"), and the absence of relationships and friends ("I can't really say I've had a friend in the last 5 years"). Combinations of these factors also included navigating the devastating effects that mental health symptoms have on relationships and friendships ("I've tried to hide my depression from my friends for years, but my best friend is so exhausted dealing with me. She must know by this point..."). For mental health symptoms, many noted that their symptoms recurred or were not well treated or that their relationship with their therapist

or medical professional was not supportive. Write-in examples for risk factors were varied and included circumstances such as cutting and self-injury and losing access to technology.

The most prevalent protective factors were positive social support (92/200, 46%), desire to get or feel better (59/200, 29.5%), access to health or mental health care (33/200, 16.5%), and engagement in activities (15/200, 7.5%). Examples of posts that indicated positive social support included support from family and friends ("A friend recommended me for a job") and worries about disappointing supportive people in their lives ("I'm so scared because my mom couldn't handle it if I weren't here, but I'm so miserable."). Those who desired to get or feel better described both the active desire to get better ("I want to believe that I'll get through these feelings") and the negative desire to not die ("I really don't want to die, but I can't keep living like this either"). In addition to these categories, our annotators also identified other protective factors mentioned by the posters, such as how their religion discouraged suicide as a solution.

Analysis of Demographic and Methods Factors

Only 43% (86/200) of posts included a discernable indication of their gender. Of the posters who mentioned their gender, 56% (48/86) were male. We noticed that gender was mentioned in the post body (eg, "hi I'm 26/M and struggling"), in the context of risk or protective factors ("there's no way I'm better than the other men she loves"), or in the mental health struggles they were currently encountering ("my gender dysphoria is very bad tonight, please help"). Of the 200 posts, 9 (4.5%) had people who identified as transgender, gender fluid, or nonbinary identities. These individuals often described being closeted for their true gender or frustrations around being misgendered.

As for age, approximately half of the posts (101/200, 50.5%) indicated the person's age group. The largest age group on the forum was young adults aged >30 years, with many of them being in college (21/200, 10.5%), and some being in high school (24/200, 12%). Many posts mentioned age in passing, with no connection to circumstances surrounding their ideation ("19yo—please help"). However, some posts often related to age as a factor for both risk ("I'm 58 and I've wasted my whole life") and protective ("I know I'll graduate [college] soon, and then it'll be easier") factors.

About 31% (62/200) of total posts described a possible mechanism for a current or past attempt at suicide. For posts that include a mechanism, these posts mention only one, and those in descending order are poisoning (23/200, 11.5%), sharp instruments and cutting (23/200, 11.5%), and self-injury via firearms (12/200, 6%). Some of these posts mention it in the context of past attempts ("I tried to drown myself") or in present or future possibilities.

Finally, we examined whether posts indicated that they were in the United States. We found that 87.5% (175/200) of posts were inferred to be from posters in the United States. As we removed location details from the posts, this category was generated mostly from inferences about context. This included current details ("I make about 30k a year") or past history of the participant ("we moved a lot between k-12"). Other contextual details were indicative for people not in the United States, such as the context of learning a new language after moving to a new and small country, or other personal indicators ("I only weigh 8 stone").

Item Correlations and Factor Analysis on 200 Posts' Ratings

First, we present the correlations of the variables in Figures 2 and 3 using tetrachoric correlations. Tetrachoric correlations are useful for measuring the strength of correlations between binary or dichotomous data. Colored or shaded cells indicate significance at the $P<.01$ level after applying the Benjamini-Yekutieli correction to account for false discovery rate.

Many variables show correlations with other risk and protective factors. Some correlations are relatively strong, such as the correlation between legal and financial concerns ($r=0.53$) and abuse and social factors ($r=0.54$). The correlations of the factors themselves are not surprising based on prior work on suicide prevention, as many factors are independently related to each other [47]. For instance, research has shown the impact of abuse and violence in childhood, commonly reframed from adverse childhood experiences, and their connections to negative outcomes in adulthood [48], such as alcohol ($r=0.27$) and substance abuse ($r=0.4$). We saw similar correlational strengths in the protective factors. There is a very strong correlation between the complementary factors of having a sense of purpose in one's life and the ability to feel better ($r=0.70$). We noted the distinctiveness of the lack of means to harm oneself, which showed no significant correlations with any other factors. We hypothesized that this factor might be distinctive from the others, and future work should explore the independence of this factor.

Figure 2. Tetrachoric correlations between risk factors.

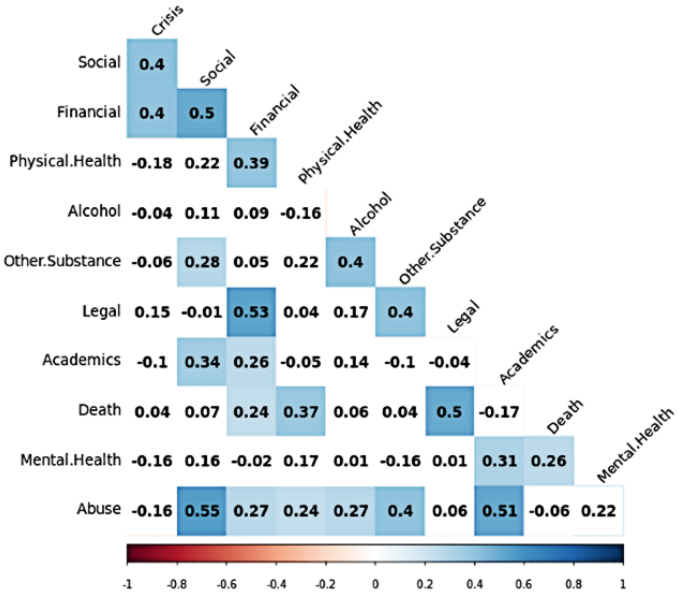
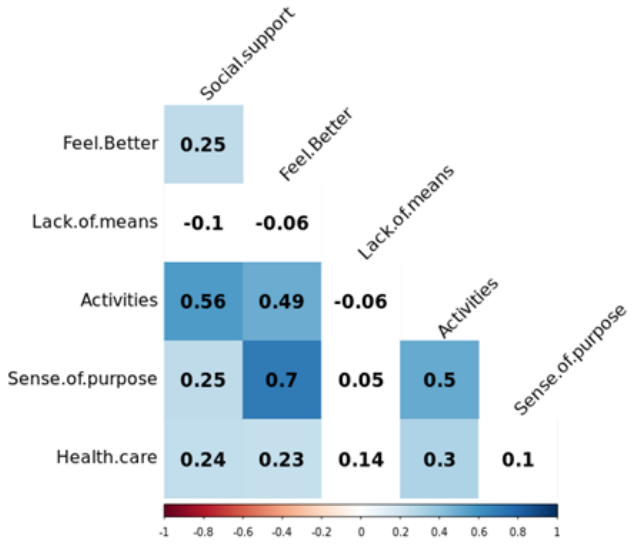


Figure 3. Tetrachoric correlations between protective factors.



Next, we present the EFA for our analysis, relying on the tetrachoric correlations we used earlier. We opted for parallel analysis rather than the scree plot of the eigenvalues as there was no distinctive *elbow* for risk factors, a common signal for effective interpretations of scree plots, as shown in Figures 4 and 5 for risk and protective factors, respectively. Parallel analysis is a complementary evaluation technique to scree plots that use simulated data to evaluate factor reduction [46]. Parallel analysis pointed to five unobserved factors for risk and 3 unobserved factors for protection. Both models have a reasonable percentage of variance explained—73% of variance explained for a reduced model of five risk factors and 57% of variance explained for a reduced model of three protective factors (we expect between 60% and 70% of variance explained, per DeVellis [46]). This aligns with our conceptual model that there is distinctiveness among the socioecological factors proposed in prior work.

Figure 4. Scree plot and parallel analysis for risk factors.

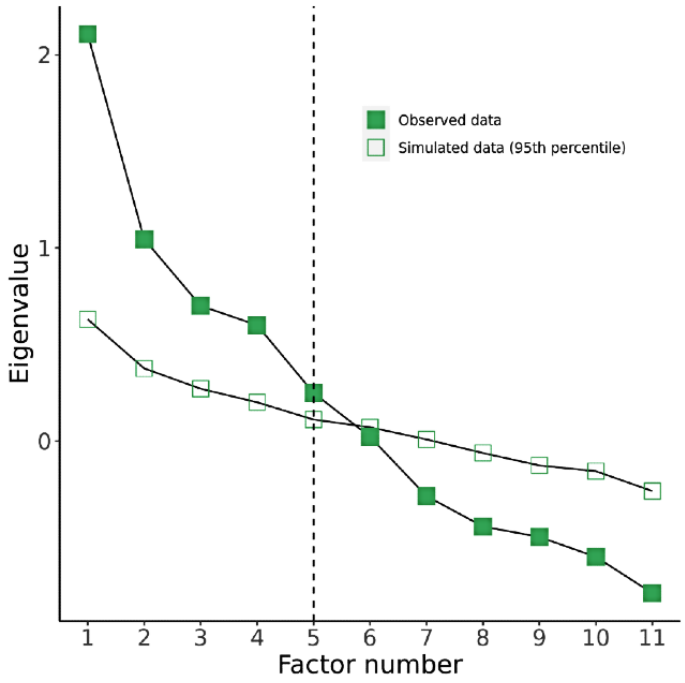
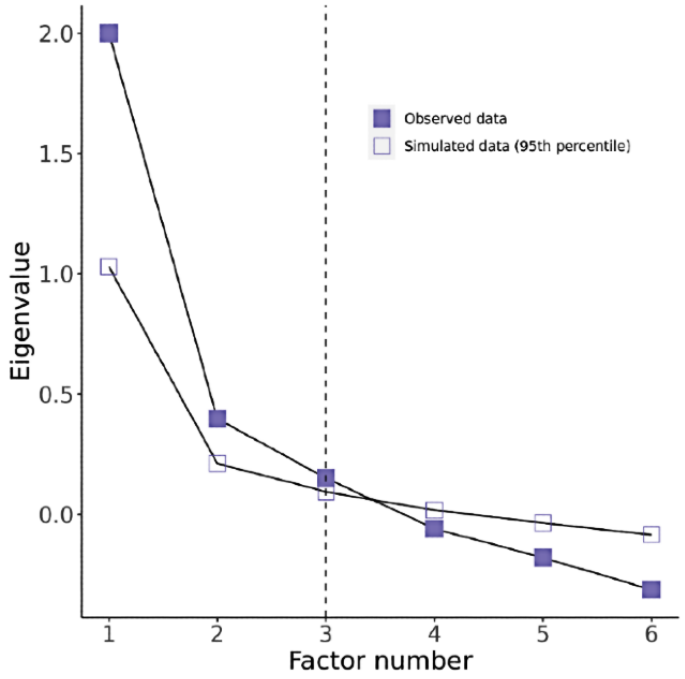


Figure 5. Scree plot and parallel analysis for protective factors.



Discussion

Principal Findings

Our work responds to recent calls within computer science to better operationalize concepts for social media analysis [25-27]. We did so through the synthesis of procedures, vocabulary, and perspectives of measurement and validity literature in linguistics [28] and psychometrics [30,31]. We have provided a more formalized, step-by-step approach to generate annotations for social media analysis for mental health.

This work affects public health and computational research by better conceptualizing risk and protective factors that influence suicidality. Leading public health data about suicidality are largely generated from suicide decedents, and information on precise circumstances and precipitants influencing individuals at risk of suicide is lacking. Large-scale information on suicide ideation is valuable as it provides information on a time point that is upstream of significant morbidity. This may help improve public health programs to prevent suicide, such as programs to enhance protective factors such as social connectedness [49].

Furthermore, this research points to improvements in the computational analysis of social media data for mental health and suicide. With more testing and a larger volume of samples, we envision that our annotation scheme can be used in semiautomated machine learning systems that screen natural

language for mention of risk and protective factors. These systems could eventually direct support and resources to those with more urgent risk or with complementary protective factors that may assist in mitigating crisis (such as shared experiences and social connections) [1].

Comparison With Previous Work

Social media data have been a fruitful data source for research on suicide. Many studies have attempted to distinguish whether an individual is suicidal [14,41,50] or may attempt suicide in the future [51]. Early research by Coppersmith et al [51] used the disclosure of a past suicide attempt to understand the pathology of risk. Another study identified 6 categories of suicide-related disclosures, such as public awareness campaigns and memorial campaigns alongside legitimate disclosures [41]. De Choudhury et al [5] studied shifts in Reddit to suicidal ideation from other mental health subreddits, and Kavuluru et al [38] designed a classification approach to detect helpful comments on r/SuicideWatch [37].

The dominant model for understanding suicide on social media has focused on risk, operationalized into categories of low, medium, or high. In early work, Homan et al [14] adopted a manual annotation process to verify the veracity and trustworthiness for clinical diagnosis of a mental disorder that is then fed to language models to understand distress and suicide risk. These annotation surveys and schemes are in service of annotation for automated or artificial intelligence systems. O'Dea et al [12] used Twitter data to develop an annotation scheme for mixed-expertise coders to annotate among *strongly concerning*, *possibly concerning*, or *safe to ignore*, which is then used for prediction. Building off shared tasks in natural language processing [52,53], both Milne et al [54] and Cohan et al [7] designed a survey that annotated with the *stoplight* system of green-amber-red-crisis, which was then fed to machine learning models to improve moderator responsiveness on ReachOut. Closest to our study, Shing et al [13] developed an ordinal risk assessment annotation with four categories for Reddit suicide crisis data, with mental health experts as annotators compared with the crowd.

An active area of research innovates in strategies for assessing mental health signals and generating labels. Some studies used trained medical professionals to generate labels [12,13]. Although this approach is promising because of its direct connection to everyday clinical practice, scaling this specialized skill to the number of posts needed for stable social media analysis is burdensome for clinicians with outside responsibilities and busy schedules [24]. Researchers have developed scalable labeling strategies for use by people other than clinicians, which have been called *proxy signals* in prior work [24,33]. Although these approaches aim to distribute labor, recent work has called into question whether these proxy signals measure what they claim to measure. Ernala et al [24] empirically compared the outcomes of proxy signals derived from prior work and found that computational models had poor external validity on verified patient data for patients with schizophrenia. Similarly, Chancellor and De Choudhury [23] found that there has been little research evaluating clinical constructs in social media data.

Together, this research points to gaps in the approaches to the annotation of suicidal behaviors. As social media data do not, by default, include clinically validated labels of suicide, processes relying on these signals must be replicable and reliable. Our study responds to and makes the first attempt at reconciling these criticisms in a labeling task designed to annotate risk and protective factors in social media data.

Considerations for Developing New Schema for Annotation

We believe our approach can be extended to other cases where teams need high-quality annotations from social media data in mental health and beyond. In this section, we provide an overview of the considerations and guiding questions to adopt this framework in new schema development.

Problem Framing and Domain Expertise

Problem framing is both the origin and evaluation point for research and practice and is a core component of construct validity. Although computer scientists are experts in technical methods and social media, they do not carry the same background, intuition, and framing expertise as psychiatrists and psychologists, researchers in medicine and psychology, social workers, or other experts. The right set of domain experts can make it clear how to instantiate certain concepts in surveys and adjust and evaluate concepts to align with notions of construct validity. Do our definitions of illness or behavior hold up to appropriate disciplinary scrutiny [23]? What, specifically, is the exact problem to evaluate? We strongly encourage working with experts in mental health as a de facto standard in work that bridges mental illness and computer science to assist with questions of construct validity.

Source of Social Media Data

In addition to problem framing, the social media data source will need to be evaluated for its capacity to provide insight into a question. Different platforms and affordances, subcommunities, and normative practices may lend themselves to answering certain kinds of questions about mental health and human behavior. Can social media data from a specific platform answer the question that the team wants to solve, or do modifications need to be made to the community data source, platform, or questions being asked?

Automated and Deliberate Filtering

Social media data are almost always processed, filtered, or curated by both the platform and the research team. Data gathering techniques may be altered by the platform, preventing the curation of a truly random sample (Twitter data streams typically provide 1%-5% of all data), and research teams may choose to remove posts that do not meet certain objective criteria for length or language patterns. For instance, we chose to remove very short and very long posts from our annotations from expert requests. What are the impacts of different kinds of filtering on the generalizability of the findings or schema?

The Tradeoff Between Complexity and Validity

A crucial balancing act will come between the complexity of the schema and the schema's validity. We anticipate that a schema that has the highest levels of construct validity will also

be time consuming given the schema's length. Simplification may be necessary to reduce time and resource costs. The same simplification can quickly become a dangerous abstraction that loses the ability to evaluate the original concept, a shadow of its original concepts, and may lead to erroneous conclusions. However, the simplification of a schema to core factors is a subfield of psychology. Methods such as EFA can assist in the process of robustly condensing schema, but this requires the appropriate use of new methods. What are the most robust strategies for managing complexity and validity?

Who Does the Labeling?

In addition to the development of schema, the actual people and groups that label social media data are just as important as the development of the schema itself. Domain experts are ideally the best to label posts; however, their time is valuable and constrained. In computer science, researchers have turned to nondomain experts and crowd workers on sites such as Mechanical Turk to quickly label large batches of data [13].

Although crowd workers have been adopted for tasks such as image labeling, their utility in subjective tasks such as mental health evaluations is a nascent area of study. Most studies use simplified versions of risk evaluations, such as the promising work of Shing et al [13] in suicide risk assessment, and do not use an expanded schema similar to ours. Who are the best sources of labels for robust schema, and how do they diverge from more accurate assessments? What thresholds of accuracy from different groups are appropriate for evaluating social media data?

Resource Management

In ideal production scenarios, there would be plenty of time to develop a robust schema and test and label thousands of posts. However, this is infeasible in practice, especially in professional environments where the time of experts is limited, and costs may drive decision-making. What resource tradeoffs are appropriate for maintaining quality standards? How many high-quality annotations can be generated, and does this ensure that a model or finding is robust? What emerging ethical and moral questions arise from the tradeoffs required for resource management?

Software for Annotation Schemas

There are many ways to format and deploy a schema through software that may have secondary impacts on time to completion, perceived complexity, and accuracy. Work in the field of human-computer interaction and crowd working considers how technical design tradeoffs may affect these variables. For example, our raters found Google Forms burdensome in prepiloting; hence, we abandoned it. We encourage mindfulness of these methods and concerns to avoid unforeseen interactions between tools and technology with the annotation schema itself.

Ethics and Privacy Considerations

We believe that all researchers have obligations to protect individuals in their data sets from harm, no matter the source of the data or protections or exemptions from ethics boards [33]. We followed emerging professional and research norms of care for social media data in sensitive contexts [33,39]. However, careful research protections do not inherently guarantee that the participants will not be reidentified, that the research process is human-centered, or that the implications will generate just outcomes for individuals whose data are analyzed [36]. Tensions in scientific reproducibility, moral imperatives for intervention, professional ethics obligations, and other factors emerge when dealing with challenging areas such as suicide prevention [15,33].

One tension in development is the balance between the inclusion of correct gender identities and risks of harm from a small sample size. Understanding a person's gender identity is an important facet of suicide prevention, as suicidal ideation disproportionately affects LGBTQ+ individuals [55]. We considered including more inclusive gender categories recommended by experts [56], including nonbinary and genderfluid identities. On manual inspection of the data set, there were very few individuals who self-described as nonbinary or genderfluid—<2% of all posts (2-3 posts of 200). We worried that our research could have a spotlighting effect on their behavior if we chose to isolate these gender categories and present comparisons of these individuals because of the small data set size. This attention risks harming individuals who may already be vulnerable for reasons related to their gender identity and poor social support from others. Therefore, we opted to bundle the identity categories together and label individuals who were genderqueer, genderfluid, and transgender as one category.

Limitations and Future Work

The primary limitation of this approach is that external validation of this annotation scheme is needed. This includes robust confirmatory factor analysis, reduction of factors into more generalized concepts, and a scale evaluation and deployment with new raters and a new data set. This facet of generalizability is important for benchmarking the performance of downstream applications. Our study focuses on the critical first step of establishing construct validity in annotation development, and our immediate and future work will focus on demonstrating external validity with new annotators and communities discussing suicide crises.

Other data concerns that might limit generalizability are connected to this concern. Although it is anchored in broader concepts for suicide risk and protective factors, this scheme was developed for the unique context of Reddit suicide crisis posts. Reddit demographics will skew toward younger audiences, may be biased for US contexts, and may miss crucial demographics at risk for suicide. We also expect explicit requests for assistance through a suicide crisis to influence our annotation schema and not translate as well to more subtle disclosures. We do not yet have a sufficient sample size to present a generalizable analysis of differences in risk and protective factors differing between different demographic factors or across cultural differences in expressions of mental illness. Future work will additionally need to extend and verify this scheme on new communities and platforms such as mental illness subreddits more broadly (eg, r/depression, r/selfharm, and r/madeofstyrofoam) or general-purpose social media sites.

However, these limitations are opportunities for future research. As a crucial next step, external validity should be added to this pipeline to develop our approach. New individuals could be involved as raters, including novices and computer scientists with domain expertise but no medical training, crowd workers, and other experts and stakeholders in this space. This would complement previous studies by comparing experts and nonexperts [13,14]. In addition, we could consider validating these data against other sources of information about suicide, such as public health data sets or alternate social media sites and communities.

Conclusions

We report the development and first validation of an annotation scheme that evaluates risk and protective factors in suicide crisis forums on Reddit. By using the socioecological model of suicide prevention, our approach expanded state-of-the-art processes by moving beyond categorical or ordinal scales to evaluate only risk. Moreover, by adding protective factors to the scheme, we provided key insights into behavior that better represents the constellation of support needs for suicide prevention. Aligning with the metrics of construct validity, we demonstrated strong substantive and structural agreement among the research team. By explicating our processes, logic, and decision-making, we not only hope to enable replicability in social media annotation of suicide but also raise awareness for annotation development.

Acknowledgments

SC and MDC were supported in part by National Institutes of Health grant #R01GM112697-01. SC completed much of this work while she was at the Georgia Institute of Technology and Northwestern University. The authors would like to thank Michael Xiao for the early piloting of the project, Harman Kaur and Jacob Thebault-Spieker for feedback on early drafts, and the University of Minnesota Institute for Research in Statistics and its Applications for assistance with factor analysis and correlations.

The findings and conclusions in this paper are those of the authors and do not necessarily represent the official position of the US Centers for Disease Control and Prevention.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Extended details of the annotation schema developed for this study.
[DOCX File , 19 KB - mental_v8i11e24471_app1.docx]

References

1. Stone DM, Simon TR, Fowler KA, Kegler SR, Yuan K, Holland KM, et al. Vital signs: trends in state suicide rates - United States, 1999-2016 and circumstances contributing to suicide - 27 states, 2015. MMWR Morb Mortal Wkly Rep 2018 Jun 08;67(22):617-624 [FREE Full text] [doi: 10.15585/mmwr.mm6722a1] [Medline: 29879094]
2. WISQARS™ — Web-based Injury Statistics Query and Reporting System. Centers for Disease Control and Prevention, National Center for Injury Prevention and Control. 2020. URL: http://www.cdc.gov/injury/wisqars/ [accessed 2021-09-02]
3. Stone D, Holland K, Bartholow B, Crosby A, Davis S, Wilkins N. Preventing suicide: a technical package of policies, programs, and practices. Suicide Prevention Resource Center.: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention; 2017. URL: https://www.sprc.org/resources-programs/preventing-suicide-technical-package-policies-programs-practices [accessed 2021-09-02]
4. 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]
5. De Choudhury M, Kıcıman E. The language of social support in social media and its effect on suicidal ideation risk. Proc Int AAAI Conf Weblogs Soc Media 2017;2017:32-41. [Medline: 28840079]
6. 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:85-94 [FREE Full text] [doi: 10.1145/2700171.2791026] [Medline: 28713876]
7. Cohan A, Young S, Yates A, Goharian N. Triaging content severity in online mental health forums. J Asso Inform Sci Technol 2017;68(11):2675-2689. [doi: 10.1002/asi.23865]
8. Soldaini L, Walsh T, Cohan A, Han J, Goharian N. Helping or hurting? Predicting changes in users? Risk of self-harm through online community interactions. In: Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic. 2018 Presented at: Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic; June 2018; New Orleans, LA p. 194-203. [doi: 10.18653/v1/w18-0621]

9. Yates A, Cohan A, Goharian N. Depression and self-harm risk assessment in online forums. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2018 Presented at: Conference on Empirical Methods in Natural Language Processing; September 2017; Copenhagen Denmark p. 2968-2978. [doi: 10.18653/v1/d17-1322]
10. Sharma A, Choudhury M, Althoff T, Sharma A. Engagement patterns of peer-to-peer interactions on mental health platforms. arXiv. 2020. URL: https://arxiv.org/pdf/2004.04999.pdf [accessed 2021-09-02]
11. O'Dea B, Larsen ME, Batterham PJ, Calear AL, Christensen H. A linguistic analysis of suicide-related twitter posts. Crisis 2017;38(5):319-329. [doi: 10.1027/0227-5910/a000443] [Medline: 28228065]
12. O'Dea B, Wan S, Batterham PJ, Calear A, Paris C, Christensen H. Detecting suicidality on Twitter. Internet Interven 2015 May;2(2):183-188 [FREE Full text] [doi: 10.1016/j.invent.2015.03.005]
13. Shing HC, Nair S, Zirikly A, Friedenberg M, Daumé III H, Resnik P. Expert, crowdsourced, and machine assessment of suicide risk via online postings. In: Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic. 2018 Presented at: Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic; June, 2018; New Orleans, LA p. 25-36 URL: https://aclanthology.org/W18-0603/
14. Homan C, Johar R, Liu T, Lytle M, Silenzio V, Alm CO. Toward macro-insights for suicide prevention: analyzing fine-grained distress at scale. In: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality.: Association for Computational Linguistics; 2014 Presented at: Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; June, 2014; Baltimore, Maryland, USA p. 107-117 URL: https://aclanthology.org/W14-3213/
15. Coppersmith G, Leary R, Crutchley P, Fine A. Natural language processing of social media as screening for suicide risk. Biomed Inform Insights 2018;10:1178222618792860 [FREE Full text] [doi: 10.1177/1178222618792860] [Medline: 30158822]
16. Gaur M, Alambo A, Sain JP, Kursuncu U, Thirunarayan K, Kavuluru R, et al. Knowledge-aware assessment of severity of suicide risk for early intervention. In: Proceedings of the World Wide Web Conference. 2019 Presented at: The World Wide Web Conference; May 13 - 17, 2019; San Francisco CA USA p. 514-525. [doi: 10.1145/3308558.3313698]
17. Nobles AL, Glenn JJ, Kowsari K, Teachman BA, Barnes LE. Identification of imminent suicide risk among young adults using text messages. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18: Conference on Human Factors in Computing Systems; April 21 - 26, 2018; Montreal QC Canada p. 1-11. [doi: 10.1145/3173574.3173987]
18. Risk and protective factors. Centers for Disease Control and Prevention, National Center for Injury Prevention and Control. 2020. URL: https://www.cdc.gov/violenceprevention/suicide/riskprotectivefactors.html [accessed 2021-09-02]
19. Rutter PA, Freedenthal S, Osman A. Assessing protection from suicidal risk: psychometric properties of the suicide resilience inventory. Death Stud 2008;32(2):142-153. [doi: 10.1080/07481180701801295] [Medline: 18693384]
20. Linehan MM, Goodstein JL, Nielsen SL, Chiles JA. Reasons for staying alive when you are thinking of killing yourself: The Reasons for Living Inventory. J Consult Clin Psychol 1983;51(2):276-286. [doi: 10.1037/0022-006x.51.2.276]
21. Franklin JC, Ribeiro JD, Fox KR, Bentley KH, Kleiman EM, Huang X, et al. Risk factors for suicidal thoughts and behaviors: a meta-analysis of 50 years of research. Psychol Bull 2017 Feb;143(2):187-232. [doi: 10.1037/bul0000084] [Medline: 27841450]
22. Ribeiro JD, Franklin JC, Fox KR, Bentley KH, Kleiman EM, Chang BP, et al. Self-injurious thoughts and behaviors as risk factors for future suicide ideation, attempts, and death: a meta-analysis of longitudinal studies. Psychol Med 2015;46(2):225-236 [FREE Full text] [doi: 10.1017/s0033291715001804]
23. Chancellor S, De Choudhury M. Methods in predictive techniques for mental health status on social media: a critical review. NPJ Digit Med 2020;3:43. [doi: 10.1038/s41746-020-0233-7] [Medline: 32219184]
24. Ernala SK, Birnbaum ML, Candan KA, Rizvi AF, Sterling WA, Kane JM, et al. Methodological gaps in predicting mental health states from social media: triangulating diagnostic signals. New York: Association for Computing Machinery; 2019 Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4 - 9, 2019; Glasgow Scotland p. 1-16. [doi: 10.1145/3290605.3300364]
25. Levonian Z, Erikson DR, Luo W, Narayanan S, Rubya S, Vachher P, et al. Bridging qualitative and quantitative methods for user modeling: tracing cancer patient behavior in an online health community. In: Proceedings of the International AAAI Conference on Web and Social Media. 2020 Presented at: Fourteenth International AAAI Conference on Web and Social Media; Jun 8-11, 2019; Online URL: https://ojs.aaai.org/index.php/ICWSM/article/view/7310/7164
26. Geiger RS, Yu K, Yang Y, Dai M, Qiu J, Tang R, et al. Garbage in, garbage out?: do machine learning application papers in social computing report where human-labeled training data comes from? In: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. 2020 Presented at: Conference on Fairness, Accountability, and Transparency; Jan 27 - 30, 2020; Barcelona Spain p. 325-326. [doi: 10.1145/3351095.3372862]
27. Jacobs AZ, Wallach H. Measurement and fairness. In: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. 2021 Mar Presented at: FAccT '21: 2021 ACM Conference on Fairness, Accountability, and Transparency; March 3 - 10, 2021; Virtual Event Canada p. 375-385. [doi: 10.1145/3442188.3445901]
28. Hovy E, Lavid J. Towards a ‘science’ of corpus annotation: a new methodological challenge for corpus. Int J Translation 2010;22(1):13-36 [FREE Full text]

29. Snow R, O'Connor B, Jurafsky D, Ng AY. Cheap and fast---but is it good?: evaluating non-expert annotations for natural language tasks. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. 2008 Presented at: EMNLP '08: Conference on Empirical Methods in Natural Language Processing; Oct 25 - 27, 2008; Honolulu Hawaii p. 254-263. [doi: 10.3115/1613715.1613751]
30. Simms LJ. Classical and modern methods of psychological scale construction. Soc Personal Psychol Compass 2008;2:414-433 [FREE Full text] [doi: 10.1111/j.1751-9004.2007.00044.x]
31. Simms LJ, David W. The construct validation approach to personality scale construction. In: Handbook of Research Methods in Personality Psychology. New York City: The Guilford Press; Dec 01, 2007:240-258.
32. Shadish WR, Cook TD, Campbell DT. Experimental and Designs for Generalized Causal Inference. Boston, US: Houghton Mifflin; 2002.
33. Chancellor S, Birnbaum ML, Caine ED, Silenzio V, Choudhury MD. 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 USA p. 79-88. [doi: 10.1145/3287560.3287587]
34. Bronfenbrenner U. Toward an experimental ecology of human development. Am Psychol 1977;32(7):513-531. [doi: 10.1037/0003-066x.32.7.513]
35. Caine ED. Forging an agenda for suicide prevention in the United States. Am J Public Health 2013 May;103(5):822-829. [doi: 10.2105/AJPH.2012.301078] [Medline: 23488515]
36. Chancellor C, Baumer EP, De Choudhury M. Who is the "Human" in human-centered machine learning: the case of predicting mental health from social media. Proc ACM Hum Comput Interact 2019;3(CSCW):1-32. [doi: 10.1145/3359249]
37. Gaur M, Alambo A, Sain JP, Kursuncu U, Thirunarayan K, Kavuluru R, et al. Knowledge-aware assessment of severity of suicide risk for early intervention. In: Proceedings of The World Wide Web Conference. 2019 Presented at: The World Wide Web Conference; May 13 - 17, 2019; San Francisco CA USA p. 514-525. [doi: 10.1145/3308558.3313698]
38. Kavuluru R, Williams AG, Ramos-Morales M, Haye L, Holaday T, Cerel J. Classification of helpful comments on online suicide watch forums. In: Proceedings of the 7th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics. 2016 Oct Presented at: BCB '16: ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics; October 2 - 5, 2016; Seattle WA USA p. 32-40. [doi: 10.1145/2975167.2975170]
39. Benton A, Coppersmith G, Dredze M. Ethical research protocols for social media health research. In: Proceedings of the First ACL Workshop on Ethics in Natural Language Processing.: Association for Computational Linguistics; 2017 Presented at: First ACL Workshop on Ethics in Natural Language Processing; April, 2017; Valencia, Spain p. 94-102. [doi: 10.18653/v1/W17-1612]
40. World Health Organization. Preventing Preventing Suicide: A Global Imperative. Geneva, Switzerland: World Health Organization; 2014.
41. Burnap P, Colombo W, Scourfield J. Machine classification and analysis of suicide-related communication on Twitter. In: Proceedings of the 26th ACM Conference on Hypertext & Social Media. 2015 Presented at: HT '15: 26th ACM Conference on Hypertext and Social Media; September 1 - 4, 2015; Guzelyurt, Northern Cyprus p. 75-84. [doi: 10.1145/2700171.2791023]
42. Risk and protective factors. Suicide Prevention Resource Center. URL: https://www.sprc.org/about-suicide/risk-protective-factors [accessed 2021-09-02]
43. McDonald N, Schoenebeck S, Forte A. Reliability and inter-rater reliability in qualitative research. Proc ACM Hum-Comput Interact 2019 Nov 07;3(CSCW):1-23. [doi: 10.1145/3359174]
44. Eugenio BD, Glass M. The Kappa statistic: a second look. Computat Linguist 2004 Mar;30(1):95-101 [FREE Full text] [doi: 10.1162/089120104773633402]
45. Hallgren KA. Computing inter-rater reliability for observational data: an overview and tutorial. Tutor Quant Methods Psychol 2012;8(1):23-34 [FREE Full text] [doi: 10.20982/tqmp.08.1.p023] [Medline: 22833776]
46. DeVellis RF. Scale Development: Theory and Applications. Thousand Oaks, California: SAGE Publications Ltd; 2016.
47. Wilkins N, Tsao B, Hertz M, Davis R, Klevens J. Connecting the Dots: An Overview of the Links Among Multiple Forms of Violence. Oakland, CA: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention; 2014.
48. Dube SR, Felitti VJ, Dong M, Chapman DP, Giles WH, Anda RF. Childhood abuse, neglect, and household dysfunction and the risk of illicit drug use: the adverse childhood experiences study. Pediatrics 2003;111(3):564-572 [FREE Full text] [doi: 10.1542/peds.111.3.564]
49. Department of Health and Human Services. 2012 National Strategy for Suicide Prevention: Goals and Objectives for Action: A Report of the U.S. Surgeon General and of the National Action Alliance for Suicide Prevention. Washington (DC): US Department of Health & Human Services (US); 2012.
50. Aladağ AE, Muderrisoglu S, Akbas NB, Zahmacioglu O, Bingol HO. Detecting suicidal ideation on forums: proof-of-concept study. J Med Internet Res 2018;20(6):e215 [FREE Full text] [doi: 10.2196/jmir.9840]
51. Coppersmith G, Ngo K, Leary R, Anthony W. Exploratory analysis of social media prior to a suicide attempt. In: Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology.: Association for Computational Linguistics;

2016 Presented at: Third Workshop on Computational Linguistics and Clinical Psychology; Jun 16, 2016; San Diego US p. 106-117. [doi: 10.18653/v1/w16-0311]
52. Milne DN, Pink G, Hachey B, Calvo RA. CLPsych 2016 shared task: triaging content in online peer-support forums. In: Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology. 2016 Presented at: Third Workshop on Computational Linguistics and Clinical Psychology; June, 2016; San Diego, CA, USA p. 118-127. [doi: 10.18653/v1/w16-0312]
53. Zirikly A, Resnik P, Uzuner O, Hollingshead K. CLPsych 2019 shared task: predicting the degree of suicide risk in Reddit posts. In: Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology.: Association for Computational Linguistics; 2019 Presented at: Sixth Workshop on Computational Linguistics and Clinical Psychology; June, 2019; Minneapolis, Minnesota p. 24-33. [doi: 10.18653/v1/W19-3003]
54. Milne DN, McCabe KL, Calvo RA. Improving moderator responsiveness in online peer support through automated triage. J Med Internet Res 2019;21(4):e11410 [FREE Full text] [doi: 10.2196/11410] [Medline: 31025945]
55. Kann L, McManus T, Harris WA, Shanklin SL, Flint KH, Queen B, et al. Youth risk behavior surveillance - United States, 2017. MMWR Surveill Summ 2018;67(8):1-114. [doi: 10.15585/mmwr.ss6708a1] [Medline: 29902162]
56. HCI guidelines for gender equity and inclusivity. Semantic Scholar. May 21. URL: https://www.semanticscholar.org/paper/HCI-Guidelines-for-Gender-Equity-and-Inclusivity-Scheuerman-Spiel/88344011a8c4fbe3823c1a967fce13859a5cb0bb [accessed 2021-09-02]

Abbreviations

EFA: exploratory factor analysis
NVDRS: National Violent Death Reporting System
r/SW: r/SuicideWatch

Edited by G Eysenbach; submitted 21.09.20; peer-reviewed by P Chikersal, J Han, K Harrigian; comments to author 23.12.20; revised version received 17.03.21; accepted 03.06.21; published 08.11.21.

Please cite as:
Chancellor S, Sumner SA, David-Ferdon C, Ahmad T, De Choudhury M
Suicide Risk and Protective Factors in Online Support Forum Posts: Annotation Scheme Development and Validation Study
JMIR Ment Health 2021;8(11):e24471
URL: https://mental.jmir.org/2021/11/e24471
doi:10.2196/24471
PMID:34747705

Review

Studies of Depression and Anxiety Using Reddit as a Data Source: Scoping Review

Nick Boettcher[1], BA

Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

Corresponding Author:
Nick Boettcher, BA
Department of Community Health Sciences
Cumming School of Medicine
University of Calgary
3D10
3280 Hospital Drive NW
Calgary, AB, T2N 4Z6
Canada
Phone: 1 (403) 220 4286
Email: nkboettc@ucalgary.ca

Abstract

Background: The study of depression and anxiety using publicly available social media data is a research activity that has grown considerably over the past decade. The discussion platform Reddit has become a popular social media data source in this nascent area of study, in part because of the unique ways in which the platform is facilitative of research. To date, no work has been done to synthesize existing studies on depression and anxiety using Reddit.

Objective: The objective of this review is to understand the scope and nature of research using Reddit as a primary data source for studying depression and anxiety.

Methods: A scoping review was conducted using the Arksey and O'Malley framework. MEDLINE, Embase, CINAHL, PsycINFO, PsycARTICLES, Scopus, ScienceDirect, IEEE Xplore, and ACM academic databases were searched. Inclusion criteria were developed using the participants, concept, and context framework outlined by the Joanna Briggs Institute Scoping Review Methodology Group. Eligible studies featured an analytic focus on depression or anxiety and used naturalistic written expressions from Reddit users as a primary data source.

Results: A total of 54 studies were included in the review. Tables and corresponding analyses delineate the key methodological features, including a comparatively larger focus on depression versus anxiety, an even split of original and premade data sets, a widespread analytic focus on classifying the mental health states of Reddit users, and practical implications that often recommend new methods of professionally delivered monitoring and outreach for Reddit users.

Conclusions: Studies of depression and anxiety using Reddit data are currently driven by a prevailing methodology that favors a technical, solution-based orientation. Researchers interested in advancing this research area will benefit from further consideration of conceptual issues surrounding the interpretation of Reddit data with the medical model of mental health. Further efforts are also needed to locate accountability and autonomy within practice implications, suggesting new forms of engagement with Reddit users.

(JMIR Ment Health 2021;8(11):e29487) doi:10.2196/29487

KEYWORDS

depression; anxiety; mental health; Reddit; social media; review

Introduction

Background

Interest in studying depression and anxiety using publicly available social media data has grown considerably in the past decade as widespread social media use has dovetailed with the rising global incidence of mental disorders [1,2]. The discussion platform Reddit has become a popular social media data source in this area of study, in part because of the unique ways in which the platform is facilitative of research. This scoping review is about understanding the landscape of research using the social

media platform Reddit as a primary data source for studying depression and anxiety. Approach to this research area with a scoping review is supported by 2 points of rationale. First, no work has been done to synthesize the existing research on depression and anxiety using Reddit outside of a small selection of review articles that included Reddit-focused studies under broader topics of social media data and mental health [3-5]. The Reddit platform affords researchers unique methodological opportunities for studying depression and anxiety and warrants a focus that is exclusive of other social media platforms. Second, timely resources are needed to cultivate informed deliberation about the conceptual and ethical dimensions of using publicly accessible social media in research on sensitive health topics, such as mental illness [6], and in public health research and practice more broadly [7]. By illuminating the scope and nature of the research landscape using Reddit data to study depression and anxiety, this review contributes to advancing understandings at the intersection of social media, research practice, and public mental health. The remainder of the *Introduction* section expands on the methodological characteristics of the Reddit platform and grounds the idea of using publicly accessible social media data to study mental health from a historical perspective.

Methodological Characteristics of the Reddit Platform

Reddit is a social media platform comprising single topic communities called subreddits that are formed, maintained, and participated in by pseudonymous users. Within subreddits, users can submit posts, respond to posts with comments, and reply to comments. Users can also engage with content by granting *upvotes* and *downvotes*, which subsequently inform the default visibility of content to other users. Content moderation is performed by volunteer moderators dually tasked with upholding individual subreddit rules and platform-wide content policies. Reddit harbors a variety of mental health subreddits. Examples relevant to this review, which are established in terms of longevity, size, and user activity, include r/mentalhealth, r/depression, and r/anxiety. Reddit distinguishes itself from Facebook, Twitter, and other health forums with its pseudonymous user system and generous length allowance for posts, comments, and replies. For years, researchers have been attracted to the way these attributes facilitate candid naturalistic expressions and exchanges of mental health information. In turn, Reddit-based studies have a definitive place in the history of studying informal web-based mental health communities [3].

At the outset of this review, it was clear that Reddit facilitates a variety of research approaches and scales of inquiry for studying depression and anxiety. Among the options for collecting naturalistic data, researchers can simply use the search function within the Reddit platform to identify posts, comments, and replies with specified keywords. For researchers interested in larger and more comprehensive data sets, Reddit's publicly accessible application programming interface (API) can be used to gather batches of data at a time according to the parameters specified in the code. For example, researchers might use the Reddit API to gather all posts, comments, and replies made to a particular subreddit over a defined period and contain a specified keyword. The Reddit API also grants researchers access to select metadata associated with posts, comments, and replies, such as the time of submission and the number of upvotes and downvotes received [8]. A widely used alternative means of accessing Reddit data is called Pushshift, a service created by developer Jason Baumgartner and designed to ingest and archive the entirety of Reddit data on an ongoing basis. As an alternative to the Reddit API, Pushshift offers researchers 2 primary benefits. First, Pushshift allows querying and retrieving historical data that are unavailable via the Reddit API, which enables, for example, the study of a community that was banned from Reddit in the past [9]. Second, Pushshift allows researchers more requests per minute than the Reddit API, shortening the time required to gather large data sets [8,9]. Aside from the Reddit API and Pushshift, researchers interested in studying depression and anxiety using Reddit data may also obtain access to data sets made available by other researchers who have already compiled the relevant data. Having briefly delineated these key methodological entry points for studying depression and anxiety using Reddit data, the remainder of the *Introduction* section provides a general historical contextualization for the study of mental health using naturalistic social media data.

Studying Mental Health Disorders Using Social Media Data

From a historical perspective, public exchanges of mental health information on social media platforms appeared nearly simultaneously alongside the opportunities for researchers to view and study these exchanges. To illustrate, communications scholar Lomborg [10] has traced the emergence of contemporary social media, both as a *communicative phenomenon* and a *research object*, to 2010. The Lomborg [10] demarcation marks the influx of a broad range of published research examining the intersections of mental disorders and social media, for example, by studying the associations between frequency of social media use and symptoms of depression and anxiety [11]. Within this wider range of inquiry, a research subgenre specifically focused on classifying mental health states using naturalistic data from Twitter and Facebook first appeared around 2013 [4], and comparable studies using Reddit data followed in 2014 [5]. The act of classifying written expressions from social media to make inferences about the mental states of users broke new ground at this time by combining disciplinary orientations and techniques from data science, psychology, and clinical psychiatry. The enthusiasm surrounding novel data sources and methodological possibilities ushered in by contemporary social media also extended to public health more broadly. For example, a 2012 commentary titled *How Social Media Will Change Public Health* predicted public health expanding its scope of practice as it began incorporating new streams of health data, including those related to mental health, which were now *on full display* because of social media [12].

The excitement that marked the early 2010s is in many ways still behind the research efforts to study mental health using social media data. The opportunities for mental illness prevention and mental health promotion remain promising, and the capacities of machine learning (ML) for generating insights from large social media data sets have expanded. However, 2 important sources of temperance arrived later in the decade, with respective origins from social media companies and the academic community. The first is our current *post–public-API age*, referring to the restrictions on public access to social media

data via the Facebook and Twitter APIs, which occurred in the wake of the 2016 Cambridge Analytica scandal [13]. Although Reddit has yet to enact comparable restrictions to API access [9], the post–public-API age has renewed emphasis on the foundations of independent social media research by highlighting that access to data is contingent on what private social media companies decide to make or keep available. Second, compelling scholarly arguments for a deeper examination of the conceptual foundations and ethical concerns surrounding the study of sensitive health issues such as mental health using social media data have increased in presence in the latter half of the 2010s [4,10,14]. These key points of historical inflection provide context on how this scoping review of studies on depression and anxiety using Reddit data enters into the brief, fast-moving history of social media's emergence as an object of dedicated research inquiry.

Objective and Research Questions

The objective of this scoping review is to determine the scope and nature of research conducted using Reddit as a primary data source for studying depression and anxiety. Specifically, this review proposes to answer the following broad research questions:

1. To what extent have depression and anxiety been studied using the data from Reddit?
2. What are the prevailing analytical practices observed in the included studies?
3. What recommendations for practice have been made by the authors of the included studies?

Methods

Overview

Scoping reviews are exploratory review studies conducted to better understand a research area. The unifying activity of scoping reviews can be thought of as *mapping* sources of evidence and key concepts [15]. Similar to the systematic review methodology, scoping reviews require careful planning in advance of collecting the literature. Researchers must articulate a research objective and questions that thread through the search strategy and criteria for assessing sources. However, the purposes for undertaking scoping reviews are typically more flexible than those motivating systematic reviews. Procedures followed for this scoping review were guided by the influential Arksey and O'Malley [16] methodological framework, the Levac [15] elaborations to the Arksey and O'Malley framework, a guidance article from the Joanna Briggs Institute Scoping Review Methodology Group [17], and the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews) checklist [18] for reporting the results of scoping reviews.

Search Strategy

To identify relevant studies, the search strategy encompassed health science databases MEDLINE, Embase, CINAHL, PsycINFO, and PsycARTICLES and general science databases Scopus and ScienceDirect. Computer science databases ACM Digital Library and IEEE Xplore Digital Library were also searched to capture refereed conference proceedings that may not have been indexed in other science databases [19]. An initial search was performed on October 22, 2020, and the results were imported into Endnote (Clarivate Analytics, Inc) and sorted by databases. To update the review before submitting for publication, a follow-up search was performed on January 22, 2021, with identical search terms and databases. New sources yielded from the follow-up search were identified following the Bramer [20] technique for updating systematic literature searches using Endnote's deduplication feature. The full Ovid MEDLINE database search strategy is provided in Multimedia Appendix 1.

Studies were also gathered from 2 information sources outside of the academic databases. First, the reference lists of recent review articles, including Reddit-focused studies under the broader topics of social media data and mental health, were hand-searched before the initial database search [3,4,21]. Second, the author set a weekly Google Scholar alert that ran from the day of the initial search to that of the follow-up search and identified newly indexed studies with the word *Reddit* in the title. Studies identified by the weekly alerts were imported into a separate database folder in Endnote during the follow-up search. Google Scholar was restricted to a supplementary search resource because of its limited capabilities for structured searches and issues with transparency and reproducibility of search results [22].

Eligibility Criteria

Overview

Eligible studies used naturalistic data from Reddit as a primary data source and featured an analytic focus on depression or anxiety. Studies were limited to those published in English. To be included at the abstract screening stage, studies had to use the term *Reddit* in their title or abstract and *depression* or *anxiety* in the title or abstract. As familiarity with the subject matter increased, inclusion criteria were further articulated following the *participants concept context* framework suggested by Peters et al [17]. These criteria have been defined in the following subsections.

Participants

Participants of the included studies were Reddit users whose publicly available posts, comments, and replies were unobtrusively analyzed as a primary data source for studying depression and anxiety.

Concept

Included studies examined depression or anxiety as core concepts, meaning that a conceptual focus on depression or anxiety was specified at the level of methodology. Studies that did not make this methodological specification were excluded even if findings related to depression or anxiety were reported. Aside from the criteria of methodological specification, the concepts of *depression* and *anxiety* were treated inclusively throughout the screening and review steps.

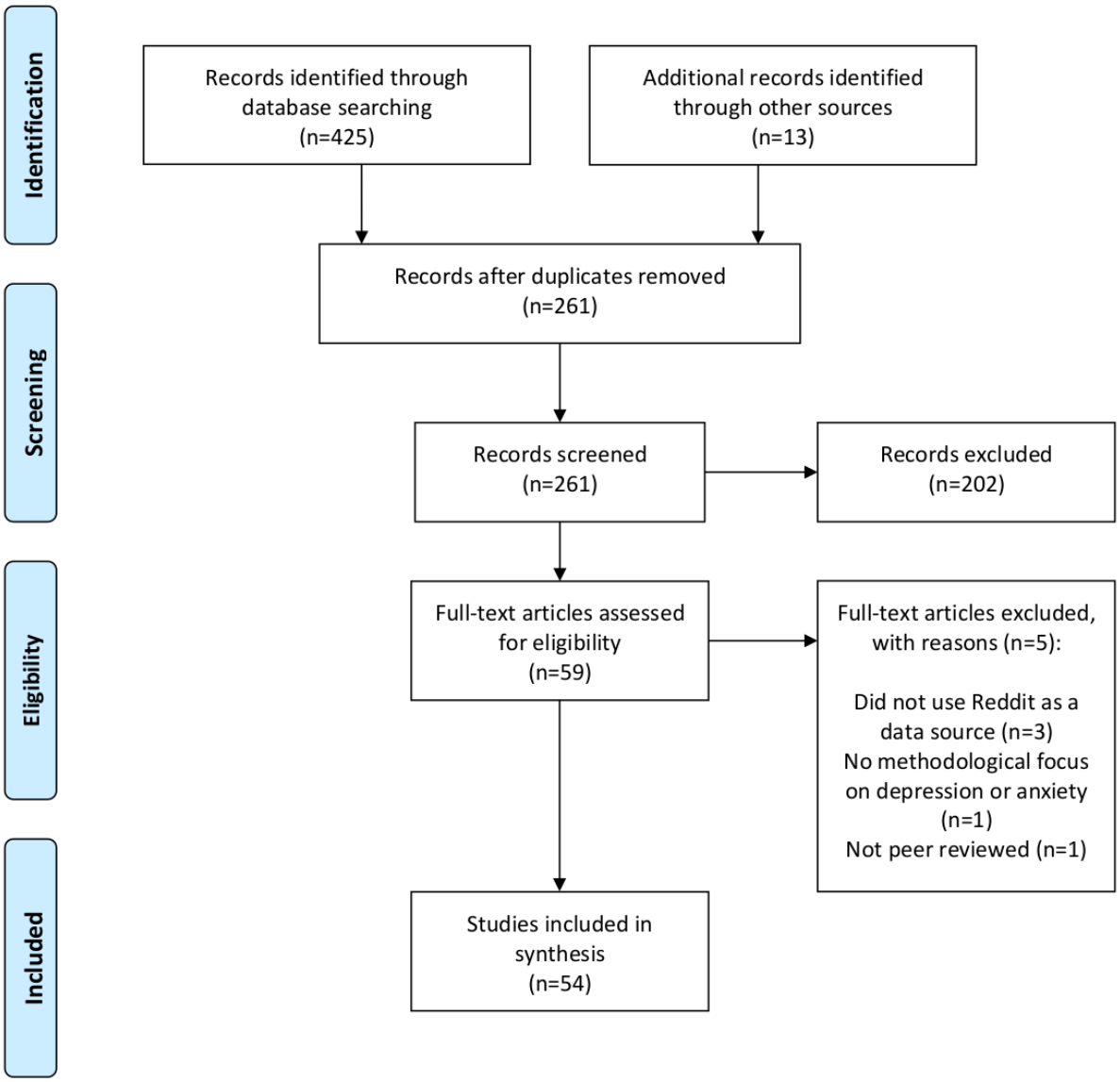
Context

The contextual focus of this review is naturalistic data from Reddit. Studies using Reddit to recruit participants for surveys or interviews fall outside this context.

Study Selection

Covidence (Veritas Health Innovation, Ltd) reference management software was used to screen abstracts, review full texts, and chart data from included studies. After duplicates were removed, the author and a second reviewer collaboratively screened the first 10% of abstracts. Then, the author and the second reviewer screened the second 10% of abstracts independently. They then met again to check in and resolve conflicts before independently screening the remaining 80% of the abstracts. This approach resulted in a high degree of agreement between the 2 reviewers (95%), and conflicting screening decisions were resolved by revisiting the abstracts in question and discussing them alongside the study objective and inclusion criteria. A similar approach was taken for the full-text review stage, in which there were no conflicts. A total of 60 additional references were captured in the follow-up search executed on January 22, 2021, and these additional studies were processed through abstract screening and full-text review during collaborative sessions. Data charting for the additional studies was performed using the same form as in the initial studies. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram presented in Figure 1 was backward-corrected to reflect all included sources, including those collected in the follow-up search [20].

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



Data Charting

The 2 reviewers (the author and the second reviewer mentioned above) initially met to draft the data charting form. The reviewers used the first iteration of the data charting form to independently chart data from 5 papers before meeting again to revise the form. The reviewers then used the revised form to chart the remainder of the included studies. Information charted from each paper included (1) publication details, (2) study design (objectives and conceptualization of depression or

anxiety), (3) methodology (data collection techniques, data set characteristics, data preprocessing, analysis methods, and outcome measures), and (4) results (summary of findings, implications, and recommendations for future research). Early data charting revealed an abundance of studies using ML models, and additional data were charted for these, including the overall type of ML approach, feature extraction methods, and classifier types. For a subset of ML-based studies that were distinctly focused on making predictive classifications of user mental health states, techniques for supporting the ground truth of classification decisions were also charted following a typology of common methods identified by Chancellor and De Choudhury [4]. A formal appraisal of the quality and sources of bias among the included studies was not meaningful to the objective of this scoping review.

Summarizing and Reporting Results

Once all the data were charted, the included studies were first descriptively summarized according to the date, type of publication (journal article vs conference proceeding), and institutional location of the first author. Then, 4 tables were created to organize the included studies according to the key categories of interest in response to the research objective and questions. Finally, accompanying text was written to support the information contained in the tables through narrative elaboration, illustration, and context. Throughout the reporting process, the imperative of the scoping review methodology to "establish how a particular term is used in what literature by whom, and for what purpose" [23] remained a central concern with respect to depression and anxiety to better understand the broader implications of studying these concepts using data from Reddit.

Results

Overview

We screened 425 abstracts for possible inclusion, resulting in 59 (13.9%) full-text articles being assessed for eligibility. Of the 59 papers, 5 (8%) full texts were excluded at the full-text stage, leaving 54 (92%) studies to be included in the review (Figure 1) [24-77]. The included 54 studies were conducted between 2014 and 2020 and comprised of 31% (17/54) journal articles and of 69% (37/54) refereed conference proceedings. The claims of the fast-advancing nature of this research area are not hyperbole, as 17% (9/54) of the included studies were published and newly indexed in the 3-month interval between the initial and follow-up database searches. The first authors of the included studies were associated with institutions from 21 different countries, and the United States was the most represented among these, with 30% (16/54) of the studies.

A group of 16 included studies was unique in its association with the Conference and Labs of the Evaluation Forum (CLEF), an annual independent peer-reviewed information systems conference organized around experimentation on shared tasks. In 2017, CLEF introduced the *eRisk* shared task, which invites entrants to create experimental models for predicting mental health risks using social media data sets supplied by the conference organizers [78]. Owing to the fact that CLEF eRisk tasks across the years 2017-2020 used data sets collected from Reddit and featured a focus on depression detection, our search strategy captured the studies associated with eRisk tasks held in 2017 (8/54, 15%), 2018 (5/54, 9%), 2019 (1/54, 2%), and 2020 (1/54, 2%). This group has been referred to as the *CLEF eRisk studies* throughout the *Results* and *Discussion* sections as further observations have been made about the distinct contributions of the eRisk shared tasks to the overall landscape of research using Reddit data to study depression and anxiety. Following the Figure 1 PRISMA flow diagram, tables and corresponding results have been presented according to 4 key study characteristics: (1) mental health conditions of focus, (2) data collection approaches, (3) analytic focus, and (4) practical implications. The first of these study characteristics has been presented in Table 1, which has organized the included studies according to their depression versus anxiety focus while also conveying an overall picture of the mental health conditions studied.

Table 1. Mental health conditions of focus (N=54).

Mental health condition of focus	Number of included studies with focus on condition, n (%)	References
Depression (n=32)		
No additional conditions studied	20 (63)	[24-27,29,30,34-38,45,46,57,58,63,64,69,70,74]
Eating disorders	5 (16)	[28,31-33,39]
Bipolar disorder	4 (13)	[41,59,65,77]
Context: postpartum depression	3 (9)	[56,67,75]
Schizophrenia and psychosis	1 (3)	[59]
Self-harm	1 (3)	[76]
Depression and anxiety (n=19)		
Bipolar disorder	14 (74)	[40,42,44,47,49,50,53-55,62,66,71-73]
Suicidal ideation	8 (42)	[44,47,49,50,55,62,66,73]
Borderline personality disorder	8 (42)	[49,50,53-55,62,66,73]
Neurodevelopmental conditions[a]	8 (42)	[49-51,54,55,71-73]
Schizophrenia and psychosis	7 (37)	[42,49,50,53-55,66]
Posttraumatic stress disorder	6 (32)	[28,42,51,53,60,66]
Self-harm	5 (26)	[42,49,50,62,66]
Eating disorders	4 (21)	[43,53,62,73]
Depersonalization and derealization disorder	3 (16)	[42,62,66]
Substance use	4 (21)	[44,55,62,73]
Dementia	2 (11)	[49,65]
Context: rheumatoid arthritis	1 (5)	[61]
Anxiety (n=3)		
No additional conditions studied	2 (67)	[52,68]
Bipolar disorder, borderline personality disorder, neurodevelopmental conditions[a] schizophrenia, self-harm, and substance use	1 (33)	[48]

[a]The category *neurodevelopmental conditions* refers to studies in which authors specify a focus on autism, attention deficit hyperactivity disorder, and/or Asperger syndrome.

Mental Health Conditions of Focus

Of the 54 studies, depression was researched in 51 (94%) studies, including 20 (37%) studies that exclusively focused on depression. Meanwhile, anxiety was researched in 43% (23/54) of studies, and only 4% (2/54) of studies examined anxiety exclusively. The relative prominence of depression-focused studies can be partially explained by the collection of 16 CLEF eRisk studies, all of which focused on the detection of depression. Table 1 also captures 31% (5/16) of the CLEF eRisk studies that reported participation in a separate shared task for the early detection of anorexia in addition to the 2018 shared task for depression detection. Overall, the abundance of conditions studied in addition to depression and anxiety signals research interest in comparative studies of mental health conditions using Reddit data. Many of the additional conditions were related to depression and anxiety, as exemplified by the inclusion of bipolar disorder in 35% (19/54) of the studies. However, a particularly wide breadth of conditions of focus was noted in some studies that used Reddit data to investigate mental health phenomena as diverse as autism and dementia within a single study while also focusing on depression and anxiety. In summary, depression was researched to a greater extent than anxiety by a wide margin, and it is clear that the Reddit platform facilitates research methodologies designed for simultaneous examinations of multiple mental health conditions. In order to better understand the methodological execution of included studies, the approaches taken to data collection have been presented in Table 2.

Table 2. Data collection approaches (N=54).

Data collection approaches (data set category) and data source	Included studies, n (%)	References
Studies using original data sets (n=24)		
Reddit API[a]	12 (50)	[42,44,46,52,60,62,63,68,71,72,75,77]
Pushshift	10 (42)	[47,49,50,53-55,58,59,66,76]
Reddit search function	1 (4)	[61]
Google search	1 (4)	[56]
Studies using a premade data set (n=21)		
CLEF[b] eRisk 2017 data set	8 (38)	[24-26,29,34-37]
CLEF eRisk 2018 data set	5 (24)	[27,31-33,39]
CLEF eRisk 2019 data set	1 (5)	[38]
CLEF eRisk 2020 data set	1 (5)	[30]
Multiple CLEF eRisk data sets	1 (5)	[28]
Data set from Yates et al [76]	3 (14)	[45,69,74]
Data set from Gkotsis et al [50]	1 (5)	[48]
Data set from Pirina and Çöltekin [63]	1 (5)	[70]
Studies with multiple data collection approaches (n=4)		
Reddit API and Pushshift.io	1 (25)	[57]
Reddit search function and Reddit API	1 (25)	[67]
Reddit API plus data set from Pavalanathan and De Choudhury [62]	1 (25)	[41]
CLEF eRisk 2017 data set and data set from Yates et al [76]	1 (25)	[64]
Studies with unclear data collection approach (n=5)		
Data collection not clearly described	5 (100)	[40,43,51,65,73]

[a]API: application programming interface.

[b]CLEF: Conference and Labs of the Evaluation Forum.

Data Collection Approaches

Table 2 organizes the included studies according to how researchers approached data collection. Approximately 44% (24/54) of studies reported creating original data sets, and 22% (12/54) of studies did so by accessing Reddit's official API. Another 19% (10/54) of studies created original data sets by accessing Pushshift, an archived corpus of Reddit data managed by the developer Jason Baumgartner and sometimes signaled within studies as the Reddit *data repository* [47] or *data dump* [49]. Perhaps unsurprisingly, the 4% (2/54) of studies that collected original data sets using neither the Reddit API nor Pushshift were qualitative studies using smaller data sets. To illustrate, Park et al [61] used the Reddit search function to collect a sample of 81 discussion threads for qualitative descriptive analysis, whereas the Maxwell [56] qualitative study used Google Search to identify a single discussion thread consisting of 1 post and 294 comments, with which Maxwell et al [56] performed an in-depth thematic analysis.

Researchers used premade data sets in 39% (21/54) of the studies. Of the 54 studies, data sets for all 16 (30%) included CLEF eRisk studies were categorized as premade as eRisk participants were supplied with Reddit-based data sets created by competition organizers. A total of 2 observations bear mentioning here about the nature of these CLEF eRisk data sets. First, year-by-year iteration was reflected in the 2018, 2019, and 2020 CLEF eRisk data sets using portions of data sets from the preceding year as testing data for participants to tune their classification systems. Second, 9% (5/54) of studies identified using CLEF eRisk data sets after the official timeline of respective annual tasks. The availability and continued uptake of CLEF eRisk data sets is notable as it allows researchers to take their time to study and attempt to outperform the best-performing entries. Aside from the CLEF eRisk studies, the existing data sets accessed by researchers were made available through previously published studies. The most influential among these was the *Reddit Self-Reported Depression Diagnosis* data set, introduced in an included 2017 study by Yates et al [76] and subsequently cited as a primary data source in 7% (4/54) of other studies.

Of the 54 studies, 5 (9%) studies collected data from Reddit using multiple means. An inventive example of multiple approaches to data collection was noted in the study by Shatte et al [67], which used Reddit data to examine social media markers of postpartum depression in fathers. Shatte et al [67] began data collection using the Reddit search function to identify

a cohort of 365 Reddit users who made birth announcements in a subreddit for fathers from 2016 to 2018. Then, these authors accessed the Reddit API to collect all posts and comments made by these 365 users in the 6 months before and after each user's birth announcement, resulting in a data set of 67,796 posts and comments. Finally, they analyzed changes in depressive language and discussion topics following each user's birth announcement [67]. Combinatory approaches to data collection, such as those deployed by Shatte et al [67], reflect the unique methodological possibilities afforded by the Reddit platform.

Finally, 9% (5/54) of studies were inexact in describing how the data from Reddit were accessed. For example, the De Alva et al [40] study described *selecting* 32 posts containing keywords from 6 mental health subreddits without stating how the researchers searched for the keywords. Other examples included references to data being *crawled* [73], *downloaded* [65], and *collected* [43] without further specification and an uncited reference to a previously used data set [51]. Having conveyed an overview of data collection approaches, Table 3 presents the analytic focus of the included studies to better understand how researchers used their data sets to study depression and anxiety.

Table 3. Analytic focus of included studies (N=54).

Analytic focus (general focus and specific focus)	Included studies, n (%)	References
Focus on predictive mental health classification (n=36)		
Binary classification of user mental health	27 (75)	[24-29,31-37,39,45,51,63,64,67-72,74-76]
Severity-focused classification of user mental health	2 (6)	[30,38]
Disclosure-focused classification of user mental health	1 (3)	[42]
Multilabel classification of user mental health	4 (11)	[50,53,54,65]
Subreddit-level mental health classification	2 (6)	[41,48]
Focus on mental health language and interactions (n=18)		
Subreddit-level analysis	12 (67)	[43,44,46,49,55,57,60,66,73,77]
User-level analysis	3 (17)	[52,58,62]
Discussion-level qualitative analysis	3 (17)	[40,56,61]
Multilevel analysis	2 (11)	[47,59]

Analytic Focus

Table 3 presents the primary analytic focus of the included studies organized broadly into studies that used Reddit data to (1) make predictive classifications of user mental health states and (2) analyze mental health language and interactions. Within the 2 general categories, studies were further organized into subcategories according to conceptual focus and scale of inquiry.

Studies Focused on User Mental Health Classification

Approximately 67% (36/54) of studies used data from Reddit as the basis for predictive mental health classification using ML. The most popular ML approaches were deep learning (DL; 11/36, 30%), supervised ML (6/36, 17%), and combinations of DL and supervised machine (9/36, 25%).

Researchers took a range of approaches to process naturalistic Reddit data into features for use in their ML systems and, in many cases, combined and compared the feature sets within a single study. The most represented features among the classification studies were based on categories of words from pre-existing mental health lexicons and data sets (14/36, 39%). Other common features included n-grams (9/36, 25%), bag-of-words (9/36, 25%), and term frequency–inverse document frequency vectors (9/36, 25%). ML classification was most frequently performed with a neural network (17/36, 47%), support vector machine (10/36, 28%), random forest (7/36, 19%), and logistic regression (6/36, 17%) classifiers. Of the 36 classification studies, 26 (72%) reported on binary classification, with decisions about the mental health states of Reddit users framed in terms of *yes or no* (12/36, 33%) and *at risk or not at risk* (14/36, 39%). Approximately 11% (4/36) of studies performed multilabel mental health classification with DL models that predicted *which* mental health problem was represented in the Reddit user text. Multilabel classification appears to be a forerunning analytic focus, as all 4 studies in this category were published in 2020.

A focus on the severity of depressive symptoms was observed in 6% (2/36) of the classification studies associated with the 2019 CLEF eRisk shared task on early depression risk detection. The authors of these severity-focused experiments were given a data set of Reddit posts and asked to ordinally classify users' depression as mild, moderate, or severe by transposing inferred signs of depression into responses to the Beck's Depression Inventory questionnaire [79]. Another approach to user classification was found in the Balani and De Choudhury [42] study in which user posts from mental health subreddits and control subreddits were classified according to self-disclosure, defined as the degree to which users revealed personal information and vulnerable thoughts, beliefs, and experiences. Of the 36 classification studies, 2 (6%) studies endeavored to perform mental health classification at the subreddit level. These included the Gaur et al [48] study, which mapped the aggregate content of 15 mental health subreddits to diagnostic categories from the Fifth Edition of the Diagnostic and Statistical Manual of Mental Disorders, and the study by Bagroy et al [41], which gathered 43,468 posts from *mental health* and *control* subreddits to train and test a model for classifying broad trends in

expressions of mental health and distress in 109 university subreddits.

Ground Truth in User Mental Health Classification Studies

Among the 36 mental health classification studies, broad trends were noted in the provision of ground truth, or in other words, the baseline by which predictive classifications of the mental health states of Reddit users could be considered valid [14]. Although the following trends are outlined here individually, they usually appeared in some combination within single studies. Researchers targeted posts from mental health subreddits in 75% (27/36) of studies and collected user posts deemed to express self-disclosure of a mental health diagnosis in 69% (25/36) of studies. The ground truth of predictive claims was supported with control data in 78% (28/36) of studies. For example, the Shen and Rudzicz [68] study used ML to classify anxiety in user posts with a data set of 9971 posts from 4 anxiety subreddits and 12,837 posts from 25 control subreddits deemed unrelated to mental health. Human annotators contributed to labeling data in 75% (27/36) of studies and were variously described as layperson annotators [76], raters familiar with Reddit and its mental health communities [42], Amazon Mechanical Turk workers [65], 2 mental health domain experts [48] a clinical psychologist [67], a social media expert and clinical psychologist duo [41], and simply human annotators [53,66]. Of the 36 mental health classification studies, 14 (39%) studies incorporated external mental health data sets into data labeling procedures to support the ground truth of classification. External data set sources ranged from Wikipedia [36], Twitter [37], and AskAPatient [65] to formalized medical sources, including the Unified Medical Language System [31], the International Classification of Diseases, 10th Revision [48], and the Fifth Edition of the Diagnostic and Statistical Manual of Mental Disorders [48,69].

In the 16 CLEF eRisk studies, ground truth relied on the data sets and parameters of analysis provided by eRisk shared task organizers and featured a combination of the community participation, self-disclosure, and control data trends outlined above. Notably, the 2019 and 2020 shared tasks took an additional step in which CLEF organizers contacted Reddit users whose posts and comments were collected for the shared task data set and asked them to fill out a validated depression questionnaire. User-completed questionnaires were then used as ground truth for assessing the performance of entrants who attempted to fill out the same questionnaires using only a curated history of Reddit posts and comments from each user [80]. Although user-completed depression questionnaires supply a more traditionally valid conception of ground truth than annotated disclosures of depression diagnoses in Reddit posts, the solicitation of these questionnaires marks a disjuncture from the passive analytic focus that otherwise characterized the mental health classification studies.

Studies Focused on Mental Health Language and Interactions

In distinction to studies focused on classifying the mental states of users, 33% (18/54) of studies used a range of methods to study depression or anxiety with an analytic focus on language, discussions, and user interactions. Of these, 39% (7/18) of studies used ML, all of which entailed the use of unsupervised models with the exception of the Sharma and De Choudhury [66] study, which used supervised ML models to classify the degree of support exhibited in comments. Latent Dirichlet Allocation, an unsupervised ML model used to generate topics from large data sets of naturalistic texts, was most represented in 22% (4/18) of studies. A further 44% (8/18) of studies, although not using ML models, applied natural language processing (NLP) techniques to generate features or topics for analyzing mental health language and interactions. Among the 15 studies in this category that used ML and/or NLP, Reddit text was most often processed into features through assignment into categories from the Linguistic Inquiry and Word Count mental health lexicon (7/18, 39%) by using term frequency–inverse document frequency vectors (4/18, 22%) and by processing text into readability metrics (2/18, 11%). Of the 18 studies, 3 (17%) purely qualitative studies examined depression and anxiety in the context of Reddit discussion threads about rheumatoid arthritis [61], fathers' experiences with postpartum depression [56], and the effectiveness of mobile mental health apps [40].

The analytic focus of studies that focused on language and interactions also ranged in scale. Approximately 67% (12/18) of studies analyzed subreddit-level language phenomena, including the Chakravorti et al [43] comparative study of trends in discussion topics in r/depression, r/anxiety, and r/suicidewatch from 2012 to 2018 and the Low et al [55] study, which used the COVID-19 pandemic as a point of reference for examining changes in discussion topics across 15 mental health subreddits from *prepandemic* versus *midpandemic* periods. Of the 18 studies, 3 (17%) demonstrated a user-level analytic focus on language. An illustrative example of this category is the Ireland et al [52] study of how 1409 users of the r/anxiety subreddit exhibited differences in language use when posting in r/anxiety versus other subreddits compared with a control group of users with no history of participation in r/anxiety. Finally, 11% (2/18) of studies analyzed language at multiple levels. These included the Park and Conway [59] study of written communication challenges encountered in Reddit's mental health subreddits, which featured a subreddit-level linguistic analysis in addition to a longitudinal user-level analysis. Moving on from the analytic focus of included studies, Table 4 summarizes the implications for practice gathered from the discussion and conclusion sections of the included studies.

Table 4. Summarized practice implications (N=30).

Practice implication category and types	Included studies, n (%)	References
Professional-focused implications (n=19)		
Interventions from health professionals	18 (95)	[42-44,48,52,63-65,68-70,72,74-77]
Interventions from university counselors	1 (5)	[41]
User-focused implications (n=9)		
Interventions to inform or direct users to information	4 (44)	[42,55,60,66]
Interventions to prompt users to assess mental health risk	3 (33)	[54,58,67]
Interventions to encourage comfort in discussing sensitive health topics	2 (22)	[62,73]
Intervention to alter user text for improved readability	1 (11)	[59]
Moderator-focused implications (n=6)		
Moderator-focused tool for sorting information about posts and users	6 (100)	[42,49,50,58,59,66]
Patient education programs (n=3)		
Creation of formal patient education and support	3 (100)	[56,60,61]

Practice Implications

Approximately 56% (30/54) of studies built on findings with one or more practice implications, most of which 63% (19/30) suggested incorporating insights from Reddit data into professional mental health practice. A demonstration of this idea was found in the Rao et al [64] study, which applied a neural network model to classify Reddit users as depressed. In discussing the implications of their study, Rao et al [64] highlighted the future possibility of "sensitive applications in combining clinical care with users' online activities" [64]. Other studies were more specific in envisioning practice implications for clinicians as they described a future "clinical tool" [72] or a "diagnostic aid" [63] designed to bring Reddit data into professional contexts. However, others used the more generalized language of "mechanisms" [70], "automated processes" [74], "services" [42], and "resources" [43] in presenting implications motivated by enhancing professional mental health practice with the results of studying depression and anxiety using Reddit data.

Of the 30 studies, 9 (30%) suggested practice implications focused on Reddit users. Common among the user-focused implications were suggestions for interventions to direct users to information. The Sharma and De Choudhury [66] study, for example, used NLP techniques to measure concepts of accommodation and support in discussions on mental health subreddits. Sharma and De Choudhury [66] suggested that posts determined as receiving exemplary levels of accommodation and support be embedded into community guidelines for promoting "subconscious learning of the linguistic style of the community" [66] to assist users as they consider posting. Another strand of user-focused implications was proposed in the Park et al [60] study, in which Park et al [60] envisioned functionality designed to help users of Reddit's mental health subreddits connect with other users discussing similar mental health issues and who share appropriate "contextual elements of experience" [60]. The Low et al [55] study similarly recommended an intervention to guide users identified as expressing mental distress to subreddit communities known for high levels of support or moderator activity.

Some studies pictured future applications of ML classification that are designed to enable Reddit users to self-assess their mental state. For example, in the discussion following the results of the Kim et al [54] multilabel classification study, Kim et al [54] envisioned a service which, with user consent, accesses a user's post history to "provide the probabilities of each mental disorder" [54]. The Park et al [58] study of changes in language use among long-term users of the r/depression subreddit recommended a similar automated process that would continuously monitor the writing of individual users to detect "undesirable linguistic changes" and subsequently intervene to "raise self awareness of their changes of linguistic or emotional state" [58].

Of the 30 studies, 6 (20%) suggested practice implications for moderators of Reddit's mental health subreddit communities, and ideas for improving moderation workflow through automated means were common among these. For example, Gkotsis et al [49] suggested the assignment of urgency markers for posts in need of timely moderator attention, and Sharma et al [66] recommended new tools to help the moderators of mental health subreddits "efficiently and quickly navigate the stream of incoming requests" [66]. Finally, 10% (3/30) of studies featuring practice implications suggested incorporating the study findings into formal patient education. To illustrate, the Park et al [61] qualitative study of depression and anxiety in the context of rheumatoid arthritis positioned study findings as potentially contributing to improved practical recommendations to guide health care within rheumatology. In summary, 56% (30/54) of the included studies mentioned practice implications that went beyond theoretical implications or recommendations for future research. Taken together, these practice implications give a lens through which the goals and imagined futures of this research area come into focus.

Discussion

Principal Findings

To the author's knowledge, this scoping review is the first to map academic literature focused on the study of depression and anxiety using data from Reddit. The objective was to better understand the scope and nature of this research space by detailing its mental health conditions of focus, data collection approaches, analytic focus, and practice implications. The results showed comparatively more research attention directed to depression versus anxiety, an even split of original and premade data sets, a favored analytic focus on classifying the mental health states of Reddit users, and practical implications that frequently recommended new professionally driven monitoring and outreach for Reddit users. Researchers interested in advancing the study of depression and anxiety using Reddit data will benefit from further consideration of key insights and tensions contained within the main results, which are elaborated in the following 2 sections: (1) conceptual issues surrounding the interpretation of Reddit data with the medical model of mental health and (2) the importance of locating accountability and autonomy in practice implications suggesting new forms of engagement with Reddit users.

Depression, Anxiety, and the Medical Model

Observations about the ways in which depression and anxiety were studied using Reddit data rest on 2 premises established at the outset of this review. First, the circulation of mental health information on Reddit is user driven and conceptually distinct from the domain of working medical professionals. Second, research practice works to transform Reddit from the naturalistic *communicative phenomenon* of social media into a *research object* through the application of research methodologies and the accumulation of academic knowledge [10]. This review shows the concepts of depression and anxiety on Reddit emerging as part of a research object that favors interpreting mental health problems through the medical model of practice and explanation [81]. The medical model was readily noticeable in generalizations about the nature of Reddit's mental health subreddits through references to r/depression and r/anxiety as "clinical subreddits" [72] and the depictions of users accessing mental health subreddits as "diagnostic groups" [53], and "patients" [48,71]. Although these broad strokes of medical terminology are conjecture, they illustrate the medical model being put to work as an *interpretive frame* circumscribing grammar, conceptual boundaries, and claims to relevance in research [82]. The medical model has thus far proven to be influential as an interpretive frame; however, its influence appears to subsist in the absence of wider debate and negotiation. To bring understandings of depression and anxiety on Reddit into maturation, constructive thinking and discussion about the medical model as an interpretive frame will be needed.

A thought-provoking exception to implicit assumptions about the medical nature of Reddit's mental health communities was noted in the Park et al [60] longitudinal study of thematic similarities and differences between r/depression, r/anxiety, and r/PTSD subreddits. Park et al [60] combined a topic modeling algorithm with qualitative analysis to analyze a total of 7410 posts and 132,599 comments made between January 2011 and December 2015. On the basis of comparative findings on the r/depression subreddit, such as its larger size and less active userbase, Park et al [60] hypothesized that "the word 'depression' perhaps has a larger set of connoted meanings, some clinical and others not; and thus, those who participate in this subreddit may be a more diffuse and transient group" [60]. With this observation, Park et al [60] suggested an alternative conceptualization of depression on Reddit that more resembles a communicative phenomenon and likely possesses far less clinical relevance. In recognizing this more expansive meaning of depression, Park et al [60] illustrated the medical model of mental health as just one way to approach the concepts of depression and anxiety in the context of social media data [4].

Although depression and anxiety can refer to categories of disorder diagnoses, they can also refer to symptoms of other mental health conditions, transient emotional expressions, or something else entirely. Borrowing an idea from the early 20th-century philosopher Wittgenstein [83], perhaps depression and anxiety on Reddit are best understood as *family resemblance concepts* with plural meanings that overlap, diverge, and shift over time on the scale of individual user expressions, discussion threads, and entire subreddit communities. Not only do these meanings elude clean division along the lines of professional versus lay knowledge, but it is also possible that ambiguity surrounding meanings of depression and anxiety can serve as a rhetorical resource in user-led discussions [84]. To this end, researchers seeking a broader understanding of what depression and anxiety mean for mental health information seekers on Reddit would benefit by incorporating openness to plural meanings of these terms into methodological choices. Although looking beyond the circumscriptions of the medical model may involve departing from the goal of accurately classifying user mental health states, it does not imply adversity to the analysis of big data sets using ML and DL. This difference was illustrated by some techniques charted in this review, such as Latent Dirichlet Allocation [44,46,55,57] and a relationship modeling network [47], which were used to investigate mental health language by processing naturalistic text in an unsupervised or bottom-up fashion, meaning that data from Reddit did not converge with any external lexicons or prelabeled data in processing. Unsupervised systems still reflect researcher choices and perspectives at other methodological decision points, such as the data collection approach, the tuning of algorithm parameters, and the interpretation of output; however, it is notable that unsupervised techniques avoid freighting the concepts of depression and anxiety with external information while still leveraging the analytic insight of ML. Qualitative research designs offer another methodological path to broaden the conceptualizations of depression and anxiety while also introducing limits to the scope and depth of analysis feasible for the *human instrument* in a single study. It stands that the qualitative designs included in this review were oriented to smaller, context-driven analyses, for example, understanding the meanings of depression in the context of other conditions such as rheumatoid arthritis [61].

In summary, provisional application of the medical model of mental health may be appropriate in certain approaches for

studying depression and anxiety using Reddit data. One included study paraphrased the goal of preventing mental illness through predictive classification of social media data as wider progress toward "an unfulfilled promise of clinical science" [72]; this may be possible if ML approaches are carefully integrated with not only the terminology but also the human expertise and ground truth standards of the medical model. A likely explanation for the current dearth of substantive discussion about the conceptual foundations of the medical model in this research area is the interweaving of the medical model of mental health with the language and concepts of ML classification systems. The contingent of researchers potentially interested in wider conceptual engagement with the medical model extends fairly broadly into medical and social sciences; however, the specialized language of ML may act as a buffer against such engagement in this research niche [85]. The results of this review are hopefully encouraging, as they show that there is space for further examining the daring conceptual feat of importing established medical terminology into the novel algorithmic models of ML-driven studies of naturalistic Reddit data. More generally, the maturation of this research space will be served by putting the medical model into perspective as just one framework among many from which to comprehend the meanings of depression and anxiety with data from Reddit. The final section of the *Discussion* connects the conceptual issues arising from adherence to the medical model to ambitions for materializing tangible impacts on the mental health of Reddit users.

Locating Accountability and Autonomy in Practical Implications

At the methodological center of the studies included in this review is a one-way flow of information, as researchers unobtrusively gather and analyze data sets of user expressions from Reddit related to depression and anxiety. User expressions eventually become the substrate of academic publications that leave no footprints in the web-based environments in which Reddit users originally participated. Given the ethical delicacy of this achievement, it is notable that the discussion sections of included studies frequently harbored the ambition to cross from a passive research practice into engagement with Reddit users through new professional digital outreach initiatives entailing monitoring and intervention [86]. In one sense, the horizon of eventual intervention lends purpose and a promise of future impact to studies of depression and anxiety using Reddit data. However, the digital outreach initiatives proposed in the included studies were understated in the positioning of Reddit—both the user base and the company—within as of yet unrealized sociotechnical configurations of academic researchers, ML systems designed to classify mental states, clinicians, and the larger digital wellbeing industry [87]. For researchers imagining new expressions of digital public mental health for Reddit users related to depression and anxiety, it will be useful to provide additional ideas for navigating the piecemeal, international, and commercially inclusive structure of the sociotechnical configurations involved. The idea of accountability in digital public health, summarized by Hoeyer et al [88] as "defining who needs to know—and do—what, and for and to whom" [88], would be a helpful anchor to ask the questions needed for advancing practice implications in this research area. For example, how would the scope of practice and incentive structure of professionals associated with a digital outreach initiative intersect with the accountability for the mental health needs and preferences of Reddit users? Relatedly, what would be the best way to incorporate algorithmic accountability of ML systems designed to initiate engagement with Reddit users or health professionals based on the classification of mental health states [89]? To work toward a more ethically sensitive foundation of accountability in decision-making about digital outreach, it will be necessary to consider these questions at the level of sociotechnical configurations. At a narrower level, researchers would also benefit from wading into the ethical dimensions of autonomy in the context of users seeking and sharing sensitive mental health information on Reddit.

The pseudonymous character of the user experience on the Reddit platform clearly grants a measure of autonomy for users to make relatively informed decisions about expressing sensitive mental health information. For researchers envisioning digital outreach for depression or anxiety based on monitoring the activity of Reddit users, valuing autonomy requires taking up the fundamental issue of whether Reddit users should be made aware that the signals of their participation in Reddit communities are being used as input features for the inference and assessment of mental health states. At stake is the very trust and disinhibition of Reddit users, which researchers identify as a merit of the platform as a mental health information environment [62]. Concerns related to monitoring are particularly consequential for Reddit users, as mental health states would be classified primarily through firsthand written expressions. As researchers tune the parameters of systems designed to monitor and classify written expressions of depression and anxiety that are actionable for some kind of outreach, they also inherit responsibility for deciding what constitutes proficient and mentally healthy written text. Digital outreach acting on assessments of naturalistic user text will inevitably be rooted in adherence to predetermined, standardized forms of communication that may not be applicable to users' diverse communicative choices, abilities, and styles [90]. Furthermore, it would be wise not to underestimate the potential harms to individuals who become aware that a digital outreach system has labeled them as experiencing depression and anxiety. For example, a system for monitoring user text with a threshold for initiating digital outreach that is highly sensitive will generate false positives resulting in unknown harms to the users who become recipients of inappropriately deployed digital outreach [91]. It is also likely that harms would be sustained to some users whose mental state has been accurately classified as depressive or anxious but for whom digital outreach would be unwelcomed.

Studies of depression and anxiety using Reddit data have yet to amount to any tangible impacts for Reddit users; however, the intention to shift toward professional-facing and user-facing digital outreach was a common theme among the practice implications of the included studies. For researchers invested in realizing this variety of practice implications, there is a need to define accountability and locate it within the novel

sociotechnical configurations that would be mobilized to deliver digital outreach. Stemming from the broad issue of accountability, practice implications in this research area would benefit from more substantial explorations of the meanings of autonomy in the context of depression and anxiety on the Reddit platform. Considering user autonomy from various angles will improve understandings of the nature of Reddit as an informational and social resource for depression and anxiety and, in turn, inform ethical deliberation regarding if, how, and by whom new digital outreach should be introduced.

Strengths and Limitations

There were several strengths to this review. The broad inclusion criteria led to a relatively complete picture of the different ways in which depression and anxiety are currently being studied using Reddit data. Another strength was that the findings were synthesized and discussed with the intention of heightening awareness of the conceptual and ethical issues that can be challenging to apprehend in the context of individual studies but are, nonetheless, unfolding in relation to this research area. A key limitation of this review was that it excluded studies without an explicit methodological focus on depression and anxiety, and this decision was made for reasons of feasibility in addition to conceptual coherence. Therefore, studies with an exclusive focus on topics related to depression and anxiety, such as suicidality and eating disorders, were not considered in the findings. Another limitation was that this review captured many but not all studies related to the CLEF eRisk shared tasks on early depression detection. Researchers seeking to comprehensively capture all studies of depression using Reddit data in the future should include a *CLEF eRisk* concept in their systematic search strategy, as it appears that not all CLEF eRisk studies mention Reddit in the title or abstract. Finally, in accordance with the objective of this review, no formalized risk of bias assessment or quality appraisal steps were conducted. Such steps may be more appropriate in a systematic review focused on studies featuring a shared analytic focus or those using a specific method to study depression or anxiety using Reddit data.

Conclusions

The objective of this review was to build an understanding of the scope and nature of research conducted using Reddit as a primary data source for studying depression and anxiety. A total of 54 studies were included for the review, and key features of methodological interest were communicated through tabular and descriptive means. The results demonstrated that studies of depression and anxiety using Reddit data are currently bound to a prevailing methodology that favors a technical, solution-based orientation. The discussion sheds perspective on this trajectory by highlighting the conceptual issues related to the medical model of mental health and the ethical issues pertaining to new forms of professional engagement with Reddit users for mental health prevention and treatment. This scoping review serves as a point of orientation as researchers navigate and build upon the landscape of research on depression and anxiety using Reddit data.

Acknowledgments

This paper was supported by a doctoral scholarship from the Social Sciences and Humanities Research Council of Canada. The opinions, results, and conclusions reported in this paper are those of the author and are independent of the funding source. The author thanks Morgan Jenkins for contributing as the second reviewer in the screening, full-text review, and data charting stages of this review.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Ovid MEDLINE full search strategy.
[DOCX File , 19 KB - mental_v8i11e29487_app1.docx]

References

1. 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;5(3):245-257 [FREE Full text] [doi: 10.1007/s41347-020-00134-x] [Medline: 33415185]
2. World Health Organization. Depression and Other Common Mental Disorders: Global Health Estimates. Geneva: World Health Organization; 2017:1-24.
3. Guntuku SC, Yaden DB, Kern ML, Ungar LH, Eichstaedt JC. Detecting depression and mental illness on social media: an integrative review. Curr Opin Behav Sci 2017 Dec;18:43-49. [doi: 10.1016/j.cobeha.2017.07.005]
4. Chancellor S, De Choudhury M. Methods in predictive techniques for mental health status on social media: a critical review. NPJ Digit Med 2020 Mar 24;3(1):1-11. [doi: 10.1038/s41746-020-0233-7]
5. Conway M, Hu M, Chapman WW. Recent advances in using natural language processing to address public health research questions using social media and ConsumerGenerated data. Yearb Med Inform 2019 Aug;28(1):208-217 [FREE Full text] [doi: 10.1055/s-0039-1677918] [Medline: 31419834]

6. Naslund JA, Aschbrenner KA. Risks to privacy with use of social media: understanding the views of social media users with serious mental illness. Psychiatr Serv 2019 Jul 01;70(7):561-568 [FREE Full text] [doi: 10.1176/appi.ps.201800520] [Medline: 30947635]
7. Hunter RF, Gough A, O'Kane N, McKeown G, Fitzpatrick A, Walker T, et al. Ethical issues in social media research for public health. Am J Public Health 2018 Mar;108(3):343-348. [doi: 10.2105/AJPH.2017.304249] [Medline: 29346005]
8. API. GitHub. URL: https://github.com/reddit-archive/reddit/wiki/API [accessed 2021-11-10]
9. Baumgartner J, Zannettou S, Keegan B, Squire M, Blackburn J. The pushshift reddit dataset. ArXiv.org. 2020. URL: https://arxiv.org/abs/2001.08435 [accessed 2021-11-10]
10. Lomborg S. A state of flux: histories of social media research. Eur J Commun 2016 Dec 1;32(1):6-15. [doi: 10.1177/0267323116682807]
11. Seabrook EM, Kern ML, Rickard NS. Social networking sites, depression, and anxiety: a systematic review. JMIR Ment Health 2016 Nov 23;3(4):e50 [FREE Full text] [doi: 10.2196/mental.5842] [Medline: 27881357]
12. Dredze M. How social media will change public health. IEEE Intell Syst 2012;27(4):81-84. [doi: 10.1109/mis.2012.76]
13. Walker S, Mercea D, Bastos M. The disinformation landscape and the lockdown of social platforms. Inf Commun Soc 2019 Aug 29;22(11):1531-1543. [doi: 10.1080/1369118x.2019.1648536]
14. 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: The Conference on Fairness, Accountability, and Transparency; Jan 29 - 31, 2019; Atlanta GA USA. [doi: 10.1145/3287560.3287587]
15. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
16. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
17. Peters M, Marnie C, Tricco A, Pollock D, Munn Z, Alexander L, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth 2020 Oct;18(10):2119-2126. [doi: 10.11124/JBIES-20-00167] [Medline: 33038124]
18. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
19. Grudin J. Journal-conference interaction and the competitive exclusion principle. Interactions 2013;20(1):68-73. [doi: 10.1145/2405716.2405732]
20. Bramer W, Bain P. Updating search strategies for systematic reviews using EndNote. J Med Libr Assoc 2017 Jul;105(3):285-289 [FREE Full text] [doi: 10.5195/jmla.2017.183] [Medline: 28670219]
21. Conway M, Hu M, Chapman WW. Recent Advances in Using Natural Language Processing to Address Public Health Research Questions Using Social Media and ConsumerGenerated Data. Yearb Med Inform 2019;28(01):208-217. [doi: 10.1055/s-0039-1677918]
22. Gusenbauer M, Haddaway NR. Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. Res Synth Methods 2020 Mar;11(2):181-217 [FREE Full text] [doi: 10.1002/jrsm.1378] [Medline: 31614060]
23. Anderson S, Allen P, Peckham S, Goodwin N. Asking the right questions: scoping studies in the commissioning of research on the organisation and delivery of health services. Health Res Policy Syst 2008 Jul 09;6:7 [FREE Full text] [doi: 10.1186/1478-4505-6-7] [Medline: 18613961]
24. Almeida H, Briand A, Meurs M. Detecting early risk of depression from social media user-generated content. Working Notes of CLEF 2017. URL: http://ceur-ws.org/Vol-1866/paper_127.pdf [accessed 2021-11-10]
25. Benamara F, Moriceau V, Mothe J, Ramiandrisoa F, He Z. Automatic detection of depressive users in social media. In: Proceedings of the Conférence francophone en Recherche d'Information et Applications (CORIA). 2020 Presented at: Conférence francophone en Recherche d'Information et Applications (CORIA); May 2018; Rennes, France. [doi: 10.24348/coria.2018.paper4]
26. Briand A, Almeida H, Meurs M. Analysis of social media posts for early detection of mental health conditions. In: Advances in Artificial Intelligence. Cham: Springer; 2018.
27. Bucur A, Dinu L. Detecting early onset of depression from social media text using learned confidence scores. In: Proceedings of the Seventh Italian Conference on Computational Linguistics CLiC-it. 2020 Presented at: The Seventh Italian Conference on Computational Linguistics; Nov 30-Dec 2, 2020; Bologna, Italy. [doi: 10.4000/books.aaccademia.8305]
28. Burdisso SG, Errecalde M, Montes-y-Gómez M. τ-SS3: a text classifier with dynamic n-grams for early risk detection over text streams. ArXiv.org. 2020. URL: https://arxiv.org/abs/1911.06147 [accessed 2021-11-10]
29. Farías-Anzalduá A, Montes-Y-Gómez M, Pastor LA, González-Gurrola L. UACH-INAOE participation at eRisk2017. Working Notes of CLEF 2017. 2017. URL: http://ceur-ws.org/Vol-1866/paper_136.pdf [accessed 2021-11-10]
30. Madani A, Boumahdi F, Boukenaoui A, Kritli M, Hentabli H. USDB at eRisk 2020: deep learning models to measure the severity of the signs of depression using reddit posts. Working Notes of CLEF 2020. 2020. URL: https://www.

semanticscholar.org/paper/USDB-at-eRisk-2020%3A-Deep-Learning-Models-to-Measure-Madani-Boumahdi/152e0e407bdaf0312e4bd6c60ef30fc9ac022067 [accessed 2021-11-10]
31. Paul S, Kalyani J, Basu T. Early detection of signs of anorexia and depression over social media using effective machine learning frameworks. Working Notes of CLEF 2018. 2018. URL: https://www.semanticscholar.org/paper/Early-Detection-of-Signs-of-Anorexia-and-Depression-Paul-Jandhyala/55a8a03b19286454f26a1af62a1a1ed95c0e497a [accessed 2021-11-10]
32. Ragheb W, Moulahi B, Azé J, Bringay S, Servajean M. Temporal mood variation: at the CLEF eRisk-2018 tasks for early risk detection on the internet. Working Notes of CLEF 2018. 2018. URL: https://hal.archives-ouvertes.fr/lirmm-01989632/ [accessed 2021-11-10]
33. Ragheb W, Azé J, Bringay S, Servajean M. Language modeling in temporal mood variation models for early risk detection on the internet. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction. Cham: Springer; 2019.
34. 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:- [FREE Full text] [Medline: 29075167]
35. Sadeque F, Xu D, Bethard S. Measuring the latency of depression detection in social media. In: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. 2018 Presented at: ACM International Conference on Web Search and Data Mining; Feb 5-9, 2018; Marina Del Rey, USA. [doi: 10.1145/3159652.3159725]
36. Shah F, Ahmed F, Joy S. Early depression detection from social network using deep learning techniques. In: Proceedings of the 2020 IEEE Region 10 Symposium (TENSYMP). 2020 Presented at: 2020 IEEE Region 10 Symposium; Jun 5-7, 2020; Dhaka, India. [doi: 10.1109/tensymp50017.2020.9231008]
37. Stankevich M, Isakov V, Devyatkin D, Smirnov I. Feature engineering for depression detection in social media. In: Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods - ICPRAM. 2018 Presented at: International Conference on Pattern Recognition Applications and Methods; Jan 16-18, 2018; Funchal, Portugal. [doi: 10.5220/0006598604260431]
38. Van Rijen P, Teodoro D, Naderi N. A data-driven approach for measuring the severity of the signs of depression using reddit posts. In: Working Notes of CLEF 2019 - Conference and Labs of the Evaluation Forum. 2019 Presented at: Conference and Labs of the Evaluation Forum (CLEF); Sep 9-12, 2019; Lugano, Switzerland.
39. Wang Y, Huang H, Chen H. A neural network approach to early risk detection of depression and anorexia on social media text. In: Proceedings of the Conference and Labs of the Evaluation Forum (CLEF). 2018 Presented at: Conference and Labs of the Evaluation Forum (CLEF); Sep 10-14; Avingnon, France.
40. Alva F, Wadley G, Lederman R. It feels different from real life: users' opinions of mobile applications for mental health. In: Proceedings of the Annual Meeting of the Australian Special Interest Group for Computer Human Interaction. 2015 Presented at: The Annual Meeting of the Australian Special Interest Group for Computer Human Interaction; Dec 7-10, 2015; Parkville, Australia. [doi: 10.1145/2838739.2838806]
41. Bagroy S, Kumaraguru P, Choudhury M. A social media based index of mental well-being in college campuses. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI '17: CHI Conference on Human Factors in Computing Systems; May 6-11, 2017; Denver, USA. [doi: 10.1145/3025453.3025909]
42. Balani S, De Choudhury M. Detecting and characterizing mental health related self-disclosure in social media. In: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems. 2015 Presented at: CHI '15: CHI Conference on Human Factors in Computing Systems; Apr 18-23, 2015; Seoul, Korea. [doi: 10.1145/2702613.2732733]
43. Chakravorti D, Law K, Gemmell J, Raicu D. Detecting and characterizing trends in online mental health discussions. In: Proceedings of the 2018 IEEE International Conference on Data Mining Workshops. 2018 Presented at: IEEE International Conference on Data Mining Workshops; Nov 17-20, 2018; Singapore. [doi: 10.1109/icdmw.2018.00107]
44. De Choudhury M, De S. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In: Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media. 2014 Presented at: The Eighth International AAAI Conference on Weblogs and Social Media; Jun 1-4, 2014; Ann Arbor, USA.
45. Cong Q, Feng Z, Li F, Xiang Y, Rao G, Tao C. X-A-BiLSTM: a deep learning approach for depression detection in imbalanced data. In: Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine. 2018 Presented at: 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); Dec 3-6, 2018; Madrid, Spain. [doi: 10.1109/bibm.2018.8621230]
46. Feldhege J, Moessner M, Bauer S. Who says what? Content and participation characteristics in an online depression community. J Affect Disord 2020 Feb 15;263:521-527. [doi: 10.1016/j.jad.2019.11.007] [Medline: 31780138]
47. Fraga B, Silva A, Murai F. Online social networks in health care: a study of mental disorders on reddit. In: Proceedings of the IEEE WIC ACM International Conference on Web Intelligence. 2018 Presented at: 2018 IEEE/WIC/ACM International Conference on Web Intelligence; Dec 3-6, 2018; Santiago, Chile. [doi: 10.1109/wi.2018.00-36]
48. Gaur M, Kursuncu U, Alambo A. "Let Me Tell You About Your Mental Health!": contextualized classification of reddit posts to DSM-5 for web-based intervention. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management. 2018 Presented at: CIKM '18: The 27th ACM International Conference on Information and Knowledge Management; Oct 22-26, 2018; Torino, Italy. [doi: 10.1145/3269206.3271732]

49. Gkotsis G, Oellrich A, Hubbard T. The language of mental health problems in social media. In: Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology. 2016 Presented at: Third Workshop on Computational Linguistics and Clinical Psychology; June, 2016; San Diego, USA p. 63-73. [doi: 10.18653/v1/w16-0307]
50. Gkotsis G, Oellrich A, Velupillai S, Liakata M, Hubbard TJ, Dobson RJ, et al. Characterisation of mental health conditions in social media using Informed Deep Learning. Sci Rep 2017 Mar 22;7:45141 [FREE Full text] [doi: 10.1038/srep45141] [Medline: 28327593]
51. Liu T, Liang Y, Yu Z. The influence of text length on text classification model. In: Communications in Computer and Information Science. Singapore: Springer; 2020.
52. Ireland M, Iserman M. Within and between-person differences in language used across anxiety support and neutral reddit communities. In: Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic. 2018 Presented at: Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic; Jun 5, 2018; New Orleans, USA. [doi: 10.18653/v1/w18-0620]
53. Jiang Z, Levitan S, Zomick J, Hirschberg J. Detection of mental health from reddit via deep contextualized representations. In: Proceedings of the 11th International Workshop on Health Text Mining and Information Analysis. 2020 Presented at: 11th International Workshop on Health Text Mining and Information Analysis; Nov 20, 2020; Online. [doi: 10.18653/v1/2020.louhi-1.16]
54. Kim J, Lee J, Park E, Han J. A deep learning model for detecting mental illness from user content on social media. Sci Rep 2020 Jul 16;10(1):1-6. [doi: 10.1038/s41598-020-68764-y]
55. Low DM, Rumker L, Talkar T, Torous J, Cecchi G, Ghosh SS. Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on reddit during COVID-19: observational study. J Med Internet Res 2020 Oct 12;22(10):e22635 [FREE Full text] [doi: 10.2196/22635] [Medline: 32936777]
56. Maxwell D, Praetorius RT, Keaton C, Alam K. Beyond bath bombs: partner's response to postpartum depression. Journal of Family Studies 2020:1-13. [doi: 10.1080/13229400.2020.1827007]
57. Moßburger L, Wende F, Brinkmann K, Schmidt T. Exploring Online Depression Forums via Text Mining: A Comparison of Reddit and a Curated Online Forum. In: Proceedings of the Fifth Social Media Mining for Health Applications Workshop & Shared Task. 2020 Presented at: 5th Social Media Mining for Health Applications (#SMM4H) Workshop & Shared Task; Dec 12, 2020; Barcelona, Spain.
58. Park A, Conway M. Longitudinal changes in psychological states in online health community members: understanding the long-term effects of participating in an online depression community. J Med Internet Res 2017 Mar 20;19(3):e71 [FREE Full text] [doi: 10.2196/jmir.6826] [Medline: 28320692]
59. Park A, Conway M. Harnessing reddit to understand the written-communication challenges experienced by individuals with mental health disorders: analysis of texts from mental health communities. J Med Internet Res 2018 Apr 10;20(4):e121 [FREE Full text] [doi: 10.2196/jmir.8219] [Medline: 29636316]
60. Park A, Conway M, Chen AT. Examining thematic similarity, difference, and membership in three online mental health communities from reddit: a text mining and visualization approach. Comput Human Behav 2018 Jan;78:98-112. [doi: 10.1016/j.chb.2017.09.001] [Medline: 29456286]
61. Park JYE, Howren AM, Davidson E, De Vera MA. Insights on mental health when living with rheumatoid arthritis: a descriptive qualitative study of threads on the Reddit website. BMC Rheumatol 2020;4(1):1-9. [doi: 10.1186/s41927-020-00163-2]
62. Pavalanathan U, De Choudhury M. Identity management and mental health discourse in social media. In: Proceedings of the 24th International Conference on World Wide Web. 2015 Presented at: WWW '15: 24th International World Wide Web Conference; May 18-22, 2015; Florence, Italy. [doi: 10.1145/2740908.2743049]
63. Pirina I, Çöltekin Ç. Identifying depression on reddit: the effect of training data. In: Proceedings of the 2018 EMNLP Workshop SMM4H: The 3rd Social Media Mining for Health Applications Workshop & Shared Task. 2018 Presented at: The 3rd Social Media Mining for Health Applications Workshop & Shared Task; October, 2018; Brussels, Belgium. [doi: 10.18653/v1/w18-5903]
64. Rao G, Zhang Y, Zhang L, Cong Q, Feng Z. MGL-CNN: a hierarchical posts representations model for identifying depressed individuals in online forums. IEEE Access 2020;8:32395-32403. [doi: 10.1109/access.2020.2973737]
65. Scepanovic S, Martin-Lopez E, Quercia D, Baykaner K. Extracting medical entities from social media. In: Proceedings of the ACM Conference on Health, Inference, and Learning. 2020 Presented at: ACM Conference on Health, Inference, and Learning; Apr 2-4, 2020; Toronto, Canada. [doi: 10.1145/3368555.3384467]
66. Sharma E, Choudhury M. Mental health support and its relationship to linguistic accommodation in online communities. In: Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. 2018 Presented at: CHI '18: the 2018 CHI Conference on Human Factors in Computing Systems; Apr 21-26, 2018; Montréal, Canada. [doi: 10.1145/3173574.3174215]
67. Shatte AB, 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]

68. Shen J, Rudzicz F. Detecting anxiety through reddit. In: Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology. 2017 Presented at: Fourth Workshop on Computational Linguistics and Clinical Psychology; Aug 3, 2017; Vancouver, Canada. [doi: 10.18653/v1/w17-3107]
69. Song H, You J, Chung J, Park J. Feature attention network: interpretable depression detection from social media. In: Proceedings of the 32nd Pacific Asia Conference on Language, Information and Computation. 2018 Presented at: 32nd Pacific Asia Conference on Language, Information and Computation; Dec 1-3, 2018; Hong Kong, China.
70. Tadesse MM, Lin H, Xu B, Yang L. Detection of depression-related posts in reddit social media forum. IEEE Access 2019 Apr 4;7:44883-44893. [doi: 10.1109/access.2019.2909180]
71. Tariq S, Akhtar N, Afzal H, Khalid S, Mufti MR, Hussain S, et al. A novel co-training-based approach for the classification of mental illnesses using social media posts. IEEE Access 2019 Nov 14;7:166165-166172. [doi: 10.1109/access.2019.2953087]
72. Thorstad R, Wolff P. Predicting future mental illness from social media: a big-data approach. Behav Res Methods 2019 Aug;51(4):1586-1600. [doi: 10.3758/s13428-019-01235-z] [Medline: 31037606]
73. Toulis A, Golab L. Social media mining to understand public mental health. In: Data Management and Analytics for Medicine and Healthcare. Cham: Springer; 2017.
74. Trifan A, Antunes R, Matos S, Oliveira J. Understanding depression from psycholinguistic patterns in social media texts. In: Advances in Information Retrieval. Cham: Springer; 2020.
75. Trifan A, Semeraro D, Drake J, Bukowski R, Oliveira JL. Social Media Mining for Postpartum Depression Prediction. Amsterdam, Netherlands: IOS Press; 2020.
76. Yates A, Cohan A, Goharian N. Depression and self-harm risk assessment in online forums. In: Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. 2017 Presented at: Conference on Empirical Methods in Natural Language Processing; September 7-11, 2017; Copenhagen, Denmark p. 2968-2978. [doi: 10.18653/v1/d17-1322]
77. Yoo M, Lee S, Ha T. Semantic network analysis for understanding user experiences of bipolar and depressive disorders on Reddit. Inf Process Manag 2019 Jul;56(4):1565-1575. [doi: 10.1016/j.ipm.2018.10.001]
78. Losada D, Crestani F, Parapar J. eRISK 2017: CLEF Lab on early risk prediction on the internet: experimental foundations. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction. Cham: Springer; 2017.
79. Beck AT, Steer RA, Carbin MG. Psychometric properties of the Beck Depression Inventory: twenty-five years of evaluation. Clin Psychol Rev 1988;8(1):77-100. [doi: 10.1016/0272-7358(88)90050-5]
80. Losada D, Crestani F, Parapar J. Overview of eRisk 2020: early risk prediction on the internet. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction. Cham: Springer; 2020.
81. Huda AS. The medical model and its application in mental health. Int Rev Psychiatry 2021 Aug;33(5):463-470. [doi: 10.1080/09540261.2020.1845125] [Medline: 33289576]
82. Lindgren S, Holmström J. Social science perspective on artificial intelligence. J Digit Soc Res 2020 Nov 09;2(3):1-15. [doi: 10.33621/jdsr.v2i3.65]
83. Wittgenstein L. Philosophical Investigations. West Sussex, UK: Blackwell; 2009.
84. Defossez E. No magic pills: a Burkean view on the ambiguity of mild depression. Rhetoric Rev 2020 Sep 07;39(3):330-343. [doi: 10.1080/07350198.2020.1764750]
85. Frank MR, Wang D, Cebrian M, Rahwan I. The evolution of citation graphs in artificial intelligence research. Nat Mach Intell 2019 Feb 11;1(2):79-85. [doi: 10.1038/s42256-019-0024-5]
86. Brownlie J. Looking out for each other online: digital outreach, emotional surveillance and safe(r) spaces. Emotion Space Soc 2018 May;27:60-67. [doi: 10.1016/j.emospa.2018.02.001]
87. Burr C, Morley J, Taddeo M, Floridi L. Digital psychiatry: risks and opportunities for public health and wellbeing. IEEE Trans Technol Soc 2020 Feb 28;1(1):21-33. [doi: 10.1109/tts.2020.2977059]
88. Hoeyer K, Bauer S, Pickersgill M. Datafication and accountability in public health: introduction to a special issue. Soc Stud Sci 2019 Aug;49(4):459-475 [FREE Full text] [doi: 10.1177/0306312719860202] [Medline: 31382859]
89. Klinger U, Svensson J. The end of media logics? On algorithms and agency. New Media Soc 2018 Jun 25;20(12):4653-4670. [doi: 10.1177/1461444818779750]
90. Duque C, Lashewicz B. Reframing less conventional speech to disrupt conventions of "compulsory fluency": a conversation analysis approach. Disabl Stud Q 2018 May 31;38(2):-. [doi: 10.18061/dsq.v38i2.5821]
91. Radden J. Public mental health and prevention. Pub Health Ethics 2018 Jul;11(2):126-138. [doi: 10.1093/phe/phx011]

Abbreviations

API: application programming interface
CLEF: Conference and Labs of the Evaluation Forum
DL: deep learning
ML: machine learning
NLP: natural language processing
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Extension for Scoping Reviews

Edited by J Torous; submitted 13.04.21; peer-reviewed by L Costantini, A Lamer; comments to author 25.05.21; revised version received 20.07.21; accepted 15.08.21; published 25.11.21.

Please cite as:
Boettcher N
Studies of Depression and Anxiety Using Reddit as a Data Source: Scoping Review
JMIR Ment Health 2021;8(11):e29487
URL: https://mental.jmir.org/2021/11/e29487
doi:10.2196/29487
PMID:34842560

Original Paper

A Smartphone-Based Self-Management Intervention for Individuals with Bipolar Disorder (LiveWell): Qualitative Study on User Experiences of the Behavior Change Process

Geneva K Jonathan[1], MS; Cynthia A Dopke[1], PhD; Tania Michaels[2], MD; Clair R Martin[1], BA; Chloe Ryan[3], MSW; Alyssa McBride[4], BA; Pamela Babington[1], BA; Evan H Goulding[1], MD, PhD

[1]Department of Psychiatry and Behavioral Sciences, Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

[2]Department of Pediatrics, Loma Linda Children's Hospital, Loma Linda, CA, United States

[3]Department of Social Work, UPMC Western Psychiatric Hospital, Pittsburgh, PA, United States

[4]Feinberg School of Medicine, Northwestern University, Chicago, IL, United States

Corresponding Author:
Evan H Goulding, MD, PhD
Department of Psychiatry and Behavioral Sciences
Feinberg School of Medicine
Northwestern University
680 N Lake Shore Dr Suite 1520
Chicago, IL, 60660
United States
Phone: 1 312 503 1189
Email: e-goulding@fsm.northwestern.edu

Abstract

Background: Bipolar disorder is a severe mental illness characterized by recurrent episodes of depressed, elevated, and mixed mood states. The addition of psychotherapy to pharmacological management can decrease symptoms, lower relapse rates, and improve quality of life; however, access to psychotherapy is limited. Mental health technologies such as smartphone apps are being studied as a means to increase access to and enhance the effectiveness of adjunctive psychotherapies for bipolar disorder. Individuals with bipolar disorder find this intervention format acceptable, but our understanding of how people utilize and integrate these tools into their behavior change and maintenance processes remains limited.

Objective: The objective of this study was to explore how individuals with bipolar disorder perceive and utilize a smartphone intervention for health behavior change and maintenance.

Methods: Individuals with bipolar disorder were recruited via flyers placed at university-affiliated and private outpatient mental health practices to participate in a pilot study of LiveWell, a smartphone-based self-management intervention. At the end of the study, all participants completed in-depth qualitative exit interviews. The behavior change framework developed to organize the intervention design was used to deductively code behavioral targets and determinants involved in target engagement. Inductive coding was used to identify themes not captured by this framework.

Results: In terms of behavioral targets, participants emphasized the importance of managing mood episode–related signs and symptoms. They also discussed the importance of maintaining regular routines, sleep duration, and medication adherence. Participants emphasized that receiving support from a coach as well as seeking and receiving assistance from family, friends, and providers were important for managing behavioral targets and staying well. In terms of determinants, participants stressed the important role of monitoring for their behavior change and maintenance efforts. Monitoring facilitated self-awareness and reflection, which was considered valuable for staying well. Some participants also felt that the intervention facilitated learning information necessary for managing bipolar disorder but others felt that the information provided was too basic.

Conclusions: In addition to addressing acceptability, satisfaction, and engagement, a person-based design of mental health technologies can be used to understand how people experience the impact of these technologies on their behavior change and maintenance efforts. This understanding may then be used to guide ongoing intervention development. The participants' perceptions aligned with the intervention's primary behavioral targets and use of a monitoring tool as a core intervention feature. Participant feedback further indicates that developing additional content and tools to address building and engaging social support may be

an important avenue for improving LiveWell. A comprehensive behavior change framework to understand participant perceptions of their behavior change and maintenance efforts may help facilitate ongoing intervention development.

(JMIR Ment Health 2021;8(11):e32306) doi:10.2196/32306

KEYWORDS

behavioral intervention technology; mHealth; bipolar disorder; depression; illness management; smartphone; behavior change; early warning signs; self-management; qualitative; behavior; intervention; management; user experience; perception; utilization

Introduction

Bipolar disorder (BD) is a severe mental illness characterized by recurrent episodes of depressed, elevated, and mixed mood states [1]. Episode recurrence, prolonged episodes, and interepisode symptoms often adversely impact psychosocial functioning and quality of life [2-7]. The addition of psychotherapy to pharmacological management has been shown to decrease symptoms, lower relapse rates, and improve quality of life [8-16]. Unfortunately, empirically supported adjunctive psychotherapies for BD can be hard to access because of barriers such as limited provider availability, clinic location, and financial burden [17,18]. These barriers stress the need for more cost-effective and accessible treatment modalities. Mental health technologies (MHTs) such as smartphone- and web-based interventions may be well-suited to increasing access to and enhancing the functionality of adjunctive psychotherapy for individuals with BD.

Over the last decade, MHTs have been developed for various mental health challenges [19-21]. Among individuals with BD, research indicates high rates of smartphone ownership [22] as well as interest and willingness to access BD-related information via technology [23]. To address access barriers and enhance treatment for individuals with BD, smartphone apps that port self-management strategies from empirically supported psychotherapies have been developed and individuals that use these apps report high levels of satisfaction [24-30]. Despite the emergence of these technology-delivered interventions, we still have limited knowledge of how individuals with BD experience these treatment formats [27].

Studies that use qualitative methods to evaluate individuals' lived experiences while using and applying these interventions in day-to-day activities can highlight the potential benefits and disadvantages of intervention components [31]. Despite the potential of qualitative methods to elucidate factors influencing behavior change and maintenance processes [32,33], only a small number of BD MHT studies have explored how users perceive MHT use for stimulating these processes. Of the existing studies, individuals with BD reported finding MHTs usable and useful for disease management [34,35]. More specifically, they often report that mood and activity monitoring using a smartphone can help increase insight and behavior change [36]. However, current studies have not yet comprehensively examined how MHT use influences behavior change and maintenance processes related to the multiple targets and approaches proposed to underlie living well with BD.

The current paper describes a thematic analysis of in-depth exit interviews initiated immediately after participants completed a field trial for LiveWell, a smartphone-based self-management intervention for individuals with BD. The analysis presented here focuses on how individuals with BD perceive and utilize this smartphone-based intervention for health behavior change and maintenance.

Methods

Participants

The study was reviewed and approved by the Northwestern University Institutional Review Board. Participants were recruited via flyers describing the smartphone intervention and eligibility criteria. Flyers were placed at university-affiliated and private outpatient mental health practices. Eligible participants were 18 to 65 years old and had a Diagnostic and Statistical Manual of Mental Disorders-IV diagnosis of BD 1 with a minimum of two acute mood episodes within 2 years of enrollment. Individuals were excluded if they: (1) were not in current psychiatric care; (2) met criteria for a substance use disorder within the last 6 months; (3) met criteria for another psychiatric diagnosis or had symptoms for which participation in the study was either inappropriate or dangerous, including current severe suicidal ideation or a serious suicide attempt in the last 12 months; (4) were pregnant or planned to become pregnant; (5) had visual, hearing, voice, or motor impairment that would prevent completion of the study procedures or limit smartphone use; (6) were unable to speak or read English; or (7) were in a current mood episode at the baseline assessment.

Individuals who were interested in participation were encouraged to call the research team or contact the team via the study's website. Before the initial telephone screening, participants provided informed consent for online or telephone screening. The initial telephone screening was conducted to establish a BD diagnosis using the Mini International Neuropsychiatric Interview [37]. If eligible, users completed a written consent form for study participation prior to engaging in a face-to-face interview with a study clinician (psychiatrist or psychologist). At the face-to-face interview, an abbreviated version of the Affective Disorders Evaluation and the Clinical Monitoring Form was used to confirm the diagnosis of BD Type 1 [38,39]. Individuals with a confirmed diagnosis at the clinic visit were scheduled for a baseline assessment. If these individuals were not in an episode at the baseline assessment, they were enrolled in the pilot study.

Participants were compensated for time and travel costs: (1) US $10 for travel costs and telephone assessment and (2) US $15 for each assessment, including the clinical assessment, baseline/monthly telephone assessment, exit interview, and app training. Eleven participants were enrolled in the pilot study.

The participants were 21 to 62 years old (mean 36 years, SD 14), including 4 men and 7 women. The majority (n=11) were non-Hispanic White. In terms of relationship status, 3 were married/living as married, 3 were divorced, and 5 were never married. With respect to education, 5 participants indicated some college, one had a college degree, and five had education beyond college. Two participants were students, 6 were employed, 1 was unemployed, and 2 were on disability.

Procedures

All participants were provided with a smartphone and a data plan and completed an 8-week pilot study. Participants had a face-to-face meeting with a coach who used a structured script and handouts to instruct them on the use of the app and the role of the coach [40]. Following the face-to-face meeting, participants completed six phone calls (weeks 1-4, 6, and 8) during which the coach used structured scripts to support app use adherence, development of personalized wellness plans, self-management strategy use, and communication with clinical care providers [40]. To provide feedback about the intervention's impact on app usability, target behavior change processes, and clinical and recovery outcomes, participants completed a structured exit interview (Multimedia Appendix 1) and an exit questionnaire after completing the pilot study [19].

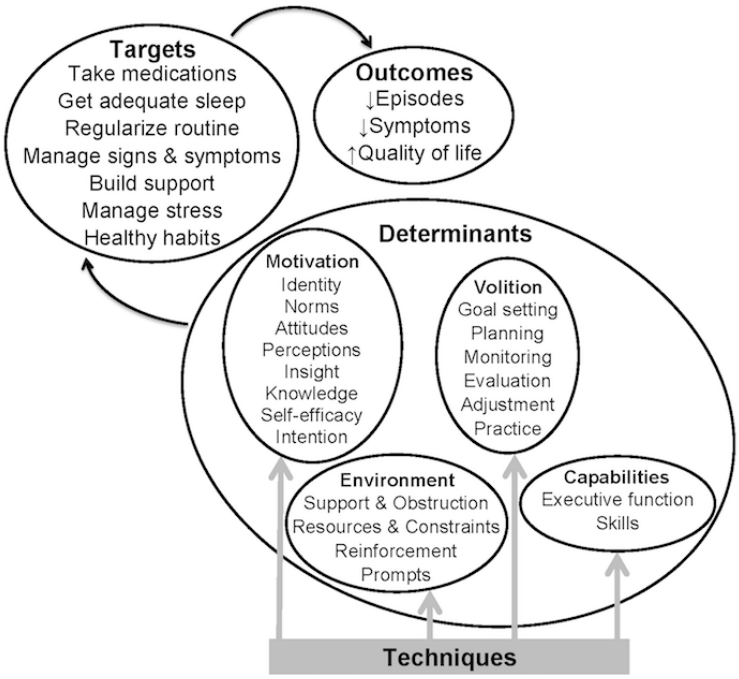
Intervention Design

The LiveWell intervention aims to assist individuals with BD in using self-management strategies to reduce relapse risk and symptom burden as well as to improve quality of life. The LiveWell intervention has technological and human support components that include a smartphone app, secure server and website, and coach. The smartphone app has five components: Foundations, Toolbox, Wellness Plan, Daily Check In, and Daily Review [19,41]. The core of the intervention is the Daily Check In, which helps participants monitor behavioral targets proposed to be important for managing BD and staying well (medication adherence, sleep duration, routine, managing signs and symptoms) [10,11]. Participants use the smartphone app to check in daily (Daily Check In) and monitor these targets. An expert system (Daily Review) provides interactive, personalized real-time feedback based on their Daily Check In data [41]. Additionally, participants have access to psychoeducational content in the Foundations and Toolbox that helps them develop a personalized Wellness Plan, which addresses lifestyle skills for reducing risk, coping skills for managing signs and symptoms, and resources essential for staying well. In addition to addressing the targets monitored with the Daily Check In, the Foundations and Toolbox also discusses attending to healthy habits concerning substance use, diet, and exercise; managing stressors; and building and using support systems to stay well. The coach supports app use adherence, self-management skill use, and clinical care communication. An initial face-to-face meeting with the coach helps participants identify personalized wellness anchors for a wellness rating scale (0 balanced, –1/+1 daily hassles/uplifts, –2/+2 prodromal/residual symptoms, –3/+3 episode, –4/+4 crisis). The wellness scale is used during the Daily Check In for monitoring signs and symptoms [40]. Screenshots of the LiveWell intervention components can be found in Multimedia Appendix 2.

Intervention Framework

LiveWell was designed using a behavior change framework that integrates user feedback with information from empirically supported psychotherapies for BD, health psychology behavior change theories, and chronic disease self-management models [19,41]. The framework proposes that (1) engaging in target behaviors improves clinical and recovery outcomes, (2) behavioral determinants govern enactment of target behaviors, and (3) exposure to behavior change technique content and tool use alter behavioral determinants (Figure 1). The framework integrates and organizes behavioral determinants defined in existing behavior change theories into four domains: motivational determinants involved in developing an intention to engage in a behavior, volitional determinants involved in enacting the behavior, environmental determinants, and capabilities that impact motivation and volition [32,33,42-58]. This framework guided the deductive coding performed during thematic analysis of participants' feedback about the impact of the intervention on their behaviors and wellness.

Figure 1. LiveWell behavior change framework.



Analysis

The exit interviews (N=11) were transcribed verbatim and used for thematic analysis [59]. Initial codes were developed using deductive coding guided by the exit interview script (Multimedia Appendix 1) and the intervention's behavior change framework [19]. Three researchers independently performed a preliminary round of coding during which transcripts were partitioned into excerpts (transcript lines conveying a codable unit) and exported to Microsoft Excel spreadsheets (Multimedia Appendix 3). Intervention subthemes, or emergent patterns that occurred within themes, were inductively coded and deductively grouped into larger themes and therein determinants. Coders used nominal group consensus, where they met with a moderator to discuss, clarify differences in coding, and finalize codes [60]. Following this process, the frequencies of all excerpts and codes were quantified. The decision to quantify the qualitative data was driven by the research team's desire to clearly identify patterns as well as inconsistencies and outliers within participant responses. Quantification of qualitative findings can help researchers recognize diversity in qualitative data [61]; provide data transparency to avoid selective cherry-picking of data [62,63]; and add precision to the presentation of findings in terms of importance, frequency, or strength of findings [61,64].

A total of 210 excerpts (mean 19, SD 8 per transcript) were given 1-5 codes (110 coded once, 65 coded twice, 23 coded three times, 10 coded four times, 2 coded five times). For each excerpt, coders identified and coded whether or not participants discussed one of LiveWell's behavioral targets (manage signs and symptoms, sleep duration, medication adherence, routine, healthy habits, build support, or none) as well as any determinants participants discussed as impacting the targets (Multimedia Appendix 3). A total of 329 codes were identified pertaining to participants' experience of the intervention impact. To account for participants who discussed a coded element frequently, a ranking score (range from 1 to 10) was assigned at each level of coding to provide a metric of how often participants discussed a given code (code count, CC) weighted by the number of participants (participant count, PC) who discussed the code:

$$\text{ranking score} = 10^{(\log[CC \times PC])/\max(\log[CC \times PC])}$$

Processing of the Excel spreadsheets to obtain counts and scores was completed using MATLAB (MathWorks). The scoring output including codes endorsed by only one or two participants is provided in Multimedia Appendix 3. Participant responses to the pilot study exit questionnaire (N=11), including questions assessing the intervention's impact on outcomes, targets, and determinants (Multimedia Appendix 4), were summarized from the 7-point response scales into two categories: disagree/strongly disagree and agree/strongly agree. Results related to app usability and user-centered development, which were also assessed during the exit interview and questionnaire analysis, have been presented elsewhere [19].

Results

Overview of Themes and Subthemes

During the exit interviews, participants discussed targets (Table 1 and Table 2) and determinants (Table 3) that aligned with LiveWell's behavior change framework. For some determinants, subthemes were also identified: monitoring (checking in, reflection, self-awareness), social support (bond, accountability, legitimacy; planning, goal-setting, monitoring, prompts), and knowledge (useful, basic) (Multimedia Appendix 3).

Table 1. Exit interview behavioral targets.

Target	Rank score (range 1-10)	Percent participants	Percent codes
Manage symptoms and signs	10.0	100.0	29.8
Routine	7.8	90.9	15.2
Sleep	7.1	81.8	11.6
Medication	6.6	63.6	13.4
Build support	6.2	81.8	9.1
Monitored[a]	4.8	63.6	5.2
Healthy habits	4.6	54.5	5.2

[a]Participants mentioned using the Daily Check In to monitor behavior but did not discuss a specific target monitored using the Daily Check In.

Table 2. Exit questionnaire targets.[a]

Behavior	Questions	Percent of participants	
		Disagree and strongly disagree	Agree and strongly agree
Manage signs and symptoms	My use of the app increased my ability to identify, monitor, and manage early warning signs and symptoms	0	82
Routine	My use of the app helped me maintain a more regular routine	0	55
Medication	My use of the app increased my medication adherence	18	45
Sleep	My use of the app helped me to get the recommended amount of sleep	0	45

[a]Only questions regarding targets are included here. Responses for two additional questions regarding outcomes and determinants are available in Multimedia Appendix 4.

Table 3. Exit interview determinants.[a]

Determinants	Rank score (range 1-10)	Percent participants	Percent codes
Volition	10.0	100.0	44.4
Monitoring	10.0	100.0	34.3
Evaluation	3.9	54.5	3.3
Adjustment	3.7	45.5	3.3
Planning	3.5	54.5	2.4
Goal setting	2.0	27.3	0.9
Environment	8.9	100.0	31.0
Social support	8.7	100.0	22.5
Constraints	3.9	63.6	3.0
Reinforcement	3.7	45.5	3.3
Prompts	2.9	36.4	2.1
Motivation	8.3	100.0	24.0
Knowledge	6.2	100.0	7.9
Intention	4.9	54.5	7.0
Insight	3.9	45.5	4.3
Self-efficacy	2.7	27.3	2.1
Attitudes	2.2	27.3	1.2
Norms	2.2	27.3	1.2

[a]Domains and determinants are included in the table if 3 or more participants discussed them in the interviews.

Targets

Participants discussed the Daily Check In targets (manage symptoms and signs, routine, sleep, medication use); building and using a support network (coach, family and friends, providers); and developing healthy habits around substance use, diet, and exercise. In terms of target behaviors, participants most frequently discussed the intervention's role in assisting them with recognition and management of their early warning signs and symptoms (Table 1).

> *[My] strongest memory is pulling my mood back from that mild up. It was a big deal. I was really glad to have that phone in my hands when I realized that was happening.* [user ID 2005, exist interview transcript start line 1118, stop line 1126; see Multimedia Appendix 3]

> *[LiveWell] helped me realize that it's okay to have mood variations, that's human, which is something I'm still dealing with determining 1s and 2s, and what's a normal variation or not.* [user ID 2016, transcript lines 62-90]

Participants also felt that the intervention impacted their efforts to keep a regular routine, get the right amount of sleep, take their medications, build a support system, and engage in healthy habits such as exercise (Table 1).

> *I definitely started trying to stay within that window of for going to bed…I started noticing when I wasn't getting the right amount of sleep or when my schedule was very off…It helps me course correct a bit faster.* [user ID 2061, transcript lines 711-732]

> *Thinking it through helped me be aware of my behaviors and my sleep patterns especially. I've been really trying to work a lot with my sleep because it helps to have it there in black and white, like this is how much I slept last night, this is how much I slept every night previous.* [user ID 2001, transcript lines 153-170]

In responding to exit questionnaires about the intervention's perceived utility on making changes in behaviors (Table 2), participants' responses aligned with the thematic interviews. Most participants reported that the intervention helped increase their ability to identify, monitor, and manage early warning signs. Additionally, some participants found that the intervention helped with developing a consistent routine and optimizing sleep duration. While some participants felt that app use helped medication adherence, two participants did not find it helpful for this target behavior. However, these individuals reported 100% adherence to their medications upon starting the intervention.

Determinants

Monitoring

In terms of determinants, participants most frequently discussed how monitoring their behaviors provided an opportunity to identify and make progress toward their behavioral target goals (Table 3). They pointed out that the Daily Check In was especially helpful for monitoring symptoms (Multimedia Appendix 3).

> *The Daily Check In [worked best for me]. Knowing that I was being monitored. That I was gonna get help or recommendations on what to do. That was great. Knowing that I was really tracking what was going on and becoming more aware of what to look for.* [user ID 2041, transcript lines 787-757]

In particular, participants expressed that monitoring helped with managing early warning signs (Multimedia Appendix 3).

> *I noticed when I was having a mild up-phase. I don't think I would've noticed it without the personalized anchors that's something that I tend to have less insight about. I'm like… I've had a couple of plus two days and this, this and this is happening… It was nice to look at [my Wellness Plan] and say okay I have some things that I can do to try to bring this down and if it doesn't go down I know that I need to make a phone call.* [user ID 2005, transcript lines 1104-1116]

Additionally, daily monitoring enabled some participants to make plans involving their supports to improve their target behaviors such as sleep:

> *If I have 2 or 3 nights with less than 6 hours of sleep, something is gonna happen so I make sure my husband is the person who takes care of the kids that night and I'll sleep in the guest bedroom.* [user ID 2086, transcript lines 190-207]

Support

After monitoring, the determinant that participants discussed most frequently as being important to behavior change was social support from coaches, family, friends, and providers (Table 3). With regard to coaching interactions, participants brought up components of the supportive accountability model, which argues that human support increases adherence through accountability to a coach who is deemed as trustworthy, benevolent, and having expertise [65]. Specifically, participants reported that they liked, trusted, or respected their coaches (bond). They also acknowledged that the coach helped to keep them responsible when they were unable to meet their mutually agreed upon goals (accountability): "I like the idea of working with the coach…Just having someone to check in with about it and kind of also be accountable to it" (user ID 2065, transcript lines 555-559).

Additionally, participants discussed the coaches' influence due to their perceived expertise on BD-related topics such as the target behaviors (Multimedia Appendix 3): "You just felt like somebody [coach] was listening and monitoring what was going on in your life and helping you figure out if you're going too far this way or too far that way" (user ID 2041, transcript lines 672-685).

Participants also shared that working with a coach helped them carry out volitional processes such as goal-setting, monitoring of signs and symptoms, and using prompts to work toward achieving target behaviors such as medication adherence: "My

conversations with [my coach] were a little more helpful in terms of figuring out exactly what to do in terms of keeping a routine and taking my meds at the right time" (user ID 2016, transcript lines 152-160).

Similarly, participants cited components of supportive accountability (legitimacy and accountability) inspired by the involvement of their providers (Multimedia Appendix 3). For example, some participants felt comforted by the idea that their providers had access to their Daily Check In data.

> *[My therapist] looked at [my clinical summary] a couple times and found it useful. I see her every other week. So she referred back to it the first time…She mentioned something like "Yeah I saw this has been your pattern" and I was like "What? Oh yeah, that's right…" It actually was kind of—can I say comforting?…There was a sense of...I don't know the word, but just that she's looking at it as well.* [user ID 2065, transcript lines 591-633]

They also noted that the intervention provided a means to share information with family members about BD and engage their support network to assist them with volitional processes such as planning and monitoring: "There was a lot of good information in there...to be more reflective of what's going on… and to involve people more directly, specifically my daughter" (user ID 2066, transcript lines 84-92).

Knowledge

The third most discussed determinant was knowledge (Table 3). Most participants found the knowledge offered about BD useful and emphasized that sharing this information with friends and family was particularly beneficial.

> *The recommendations...[are] really good stuff to know and things I could share with my family and support people, so they know what to look out for or what I'm looking out for.* [user ID 2041, transcript lines 215-234]

> *Foundations are good for people who are maybe newer to the disease or if I were to share that with friends and family.* [user ID 2086, transcript lines 166-181]

However, some participants felt that the content was a review of familiar information and wanted more advanced materials*:* "[The foundations] weren't totally new to me, because I've done a lot of DBT [dialectical behavioral therapy]...I'm someone who's been through a lot of therapy" (user ID 2061, transcript lines 92-103).

Motivation

In addition to knowledge, participants also discussed other motivational determinants, including intention, self-efficacy, insight, attitudes, and norms (Table 3). Participants discussed their intentions and sense of self-efficacy in developing more regular routines and better sleep habits, as well as managing symptoms and signs, and taking medications.

> *The thing that helped me the most was trying to stick to a routine… I needed more routine.* [user ID 2063, transcript lines 6-20]

> *I have a hard time making myself follow a routine or a structure... I'm not good at doing that.* [user ID 2066, transcript lines 112-121]

Participants also stated that the Daily Check In and Wellness Plan helped them develop insight by building their self-awareness about symptoms and encouraging daily reflection about their illness experience.

> *Personalizing the information was really helpful, like within the Wellness Plan, within triggers…again just because it made me so much more aware of myself.* [user ID 2066, transcript lines 720-730]

> *[My strongest memory was] definitely the check in and the rating of myself. That was the biggest part of the check in for me to have that time to sit down and really say like okay for the last 24 hours how was I really? I'm good now but let's think back, or I'm not doing so well, what happened in the last 24 hours? Was it situational or was it not situational?* [user ID 2086, transcript lines 109-111]

Moreover, participants discussed how using the app impacted their perceptions about medications and their attitudes regarding the importance of medications and sleep duration.

> *Medications…were my kind of thing. Not really that I had negative beliefs or anything about medications but just why they are important, and even if they don't feel like they are important one day, they are probably important the next day.* [user ID 2016, transcript lines 105-115]

Regarding identifying and managing signs and symptoms, they noted that norms about what others think and do were useful and reassuring*:* "[The Wellness Plan] kind of helped to normalize things, like, or, put things more into perspective. Like if, you know, this is what the standard you know" (user ID 2063, transcript lines 183-214).

Volition

In addition to monitoring, participants also discussed other volitional determinants, including evaluation, adjustment, planning, and goal-setting (Table 3). Participants that engaged in evaluation also discussed how this process prompted them to adjust their behaviors to improve their overall wellness*:*

> *On the few times that I was having kind of some mild depression symptoms "OK you gotta dial it up" and when I was having a small bout of hypomanic symptoms "OK…dial it down don't talk so much, slow down."* [user ID 2005, transcript lines 116-123]

Similarly, some participants noted that the Daily Check In encouraged them to evaluate patterns in their behavior and whether or not these patterns aligned with their behavioral goals.

> *The daily check in, you have that moment of looking back and seeing what happened. When I dipped down to that 2, I realized that I was going down a path. I had my early warning signs...I went to the wellness plan and did look it over.* [user ID 2086, transcript lines 219-237]

Environment

In addition to social support, participants discussed other environmental determinants, including constraints, reinforcement, and prompts (Table 3). Participants acknowledged that their physical environment such as a new job or varying school schedules constrained their ability to make changes in their target behaviors such as routine: "It's hard being a student and having a regular routine" (user ID 2016, transcript lines 597-613).

Participants also stated that wanting to obtain high percentage scores on the daily review feedback bar charts reinforced their efforts to take their medications and get the right amount of sleep: "I had a hard time remembering to take my medications and being motivated once I did forget to take them…with [LiveWell] you could at least say 100% every day on medication so that really helped" (user ID 2041, transcript lines 5-13).

Despite occasional difficulties with their surrounding environments, participants reported that LiveWell helped to identify physical stimuli that helped remind them to engage in a behavior, such as taking medications: "LiveWell was a reminder to take my meds. If I wasn't going to bed, I would remember to put them close to my bed" (user ID 2016, transcript lines 638-649).

Discussion

A person-based approach was used to explore participants' experience of a smartphone-based self-management intervention for BD. Participants' accounts highlighted how they perceived the intervention impacting their efforts to stay well. Deductive thematic analysis of participants' experiences identified behavioral targets and determinants that aligned with LiveWell's behavior change framework and several subthemes also emerged from inductive analysis.

In terms of behavioral targets, participants most frequently discussed the importance of learning about and making an effort to manage signs and symptoms, suggesting that this target's inclusion was highly valued. Most participants also expressed that keeping a regular routine, getting the right amount of sleep, taking medications as prescribed, and engaging in healthy habits (eg, proper diet and exercise) were target behaviors they felt were necessary to address. However, some participants discussed difficulties managing these target behaviors, especially balancing the maintenance of a regular routine with environmental constraints. Furthermore, two participants, who started the intervention reporting 100% adherence with their psychiatric medication use, indicated that the intervention did not help with medication adherence. Their feedback highlights the importance of recognizing that not all targets may be applicable or relevant to all participants. Thus, addressing baseline target behavior may be useful in identifying whether or not participants need support for behavior change or maintenance concerning a specific target.

Among behavioral determinants, participants felt that monitoring played a significant role in staying well. In particular, participants felt that regular monitoring enhanced their ability to identify and manage early warning signs and symptoms. In addition, about half of the participants discussed how monitoring helped them develop a regular routine, optimize sleep duration, and adhere to medication regimes. This feedback is consistent with existing smartphone intervention studies in which individuals with BD indicated that mood and activity monitoring helped to identify the relationship between mood states, sleep, exercise, and changes in behavior [36,66]. Data from empirically supported psychotherapies for BD indicate that the ability to distinguish between early warning signs and transitioning into an episode improves clinical outcomes [67,68]. This finding suggests that monitoring using a smartphone app may lead to improved clinical outcomes for individuals with BD.

Participants also expressed that monitoring using the Daily Check In led to increased reflection and awareness that helped them manage signs and symptoms and other target behaviors. This report from participants suggests that monitoring helps individuals build insight, including awareness of having BD, the presence of symptoms and their consequences, and the need for treatment. Enhanced awareness and reflection due to monitoring may have important implications for improved outcomes [69]. Higher levels of insight about BD such as better awareness of the illness, particularly awareness of the need for treatment, is associated with better medication adherence [70], higher self-reported quality of life [71], and increased potential to slow the progression of symptoms into a full-blown mood episode [72]. Participants' discussion of the impact of monitoring on aspects of insight such as self-awareness of signs and symptoms also reveals that determinants (ie, monitoring and insight) may interact with one another in addition to impacting targets. This interaction of determinants is consistent with chronic disease self-management models, which consider behavior change and maintenance processes as involving a continuous and reciprocal system in which multiple wellness outcomes, target behaviors, and behavioral determinants interact continuously and reciprocally to impact health behavior change [73,74].

In addition to monitoring, participants frequently described social support as critical to their efforts to stay well. Participants underscored bond, accountability, and legitimacy as crucial components of the coaches' influence in motivating them to use the app and self-management strategies to achieve their target goals. This finding aligns with literature suggesting that the inclusion of human support helps make interventions more personally relevant and may improve engagement and decrease attrition [19,38,40,42]. Participants also expressed that coaches helped assist them with volitional determinants such as goal-setting and monitoring. In addition, participants highlighted that support from family, friends, and health care professionals was valuable for making plans and monitoring target behaviors. This feedback is consistent with previous research indicating that a lack of social support can hinder self-management and that calling on trusted individuals for assistance is essential for chronic disease self-management [75-78]. Participants' discussion of the value of obtaining assistance from family and friends and working with their providers suggests that incorporating additional content and tools to aid participants in building support would be valued, utilized, and may improve intervention outcomes. Although the intervention contained

content about building and seeking support in BD self-management, this was not a primary target behavior. The emphasis that participants placed on the important role of social supports in managing behavioral targets and staying well suggests that increasing intervention content to assist individuals in building and seeking support may be an effective means to improve self-management interventions for BD and other mental health conditions.

The significance of monitoring and social support as valuable determinants may offer insight into how to strengthen the role of these determinants in future MHTs. First, monitoring has been discussed widely as an essential component of managing wellness for BD in MHTs and traditional face-to-face therapy settings [11,19,66,79,80]. As evidenced by the targets the participants discussed in our qualitative interviews, monitoring can apply to multiple behavioral targets. Given that the BD symptom experience at the individual level varies, MHT users may want to emphasize or focus on different behavioral targets for monitoring purposes. For instance, if an individual reports self-efficacy in medication adherence but difficulty with managing sleep duration, providing said individual with an option to opt out of medication adherence or adding a more relevant behavioral target may strengthen the user's personal connection and engagement with the MHT. Participants' identification of social support as a meaningful determinant also has important implications for how and in what capacity to involve social elements within MHTs. LiveWell did not include an opportunity for participants to engage in peer-to-peer discussion and exchange ideas. However, previous MHTs that have integrated a social component have demonstrated that the connection and support participants generate provide positive reinforcement and encouragement to engage in target behaviors [81,82]. Our research team sought to address feedback about the value of social support by enhancing the coach assistance that participants received throughout the study [40]. Future MHTs may consider integrating social support from family, friends, and providers more readily into the intervention.

One limitation of this study is that the interviews were conducted using an interview guide primarily focused on feasibility, usability, and satisfaction. Due to this interview guide format, it is possible that these qualitative findings do not represent the breadth of responses that may have been discovered had the research team asked directly about behavior change or had used more broadly open questions about the intervention's impact on day-to-day life and wellness. However, despite the lack of direct questioning about behavior change, participants still spoke in-depth about their experiences integrating the intervention in their behavior change processes. Another limitation is that the research team mainly used a deductive coding approach utilizing the intervention's underlying behavior change framework, which integrates information from empirically supported psychotherapies for BD [5,6,9,12-14,83-85], health psychology behavior change theories [32,33,42-58], and chronic disease self-management models [73,86-91]. This approach may have impacted the ability to identify novel themes from participants' responses. Nevertheless, the researchers identified several inductive subthemes such as reflection, awareness, and insight in developing their codebook. In addition, quantifying qualitative information based on themes may reduce the rich interpretation of data expressed during participant interviews [92]. Quantifying the frequency of discussion may also not capture whether participant comments are positive or negative. Finally, because of the study's small and limited sample size, the results are likely not representative of all people who may utilize a smartphone self-management intervention for BD. This limits inferences that can be drawn about the prevalence of these findings beyond the current sample.

Utilizing qualitative approaches to understand how participants perceive the impact of technologies on their behavior change and maintenance processes provides an opportunity to understand how these technologies are integrated into their daily lives. As a result, qualitative approaches may highlight how MHTs can be developed to better meet participants' needs. In this study, participants discussed the importance of several behavioral targets and determinants that the intervention aimed to address, suggesting that the intervention framework and design aligned with participants' needs and interests. However, participants also emphasized the importance of gaining support from family and friends, even though support was not emphasized or extensively developed as a target in the intervention content and tools. Using person-based development approaches to move beyond examining usability to comprehensively examine how participants perceive MHTs impacting their behavior change and maintenance efforts may thus provide new ideas about how to design and improve these technologies.

Acknowledgments

We would like to thank our study participants for their time and valuable input. The research reported in this publication was supported by the National Institute of Mental Health of the National Institutes of Health under award numbers R34MH100460 and R01MH110626.

Conflicts of Interest

EG has accepted honoraria from Otsuka Pharmaceuticals. The other authors have no conflicts of interest to declare.

Multimedia Appendix 1
LiveWell pilot study exit interview script.
[PDF File (Adobe PDF File), 116 KB - mental_v8i11e32306_app1.pdf]

Multimedia Appendix 2
LiveWell app screenshots.
[PDF File (Adobe PDF File), 1439 KB - mental_v8i11e32306_app2.pdf]

Multimedia Appendix 3
LiveWell pilot study exit interview data.
[XLSX File (Microsoft Excel File), 120 KB - mental_v8i11e32306_app3.xlsx]

Multimedia Appendix 4
LiveWell pilot study exit questionnaire data.
[XLSX File (Microsoft Excel File), 13 KB - mental_v8i11e32306_app4.xlsx]

References

1. DSM-5: Diagnostic Statistical Manual of Mental Disorders. Washington, DC: American Psychiatric Association; 2013.
2. Lam DH, Watkins ER, Hayward P, Bright J, Wright K, Kerr N, et al. A randomized controlled study of cognitive therapy for relapse prevention for bipolar affective disorder: outcome of the first year. Arch Gen Psychiatry 2003 Feb;60(2):145-152. [doi: 10.1001/archpsyc.60.2.145] [Medline: 12578431]
3. Judd LL, Akiskal HS, Schettler PJ, Endicott J, Maser J, Solomon DA, et al. The long-term natural history of the weekly symptomatic status of bipolar I disorder. Arch Gen Psychiatry 2002 Jun;59(6):530-537. [doi: 10.1001/archpsyc.59.6.530] [Medline: 12044195]
4. Judd LL, Schettler PJ, Akiskal HS, Coryell W, Leon AC, Maser JD, et al. Residual symptom recovery from major affective episodes in bipolar disorders and rapid episode relapse/recurrence. Arch Gen Psychiatry 2008 Apr;65(4):386-394. [doi: 10.1001/archpsyc.65.4.386] [Medline: 18391127]
5. Miklowitz DJ. Adjunctive psychotherapy for bipolar disorder: state of the evidence. Am J Psychiatry 2008 Nov;165(11):1408-1419 [FREE Full text] [doi: 10.1176/appi.ajp.2008.08040488] [Medline: 18794208]
6. Crowe M, Whitehead L, Wilson L, Carlyle D, O'Brien A, Inder M, et al. Disorder-specific psychosocial interventions for bipolar disorder--a systematic review of the evidence for mental health nursing practice. Int J Nurs Stud 2010 Jul;47(7):896-908. [doi: 10.1016/j.ijnurstu.2010.02.012] [Medline: 20233615]
7. World Health Organization. The World Health Report 2001: Mental Health: New Understanding, New Hope. Geneva, Switzerland: World Health Organization; 2001.
8. Goodwin GM, Consensus Group of the British Association for Psychopharmacology. Evidence-based guidelines for treating bipolar disorder: revised second edition--recommendations from the British Association for Psychopharmacology. J Psychopharmacol 2009 Jun;23(4):346-388. [doi: 10.1177/0269881109102919] [Medline: 19329543]
9. Gregory VL. Cognitive-behavioral therapy for depression in bipolar disorder: a meta-analysis. J Evid Based Soc Work 2010 Jul;7(4):269-279. [doi: 10.1080/15433710903176088] [Medline: 20799127]
10. Miklowitz DJ, Goodwin GM, Bauer MS, Geddes JR. Common and specific elements of psychosocial treatments for bipolar disorder: a survey of clinicians participating in randomized trials. J Psychiatr Pract 2008 Mar;14(2):77-85 [FREE Full text] [doi: 10.1097/01.pra.0000314314.94791.c9] [Medline: 18360193]
11. Miklowitz DJ, Scott J. Psychosocial treatments for bipolar disorder: cost-effectiveness, mediating mechanisms, and future directions. Bipolar Disord 2009 Jun;11(Suppl 2):110-122. [doi: 10.1111/j.1399-5618.2009.00715.x] [Medline: 19538690]
12. Scott J, Colom F, Vieta E. A meta-analysis of relapse rates with adjunctive psychological therapies compared to usual psychiatric treatment for bipolar disorders. Int J Neuropsychopharmacol 2007 Feb;10(1):123-129 [FREE Full text] [doi: 10.1017/S1461145706006900] [Medline: 16787554]
13. Lam DH, Burbeck R, Wright K, Pilling S. Psychological therapies in bipolar disorder: the effect of illness history on relapse prevention - a systematic review. Bipolar Disord 2009 Aug;11(5):474-482. [doi: 10.1111/j.1399-5618.2009.00724.x] [Medline: 19624386]
14. Schöttle D, Huber CG, Bock T, Meyer TD. Psychotherapy for bipolar disorder: a review of the most recent studies. Curr Opin Psychiatry 2011 Nov;24(6):549-555. [doi: 10.1097/YCO.0b013e32834b7c5f] [Medline: 21918448]
15. Olfson M, Marcus SC, Druss B, Pincus HA. National trends in the use of outpatient psychotherapy. Am J Psychiatry 2002 Nov;159(11):1914-1920. [doi: 10.1176/appi.ajp.159.11.1914] [Medline: 12411228]
16. 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]
17. Busch AB, Ling D, Frank RG, Greenfield SF. Changes in the quality of care for bipolar I disorder during the 1990s. Psychiatr Serv 2007 Jan;58(1):27-33 [FREE Full text] [doi: 10.1176/ps.2007.58.1.27] [Medline: 17215409]
18. Wang PS, Lane M, Olfson M, Pincus HA, Wells KB, Kessler RC. Twelve-month use of mental health services in the United States: results from the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):629-640. [doi: 10.1001/archpsyc.62.6.629] [Medline: 15939840]

19. Jonathan GK, Dopke CA, Michaels T, Bank A, Martin CR, Adhikari K, et al. A smartphone-based self-management intervention for bipolar disorder (LiveWell): user-centered development approach. JMIR Ment Health 2021 Apr 12;8(4):e20424 [FREE Full text] [doi: 10.2196/20424] [Medline: 33843607]
20. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct 21;16(3):287-298. [doi: 10.1002/wps.20472] [Medline: 28941113]
21. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
22. Torous J, Chan SR, Yee-Marie Tan S, Behrens J, Mathew I, Conrad EJ, et al. Patient smartphone ownership and interest in mobile apps to monitor symptoms of mental health conditions: a survey in four geographically distinct psychiatric clinics. JMIR Ment Health 2014 Dec 23;1(1):e5 [FREE Full text] [doi: 10.2196/mental.4004] [Medline: 26543905]
23. Torous J, Wisniewski H, Liu G, Keshavan M. Mental health mobile phone app usage, concerns, and benefits among psychiatric outpatients: comparative survey study. JMIR Ment Health 2018 Nov 16;5(4):e11715 [FREE Full text] [doi: 10.2196/11715] [Medline: 30446484]
24. Todd NJ, Jones SH, Hart A, Lobban FA. A web-based self-management intervention for Bipolar Disorder 'living with bipolar': a feasibility randomised controlled trial. J Affect Disord 2014 Dec;169:21-29. [doi: 10.1016/j.jad.2014.07.027] [Medline: 25129531]
25. Depp CA, Ceglowski J, Wang VC, Yaghouti F, Mausbach BT, Thompson WK, et al. Augmenting psychoeducation with a mobile intervention for bipolar disorder: a randomized controlled trial. J Affect Disord 2015 Mar 15;174:23-30 [FREE Full text] [doi: 10.1016/j.jad.2014.10.053] [Medline: 25479050]
26. Depp CA, Mausbach B, Granholm E, Cardenas V, Ben-Zeev D, Patterson TL, et al. Mobile interventions for severe mental illness: design and preliminary data from three approaches. J Nerv Ment Dis 2010 Oct;198(10):715-721 [FREE Full text] [doi: 10.1097/NMD.0b013e3181f49ea3] [Medline: 20921861]
27. Gliddon E, Barnes SJ, Murray G, Michalak EE. Online and mobile technologies for self-management in bipolar disorder: a systematic review. Psychiatr Rehabil J 2017 Sep;40(3):309-319. [doi: 10.1037/prj0000270] [Medline: 28594196]
28. Ben-Zeev D, Davis KE, Kaiser S, Krzsos I, Drake RE. Mobile technologies among people with serious mental illness: opportunities for future services. Adm Policy Ment Health 2013 Jul;40(4):340-343 [FREE Full text] [doi: 10.1007/s10488-012-0424-x] [Medline: 22648635]
29. Wenze SJ, Armey MF, Miller IW. Feasibility and acceptability of a mobile intervention to improve treatment adherence in bipolar disorder: a pilot study. Behav Modif 2014 Jul 08;38(4):497-515 [FREE Full text] [doi: 10.1177/0145445513518421] [Medline: 24402464]
30. Depp CA, Perivoliotis D, Holden J, Dorr J, Granholm EL. Single-session mobile-augmented intervention in serious mental illness: a three-arm randomized controlled trial. Schizophr Bull 2019 Jun 18;45(4):752-762 [FREE Full text] [doi: 10.1093/schbul/sby135] [Medline: 30281086]
31. Pope C, Mays N. Reaching the parts other methods cannot reach: an introduction to qualitative methods in health and health services research. BMJ 1995 Jul 01;311(6996):42-45 [FREE Full text] [doi: 10.1136/bmj.311.6996.42] [Medline: 7613329]
32. Schwarzer R. Modeling health behavior change: how to predict and modify the adoption and maintenance of health behaviors. Appl Psychol 2008 Jan;57(1):1-29. [doi: 10.1111/j.1464-0597.2007.00325.x]
33. 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]
34. Bardram J, Frost M, Szántó K, Faurholt-Jepsen M, Vinberg M, Kessing L. A field trial of the Monarca system. 2013 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 2013; Paris, France p. 2627-2636. [doi: 10.1145/2470654.2481364]
35. Frost M, Doryab A, Faurholt-Jepsen M, Kessing L, Bardram J. Supporting disease insight through data analysis: refinements of the Monarca self-assessment system. 2013 Presented at: 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 2013; Zurich, Switzerland p. 133-142. [doi: 10.1145/2493432.2493507]
36. Saunders K, Bilderbeck A, Panchal P, Atkinson L, Geddes J, Goodwin G. Experiences of remote mood and activity monitoring in bipolar disorder: a qualitative study. Eur Psychiatry 2017 Mar 27;41(1):115-121 [FREE Full text] [doi: 10.1016/j.eurpsy.2016.11.005] [Medline: 28135594]
37. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59(Suppl 20):22-33;quiz 34. [Medline: 9881538]
38. Sachs GS, Guille C, McMurrich SL. A clinical monitoring form for mood disorders. Bipolar Disord 2002 Oct;4(5):323-327. [doi: 10.1034/j.1399-5618.2002.01195.x] [Medline: 12479665]
39. Sachs GS, Thase ME, Otto MW, Bauer M, Miklowitz D, Wisniewski SR, et al. Rationale, design, and methods of the systematic treatment enhancement program for bipolar disorder (STEP-BD). Biol Psychiatry 2003 Jun 01;53(11):1028-1042. [doi: 10.1016/s0006-3223(03)00165-3] [Medline: 12788248]

40. Dopke CA, McBride A, Babington P, Jonathan GK, Michaels T, Ryan C, et al. Development of coaching support for LiveWell: a smartphone-based self-management intervention for bipolar disorder. JMIR Form Res 2021 Mar 24;5(3):e25810 [FREE Full text] [doi: 10.2196/25810] [Medline: 33759798]
41. Goulding EH, Dopke CA, Michaels T, Martin CM, Khiani MA, Karr C, et al. A smartphone-based self-management intervention for individuals with bipolar disorder (LiveWell): development of an expert system to provide adaptive user feedback. JMIR Form Res 2021:preprint [FREE Full text] [doi: 10.2196/32932]
42. Weed LL. Medical records that guide and teach. N Engl J Med 1968 Mar 21;278(12):652-657. [doi: 10.1056/NEJM196803212781204] [Medline: 5637250]
43. Zimmerman BJ. A social cognitive view of self-regulated academic learning. J Educ Psychol 1989;81(3):329-339. [doi: 10.1037/0022-0663.81.3.329]
44. Bartholomew LK, Parcel GS, Kok G. Intervention mapping: a process for developing theory- and evidence-based health education programs. Health Educ Behav 1998 Oct;25(5):545-563. [doi: 10.1177/109019819802500502] [Medline: 9768376]
45. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
46. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
47. Locke EA, Latham GP. Building a practically useful theory of goal setting and task motivation. A 35-year odyssey. Am Psychol 2002 Sep;57(9):705-717. [doi: 10.1037//0003-066x.57.9.705] [Medline: 12237980]
48. Bandura A. Health promotion by social cognitive means. Health Educ Behav 2004 Apr;31(2):143-164. [doi: 10.1177/1090198104263660] [Medline: 15090118]
49. Markland D, Ryan RM, Tobin VJ, Rollnick S. Motivational interviewing and self–determination theory. J Soc Clin Psychol 2005 Sep 21;24(6):811-831. [doi: 10.1521/jscp.2005.24.6.811] [Medline: 5637250]
50. Latham GP, Locke EA. New developments in and directions for goal-setting research. Eur Psychol 2007 Jan;12(4):290-300. [doi: 10.1027/1016-9040.12.4.290]
51. Schwarzer R, Lippke S, Luszczynska A. Mechanisms of health behavior change in persons with chronic illness or disability: the Health Action Process Approach (HAPA). Rehabil Psychol 2011 Aug;56(3):161-170. [doi: 10.1037/a0024509] [Medline: 21767036]
52. Su Y, Reeve J. A meta-analysis of the effectiveness of intervention programs designed to support autonomy. Educ Psychol Rev 2010 Sep 25;23(1):159-188. [doi: 10.1007/s10648-010-9142-7]
53. 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]
54. Ries F, Hein V, Pihu M, Armenta JMS. Self-identity as a component of the Theory of Planned Behaviour in predicting physical activity. Eur Phys Educ Rev 2012 Oct 02;18(3):322-334. [doi: 10.1177/1356336x12450792]
55. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. Implement Sci 2012 Apr 24;7:37 [FREE Full text] [doi: 10.1186/1748-5908-7-37] [Medline: 22530986]
56. Jones CJ, Smith H, Llewellyn C. Evaluating the effectiveness of health belief model interventions in improving adherence: a systematic review. Health Psychol Rev 2014;8(3):253-269. [doi: 10.1080/17437199.2013.802623] [Medline: 25053213]
57. Cane J, Richardson M, Johnston M, Ladha R, Michie S. From lists of behaviour change techniques (BCTs) to structured hierarchies: comparison of two methods of developing a hierarchy of BCTs. Br J Health Psychol 2015 Feb;20(1):130-150. [doi: 10.1111/bjhp.12102] [Medline: 24815766]
58. Berkman ET, Livingston JL, Kahn LE. Finding the “self” in self-regulation: the identity-value model. Psychol Inq 2017 Aug 18;28(2-3):77-98. [doi: 10.1080/1047840x.2017.1323463]
59. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
60. Potter M, Gordon S, Hamer P. The nominal group technique: a useful consensus methodology in physiotherapy research. New Zealand J Physio 2004;32:126-130.
61. Maxwell JA. Using numbers in qualitative research. Qual Inq 2010 Apr 15;16(6):475-482. [doi: 10.1177/1077800410364740]
62. Hannah DR, Lautsch BA. Counting in qualitative research: why to conduct it, when to avoid it, and when to closet it. J Manag Inq 2010 Sep 08;20(1):14-22. [doi: 10.1177/1056492610375988]
63. Chivanga SY. Qualitative research methodology and numbers. J Soc Sci 2017 Oct 11;47(2):119-122. [doi: 10.1080/09718923.2016.11893551]
64. Neale J, Miller P, West R. Reporting quantitative information in qualitative research: guidance for authors and reviewers. Addiction 2014 Feb 14;109(2):175-176. [doi: 10.1111/add.12408] [Medline: 24422609]
65. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]

66. Michalak EE, Morton E, Barnes SJ, Hole R, CREST.BD, Murray G. Supporting self-management in bipolar disorder: mixed-methods knowledge translation study. JMIR Ment Health 2019 Apr 15;6(4):e13493 [FREE Full text] [doi: 10.2196/13493] [Medline: 30985287]
67. Lam D, Wong G. Prodromes, coping strategies and psychological interventions in bipolar disorders. Clin Psychol Rev 2005 Dec;25(8):1028-1042. [doi: 10.1016/j.cpr.2005.06.005] [Medline: 16125292]
68. Joyce PR. Illness behaviour and rehospitalization in bipolar affective disorder. Psychol Med 1985 Aug 09;15(3):521-525. [doi: 10.1017/s0033291700031391] [Medline: 4048314]
69. Látalová K. Insight in bipolar disorder. Psychiatr Q 2012 Sep 19;83(3):293-310. [doi: 10.1007/s11126-011-9200-4] [Medline: 22101737]
70. Sajatovic M, Ignacio RV, West JA, Cassidy KA, Safavi R, Kilbourne AM, et al. Predictors of nonadherence among individuals with bipolar disorder receiving treatment in a community mental health clinic. Compr Psychiatry 2009;50(2):100-107 [FREE Full text] [doi: 10.1016/j.comppsych.2008.06.008] [Medline: 19216885]
71. Dias VV, Brissos S, Frey BN, Kapczinski F. Insight, quality of life and cognitive functioning in euthymic patients with bipolar disorder. J Affect Disord 2008 Sep;110(1-2):75-83. [doi: 10.1016/j.jad.2008.01.010] [Medline: 18272231]
72. Perry A, Tarrier N, Morriss R, McCarthy E, Limb K. Randomised controlled trial of efficacy of teaching patients with bipolar disorder to identify early symptoms of relapse and obtain treatment. BMJ 1999 Jan 16;318(7177):149-153 [FREE Full text] [doi: 10.1136/bmj.318.7177.149] [Medline: 9888904]
73. Clark NM, Gong M, Kaciroti N. A model of self-regulation for control of chronic disease. Health Educ Behav 2001 Dec 30;28(6):769-782. [doi: 10.1177/109019810102800608] [Medline: 11720277]
74. Clark NM. Management of chronic disease by patients. Annu Rev Public Health 2003;24:289-313. [doi: 10.1146/annurev.publhealth.24.100901.141021] [Medline: 12415147]
75. Bair MJ, Matthias MS, Nyland KA, Huffman MA, Stubbs DL, Kroenke K, et al. Barriers and facilitators to chronic pain self-management: a qualitative study of primary care patients with comorbid musculoskeletal pain and depression. Pain Med 2009 Oct 01;10(7):1280-1290 [FREE Full text] [doi: 10.1111/j.1526-4637.2009.00707.x] [Medline: 19818038]
76. Russell SJ, Browne JL. Staying well with bipolar disorder. Aust N Z J Psychiatry 2005 Mar 26;39(3):187-193. [doi: 10.1080/j.1440-1614.2005.01542.x] [Medline: 15701069]
77. Suto M, Murray G, Hale S, Amari E, Michalak EE. What works for people with bipolar disorder? Tips from the experts. J Affect Disord 2010 Jul;124(1-2):76-84. [doi: 10.1016/j.jad.2009.11.004] [Medline: 19969370]
78. Callahan AM, Bauer MS. Psychological interventions for bipolar disorder. Psychiatr Clin North Am 1999 Sep;22(3):675-688. [doi: 10.1016/s0193-953x(05)70102-7]
79. Miklowitz DJ, Price J, Holmes EA, Rendell J, Bell S, Budge K, et al. Facilitated integrated mood management for adults with bipolar disorder. Bipolar Disord 2012 Mar;14(2):185-197 [FREE Full text] [doi: 10.1111/j.1399-5618.2012.00998.x] [Medline: 22420594]
80. Michalak EE, Suto MJ, Barnes SJ, Hou S, Lapsley S, Scott MW, CREST BD. Effective self-management strategies for bipolar disorder: A community-engaged Delphi Consensus Consultation study. J Affect Disord 2016 Dec;206:77-86. [doi: 10.1016/j.jad.2016.06.057] [Medline: 27466745]
81. Aschbrenner KA, Naslund JA, Shevenell M, Kinney E, Bartels SJ. A pilot study of a peer-group lifestyle intervention enhanced with mHealth technology and social media for adults with serious mental illness. J Nerv Ment Dis 2016;204(6):483-486. [doi: 10.1097/nmd.0000000000000530]
82. Waring ME, Jake-Schoffman DE, Holovatska MM, Mejia C, Williams JC, Pagoto SL. Social media and obesity in adults: a review of recent research and future directions. Curr Diab Rep 2018 Apr 18;18(6):34-39. [doi: 10.1007/s11892-018-1001-9] [Medline: 29671135]
83. Bond K, Anderson IM. Psychoeducation for relapse prevention in bipolar disorder: a systematic review of efficacy in randomized controlled trials. Bipolar Disord 2015 Jun;17(4):349-362. [doi: 10.1111/bdi.12287] [Medline: 25594775]
84. Gregory VL. Cognitive-behavioral therapy for mania: a meta-analysis of randomized controlled trials. Soc Work Ment Health 2010 Oct 07;8(6):483-494. [doi: 10.1080/15332981003744388]
85. Miklowitz DJ, Efthimiou O, Furukawa TA, Scott J, McLaren R, Geddes JR, et al. Adjunctive psychotherapy for bipolar disorder: a systematic review and component network meta-analysis. JAMA Psychiatry 2021 Feb 01;78(2):141-150 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.2993] [Medline: 33052390]
86. Von Korff M, Gruman J, Schaefer J, Curry SJ, Wagner EH. Collaborative management of chronic illness. Ann Intern Med 1997 Dec 15;127(12):1097-1102. [doi: 10.7326/0003-4819-127-12-199712150-00008] [Medline: 9412313]
87. Miller WR, Lasiter S, Bartlett Ellis R, Buelow JM. Chronic disease self-management: a hybrid concept analysis. Nurs Outlook 2015 Mar;63(2):154-161 [FREE Full text] [doi: 10.1016/j.outlook.2014.07.005] [Medline: 25241136]
88. Wagner EH, Austin BT, Von Korff M. Organizing care for patients with chronic illness. Milbank Q 1996;74(4):511-544. [Medline: 8941260]
89. Lorig KR, Sobel DS, Stewart AL, Brown BW, Bandura A, Ritter P, et al. Evidence suggesting that a chronic disease self-management program can improve health status while reducing hospitalization: a randomized trial. Med Care 1999 Jan;37(1):5-14. [doi: 10.1097/00005650-199901000-00003] [Medline: 10413387]

90. Salyers MP, Godfrey JL, McGuire AB, Gearhart T, Rollins AL, Boyle C. Implementing the illness management and recovery program for consumers with severe mental illness. Psychiatr Serv 2009 Apr;60(4):483-490. [doi: 10.1176/appi.ps.60.4.483] [Medline: 19339323]
91. Salyers MP, Godfrey JL, Mueser KT, Labriola S. Measuring illness management outcomes: a psychometric study of clinician and consumer rating scales for illness self management and recovery. Community Ment Health J 2007 Oct;43(5):459-480. [doi: 10.1007/s10597-007-9087-6] [Medline: 17514504]
92. Huber GP, Van de Ven AH, editors. Longitudinal field research methods: studying processes of organizational change (Organization Science). Thousand Oaks, CA: SAGE Publications, Inc; 1995.

Abbreviations

BD: bipolar disorder
CC: code count
MHT: mental health technology
PC: participant count

Edited by J Torous, G Eysenbach; submitted 22.07.21; peer-reviewed by E Morton, D Gürtler; comments to author 15.08.21; revised version received 14.09.21; accepted 17.09.21; published 22.11.21.

Please cite as:
Jonathan GK, Dopke CA, Michaels T, Martin CR, Ryan C, McBride A, Babington P, Goulding EH
A Smartphone-Based Self-Management Intervention for Individuals with Bipolar Disorder (LiveWell): Qualitative Study on User Experiences of the Behavior Change Process
JMIR Ment Health 2021;8(11):e32306
URL: https://mental.jmir.org/2021/11/e32306
doi:10.2196/32306
PMID:34813488

Original Paper

The Influence of Gender and Age on the Outcomes of and Adherence to a Digital Interdisciplinary Mental Health Promotion Intervention in an Australasian Nonclinical Setting: Cohort Study

Geraldine Przybylko[1], BCom, MPH, MBA; Darren Morton[1], BEd, MAppSci, PhD; Jason Morton[2], BEd, MAppSci, PhD; Melanie Renfrew[1], BEd, Grad Dip

[1]Lifestyle Medicine and Health Research Centre, Avondale University, Cooranbong, Australia

[2]Faculty of Education, Business and Science, Avondale University, Cooranbong, Australia

Corresponding Author:
Geraldine Przybylko, BCom, MPH, MBA
Lifestyle Medicine and Health Research Centre
Avondale University
582 Freemans Drive
Cooranbong, 2265
Australia
Phone: 61 418574001
Email: geraldineprzybylko@eliawellness.com

Abstract

Background: The global prevalence of mental health disorders is at a crisis point, particularly in the wake of COVID-19, prompting calls for the development of digital interdisciplinary mental health promotion interventions (MHPIs) for nonclinical cohorts. However, the influence of gender and age on the outcomes of and adherence to MHPIs is not well understood.

Objective: The aim of this study was to determine the influence of gender and age on the outcomes of and adherence to a 10-week digital interdisciplinary MHPI that integrates strategies from positive psychology and lifestyle medicine and utilizes persuasive systems design (PSD) principles in a nonclinical setting.

Methods: This study involved 488 participants who completed the digital interdisciplinary MHPI. Participants completed a pre and postintervention questionnaire that used: (1) the "mental health" and "vitality" subscales from the Short Form 36 (SF-36) Health Survey; (2) the Depression, Anxiety and Stress Scale (DASS-21); and (3) Satisfaction With Life Scale (SWL). Adherence to the digital interdisciplinary MHPI was measured by the number of educational videos the participants viewed and the extent to which they engaged in experiential challenge activities offered as part of the program.

Results: On average, the participants (N=488; mean age 47.1 years, SD 14.1; 77.5% women) demonstrated statistically significant improvements in all mental health and well-being outcome measures, and a significant gender and age interaction was observed. Women tended to experience greater improvements than men in the mental health and well-being measures, and older men experienced greater improvements than younger men in the mental health and vitality subscales. Multiple analysis of variance results of the adherence measures indicated a significant difference for age but not gender. No statistically significant interaction between gender and age was observed for adherence measures.

Conclusions: Digital interdisciplinary MHPIs that utilize PSD principles can improve the mental health and well-being of nonclinical cohorts, regardless of gender or age. Hence, there may be a benefit in utilizing PSD principles to develop universal MHPIs such as that employed in this study, which can be used across gender and age groups. Future research should examine which PSD principles optimize universal digital interdisciplinary MHPIs.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12619000993190; http://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=377889 and Australian New Zealand Clinical Trials Registry ACTRN12619001009101; http://www.anzctr.org.au/ACTRN12619001009101.aspx

(JMIR Ment Health 2021;8(11):e29866) doi:10.2196/29866

KEYWORDS

age; gender; adherence; digital health; interdisciplinary; mental health; promotion; intervention; lifestyle medicine; positive psychology; multicomponent; lifestyle; outcome; cohort study

Introduction

Background

The global prevalence of depression and anxiety prior to the COVID-19 pandemic was estimated at 586 million people [1]. In Australia, 1 in 5 individuals reported a mental or behavioral condition in 2017-2018 [2], with the prevalence being higher among women (22%) than men (18%) and in the 15-24 years age range (26%). Recent population-based surveys indicate that mental health has further deteriorated since COVID-19 [3,4]. In response, there is a need for accessible and scalable mental health and well-being interventions that not only support individuals suffering from mental distress but that also promote the mental health and psychological resilience of nonclinical cohorts [5-7].

Emerging literature has shown that digital mental health promotion interventions (MHPIs) can improve mental health [8,9] and prevent the onset of mental health disorders [10,11]. Research suggests several advantages of digital MHPIs, including the potential to target individuals before they reach a diagnostic threshold [12], being more economical than face-to-face interventions [13], increased acceptability, and having the potential to disseminate on a wider scale [10]. The use of mobile apps and texting for self-guided care may help to improve physical health while reducing anxiety, stress, and depression [14]. Conversely, the literature also identifies several disadvantages of digital MHPIs, including the resources spent on individuals who do not develop adverse outcomes [13], higher dropout rates [15], and the smaller effect sizes compared with those of clinical interventions [10]. Digital interventions targeting clinical cohorts have been found to be beneficial for the treatment of acute depression, and a meta-analysis concluded that gender and age are not moderating factors of the outcomes [16]. Digital MHPIs have also been shown to be effective tools for enhancing the mental health and well-being of nonclinical cohorts, although little is known about the influence of gender and age on the responsiveness to these interventions in this context [17].

Influence of Gender on MHPIs

There is a paucity of research examining the influence of gender on the outcomes of and adherence to digital interdisciplinary MHPIs when delivered in nonclinical settings. It is well established that men and women experience mental health issues and distress differently. For example, women are more likely to talk to someone, seek professional health care, protect themselves and their offspring, and continue engaging with their social networks [18]. In contrast, men typically build up their emotions over time, which may develop into adverse behaviors such as anger, violence, and hostility that in turn can compound the mental health issue [19].

Studies also show that men are typically more hesitant than women to seek help or treatment for mental health concerns [20]. Notably, women tend to rate MHPIs as more important than men [21] and self-selected mental health interventions typically have a bias toward female participation [15,22-27]. Numerous strategies such as role models, incorporating problem-solving tasks, and portraying positive male traits have been proposed for improving the outcomes of MHPIs when targeting men [28]. The use of MHPIs in male-dominated industries has been reported to improve mental health literacy and knowledge, increase social support, provide access to treatment, and reduce absenteeism [29].

Further, men may require more encouragement than women to engage in, and adhere to, digital interventions, thus requiring better implementation of programs [30]. Promoting enjoyable activities and creating sustainable cultures that facilitate group comradery are also deemed positive strategies for MHPIs [31] as they encourage trust, reduce stigma, and normalize engagement [32]. Interestingly, an Australian Football League themed app using young male role models, psychoeducation, social connection, and applied games to target men reported 60-day improvements in flourishing and a sense of connection to the intervention community regardless of gender [33]. However, there is limited understanding of the influence of gender on the outcomes of and adherence to digital interdisciplinary MHPIs, which was the main objective of this study.

Influence of Age on MHPIs

The literature is also sparse regarding the influence of age on the outcomes of and adherence to digital interdisciplinary MHPIs in a nonclinical setting. Although young adults are commonly termed “digital natives” [34,35], this does not necessarily equate to interest and engagement with digital MHPIs and services. For example, a large web-based survey on university students revealed that those with psychological distress (26.14%, 1577/6034) reported a low utilization rate of 2.98% (47/1577) for online mental health services, despite 59.99% (946/1577) reporting a willingness to use the services [36]. Regardless, digital mental health interventions have been found to be effective for improving depression, anxiety, and psychological well-being among college students. However, further investigations are required to understand the key factors to optimize such interventions [37].

Little is known on the efficacy of digital MHPIs for older adults in a nonclinical setting. New technologies are promising tools to alleviate loneliness and social isolation [38]. Novel interventions such as virtual reality interventions have been found to improve psychological well-being in older adults in a nonclinical setting, and have the potential to foster environmental mastery, personal growth, and autonomy [39]. In addition, animatronic pets provide a promising opportunity to support healthy older adults in reducing loneliness, increasing quality of life, and improving psychological well-being [40].

A key challenge is the ability to develop user-driven, action-based mental health interventions for younger men that

shift behavior, stigma, and leverage the influence of their peers [41]. Social influences are more prominent in the adoption of mobile health services among young to middle-aged adults compared with older adults [42].

A key factor affecting adherence to digital mental health interventions among older adults is whether their expectations of the potential outcomes are realistic [43]. Managing expectations reduces the likelihood of older adults deeming the intervention to be a "waste of time" and hence increases adherence, which contributes to improved outcomes of the intervention. Interestingly, mental health interventions are rarely designed with the novice digital user in mind or in accordance with the digital guidelines of older participants [44]. In fact, an "innovativeness-needs paradox" seems to exist where the people in most need of the digital MHPI are those at the highest risk of having the least access, training, skills, adoption rates, and adherence to the intervention, therefore increasing health care inequalities [45]. Digital MHPIs have the potential to reduce the gap in health care provision for older adults as many patients have long wait times for mental health providers, need help in the prevention and management of a multitude of chronic diseases, and have limited access to mental health providers as they are unable to travel long distances [45].

Persuasive Systems Design

It is well recognized that adherence is problematic in digital MHPIs. Two decades ago, Fogg [46] coined the term "interactive technology" as the design to leverage social influence and motivate and persuade humans to change their attitudes and behaviors. Interactive technology involves features that reward people with positive feedback, model a target behavior or attitude, and provide social support. Oinas-Kukkonen and Harjumaa [47] progressed the work of Fogg by proposing persuasive systems that incorporate information software or systems that are devised to reinforce, change, or formulate attitudes and/or behaviors. In addition, the persuasive systems design (PSD) model was developed specifically to optimize engagement with digital interventions [47]. The PSD model incorporates four categories of persuasion principles: Primary Task Support, Dialogue Support, System Credibility Support, and Social Support. Each category includes 7 distinct persuasion principles, including reduction, tunneling, tailoring, self-monitoring, rewards, reminders, liking, trustworthiness, expertise, surface credibility, social learning, social comparison, and normative influence and completion. Studies have also demonstrated that PSD principles can improve the outcomes of and adherence to digitally delivered interventions [48-52].

Game-based digital mental health interventions were particularly found to increase the participants' engagement and adherence over the long term [53]. Although gaming is typically used by the younger male demographics, gender and age were not associated with frequency of play [53]. In contrast, a systematic review reported no evidence that the use of gamification was associated with increased adherence to the protocol of the program [54]. However, this may be due to most studies utilizing only one feature of gamification (eg, goal-setting, progress, feedback reward, or story/theme).

Common reasons reported for nonadherence include lack of time, disinterest in the intervention, treatment no longer needed, hardware or technical issues, perceived ineffectiveness of the intervention, life events, chose not to proceed as participants felt better after undertaking a few modules, discontent with the group assignment, holiday, work commitments, poor health, and no longer wished to participate [52].

Study Objectives

Mental health promotion is crucial for improving population-level mental health. Despite the emerging literature supporting the effectiveness of digital MHPIs [23,25], there is a paucity of research investigating the influence of gender and age in nonclinical cohorts. In this study, we aimed to investigate the influence of gender and age on the outcomes of and adherence to a digital interdisciplinary MHPI in a nonclinical cohort.

Methods

Study Design

We previously reported that a randomized controlled trial (RCT) using an MHPI, referred to as the "Live More Project" or "The Lift Project," showed significant improvements (P<.001) in all outcome measures for an intervention group, whereas no changes were observed in the control group regardless of gender and age [25]. The focus of this study was to examine the influence of gender and age on the outcomes of and adherence to a digital MHPI among a larger cohort.

Study Participants

This study combined the data of two cohorts from two independent studies that utilized the same intervention in an Australasian nonclinical setting (see Table 1), creating a total study population of 488 participants. The participants were recruited voluntarily through a faith-based organization. The study was advertised as an "emotional wellness" program through the faith-based organization's internal communications channels, including bulletins and magazines.

Cohort 1 was the treatment arm (n=168) from an unblinded RCT. The Avondale University Human Research Ethics Committee approved all procedures involving human subjects for the RCT (project number 2017:13). The trial protocol was registered at the Australian New Zealand Clinical Trials Registry (ACTRN12619000993190). For more information about the study design and intervention, refer to Przybylko et al [25].

Table 1. Overview of weekly topics and challenges for the intervention.

Week	Topic	Daily challenge	Weekly challenge
1	Speak positively	Offer a genuine compliment	Memorize an inspirational text or saying
2	Move dynamically	Spend 30 minutes of moderate exercise or 10,000 steps	20 minutes of guided resistance exercises
3	Immerse in an uplifting natural environment	Spend 30 minutes in an uplifting natural environment	Experience a sunrise
4	Immerse in a positive social environment	Do something intentional to show you care	Forgive someone who has hurt you
5	Look to the positive	Spend 15 minutes to reflect on three things that went well	Write a letter of gratitude to someone and share it with them
6	Eat nutritiously	Eat eight servings of plant-based food	Prepare a high-fiber, plant-based meal with one or more friends
7	Rest: sleep	Spend 8 hours in bed without a device	Spend an evening by firelight
8	Rest from stress	Spend 15 minutes in a quiet place, relaxing, and being mindful of surroundings	Take a day off work and a digital Sabbath (going "offline" for 24 hours to recharge)
9	Serving others	Perform a random act of kindness	Use signature strength to perform an act of service
10	What does it take to flourish?	Continue challenges found to be helpful	Continue challenges found to be helpful

Cohort 2 involved a three-arm (n=320) randomized comparative study that examined the influence of different modes of human support on the intervention. The Avondale University Human Research Ethics Committee approved the conduct of the study (project number 2018:09) and the trial was registered at the Australian New Zealand Clinical Trials Registry (ACTRN12619001009101); refer to Renfrew et al [23] for a detailed explanation of the study. The study indicated that the intervention improved the mean scores for all mental health metrics, regardless of the addition of human support. Further, the mode of human support offered in addition to the intervention had no influence on the outcomes of the intervention [55]. Hence, as the mental health outcomes were similar for all arms in the study, all participants were pooled to form Cohort 2.

Intervention

Both cohorts participated in an intervention referred to as the "The Live More Project" or "The Lift Project" [56,57]. The 10-week digital interdisciplinary MHPI used evidenced-based strategies from the disciplines of lifestyle medicine and positive psychology for improving mental health and well-being, as detailed in Table 1.

The intervention was based on Ajzen's [58] Theory of Planned Behavior and employed an experiential pedagogical framework of "Learn, Experience, Think, Share," which was accessed through an electronic learning management system. Weekly 30-minute video sessions were aimed to educate and empower participants to make positive behavior changes. Daily and weekly experiential challenges provided practical application. The following PSD principles were used in the intervention to improve adherence: gamification to increase challenge points and badges by participating in the challenges; a social forum to comment, post photos, and encourage interaction between the participants to provide accountability; and provision of reminders to watch the videos and log challenges. Refer to Multimedia Appendix 1 for website and app screenshots. An electronic book and electronic workbook were also provided to expand the participants' knowledge and to journal their experience during the intervention.

Measurements

Mental Health Outcomes

All participants in each cohort completed a self-reported questionnaire at preintervention (Week 1) and postintervention (Week 12). The questionnaire included sociodemographic characteristics such as age, gender, ethnicity, level of education, country of birth, and three validated instruments. An outline of the instruments used in the questionnaire are detailed below.

Short Form 36-Item Health Survey

The Short Form 36-item (SF-36) Health Survey is a self-reported health questionnaire appropriate for use among a general population [59]. This survey consists of 36 items that assess eight scales: general health, mental health, vitality, social function, physical function, role limitations due to physical health, role limitations due to emotional problems, and bodily pain. These subscales can be used separately; the two subscales measuring positive affect (ie, "mental health" and "vitality") [60] were used in this study. The "mental health" subscale assesses emotional well-being (5 items) and the "vitality" subscale assesses energy and fatigue (4 items) [61]. Both subscales generate a score between 0 and 100, with a higher score representing a higher level of mental health and vitality. Although exact cut-off scores have not been established for the two subscales, studies have indicated that a mental health score less than 56 is indicative of major depression [62] and a score less than 45 on the vitality subscale was classified as fatigued [63]. This study observed a Cronbach α of .86 for mental health and .88 for vitality, indicating good internal consistency.

Depression, Anxiety and Stress Scale

The 21-item Depression, Anxiety and Stress Scale (DASS-21) is a well-used assessment to measure the negative affect of

emotional states—depression, anxiety, and stress (7 items per subscale)—on both clinical and nonclinical populations [64]. The questionnaire generates a score between 0 and 21, with a higher score representing increasing levels of mental distress [65]. Individuals were considered "symptomatic" if they reached the threshold of greater than 4 for depression, greater than 3 for anxiety, and greater than 7 for stress. This study observed good internal consistency with a Cronbach α of .87 for depression, .70 for anxiety, .83 for stress, and .90 for the overall score for the 3 domains.

Satisfaction With Life Scale

The 5-item Satisfaction With Life Scale (SWLS) assesses global life satisfaction [66] and is used in numerous settings [67]. The questionnaire generates a score between 5 and 35, with a higher score representing increasing levels of life satisfaction. A score of 19 indicates an average life satisfaction. This study observed a Cronbach α of .88, indicating good internal consistency.

Measurements

Adherence

This study used the following adherence measures for the intervention that have been established previously [55].

Videos Viewed

Each week the participants were introduced to a weekly topic (see Table 1) that was presented using an educational video. The total number of weekly videos viewed was used to measure primary adherence and was measured out of a total of 10. A video was marked as "viewed" when 80% or more of the presentation had been played.

Experiential Challenge Activities

Participants were encouraged to put what they had learned each week into action by participating in experiential challenge activities. Adherence to challenges was calculated through the total weekly challenge score and the total number of weeks that the participants had completed the challenge. The daily challenge was awarded 10 points with a maximum of 70 points per week, and weekly challenges were allocated 30 points. Hence, participants had the opportunity to achieve 100 points per week, for a total of 1000 points at the end of the 10-week intervention.

Statistical Analysis

The data were analyzed using SPSS Statistics (version 25). The χ^2 test was used to examine the difference in the baseline characteristics. Descriptive statistics, involving frequencies, means, SDs, and 95% CIs, are used to present the mental health and well-being outcomes, as well as the adherence measures.

Multivariate analysis of variance (MANOVA) was used for comparisons as there were several categorical independent variables and continuous dependent variables. Data were prescreened and cleaned to ensure the robustness of the MANOVA, which was also aided by the large sample size. MANOVA, using the general linear modeling (GLM) function in SPSS, was used to test for time effects (pre to postintervention), gender and age effects, and their interactions. When significant, Bonferroni post hoc analyses were utilized to determine significant changes from pre to postintervention to compare gender and age differences, and to explore significant interactions. Pearson correlation analysis was used to evaluate the relationship between the mental health or well-being outcomes and adherence measures. Analysis of variance was used to compare differences in the outcomes and adherence measures between the age categories. Paired and independent sample *t* tests were used to explore gender differences in the mental health and well-being outcome measures. Missing data for age (n=4) were replaced with the mean age and missing data (n=14) for mental health outcomes were removed from the analysis.

Results

Participant Baseline Characteristics

A total of 488 participants completed the preintervention questionnaire (week 1) and postintervention questionnaire (week 12).

Cohorts 1 and 2 differed with regard to age (mean 49.3 years, SD 14.1 and 45.9 years, SD 14.0, respectively; P=.01), gender balance (women: 69.1%, 116/168 and 81.9%, 262/320, respectively; P=.002), and ethnic representation (White: 89.3%, 150/168 and 81.8%, 262/320, respectively; P=.05). Although a statistically significant difference was observed between the cohorts in these demographic variables, it is notable that in both cohorts there was a bias toward White women in the 35-54–year age category. No statistically significant difference was found between Cohorts 1 and 2 in the highest education obtained (tertiary education: 89.3% and 81.8%, respectively; P=.52).

There was a difference between Cohorts 1 and 2 in all baseline mental health measures except life satisfaction; however, the mean scores for both cohorts were in the nonclinical range for all measures. The baseline mental health measures for the two cohorts were as follows: mental health (75.5 and 66.2, P<.001), vitality (52.5 and 60.2, P<.001), depression (2.5 and 3.5, P=.001), anxiety (1.8 and 2.3, P=.02), stress (4.5 and 5.7, P=.001), and life satisfaction (23.9 and 23.1, P=.08). Combining these two cohorts increased the heterogeneity of the total sample, which in turn increased the generalizability of the study.

The combined cohort (N=488), which formed the population for this study, had a mean age of 47.1 years (SD 14.1) and were mostly women (77.9%, 380/488). The ethnicity of the population was largely White (83.4%, 407/488), followed by Other (5.3%, 26/488), Asian (4.5%, 22/488), Maori/Pacific Islander (3.3%, 16/488), Black/African American (2.3%, 11/488), Spanish/Hispanic/Latino (0.6%, 3/488), and Indigenous (0.6%, 3/488). The highest level of education achieved was tertiary education (84.4%, 413/488), followed by secondary/high school (15.2%, 75/488) and primary/elementary (0.4%, 2/488).

In the absence of standardized or universally accepted age categorization, the authors determined three age categories based on the age grouping system of the World Health Organization [68]: 18-34 years (younger adults: 21.7%, 106/488), 34-54 years (middle-aged adults: 47.1%, 230/488), and ≥55 years (older adults: 31.1%, 152/488). However, the World Health

Organization acknowledges that there is no conceptual justification for selecting one age standard over another [68].

There was no statistically significant difference between men and women (χ^2_{487}=1.42, *P*=.70) or the age categories (F_{487}=0.30, *P*=.69) with regard to ethnic representation (F_{487}=0.03, *P*=.98) or highest level of education obtained (F_{487}=0.23, *P*=.59). However, statistically significant differences were found between genders and between age categories in some of the preintervention psychometric measures (see Table 2). At preintervention, women reported poorer mental health metrics than men for mental health (t_{487}=4.85, *P*<.001), vitality (t_{487}=3.94, *P*<.001), depression (t_{487}=–3.13, *P*=.002), anxiety (t_{487}=–3.05, *P*=.002), and stress (t_{487}=–4.14, *P*<.001), but not life satisfaction (t_{487}=1.09, *P*=.29). However, the mean scores were found to be in the nonclinical range. The ≥55 years age category had a significantly better score than the 18-34 (*P*<.001) and 34-54 (*P*=.002) age categories for mental health; the 18-34 (*P*<.001) and 34-54 (*P*<.001) categories for vitality; the 18-34 category for depression (*P*<.001) and anxiety (*P*=.002); and the 18-34 (*P*<.001) and 34-54 (*P*<.001) categories for stress. The 35-54 age category had a significantly higher score than the 18-34 age category for depression (*P*=.03) and anxiety (*P*=.03), but not for stress (*P*=.97).

Table 2. Pre to postintervention changes in each of the outcome measures defined by the gender and age categories.

Outcome measure	Preintervention (week 1), mean (SD)	Postintervention (week 12), mean (SD)	Difference, mean (%)	*t* test (*df*)	95% CI	*P* value	Effect size (*Cohen d*)
Mental health							
Overall	69.4 (16.4)	77.9 (14.9)	8.5 (12)	–13.306 (487)	–9.72 to –7.22	<.001	0.54
Gender							
Men	76.0 (14.4)	81.3 (16.2)	5.3 (7)	–3.227 (107)	–8.59 to –2.05	.002	0.35
Women	67.5[a] (16.5)	76.9 (14.4)	9.4 (14)[a]	–14.197 (379)	–10.69 to –8.09	<.001	0.61
Age category (years)							
18-34	64.1 (16.6)	71.3 (17.8)	7.2 (11)	–1.304 (105)	–10.66 to –3.83	<.001	0.42
35-54	68.5 (15.3)	77.0 (14.1)	8.4 (12)[b]	–9.695 (229)	–10.13 to –6.71	<.001	0.57
55+	74.3[b,c] (16.7)	83.7 (11.7)	9.5 (13)[b]	–9.182 (151)	–11.50 to –7.43	<.001	0.67
Vitality							
Overall	57.6 (18.2)	86.2 (17.6)	10.5 (18)	–14.018 (487)	–12.01 to –9.06	<.001	0.59
Gender							
Men	63.5 (16.7)	69.9 (19.9)	6.4 (10)	–3.416 (107)	–10.06 to –2.67	<.001	0.35
Women	55.9[a] (18.3)	67.7 (16.8)	11.8 (21)[a]	–14.772 (379)	–13.31 to –10.19	<.001	0.67
Age category (years)							
18-34	53.9 (15.7)	63.0 (19.0)	9.1 (17)	–4.737 (105)	–12.96 to –5.31	<.001	0.53
35-54	55.6 (18.7)	66.6 (17.3)	11.1 (20)[b]	–10.647 (229)	–13.14 to –9.04	<.001	0.62
55+	63.2[b,c] (17.8)	73.9 (15.4)	10.7 (17)	–8.475 (151)	–13.21 to –8.21	<.001	0.64
Depression							
Overall	3.2 (3.3)	1.9 (2.6)	–1.3 (–39)	10.575 (487)	1.02 to 1.49	<.001	0.43
Gender							
Men	2.3 (3.0)	1.5 (2.7)	–0.8 (–34)	3.464 (107)	0.34 to 1.26	.001	0.28
Women	3.4[a] (3.3)	2.0 (2.6)	–1.4 (–40)[a]	10.121 (379)	1.12 to 1.66	<.001	0.47
Age category (years)							
18-34	4.1 (3.5)	2.6 (3.3)	–1.6 (–38)	5.675 (105)	1.03 to 2.14	<.001	0.47
35-54	3.1 (3.1)	2.0 (2.6)	–1.1 (–35)	6.562 (229)	0.76 to 1.41	<.001	0.38
55+	2.7[b] (3.2)	1.4 (1.9)	–1.3 (–49)	6.095 (151)	0.88 to 1.72	<.001	0.50
Anxiety							
Overall	2.1 (2.3)	1.3 (1.8)	–0.8 (–38)	9.242 (487)	0.64 to 0.98	<.001	0.39
Gender							
Men	1.5 (1.9)	0.9 (1.7)	–0.6 (–42)	3.896 (107)	0.32 to 0.97	<.001	0.36
Women	2.3[a] (2.4)	1.5 (1.8)	–0.9 (–35)	8.382 (379)	0.65 to 1.05	<.001	0.40
Age category (years)							
18-34	2.8 (2.7)	1.6 (2.1)	–1.2 (–42)	4.886 (105)	0.70 to 1.67	<.001	0.49
35-54	2.1[b] (2.2)	1.3 (1.9)	–0.7 (–35)	6.574 (229)	0.51 to 0.94	<.001	0.35
55+	1.8[b] (2.1)	1.1 (1.4)	–0.7 (–39)	4.516 (151)	0.38 to 0.98	<.001	0.39
Stress							
Overall	5.3 (3.4)	3.8 (3.1)	–1.5 (–28)	11.313 (487)	1.24 to 1.76	<.001	0.47
Gender							

Outcome measure	Preintervention (week 1), mean (SD)	Postintervention (week 12), mean (SD)	Difference, mean (%)	*t* test (*df*)	95% CI	*P* value	Effect size (*Cohen d*)
Men	4.2 (2.8)	3.0 (3.2)	–1.2 (–28)	4.270 (107)	0.63 to 1.73	<.001	0.39
Women	5.7[a] (3.4)	4.1 (3.0)	–1.6 (–28)	10.554 (379)	1.30 to 1.89	<.001	0.50
Age category (years)							
18-34	6.4 (3.6)	4.9 (3.8)	–1.5 (–24)	4.812 (105)	0.90 to 2.17	<.001	0.42
35-54	5.5 (3.1)	4.1 (2.9)	–1.4 (–26)	7.909 (229)	1.08 to 1.79	<.001	0.48
55+	4.3[b,c] (3.4)	2.7 (2.3)	–1.6 (–37)	6.477 (151)	1.09 to 2.04	<.001	0.55
Life satisfaction							
Overall	23.4 (6.8)	25.8 (6.4)	2.4 (10)	–11.991 (487)	–2.85 to –2.04	<.001	0.37
Gender							
Men	24.0 (6.5)	26.2 (6.0)	2.2 (9)	–5.694 (107)	–3.02 to –1.46	<.001	0.35
Women	23.2 (6.8)	25.7 (6.5)	2.5 (11)	–10.562 (379)	–2.97 to –2.04	<.001	0.38
Age category (years)							
18-34	23.3 (7.2)	25.7 (6.7)	2.4 (10)	–5.114 (105)	–3.33 to –1.47	<.001	0.35
35-54	22.9 (7.0)	25.7 (6.6)	2.8 (12)	–9.206 (229)	–3.38 to –2.19	<.001	0.41
55+	24.0 (6.2)	26.0 (5.9)	2.0 (8)	–5.864 (151)	–2.63 to –1.30	<.001	0.33

[a]Significant gender difference.

[b]Significant difference from the 18-34 age category.

[c]Significant difference from the 35-54 age category.

Mental Health Outcomes

Overall Intervention Effect

MANOVA results of the changes in the mental health and well-being outcomes from pre to postintervention indicated a statistically significant difference for gender (F_{487}=2.81, *P*=.01, Wilks Λ=0.97, η^2=0.03) and age (F_{487}=2.46, *P*=.004; Wilks Λ=0.94, η^2=0.03). A significant gender and age interaction (F_{487}=2.14, *P*=.01; Wilks Λ=0.95, η^2=0.03) was observed, with younger females experiencing greater improvements than the older females in 5 out of 6 outcome measures. This trend was not evident among the males. Table 2 shows the changes in mental health and well-being outcomes from pre to postintervention, reported for gender and the age group categories. Statistically significant improvements in all mental health and well-being measures were observed.

Influence of Gender on Mental Health Outcomes

Although women reported lower levels of mental health (ie, higher emotional distress) at preintervention, they experienced a higher mean change than men in mental health (F_{487}=13.16, *P*<.001), vitality (F_{487}=11.90, *P*=.001), and depression (F_{487}=3.89, *P*=.05), as seen in Table 2. No significant differences were observed between men and women with respect to anxiety (F_{487}=0.87, *P*=.35), stress (F_{487}=0.88, *P*=.35), or life satisfaction (F_{487}=3.53, *P*=.06).

Influence of Age on Mental Health Outcomes

Although the ≥55-year age category had higher levels of mental health (ie, lower emotional distress) at preintervention, they experienced a significantly higher mean change in the mental health scale (F_{487}=5.15, *P*=.006) than the younger age categories. However, there were no statistically significant differences between the age categories for vitality (F_{487}=2.05, *P*=.13), depression (F_{487}=0.53, *P*=.58), anxiety (F_{487}=1.53, *P*=.22), stress (F_{487}=0.32, *P*=.73), or life satisfaction (F_{487}=2.15, *P*=.12). The pre to postintervention results indicated that the 18-34–year age category had a significantly lower mean change than the 35-54 (*P*=.009) and ≥55 (*P*=.002) age categories for mental health, and the 35-54 age category for vitality (*P*=.05). Despite the 18-34 age category achieving a higher score (ie, indicating higher emotional distress) at postintervention compared to the 35-54 age category for depression (*P*=.31), anxiety (*P*=.14), and stress (*P*=.54), no statistically significant differences were observed. There were also no statistically significant differences found between the 18-34 and 35-54 age categories for life satisfaction (*P*=.06), or between the 35-54 and ≥55 age categories for any outcome measures: mental health (*P*=.37), vitality (*P*=.80), depression (*P*=.65), anxiety (*P*=.67), stress (*P*=.50), and life satisfaction (*P*=.16). Every age category for both genders showed a statistically significant improvement (ie, lower emotional distress) in mental health and well-being metrics, except for the 18-34–year age category for mental health (*P*=.13) and vitality (*P*=.13), and the men in the ≥55-year age category for stress (*P*=.09).

Adherence

Overall Intervention Effect

MANOVA results of the adherence measures indicated a statistically significant difference for age (F_{487}=2.20, *P*=.04;

Wilks Λ=0.97, η^2=0.01), but not gender (F_{487}=1.25, P=.29; Wilks Λ=0.99, η^2=0.01). No statistically significant interaction between gender and age (F_{487}=0.75, P=.61; Wilks Λ=0.99, η^2=0.01) was observed.

Influence of Gender on Adherence

As shown in Table 3, there was no statistically significant gender difference in the number of videos watched (t_{487}=–0.52, P=.61), total challenge points achieved (t_{487}=–1.44, P=.15), or the number of weeks that challenges were engaged with (t_{487}=–1.72, P=.09). Although women recorded higher mean challenge points and a higher percentage watched all 10 videos, there was no significant difference between the genders.

Table 3. Adherence outcomes by gender.

Variable	Men (n=108)	Women (n=380)	Total (N=488)	Between-group difference, *P* value
Number of videos viewed (%)				N/A[a]
10	22	33	31	
8-9	5	3	3	
5-7	52	39	43	
1-4	16	20	19	
0	5	5	5	
Number of videos viewed, mean (SD)	6.4 (2.9)	6.6 (3.2)	6.55 (3.18)	.61
Challenge, mean (SD)				
Challenge points (out of 1000)	355.6 (370.6)	412.2 (361.2)	377.5 (354.0)	.15
Number of weeks challenge scores logged (out of 10)	4.4 (3.72)	5.1 (3.63	4.8 (3.6)	.09

[a]N/A: not applicable.

Influence of Age on Adherence

A statistically significant difference was observed between the age categories in the number of videos watched (F_{487}=5.99, P=.003); however, as shown in Table 4, there was no significant difference in the total challenge points (F_{487}=2.448, P=.09) or total number of weeks that challenges were recorded (F_{487}=2.563, P=.08). The age categories of 35-54 and ≥55 years recorded the same mean number of videos watched, which was higher than that for the 18-34 years age category. Both the mean challenge score and the number of weeks that challenge scores were logged showed improvements with age, although the difference was not statistically significant.

Table 4. Adherence outcomes between age categories.

Variable	18-34 years (n=106)	35-54 years (n=230)	≥55 years (n=152)	Total (N=488)	Between-group difference, *P* value
Number of videos viewed (%)					N/A[a]
10	25	33	31	89	
8-9	1	4	3	8	
5-7	38	42	45	125	
1-4	26	18	18	62	
0	10	3	4	17	
Number of videos viewed, mean (SD)	5.6 (3.51)	6.8 (3.06)	6.8 (2.99)	6.6 (3.17)	.01
Challenge, mean (SD)					
Challenge points (out of 1000)	340.8 (346.3)	398.7 (339.0)	442.9 (372.2)	400.3 (363.7)	.09
Number of weeks challenge scores logged (out of 10)	4.4 (3.64)	5.0 (3.65)	5.4 (3.62)	5.0 (3.65)	.08

[a]N/A: not applicable.

Discussion

Principal Results

The aim of this study was to investigate the influence of gender and age on the outcomes of and adherence to a digital interdisciplinary MHPI in a nonclinical cohort. To the authors' knowledge, this study is the first to investigate the effect of gender and age on the outcomes of a digital interdisciplinary MHPI that employed an array of strategies from the disciplines of lifestyle medicine and positive psychology in a nonclinical Australasian setting. Stratification by gender and age showed significant improvements in all mental health and well-being outcomes. Hence, digital interventions such as those employed in this study are useful across gender and age groups for mental health promotion and building psychological resilience.

A female bias was observed in this study, which is consistent with the literature of positive psychology interventions as mentioned previously. Despite the population scoring in the nonclinical range for mental health and well-being, women reported significantly lower mental health scores (ie, higher emotional distress) than men at baseline. However, the women experienced greater improvements than men in the mental health, vitality, depression, and life satisfaction measures. Notably, the women experienced twice the mean change increase in the mental health and vitality subscales compared with the change reported by men, resulting in similar outcome scores to the men at postintervention. This indicates that those scoring lower in the mental health and well-being outcomes can achieve higher mean changes, presumably as there is greater potential for improvement. This is consistent with the results of our previous study using the same intervention that reported higher levels of change were experienced by those with the lowest mental health score at preintervention [57].

This study observed small to medium effect sizes for gender and age on mental health outcomes of a digital interdisciplinary MHPI, which is consistent with the literature of universal digital mental health interventions [10]. However, Tan et al [69] asserts that the impact of small effect sizes can be large when translated to a population level. Hence, digital interdisciplinary MHPIs provide a potential strategy to deliver low-cost and scalable interventions to build the psychological resilience of an individual to help them cope with the adversities experienced in life [70].

This study showed that those aged ≥55 years achieved better mental health and well-being outcomes than the younger age categories in all mental health metrics from pre to postintervention, except for stress. In addition, older men experienced greater improvements than the younger men in the mental health and vitality subscales of the SF-36. These findings are consistent with a meta-analysis of positive psychology interventions, indicating that mental health benefits increased with age [71]. However, these finding are counterintuitive, as younger adults are more frequent users of the internet than older adults (ie, "digital natives"), which could be hypothesized to influence the outcomes of a digitally delivered program [72-74]. A modulating factor might be the time availability. In a previous qualitative study [75], we reported that "time" was perceived as a major barrier to adherence for many participants, although the older participants expressed that retirement provided them with more time to adhere to the intervention. Notably, outcomes of a digital mental health intervention were shown to be related to higher levels of adherence such as higher levels of time spent on the digital platform, number of sessions completed, percentage of the program viewed, and number of activities compared to the control group [76].

Moreover, this study found no significant differences in any adherence measures across gender and age, except for older adults who watched a significantly higher mean number of videos than younger adults. This is consistent with a meta-analysis showing that age was not a predictor of adherence in 13 out of 18 trials [77]. However, this contrasts with a systematic review that found gender to be a consistent predictor of adherence, with women having a higher probability to complete the intervention compared with men [78]. Nevertheless, the authors acknowledged that higher preintervention scores for depression and low scores in anxiety were also found to predict greater adherence.

Naslund et al [79] suggested that focusing digital technologies on early intervention for younger people is key for advancing global mental health. However, Forsman et al [17] argues that the implementation and innovation of mental health promotion for older adults must not be overlooked. Mental health promotion for older adults is of particular importance for three key reasons: there is a higher mental health burden of disease for older adults, digital mental health solutions can improve the mental health care of older adults, and it is well recognized that the mental health of young individuals is strongly influenced by the well-being of their older caregivers [72].

As this study observed improvements in the mental health outcomes of and adherence to a digital interdisciplinary MHPI regardless of gender and age, the authors challenge the concept of focusing solely on mental health promotion for younger or older adults. Instead, the authors encourage developers to be strategic and design digital interdisciplinary MHPIs for all adults (ie, universal). The intervention used in this study employed strategies to increase engagement and adherence among men (ie, using male role models, portraying positive male traits, promoting enjoyable activities, and facilitating peer involvement), younger adults (ie, action-based intervention and leverage the influence of peers), and older adults (ie, designing the intervention for the novice user and managing expectations of the intervention).

In addition, principles from the established PSD categories were incorporated into the intervention. First, from the Primary Task Support category, "reduction," "tunneling," and "self-monitoring" were used to aid adoption by novice users and older adults, and to increase adherence by encouraging behavioral change through participation in a variety of challenges (ie, enjoyable activities). Second, from the Dialogue Support category, the PSD principles of "rewards," "reminders," and "liking" were incorporated in the MHPI to increase adherence in the form of alerts and personalized reminders [32,49,54]. Third, from the System Credibility Support category, the PSD principles of "trustworthiness," "expertise," and

"real-world feel" were incorporated by using an internationally recognized male role model that provided credibility to the intervention to build trust and to portray positive male traits. Lastly, from the Social Support category, the PSD principles of "social learning," "social comparison," "social facilitation," and "competition" were employed to promote peer involvement through social interaction (ie, encourage participants to write comments and post pictures in relation to the challenges) and increase adherence and accountability through the use of gamification (ie, points, badges, and the leaderboard). The culmination of the design elements [31,43,80,81] incorporated in the intervention resulted in it being effective; however, further research should investigate which elements are most beneficial and for whom.

Strengths and Limitations

The strengths of this study are outlined below. First, this study is strengthened by a large number of participants (N=488) and vast age range (18-88 years old) across geographically diverse areas. The second main strength is the MHPI's novel interdisciplinary nature that utilized multicomponent, evidenced-based strategies from the disciplines of lifestyle medicine and positive psychology. Using multicomponent strategies, rather than employing a single tactic, is also deemed to be more efficacious [56,71,82]. Third, the use of PSD principles in the intervention were both gender-responsive and age-sensitive. Increasing the number of PSD principles does not necessarily lead to better outcomes [52]. Future studies could investigate which PSD principles best optimize universal MHPIs.

There are also several limitations of the study. First, the participants were self-selected and drawn from a faith-based population. Hence, they may have entered the study with higher motivation levels and readiness for change than the general population, which may accordingly limit the generalizability of the findings. Second, there was a female bias to the study—which is often observed in positive psychology interventions—and may limit the generalizability of the intervention to male participants. Future studies could explore the use of male-centric advertising and recruitment locations to increase the number of male participants. Third, the study observed small to medium effect sizes for gender and age, which is consistent with the literature. Lastly, as the intervention was promoted as a mental well-being intervention, the sample was in the "nonclinical range" for the mental health scores. Therefore, further research will need to be undertaken to investigate the influence of gender and age on the outcomes of and adherence to digital mental health interventions that integrate strategies from positive psychology and lifestyle medicine when dealing with clinical populations.

Conclusions

The findings of this study demonstrate that a digital interdisciplinary MHPI that employed multicomponent evidence-based strategies from the disciplines of lifestyle medicine and positive psychology using PSD principles can significantly improve mental health and well-being outcome measures across gender and age categories in a nonclinical setting. There may be a benefit in utilizing PSD principles to develop universal MHPIs such as that employed in this study, which can be used across gender and age groups. Future research should examine which PSD principles optimize a universal digital interdisciplinary MHPI.

Acknowledgments

The Seventh-day Adventist Church of the South Pacific Division funded the development of the electronic learning management system and mobile app. They provided the intervention cost-free to participants and offered technical support throughout the duration of the study.

Authors' Contributions

GP, DM, and JM designed the study. GP and DM developed the intervention. GP drafted the manuscript and GP and MR coordinated the studies. GP and MR collected the data. GP analyzed the data. DM and JM provided supervision in the coordination of the study and analyzing the data. DM and JM were major contributors in providing supervision in writing the manuscript. JM, DM, and MR provided critical revision of the article. All authors read and approved the final manuscript.

Conflicts of Interest

DM operates a "profit-for-purpose" trust that administers the delivery of a version of the intervention; however, no personal remuneration is received. GP is employed by the South Pacific Division of the Seventh-day Adventist Church, which administers the intervention among members of the organization, including the participants of this study. No authors have a financial interest in the initiative and there are no other conflicts of interest to declare.

Multimedia Appendix 1
Website and app screenshots.
[PDF File (Adobe PDF File), 1621 KB - mental_v8i11e29866_app1.pdf]

References

1. Depression and other common mental disorders: Global health estimates. World Health Organization. 2017 Jun 27. URL: https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf [accessed 2020-06-27]
2. Australia's health 2020: data insights. Australia's health series no. 17. Cat. no. AUS 231. Australian Institute of Health and Welfare. Canberra; 2020 Jul 23. URL: https://www.aihw.gov.au/reports/australias-health/australias-health-2020-data-insights [accessed 2020-07-23]
3. Fisher JR, Tran TD, Hammarberg K, Sastry J, Nguyen H, Rowe H, et al. Mental health of people in Australia in the first month of COVID-19 restrictions: a national survey. Med J Aust 2020 Nov;213(10):458-464 [FREE Full text] [doi: 10.5694/mja2.50831] [Medline: 33107063]
4. 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]
5. World HO. Prevention and promotion in mental health. Mental Health: Evidence and Research Department of Mental Health and Substance Dependence World Health Organization. 2002. URL: https://www.who.int/mental_health/media/en/545.pdf [accessed 2021-03-21]
6. Kavčič T, Avsec A, Zager Kocjan G. Psychological functioning of Slovene adults during the COVID-19 pandemic: does resilience matter? Psychiatr Q 2021 Mar 17;92(1):207-216 [FREE Full text] [doi: 10.1007/s11126-020-09789-4] [Medline: 32556914]
7. Macedo T, Wilheim L, Gonçalves R, Coutinho ESF, Vilete L, Figueira I, et al. Building resilience for future adversity: a systematic review of interventions in non-clinical samples of adults. BMC Psychiatry 2014 Aug 14;14(1):227-228 [FREE Full text] [doi: 10.1186/s12888-014-0227-6] [Medline: 25266031]
8. Koydemir S, Sökmez AB, Schütz A. A meta-analysis of the effectiveness of randomized controlled positive psychological interventions on subjective and psychological well-being. Appl Res Qual Life 2020 Jan 02;16(3):1145-1185. [doi: 10.1007/s11482-019-09788-z]
9. Kalra G, Christodoulou G, Jenkins R, Tsipas V, Christodoulou N, Lecic-Tosevski D, European Psychiatry Association. Mental health promotion: guidance and strategies. Eur Psychiatry 2012 Feb;27(2):81-86. [doi: 10.1016/j.eurpsy.2011.10.001] [Medline: 22197146]
10. Deady M, Choi I, Calvo RA, Glozier N, Christensen H, Harvey SB. eHealth interventions for the prevention of depression and anxiety in the general population: a systematic review and meta-analysis. BMC Psychiatry 2017 Aug 29;17(1):310 [FREE Full text] [doi: 10.1186/s12888-017-1473-1] [Medline: 28851342]
11. Sander L, Rausch L, Baumeister H. Effectiveness of internet-based interventions for the prevention of mental disorders: a systematic review and meta-analysis. JMIR Ment Health 2016 Aug 17;3(3):e38 [FREE Full text] [doi: 10.2196/mental.6061] [Medline: 27535468]
12. Purtle J, Nelson KL, Counts NZ, Yudell M. Population-based approaches to mental health: history, strategies, and evidence. Annu Rev Public Health 2020 Apr 02;41(1):201-221. [doi: 10.1146/annurev-publhealth-040119-094247] [Medline: 31905323]
13. Greenberg MT, Abenavoli R. Universal interventions: fully exploring their impacts and potential to produce population-level impacts. J Res Educ Eff 2016 Oct 13;10(1):40-67. [doi: 10.1080/19345747.2016.1246632]
14. Rathbone AL, Prescott J. The use of mobile apps and SMS messaging as physical and mental health interventions: systematic review. J Med Internet Res 2017 Aug 24;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]
15. Gander F, Proyer RT, Ruch W. Positive psychology interventions addressing pleasure, engagement, meaning, positive relationships, and accomplishment increase well-being and ameliorate depressive symptoms: A Randomized, Placebo-Controlled Online Study. Front Psychol 2016;7:686. [doi: 10.3389/fpsyg.2016.00686] [Medline: 27242600]
16. Karyotaki E, Ebert DD, Donkin L, Riper H, Twisk J, Burger S, et al. Do guided internet-based interventions result in clinically relevant changes for patients with depression? An individual participant data meta-analysis. Clin Psychol Rev 2018 Jul;63:80-92. [doi: 10.1016/j.cpr.2018.06.007] [Medline: 29940401]
17. Forsman A, Nordmyr J, Park A, Matosevic T, McDaid D, Wahlbeck K. Technology-based interventions for mental health promotion in later life: An evidence review. Eur J Public Health 2016;26(Suppl 1):289. [doi: 10.1093/eurpub/ckw172.044]
18. Liddon L, Kingerlee R, Barry JA. Gender differences in preferences for psychological treatment, coping strategies, and triggers to help-seeking. Br J Clin Psychol 2018 Mar 09;57(1):42-58. [doi: 10.1111/bjc.12147] [Medline: 28691375]
19. Patrick S, Robertson S. Mental health and wellbeing: focus on men's health. Br J Nurs 2016 Nov 24;25(21):1163-1169. [doi: 10.12968/bjon.2016.25.21.1163] [Medline: 27882795]
20. Seaton CL, Bottorff JL, Jones-Bricker M, Oliffe JL, DeLeenheer D, Medhurst K. Men's mental health promotion interventions: a scoping review. Am J Mens Health 2017 Nov 08;11(6):1823-1837 [FREE Full text] [doi: 10.1177/1557988317728353] [Medline: 28884637]
21. Svedberg P. Attitudes to health promotion interventions among patients in mental health services--differences in relation to socio-demographic, clinical and health-related variables. J Ment Health 2011 Apr;20(2):126-135. [doi: 10.3109/09638237.2010.507684] [Medline: 20874509]

22. Kahl BL, Miller HM, Cairns K, Giniunas H, Nicholas M. Evaluation of ReachOut.com, an unstructured digital youth mental health intervention: prospective cohort study. JMIR Ment Health 2020 Oct 15;7(10):e21280 [FREE Full text] [doi: 10.2196/21280] [Medline: 33055066]
23. Renfrew ME, Morton DP, Morton JK, Hinze JS, Beamish PJ, Przybylko G, et al. A web- and mobile app-based mental health promotion intervention comparing email, short message service, and videoconferencing support for a healthy cohort: randomized comparative study. J Med Internet Res 2020 Jan 06;22(1):e15592 [FREE Full text] [doi: 10.2196/15592] [Medline: 31904578]
24. Montagni I, Tzourio C, Cousin T, Sagara JA, Bada-Alonzi J, Horgan A. Mental health-related digital use by university students: a systematic review. Telemed J E Health 2020 Feb;26(2):131-146. [doi: 10.1089/tmj.2018.0316] [Medline: 30888256]
25. Przybylko G, Morton D, Kent L, Morton J, Hinze J, Beamish P, et al. The effectiveness of an online interdisciplinary intervention for mental health promotion: a randomized controlled trial. BMC Psychol 2021 May 11;9(1):77-11 [FREE Full text] [doi: 10.1186/s40359-021-00577-8] [Medline: 33975645]
26. Bakker D, Rickard N. Engagement in mobile phone app for self-monitoring of emotional wellbeing predicts changes in mental health: MoodPrism. J Affect Disord 2018 Feb;227:432-442. [doi: 10.1016/j.jad.2017.11.016] [Medline: 29154165]
27. Davies EB, Morriss R, Glazebrook C. Computer-delivered and web-based interventions to improve depression, anxiety, and psychological well-being of university students: a systematic review and meta-analysis. J Med Internet Res 2014 May 16;16(5):e130 [FREE Full text] [doi: 10.2196/jmir.3142] [Medline: 24836465]
28. Sagar-Ouriaghli I, Godfrey E, Bridge L, Meade L, Brown JS. Improving mental health service utilization among men: a systematic review and synthesis of behavior change techniques within interventions targeting help-seeking. Am J Mens Health 2019 Jun 11;13(3):1557988319857009 [FREE Full text] [doi: 10.1177/1557988319857009] [Medline: 31184251]
29. Lee NK, Roche A, Duraisingam V, Fischer JA, Cameron J. Effective interventions for mental health in male-dominated workplaces. Mental Health Rev J 2014 Dec 02;19(4):237-250 [FREE Full text] [doi: 10.1108/MHRJ-09-2014-0034]
30. Batterham PJ, Calear AL. Preferences for internet-based mental health interventions in an adult online sample: findings from an online community survey. JMIR Ment Health 2017 Jun 30;4(2):e26 [FREE Full text] [doi: 10.2196/mental.7722] [Medline: 28666976]
31. Seaton CL, Bottorff JL, Oliffe JL, Medhurst K, DeLeenheer D. Mental health promotion in male-dominated workplaces: Perspectives of male employees and workplace representatives. Psychol Men Masc 2019 Oct;20(4):541-552. [doi: 10.1037/men0000182]
32. Robertson S, Gough B, Hanna E, Raine G, Robinson M, Seims A, et al. Successful mental health promotion with men: the evidence from 'tacit knowledge'. Health Promot Int 2018 Apr 01;33(2):334-344. [doi: 10.1093/heapro/daw067] [Medline: 27543933]
33. Cheng VWS, Davenport T, Johnson D, Vella K, Mitchell J, Hickie IB. Naturalistic evaluation of a sport-themed mental health and wellbeing app aimed at men (MindMax), that incorporates applied video games and gamification. Internet Interv 2020 Apr;20:100306 [FREE Full text] [doi: 10.1016/j.invent.2020.100306] [Medline: 32181146]
34. Judd T. The rise and fall (?) of the digital natives. Australas J Educ Technol 2018 Nov 25;34(5):99-119. [doi: 10.14742/ajet.3821]
35. Kirschner PA, De Bruyckere P. The myths of the digital native and the multitasker. Teach Teach Educ 2017 Oct;67:135-142. [doi: 10.1016/j.tate.2017.06.001]
36. Dunbar MS, Sontag-Padilla L, Kase CA, Seelam R, Stein BD. Unmet mental health treatment need and attitudes toward online mental health services among community college students. Psychiatr Serv 2018 May 01;69(5):597-600. [doi: 10.1176/appi.ps.201700402] [Medline: 29540117]
37. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. J Med Internet Res 2019 Jul 22;21(7):e12869 [FREE Full text] [doi: 10.2196/12869] [Medline: 31333198]
38. Poscia A, Stojanovic J, La Milia DI, Duplaga M, Grysztar M, Moscato U, et al. Interventions targeting loneliness and social isolation among the older people: An update systematic review. Exp Gerontol 2018 Feb;102:133-144. [doi: 10.1016/j.exger.2017.11.017] [Medline: 29199121]
39. Montana JI, Matamala-Gomez M, Maisto M, Mavrodiev PA, Cavalera CM, Diana B, et al. The benefits of emotion regulation interventions in virtual reality for the improvement of wellbeing in adults and older adults: a systematic review. J Clin Med 2020 Feb 12;9(2):500 [FREE Full text] [doi: 10.3390/jcm9020500] [Medline: 32059514]
40. Tkatch R, Wu L, MacLeod S, Ungar R, Albright L, Russell D, et al. Reducing loneliness and improving well-being among older adults with animatronic pets. Aging Ment Health 2021 Jul;25(7):1239-1245. [doi: 10.1080/13607863.2020.1758906] [Medline: 32363903]
41. 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(1):1-10. [doi: 10.1186/1471-244x-13-119]
42. Kenny G, Connolly R. Towards an inclusive worldxploring m-health adoption across generations. 2017 Jun 05 Presented at: The 25th European Conference on Information Systems (ECIS); June 2017; Guimarães, Portugal p. 5-10.

43. Pywell J, Vijaykumar S, Dodd A, Coventry L. Barriers to older adults' uptake of mobile-based mental health interventions. Digit Health 2020;6:2055207620905422 [FREE Full text] [doi: 10.1177/2055207620905422] [Medline: 32110429]
44. Wasilewski MB, Stinson JN, Cameron JI. Web-based health interventions for family caregivers of elderly individuals: a scoping review. Int J Med Inform 2017 Jul;103:109-138. [doi: 10.1016/j.ijmedinf.2017.04.009] [Medline: 28550996]
45. Seifert A, Reinwand DA, Schlomann A. Designing and using digital mental health interventions for older adults: being aware of digital inequality. Front Psychiatry 2019 Aug 9;10:568-571. [doi: 10.3389/fpsyt.2019.00568] [Medline: 31447716]
46. Fogg BJ. Persuasive technology: using computers to change what we think and do. Ubiquity 2002 Dec;2002(December):5. [doi: 10.1145/764008.763957]
47. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009;24:485-500. [doi: 10.17705/1CAIS.02428]
48. Asbjørnsen RA, Smedsrød ML, Solberg Nes L, Wentzel J, Varsi C, Hjelmesæth J, et al. Persuasive system design principles and behavior change techniques to stimulate motivation and adherence in electronic health interventions to support weight loss maintenance: scoping review. J Med Internet Res 2019 Jun 21;21(6):e14265 [FREE Full text] [doi: 10.2196/14265] [Medline: 31228174]
49. Oduor M, Oinas-Kukkonen H. Committing to change: a persuasive systems design analysis of user commitments for a behaviour change support system. Behav Inf Technol 2019 Apr 04;40(1):20-38. [doi: 10.1080/0144929x.2019.1598495]
50. Radomski AD, Wozney L, McGrath P, Huguet A, Hartling L, Dyson MP, et al. Potential reduction of symptoms with the use of persuasive systems design features in internet-based cognitive behavioral therapy programs for children and adolescents with anxiety: a realist synthesis. JMIR Ment Health 2019 Oct 23;6(10):e13807 [FREE Full text] [doi: 10.2196/13807] [Medline: 31647474]
51. Lentferink AJ, Oldenhuis HK, de Groot M, Polstra L, Velthuijsen H, van Gemert-Pijnen JE. Key components in eHealth interventions combining self-tracking and persuasive eCoaching to promote a healthier lifestyle: a scoping review. J Med Internet Res 2017 Aug 01;19(8):e277 [FREE Full text] [doi: 10.2196/jmir.7288] [Medline: 28765103]
52. Wildeboer G, Kelders SM, van Gemert-Pijnen JEWC. The relationship between persuasive technology principles, adherence and effect of web-Based interventions for mental health: A meta-analysis. Int J Med Inform 2016 Dec;96:71-85. [doi: 10.1016/j.ijmedinf.2016.04.005] [Medline: 27117057]
53. Mandryk RL, Birk MV. Toward game-based digital mental health interventions: player habits and preferences. J Med Internet Res 2017 Apr 20;19(4):e128 [FREE Full text] [doi: 10.2196/jmir.6906] [Medline: 28428161]
54. Brown M, O'Neill N, van Woerden H, Eslambolchilar P, Jones M, John A. Gamification and adherence to web-based mental health interventions: a systematic review. JMIR Ment Health 2016 Aug 24;3(3):e39 [FREE Full text] [doi: 10.2196/mental.5710] [Medline: 27558893]
55. Renfrew ME, Morton DP, Morton JK, Hinze JS, Przybylko G, Craig BA. The influence of three modes of human support on attrition and adherence to a web- and mobile app-based mental health promotion intervention in a nonclinical cohort: randomized comparative study. J Med Internet Res 2020 Sep 29;22(9):e19945 [FREE Full text] [doi: 10.2196/19945] [Medline: 32990633]
56. Przybylko G, Morton DP, Morton JK, Renfrew ME, Hinze J. An interdisciplinary mental wellbeing intervention for increasing flourishing: two experimental studies. J Pos Psychol 2021 Mar 24:1-16. [doi: 10.1080/17439760.2021.1897868]
57. Morton DP, Hinze J, Craig B, Herman W, Kent L, Beamish P, et al. A multimodal intervention for improving the mental health and emotional well-being of college students. Am J Lifestyle Med 2020 Oct 04;14(2):216-224 [FREE Full text] [doi: 10.1177/1559827617733941] [Medline: 32231487]
58. Ajzen I. The theory of planned behaviour: reactions and reflections. Psychol Health 2011 Sep;26(9):1113-1127. [doi: 10.1080/08870446.2011.613995] [Medline: 21929476]
59. Ware JE, Sherbourne CD. The MOS 36-ltem Short-Form Health Survey (SF-36). Med Care 1992;30(6):473-483. [doi: 10.1097/00005650-199206000-00002]
60. Pressman SD, Jenkins BN, Moskowitz JT. Positive affect and health: what do we know and where next should we go? Annu Rev Psychol 2019 Jan 04;70(1):627-650. [doi: 10.1146/annurev-psych-010418-102955] [Medline: 30260746]
61. Contopoulos-Ioannidis DG, Karvouni A, Kouri I, Ioannidis JPA. Reporting and interpretation of SF-36 outcomes in randomised trials: systematic review. BMJ 2009 Jan 12;338:a3006 [FREE Full text] [doi: 10.1136/bmj.a3006] [Medline: 19139138]
62. Shaw J, Treglia M, Motheral B, Coons SJ. Comparison of the depression screening characteristics of the CES-D, MHI-5, and MCS-12 in primary care. 2000 Jun Presented at: AHSR Annual meeting: Behavioural health; June 2000; Washington DC.
63. Donovan KA, Jacobsen PB, Small BJ, Munster PN, Andrykowski MA. Identifying clinically meaningful fatigue with the Fatigue Symptom Inventory. J Pain Symptom Manage 2008 Nov;36(5):480-487 [FREE Full text] [doi: 10.1016/j.jpainsymman.2007.11.013] [Medline: 18495413]
64. Osman A, Wong JL, Bagge CL, Freedenthal S, Gutierrez PM, Lozano G. The Depression Anxiety Stress Scales-21 (DASS-21): further examination of dimensions, scale reliability, and correlates. J Clin Psychol 2012 Dec;68(12):1322-1338. [doi: 10.1002/jclp.21908] [Medline: 22930477]

65. Shea TL, Tennant A, Pallant JF. Rasch model analysis of the Depression, Anxiety and Stress Scales (DASS). BMC Psychiatry 2009 May 09;9:21 [FREE Full text] [doi: 10.1186/1471-244X-9-21] [Medline: 19426512]
66. Jang S, Kim ES, Cao C, Allen TD, Cooper CL, Lapierre LM, et al. Measurement invariance of the Satisfaction With Life Scale across 26 countries. J Cross Cult Psychol 2017 Mar 22;48(4):560-576. [doi: 10.1177/0022022117697844]
67. Pavot W, Diener E, Colvin CR, Sandvik E. Further validation of the Satisfaction with Life Scale: evidence for the cross-method convergence of well-being measures. J Pers Assess 1991 Aug;57(1):149-161. [doi: 10.1207/s15327752jpa5701_17] [Medline: 1920028]
68. Ahmad O, Boschi-Pinto C, Lopez A, Murray C, Lozano R, Inoue M. Age standardization of rates: A new WHO standard. GPE Discussion Paper Series: No.31. World Health Organization. 2001. URL: https://www.who.int/healthinfo/paper31.pdf [accessed 2021-11-02]
69. Tan L, Wang M, Modini M, Joyce S, Mykletun A, Christensen H, et al. Preventing the development of depression at work: a systematic review and meta-analysis of universal interventions in the workplace. BMC Med 2014 May 09;12(1):74 [FREE Full text] [doi: 10.1186/1741-7015-12-74] [Medline: 24886246]
70. Przybylko G, Morton DP, Renfrew ME. Addressing the COVID-19 mental health crisis: a perspective on using interdisciplinary universal interventions. Front Psychol 2021;12:644337. [doi: 10.3389/fpsyg.2021.644337] [Medline: 33927669]
71. Sin NL, Lyubomirsky S. Enhancing well-being and alleviating depressive symptoms with positive psychology interventions: a practice-friendly meta-analysis. J Clin Psychol 2009 May;65(5):467-487. [doi: 10.1002/jclp.20593] [Medline: 19301241]
72. Power E, Hughes S, Cotter D, Cannon M. Youth mental health in the time of COVID-19. Ir J Psychol Med 2020 Dec;37(4):301-305 [FREE Full text] [doi: 10.1017/ipm.2020.84] [Medline: 32611470]
73. Clarke AM, Chambers D, Barry MM. Bridging the digital disconnect: Exploring the views of professionals on using technology to promote young people's mental health. Sch Psychol Int 2017 Aug;38(4):380-397 [FREE Full text] [doi: 10.1177/0143034317700937] [Medline: 29081558]
74. Estacio EV, Whittle R, Protheroe J. The digital divide: examining socio-demographic factors associated with health literacy, access and use of internet to seek health information. J Health Psychol 2019 Oct;24(12):1668-1675. [doi: 10.1177/1359105317695429] [Medline: 28810415]
75. Renfrew ME, Morton DP, Northcote M, Morton JK, Hinze JS, Przybylko G. Participant perceptions of facilitators and barriers to adherence in a digital mental health intervention for a nonclinical cohort: content analysis. J Med Internet Res 2021 Apr 14;23(4):e25358 [FREE Full text] [doi: 10.2196/25358] [Medline: 33851925]
76. Enrique A, Palacios JE, Ryan H, Richards D. Exploring the relationship between usage and outcomes of an internet-based intervention for individuals with depressive symptoms: secondary analysis of data from a randomized controlled trial. J Med Internet Res 2019 Aug 01;21(8):e12775 [FREE Full text] [doi: 10.2196/12775] [Medline: 31373272]
77. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: A systematic and meta-analytic review. J Consult Clin Psychol 2020 Jan;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]
78. Garrido S, Millington C, Cheers D, Boydell K, Schubert E, Meade T, et al. What works and what doesn't work? A systematic review of digital mental health interventions for depression and anxiety in young people. Front Psychiatry 2019;10:759. [doi: 10.3389/fpsyt.2019.00759] [Medline: 31798468]
79. Naslund JA, Gonsalves PP, Gruebner O, Pendse SR, Smith SL, Sharma A, et al. Digital innovations for global mental health: opportunities for data science, task sharing, and early intervention. Curr Treat Options Psychiatry 2019 Dec;6(4):337-351 [FREE Full text] [doi: 10.1007/s40501-019-00186-8] [Medline: 32457823]
80. Zubala A, MacGillivray S, Frost H, Kroll T, Skelton DA, Gavine A, et al. Promotion of physical activity interventions for community dwelling older adults: A systematic review of reviews. PLoS One 2017;12(7):e0180902 [FREE Full text] [doi: 10.1371/journal.pone.0180902] [Medline: 28700754]
81. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JEWC. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
82. Robertson S, White A, Gough B, Robinson R, Seims A, Raine G. Promoting mental health and wellbeing with men and boys: What works? Project report. Leeds: Centre for Men's Health, Leeds Beckett University. 2015. URL: https://eprints.leedsbeckett.ac.uk/id/eprint/1508/1/Promoting_MentalHealth [accessed 2021-04-30]

Abbreviations

DASS: Depression, Anxiety and Stress Scales
MANOVA: multivariate analysis of variance
MHPI: mental health promotion intervention
PSD: persuasive system design
RCT: randomized controlled trial
SF-36: Short Form 36 Health Survey

SWLS: Satisfaction With Life Scale

Edited by J Torous; submitted 24.04.21; peer-reviewed by P Batterham, B Kahl; comments to author 14.05.21; revised version received 21.06.21; accepted 03.10.21; published 11.11.21.

Please cite as:
Przybylko G, Morton D, Morton J, Renfrew M
The Influence of Gender and Age on the Outcomes of and Adherence to a Digital Interdisciplinary Mental Health Promotion Intervention in an Australasian Nonclinical Setting: Cohort Study
JMIR Ment Health 2021;8(11):e29866
URL: https://mental.jmir.org/2021/11/e29866
doi:10.2196/29866
PMID:34762058

Original Paper

Clinician Perspectives on Using Computational Mental Health Insights From Patients' Social Media Activities: Design and Qualitative Evaluation of a Prototype

Dong Whi Yoo[1], MFA, MS; Sindhu Kiranmai Ernala[1], MS; Bahador Saket[1], PhD; Domino Weir[1], MS; Elizabeth Arenare[2], BA; Asra F Ali[2], MA; Anna R Van Meter[2,3,4], PhD; Michael L Birnbaum[2,3,4], MD; Gregory D Abowd[1,5], PhD; Munmun De Choudhury[1], PhD

[1]School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA, United States

[2]The Zucker Hillside Hospital, Northwell Health, Glen Oaks, NY, United States

[3]The Feinstein Institutes for Medical Research, Manhasset, NY, United States

[4]The Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY, United States

[5]College of Engineering, Northeastern University, Boston, MA, United States

Corresponding Author:
Dong Whi Yoo, MFA, MS
School of Interactive Computing
Georgia Institute of Technology
756 W Peachtree St NW
Atlanta, GA, 30308
United States
Phone: 1 4043858603
Email: yoo@gatech.edu

Abstract

Background: Previous studies have suggested that social media data, along with machine learning algorithms, can be used to generate computational mental health insights. These computational insights have the potential to support clinician-patient communication during psychotherapy consultations. However, how clinicians perceive and envision using computational insights during consultations has been underexplored.

Objective: The aim of this study is to understand clinician perspectives regarding computational mental health insights from patients' social media activities. We focus on the opportunities and challenges of using these insights during psychotherapy consultations.

Methods: We developed a prototype that can analyze consented patients' Facebook data and visually represent these computational insights. We incorporated the insights into existing clinician-facing assessment tools, the Hamilton Depression Rating Scale and Global Functioning: Social Scale. The design intent is that a clinician will verbally interview a patient (eg, How was your mood in the past week?) while they reviewed relevant insights from the patient's social media activities (eg, number of depression-indicative posts). Using the prototype, we conducted interviews (n=15) and 3 focus groups (n=13) with mental health clinicians: psychiatrists, clinical psychologists, and licensed clinical social workers. The transcribed qualitative data were analyzed using thematic analysis.

Results: Clinicians reported that the prototype can support clinician-patient collaboration in agenda-setting, communicating symptoms, and navigating patients' verbal reports. They suggested potential use scenarios, such as reviewing the prototype before consultations and using the prototype when patients missed their consultations. They also speculated potential negative consequences: patients may feel like they are being monitored, which may yield negative effects, and the use of the prototype may increase the workload of clinicians, which is already difficult to manage. Finally, our participants expressed concerns regarding the prototype: they were unsure whether patients' social media accounts represented their actual behaviors; they wanted to learn how and when the machine learning algorithm can fail to meet their expectations of trust; and they were worried about situations where they could not properly respond to the insights, especially emergency situations outside of clinical settings.

Conclusions: Our findings support the touted potential of computational mental health insights from patients' social media account data, especially in the context of psychotherapy consultations. However, sociotechnical issues, such as transparent

algorithmic information and institutional support, should be addressed in future endeavors to design implementable and sustainable technology.

(JMIR Ment Health 2021;8(11):e25455) doi:10.2196/25455

KEYWORDS

mental health; social media; information technology

Introduction

Background

Mental health treatment relies heavily on what the patient tells their clinician during in-person consultations. However, issues of retrospective recall bias [1,2], impression management goals [3], and social desirability bias [4] have motivated mental health clinicians to augment patient reports using *collateral information* [5], such as those obtained from patients' friends and family members. According to the George Engel biopsychosocial model of care [6], such information provides a complementary and adjuvant perspective on the patient's condition, which the clinician can use to tailor treatment decisions, regulate the quality of care, and support the patient on the road to recovery [5].

The ubiquity and increasing use of digital technology have opened up new opportunities for clinicians to gather complementary sources of collateral information, which can be diverse in scope and gathered in the natural contexts of the patients [7]. Patient-generated health data of patients with irritable bowel syndrome, such as food intake and abdominal pain, have been explored in provider-patient collaboration [8]. The providers saw that self-monitoring data could support provider-patient communication; parallelly, they were also worried about insufficient time to review the data or not having meaningful results from such investments. Kim et al [9] developed DataMD, a clinician-facing patient-generated health data dashboard, by conducting design workshops with clinicians. They found that DataMD helped clinicians to improve counseling skills and facilitated in-depth communication between a clinician and patient.

Among the different types of data sources that can provide collateral information, patients' social media activities have been investigated in diverse settings such as healthy eating [10] and forensic mental health evaluations [11,12]. Researchers have suggested that social media platforms have emerged as low-cost and unobtrusive means to gather insights about behaviors, mood [13], psychological traits [14], social interactions [15], and even the mental health states of individuals [16,17]. As these platforms provide an unprompted medium through which individuals can voice their feelings and daily experiences, digital traces left behind by people on these platforms provide opportunities for clinicians to gain another layer in their understanding of patients [7,18].

In the wake of these opportunities, clinicians have expressed interest in exploring the use of patients' social media as clinically relevant information [19]. At the same time, they have been keen to weigh the benefits and drawbacks of doing so [20]. Various studies have suggested ethical guidelines, such as professional boundary management and informed consent, when incorporating social media into clinical settings [21-25]. Even if patients are fully informed, it is unclear how they will share their social activities and to what extent and how the sharing will inform clinicians' decision-making processes [26,27]. It is also possible that fully informed consented patients may alter their behaviors, which weakens the usefulness of the collateral information from social media data [28]. Moreover, the collateral information derived from patients' social media should be relevant to the clinical context and provided in a way that clinicians can access their current workflows [29].

Therefore, further research is required to create social media–based technologies that can empower clinician-patient collaboration as collateral information while preventing such technologies from exacerbating ethical concerns. Future technologies need to be able to protect professional boundaries when clinicians and patients collaborate using social technologies. In addition, patients' privacy must be respected even if the patients have consented to share their social media posts. One of the potential solutions is to computationally translate patients' social media posts into clinically meaningful insights such as the intensity of certain symptoms [30] and the possibility of relapse [31]. By only showing possibilities or indexes calculated from social media posts, some of the ethical concerns mentioned above can be assuaged; clinicians will not read what the patient posted but will be able to glean important information such as indicators of exacerbation of their symptoms. However, this approach creates other questions: What are the relevant and useful information derived from patients' social media data? How would clinicians incorporate this information into current work practices? How would new technologies be salient in addressing the ethical concerns of using sensitive personal information in a clinical context?

Objectives

To examine how collateral information computationally derived from patients' social media can support or hinder mental health therapy, we developed a clinician-facing prototype that visually represents patients' social media data. We focused on patients with mood disorders because the collateral information that can be distilled from patients' social media is relevant to patients with mood disorder [32]. We further left our target condition broad because of the early and exploratory nature of this study.

The prototype was qualitatively evaluated by 15 mental health clinicians. The evaluation study accomplished 2 goals: (1) it helped us understand whether and for what purpose clinicians could incorporate the prototype and social media insights gleaned from patients' data into their work practices and (2) it revealed concerns and potential harms in its use and adoption in real-world clinical settings. In this study, we present the

findings from the user study sessions with mental health clinicians and the implications for future mental health technologies as well as ethical considerations of using patients' social media data in the mental health context.

Methods

Overview

On the basis of the low-fidelity prototypes designed by the research team [29], we developed a prototype with Facebook data of consented patients with mood disorder in treatment at a large health center in the northeast of the United States. The prototype was qualitatively evaluated by clinicians via interviews (n=15) and focus groups (n=13) at this location. The following subsections explain the details of the prototype and the evaluation methods.

Prototype

Overview

As clinicians may be unaccustomed to the concept of computational mental health insights from social media data, we decided to design a prototype that can help clinicians understand this concept and envision its future. The design of our prototype is based on our previous codesign work to understand how computational social media analyses can be visually represented by clinicians [29]. We extended the previous low-fidelity prototypes in 2 ways: first, we used actual patient social media data in the design of the prototype because the insights generated from the actual data and deidentified vignette of the patient can help our clinician participants to evaluate the opportunities of the computational approaches; second, we situated the computational mental health insights as a part of existing clinician-facing assessments because those assessments provide our participants with a familiar base of understanding. A detailed explanation of patient social media data and the design of the prototype are provided below.

Patient Facebook Data

In general, this study draws on data from a larger study, some of which have been reported in the studies by Saha et al [30], Birnbaum et al [31], and Ernala et al [33]. In this study, the Facebook archives of a set of clinically diagnosed patients with mood disorder were downloaded following informed consent from the patients and after approval by the institutional review boards of the relevant institutions. From 110 patients who contributed their data following informed consent, we selected an exemplar set of 8 patients with mood disorder, who had the highest activity on Facebook, to build the prototype. Overall, patients had an average of 7143.4 (SD 3209.1) timeline posts and 21,043.6 (SD 16,761.6) messages spanning between 1 and 10 years (mean 6.5, SD 3.6) on Facebook. In particular, the following types of data were used for the specific purposes of our prototype: self-posts and self-comments (posts, comments, and interpersonal messages posted by the patient), including their time of posting, check-ins, friending and cotagging activities, and volume of interpersonal social interactions. In addition to Facebook data, we also accessed their primary diagnosis and hospitalization information (eg, admission and discharge dates) from their medical records.

We used a number of computational analyses on the Facebook data of patients, grounded in the symptomatic and functional impairments associated with mood disorders [29]. To identify posts indicative of depressed mood and suicidal ideation, we used machine learning classifiers (bag-of-words–based -gram models) from prior research [30]. The depression classifier showed an accuracy of 0.82, and the suicidal ideation classifier had an accuracy of 0.91. To capture insomnia, we calculated the number of Facebook posts that were posted during regular sleep hours (between midnight and 5 AM). For diurnal variation in association to mood, we calculated the number of depression-indicative posts (as predicted by the classifier) that were posted at different times of the day (morning, noon, night, and midnight defined between 5 AM and noon, noon and 5 PM, 5 PM and 10 PM, and 10 PM and 5 AM, respectively) [17]. Next, as a measure of new friendships, we calculated the number of accepted friend requests on Facebook. To operationalize social ties, we calculated the number of distinct people the patient messaged on Facebook and the total number of messages exchanged [34]. Finally, to measure the frequency of offline social interactions, we determined the number of posts that had location check-ins or cotagging with other people [31]. In general, we prioritized these specific items because they are well-validated and well-supported in the literature [31,33] in terms of revealing meaningful mental health insights from a clinical standpoint and in a clinical patient population.

Augmenting Existing Assessment Tools

We adopted 2 existing psychiatric assessment tools for the design of our prototype—the Hamilton Depression Rating Scale (HAM-D) [35] and the Global Functioning: Social (GF:S) [36]—because they are well-established tools that help clinicians track symptoms and the emotional, social, educational, and vocational functioning of patients. Typically administered in the form of interview-based assessments, there are 24 items in the HAM-D (eg, depressed mood and feelings of guilt) and 8 items in the GF:S (eg, Do you ever have problems or fallings out with friends?). For each item, clinicians ask questions to patients in person and observe their behaviors during the interview to assess their patients.

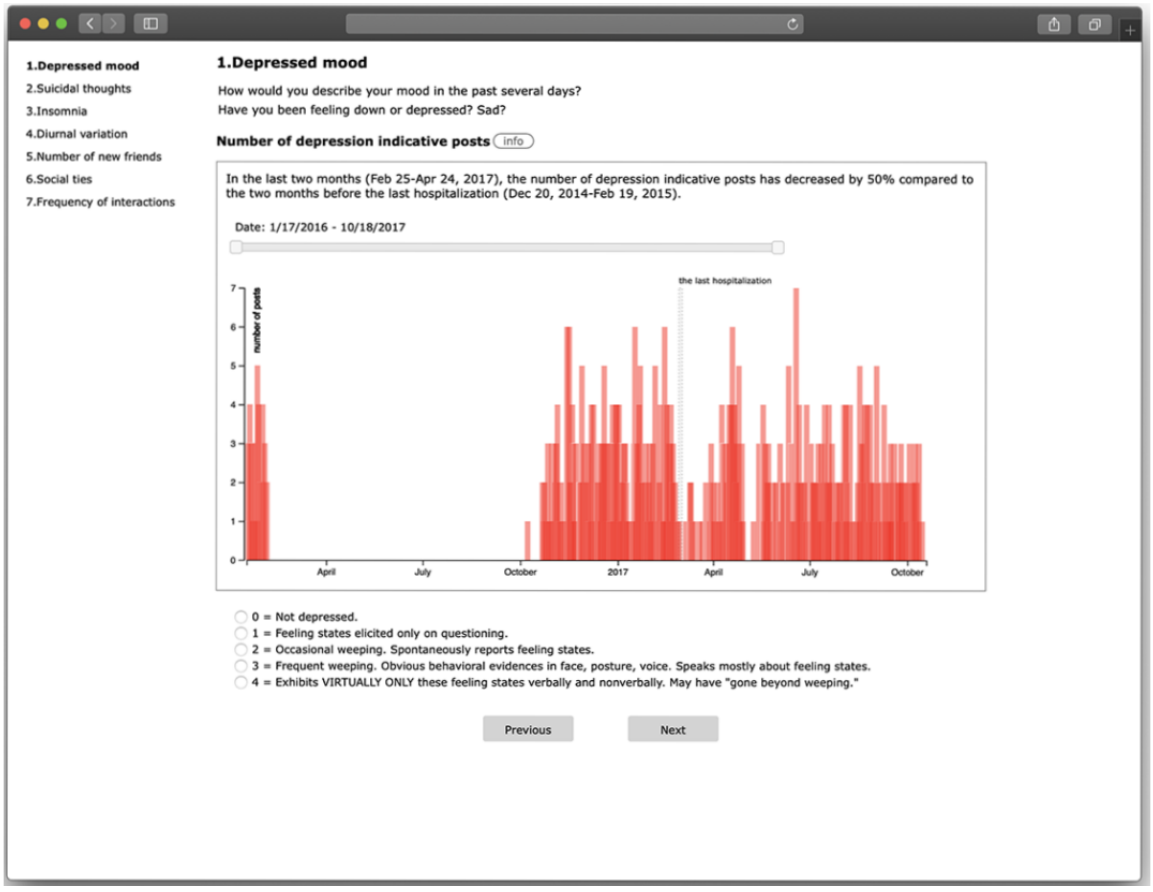
Although the assessment tools focus largely on offline behaviors—aspects that may not be covered in an individual's Facebook activities—we designed a prototype that would enable clinicians to quickly assess social media–derived insights as an additional layer of collateral information on top of what might be accessible through the assessment tools. Our clinical collaborators felt that such complementary information gathered from patients' social media activities can be useful. After deliberation and considering the social and emotional affordances of Facebook, we picked the 4 items from the HAM-D and 3 items from the GF:S (Table 1) that can be most reliably mapped to an analysis of patients' Facebook data described earlier.

Table 1. Items, interview guides, and social media analysis in the prototype.

Items	Interview guide	Social media analysis
Depressed mood	How would you describe your mood in the past several days?	Number of depression-indicative posts
Suicidal thoughts	In the past several days, have you felt that life was not worth living, or that you would be better off dead?	Number of suicidal thought indicative posts
Insomnia	How have you been sleeping in the past several days?	Number of Facebook posts between midnight and 5 AM
Diurnal variation	In the past several days, have you noticed feeling worse at any particular time of day—such as in the morning or evening?	Number of depression-indicative posts by 4 time frames
New friends	Tell me about your social life. Do you have friends? If yes, how many friends would you say you have?	Number of accepted friend requests
Social ties	Are they casual or close friends?	Number of messages and recipients
Frequency of social interactions	How often do you see friends?	Number of posts with location and tagging

We created an electronic version of a clinician-facing assessment dashboard that was augmented with social media analysis (Figure 1). We provided interview questions from the assessments on the top of the main page, so that clinicians can initiate the interview process. At the bottom of the screen, we placed anchored rating scales for the item, which come from either the HAM-D or the GF:S. Between the interview questions and the rating scales, the relevant social media analysis is displayed as collateral information. For example, for the depressed mood item, we added the number of depression-indicative posts from the patient's Facebook data, visualized as a time series bar graph. The y-axis of this graph is the number of posts, and the x-axis of the graph represents time, that is, the time from account creation to the most recent activity. We added a range slider for the clinicians to adjust the time frame. Next to the title of the social media analysis, we added an information button that shows how we calculated the number of posts (eg, depression classifier for depression-indicative posts). Finally, we added a comparison between the last 2 months and 2 months before the last hospitalization of the same patient in plain text. This was to help clinicians find patterns that could indicate a change in symptoms. Additional screenshots of the prototype (eg, the suicidal ideation view) can be accessed in Multimedia Appendix 1.

Figure 1. One view of the prototype.



Qualitative Evaluation of the Prototype

Recruitment

To formatively evaluate the prototype powered by actual patients' deidentified data, we used purposive sampling strategies to recruit mental health clinicians from a large, urban, behavioral health center located in the northeast of the United States. This research was approved by the institutional review boards of the relevant organizations.

To facilitate the recruitment process and to compensate for their participation, we provided a raffle for an iPad mini. We recruited 15 clinicians for individual interviews, and 13 of the 15 participated in a set of subsequent focus group sessions (3 sessions with 4-5 participants per session). In total, we had 8 psychiatrists, 5 clinical psychologists, and 2 licensed clinical social workers (Table 2). We grouped the focus group participants based on their availability. We recruited a heterogeneous group of clinicians because psychiatrists, psychologists, and social workers are highly collaborative in our study site.

Table 2. Participant demographics with their experience and gender. All focus group participants joined individual interviews before their focus group sessions.

Participants and title		Experience (years)	Gender
Focus group 1			
P1	Psychiatrist (MD)	8	Female
P2	Clinical psychologist (PhD)	8	Female
P3	Licensed clinical social worker (MS)	6	Female
P4	Clinical psychologist (PhD)	7	Female
P5	Clinical psychologist (PhD)	11	Female
Focus group 2			
P6	Psychiatrist (fellow, MD)	4	Male
P7	Psychiatrist (fellow, MD)	6	Female
P8	Clinical psychologist (PhD)	20	Female
P9	Licensed clinical social worker (MS)	30	Female
Focus group 3			
P10	Psychiatrist (resident, MD)	1	Female
P11	Psychiatrist (resident, MD)	2	Female
P12	Psychiatrist (resident, MD)	2	Female
P13	Psychiatrist (resident, MD)	3	Female
Only interview			
P14	Psychiatrist (resident, MD)	3	Male
P15	Clinical psychologist (trainee, MS)	5	Female

Procedure

First, to familiarize the participants with the prototype, we conducted an individual interview, where the participant explored the prototype and gave their feedback on it. Second, to envision future uses of the prototype, we conducted 3 focus group sessions where participants discussed the values and barriers of the prototype. These sessions were conducted between June and August 2019. The interview and focus group protocols are included in the Multimedia Appendix 2.

We interviewed participants in the same offices where they met their patients. Before obtaining informed consent, we provided an overview of the study. After the participants signed the informed consent form, they completed a demographic survey. We then asked questions about their work practices and their experiences with the patients' social media. We demonstrated the prototype on a laptop. Next, the participants were asked to freely explore the prototype using a think-aloud protocol. We provided a vignette of a real patient, which was deidentified and edited for the study; the prototype displayed the same patient's Facebook data. The clinician participants and the patients in the prototype were from the same behavioral health center; however, we did not check whether the participants had actually met the patients. Following the exploration, the participants answered follow-up questions regarding general feedback, compatibility with their work practices, and concerns and thoughts. The duration of the sessions ranged from 45 minutes to 80 minutes.

The subsequent focus group sessions were held for 55 minutes to 70 minutes. We explained the purpose of the focus groups and obtained informed consent from the participants. To refresh their memory of the prototype through a brief reintroduction, we asked about participants' general feedback on the prototype.

After that, the participants led the discussion comprising topics such as their willingness to use the prototype or their concerns.

Analysis of Qualitative Data

The sessions were audio-recorded with the participants' permission, and the recordings were transcribed. The transcribed data were analyzed using thematic analysis [37]. The research team (DWY and BS) first deductively coded the data based on our research aims: (1) whether and for what purpose clinicians could incorporate the prototype into their work practices and (2) concerns and potential harms in its use and adoption in real-world clinical settings. We then inductively analyzed the data to identify particular patterns in the data. The initial codebook was iteratively reviewed during regular team meetings (DWY, MDC, MLB, and ARVM) until the team reached a consensus. The list of codes, their descriptions, and examples for each are included in the Multimedia Appendix 3.

Results

Overview

Participants considered the computational insights from patients' social media to be helpful for clinician-patient collaboration. However, they also pointed out potential negative consequences and concerns that should be addressed in future technologies.

Clinician-Patient Collaboration

While freely exploring the prototype, our participants voluntarily explained how they would like to use it to have a better conversation with their patients. In particular, they mentioned diverse collaborative situations where the prototype can be useful, such as when they explore directions for consultation and track changes in patients' symptoms.

Collaborative Agenda-Setting

Our participants considered the prototype to be useful for them to collaboratively set an agenda with their patient—an approach strongly advocated in patient-centered and collaborative care models [38]. P11 suggested that reviewing the prototype with the patient at the beginning of the consultations could create awareness and concern about salient issues in a collaborative manner and the patient could feel that there is a more welcoming space where they can now "open up" and be "a contributor in [their] own treatment" (P5). The participants considered this a valuable outcome for enhancing their therapeutic alliance [39]:

> *If we have the patient in the office and were like, "Let's spend 5 minutes and go through your data together." And we look at the graphs together. And then point out, oh, it looks like at this period, you were posting a lot at nighttime, what was going on? And you just use it as a way to explore if something didn't come up in the session. So obviously it could be used in a therapeutic way to enhance the therapeutic relationship.* [P11]

Participants pointed out that if they could have time to review the prototype before the consultations, it would help set the stage for what is to come during the session, and could affect the course, direction, and quality of care, including treatment decisions. In fact, such an approach could result in fewer *hidden* concerns at the end of the consultation (P9). For instance, clinicians could look for atypical or concerning patterns that might stand out when reviewing the prototype. If appropriate, during the session, they would then actively seek to know what happened during that time or why the patient posted something in particular on social media:

> *I could see myself using it this way: so if I'm meeting them on October 17 for an appointment, I'll say, "Oh, have you had any suicidal thoughts in the last week?" And they say, "No." And then, "I'm seeing on three occasions it looks like that on social media you were expressing something that maybe was concerning for suicidal thoughts, can you tell me about what these were if you remember them?" So I might use it to hone in on specific instances of suicidal thoughts.* [P11]

Communicating Symptoms

Communicating symptoms is one of the most important parts of clinician-patient collaboration [40]; however, it is often challenging because of the subjective nature of most mental illnesses and a lack of efficacious ways to monitor them longitudinally and in a fine granularity [41]. Therefore, our participants consistently pointed out that being able to gather more objective information regarding patients' symptoms with the prototype could help both clinicians and patients to communicate symptoms:

> *Sometimes patients forget or don't recall clearly for how long they had the symptoms, or how long they thought they have been secluding themselves in their room. Sometimes they don't recall a rough timeline. But if you have the data in front of you in terms of how often have they been going out, and if you can clearly see a drop if they are a social person, but there has not been even a single tag, or they have not gone outside for a long while, you know you have to check into the situation.* [P7]

Participants further stated that sometimes patients may *minimize* certain symptoms, or as P11 noted, they can struggle to "recognize small changes," in which case the prototype can help learn about the patients' mental state. Difficulties in recognizing and communicating symptoms can be seen under certain conditions:

> *So maybe in a bipolar patient that would be more helpful. If they're saying, "Oh yeah, I'm sleeping well" but they're posting throughout the whole night, then you could see that their sleep patterns are off.* [P12]

Participants said that if the clinicians encountered such discrepancies, they would like to cautiously bring up the information from the social media analyses, as long as the patient is comfortable. In addition, they would explore the opportunity to address the gaps by having deeper conversations, such as by asking patients to unpack the foundations of this contradiction. P7 reported that patients "often live in denial" as and when they feel better intermittently. In that case, based on the prototype, it might be a meaningful psychotherapy probe to

know "why the patient is not sharing what they have not shared," that is, "was it just forgetfulness, or was it intentional on the part of the patient to hide certain things?" (P7).

However, participants also emphasized that the early moments of clinical interactions are important, as the tone of voice used by a clinician early in the visit is known to be indicative of satisfaction and compliance with treatment recommendations [38]. Therefore, they mentioned that there should be inconsistencies between what the patient says and what the prototype shows, they would approach this in a nonconfrontational manner, "do [so] subtly and bring it up to [the patients'] attention" (P3) and negotiate the appropriate time when this discussion may be timely.

As family members often engage in tracking symptoms and communicating with clinicians [42], our participants mentioned that the prototype may be useful in resolving conflicts between patients' self-reports and collateral information from their family members:

> *Sometimes there is a tendency, by parents, when they do not recall clearly, of overgeneralizing things, like the patient has not talked to anybody in the last two or three months, has been really doing bad. But when you explore it clinically, the patient may tell you otherwise, even though the parent might deny it. But when you can see from their Facebook as well that they have been going out, they have been enjoying things that they had in the past, this can definitely correlate that fact.* [P7]

We considered the information from patients' social media as another type of collateral information rather than a type of information that can replace any of the current information that clinicians may use. Our expectation is that clinicians will collectively consider every type of information available, including discreet conversations with patients. Our participants confirmed their interests in including information from social media in their decision-making process when appropriate.

Incorporating the Tool in Current Work Practices

Different types of clinicians envisioned various ways to incorporate the prototype into their work practices. Clinical psychologists and licensed social workers (and some psychiatrists) mentioned that they would like to use the prototype similar to *homework assignments* [43]—cognitive behavioral therapy strategies suggest that homework assignments can help patients practice coping strategies and restructure dysfunctional beliefs. As our clinicians already discussed patients' assignments at length during therapy sessions, they envisioned that the prototype could provide additional interesting discussion points. Although clinical psychologists and social workers preferred using the prototype to navigate their conversations, psychiatrists mentioned that they would like to check whether there were sudden changes after they modified some medication treatments. One of the psychiatrist participants, P14, mentioned that "being able to input when I started a medication would be very useful. And being able to even just note dose changes would be cool."

Second, some participants suggested that reviewing the technology before consultations might be better from the perspective of patient-clinician engagement during consultations. This will ensure that conversations are not negatively disrupted with technology use, and it will prevent patients from "feeling neglected" (P10):

> *I think this is very valuable and I think there's a very good role for that being incorporated in treatment. Personally, I might like having it here, something I review beforehand and then as needed or do a check in at a portion of the session where I'm like, "Let's look together." I just don't know logistically if I'd want to keep [the prototype] in front of me the whole session. The patient might think, I don't like how now my doctor is standing at a computer typing instead of talking to me.* [P8]

A third potential use of the prototype that some participants brought up included the possibility to learn about or keep track of specific patients' symptom improvements or downturns when patients miss an appointment. P11 cited a case in which the prototype could provide timely feedback to the patient to enable them to self-reflect and be self-aware:

> *It's often the case where patients don't recognize small changes as much as maybe other people around them. So things like, they're smiling more. They're brighter, they're more interactive, they're talking for longer during the session. Those might be signs that their mood is improving. They might not notice it. So if there was some feedback I could give them, like "I notice that you're looking a little bit brighter today or you're a little bit better. And in fact, based on your social media use, it looks like you've been posting more positive things." That would be a great way to show it.* [P11]

P13, on the other hand, found that they could use the tool to connect with the patient in a timely fashion, even if an in-person consultation was not possible:

> *The irony of it is that when patients actually get sick is when they don't come to see you. But if you are able to check in on them, like with this tool, that can be cool.* [P13]

Potential Negative Consequences

Collaboration Versus Monitoring

Another conspicuous theme throughout the participants' accounts concerns the potential negative consequences of the prototype. They pointed out that there will be a subtle line between collaboration and monitoring, and some patients might be negatively affected by the prototype. In addition, our participants expressed their negative opinions regarding the additional workload that the prototype may bring to clinicians.

Although participants voluntarily mentioned that the prototype can support collaboration and engagement during the consultations, some participants also expressed concerns; they provided scenarios where the patient may not choose to be an amicable party to the process. For instance, P9 mentioned that

the mental illness experiences of certain patients may prevent their participation in the use of this prototype or there could be negative consequences:

> *The concern that everything that they do or even if they're being monitored, big brother's watching and even though consents are signed, I mean, paranoia is what, an irrational fear and they're very vulnerable. So, it can go the other way too.* [P9]

Participants were concerned that the prototype's abilities and clinical usefulness might be undermined by the Hawthorne effect [44], wherein patients may stop posting or begin to self-censor themselves on social media, knowing clinicians' awareness of and access to this information:

> *It would be interesting how this will, this would modify their behavior on their social media considering the fact that they know now that, even their social media post has been given access to their clinicians. So like it's being monitored. So like that might modify their behavior on the social media as well, either positively or negatively, depending on if they are seeking help or if they are seeking attention, in one way or the other. They might post more or they might start posting less.* [P15]

They further conjectured that patients who are less open and engaged during consultations would not consent to provide their social media data to the prototype, which is an important concern because openness and trust are critical to therapeutic alliance:

> *I feel like for the patients I'd want it, it's those that I don't trust and they're not going to necessarily trust everyone with stuff on social media, and also kind of trust me to go look at their data, like to give me permission. So it would probably be looking at a lot of data from patients that we could just ask them the questions and they'll be honest with us.* [P4]

Workload Issues

First, despite acknowledging it as a "technicality" (P7), participants were worried about the potential burden on their workloads. P11 felt that it may not always be feasible to review patients' social media information before consultations in some programs, because in some clinical settings, patient loads are exceedingly high. In fact, the participants felt that reviewing additional data from the prototype might increase work and call for advanced training, either of which is likely to be impractical without adequate support from their institutions.

Participants also considered other areas of concern, such as general management of the prototype, explaining the scope of the prototype to (new) patients and its functioning to clinicians, maintaining informed consent from patients, getting help from the information technology staff to allow sustained use of the prototype, and ensuring that it is seamlessly integrated with other pieces of clinical information gathered by the institution—all of which they thought could lead to an increase in clinician burden. For instance, they pondered on who would educate the patients about this technology and manage issues, both technically related and patient-related. To this end, they thought they may be more willing to use this system when they are employees in a large hospital where someone else can handle the aspects surrounding the functioning of the system.

Ethical Concerns

In addition to the potential negative consequences, our participants pointed out concerns that need to be addressed before this technology could be introduced in the clinical context.

Patient Privacy

When we introduced our prototype, we explained the privacy-related settings for it. The prototype's data were collected with the patient's consent for research purposes, and we envision that future technologies will actively seek patients' consent to use their social media information in their treatment. Some participants mentioned that they were worried about patients' privacy; however, they considered achieving patients' consent to be the first step toward addressing such issues:

> *Also, the idea of someone being able to have their privacy of being able to poach these things without them having to have their doctors know about it all the time, but I guess if they're agreeing to it, and that means they don't mind.* [P1]

Our participants also provided keen insights regarding the sharing preferences of patients in a clinical context. P7 pointed out that even if the post is public, it is not clear whether the patient will be fine while sharing content with their clinicians. This idea opens up new questions about the difference between sharing a post with their friends and sharing a post with their clinicians. More importantly, this indicates that future consent procedures should be thorough in communicating the implications of sharing patients' social media data with their clinicians. These ethical implications are explored in the Discussion section.

One of our measures to respect patients' privacy, the design decision to not show the actual post in the visual representation, was well-received by clinicians. They pointed out that "not having a specific post is a little bit less invasive to the person's privacy" (P8). However, it also raised a question about the trade-off between having the ability to review what the patient wrote and to protect patients' privacy. This trade-off can be important, especially when they find a trend or pattern that might be relevant to the patient or their treatment. Multiple participants mentioned that they would like to read the post if the posts were flagged as suicidal ideation–indicative or depression-indicative, and they felt the pattern was important in the patient's treatment. We envision that this tension regarding the granularity of shared data should be considered in future designs such as specific customization options for both clinicians and patients to decide the level of details that will be shared between them.

Credibility

There were 2 dominant credibility issues that participants thought could diminish trust in the prototype: if "Facebook posts and friend requests and everything correlate to actual life" (P12) and if an algorithm applied on top of these data can distinguish different contexts and intents behind specific posts. For instance,

1 participant repeatedly pointed out several times that they wanted to learn how the different algorithms worked and when they failed. Another participant mentioned a concrete scenario in which the algorithm may not be able to provide clinically meaningful insights because of the underlying gaps in psychometric validity:

> *And how does that algorithm delineate that certain posts are more likely to be related to depression versus others? Like if somebody had just posted they're listening to some dark music, would it automatically pick that they're suffering from low mood, that's why they're. Because sometimes people just write that on their Facebook post, they're listening to this, and, like, that's a part of dark music, or in general, sad songs.* [P7]

Another participant similarly mentioned that as the prototype does not distinguish between active and passive suicidal ideation, it cannot hint at the specific circumstances under which a patient may have shared a suicidal ideation–indicative post. A lack of this context may prevent clinicians from adjusting their treatment decisions based on the prototype, and it can be particularly difficult to ensure that the data augments patients' and caregivers' accounts, instead of eroding them:

> *Unless there's behavioral action to back up what the person may have posted, I feel like it's unclear how much conviction they had, and what they were saying, or whether it was just for attention. I think in terms of what is said, maybe more active stuff like, "I want to die; Life isn't worth living anymore" will be more useful.* [P10]

Another participant further questioned relying on Facebook as the sole source of collateral information, as "people might [be] on a Facebook break" (P12). Ultimately, without subverting the utility of the prototype, participants said that, in the absence of an implicit level of trust or transparency in the functioning of the algorithms, they would consider the social media insights with caution:

> *I would probably trust the patient's report more than I would trust the data from Facebook. I mean if the patient's saying they're doing totally fine and then they're having a bunch of suicidal posts, I guess it's a thing to bring up. But I wouldn't necessarily feel that they're suicidal because their posts say they are. I guess I would want more information, I wouldn't just take it at face value. Because I know people post things for all different kinds of reasons. So I guess that's my hesitancy, is this. . . and I got to trust what I see here.* [P11]

Liability

Liability issues may arise when clinicians have access to patients' social media information via the prototype, which indicates an exacerbation of their symptoms, such as a crisis, but clinicians are not in a position to take any appropriate action:

> *There are posts which can be very critical. For example, "Oh, I'm going to kill myself now." And then, if you don't see this post, even though you have access to this information and you can access it at any time, are you responsible? So, there's a lot more questions that come up if you have unlimited access to patient information at any time because the computer can then flag it. Right? That's why I'm kind of my concerned more with like the legal and ethical stuff of how much you can/cannot be held responsible for.* [P6]

To this end, our participants wanted to clarify whether the Facebook data collection, and the analytics on top of it, happens in real time or if it is an episodic event that only happens when they meet their patients. On hearing that we intend the prototype to be used only when the clinician meets with the patient, participants thought the very act of volunteering their Facebook data may lead patients to think they are receiving 24/7 care; they may expect crisis mitigation resources all the time, outside of periodic clinician consultations. Participants felt that this could not only impact how clinicians currently manage crisis scenarios but also negatively impact their therapeutic relationship when patients' concerns are not addressed as they occur. Consequently, participants highlighted the need for ethical and legal help from their institution:

> *I like to welcome that idea but I think if I'm in my private clinic or if I'm the only clinician then I would think about whether I would apply this, given various legal and ethical questions. Perhaps I would be more comfortable using it in the larger institution like here, in this hospital. Because they have legal rights and experts, and then if they say, "Okay, you can use it," then, I'll probably be more comfortable using it.* [P6]

Clinicians also brainstormed the liability around the aforementioned possible use case where they accessed patients' social media information outside of consultations, such as when patients missed their consultations. We envision that, if the proposed technology is implemented in a real-world setting, it will be imperative to delineate when and in what circumstances (eg, during or outside of consultations) accessing these data is acceptable to the patients.

Discussion

Implications for Future Mental Health Technologies

Our work raises a vital question—how do we expect mental health treatment to be shaped in the future if a technology such as our prototype were to be used by clinicians?

The Future of Clinicians' Work

Our findings reveal that our prototype can be a step forward in developing clinician-facing technologies that harness voluntarily shared patient social media data in mental health care delivery—a possibility advocated in prior work [33]. We found our prototype to be capable of providing a nuanced understanding of a patient's unique illness course and clinical needs over time. Augmenting the short infrequent visits of today with our prototype, clinicians felt they could distill a stream of fluctuations in symptoms for a patient, calibrated against their baseline behaviors, and quantified against their past trends to detect subtle changes. Clinicians also appreciated the

opportunity to correlate, corroborate, and contrast a patient's clinical presentation with their behavior outside of the visit setting—a capability that can be particularly meaningful when a patient, because of cognitive impairment, has difficulty articulating their condition.

At the same time, clinicians expressed concerns regarding the credibility of the computational approaches that power the prototype. This was largely attributed to the fact that the acquisition of patient data was opaque to the clinicians and because providers thus far have not acquired patients' social media data in the past. This is perhaps also unsurprising because many mental health clinicians are not used to using algorithm-generated information in their day-to-day work. Our participants repeatedly asked if social media reflected the patients' actual mental health state, how the algorithms work, whether they were tested in a real treatment scenario, and when the algorithms failed. We noted that previous work has shown that people's social media activities represent their actual selves [45], and we used previously validated social media measures of mental health in powering this tool [30,31,33]. However, we acknowledge that the prototype needs to persuade our potential users rather than relying on the support of previous research.

Technology has been reshaping the future of work in many domains [46,47]. Clinical work on mental health is no exception. These identified general needs for trustworthy algorithms are a core aspect of the future of clinicians' work, which resonates with recent studies highlighting the importance of explainable, interpretable artificial intelligence and machine learning in health care [48,49]. Alongside these efforts, our findings also emphasize the need to consider structural changes in the future of clinicians' work, specifically educating clinicians about the technology, not only to reduce the negative impact and potential harm attributed to poor credibility but also to make such technology more accessible to clinicians who may be conservative about new technologies.

We suggest the following calls to action accordingly:

1. Include resources to clarify the data collection process, when the system acquires the data, where the data are stored, and how the system accesses patient social media accounts, before both clinicians and patients experience this technology.
2. Provide the details of the algorithms both on-demand and contextualized in their demonstrated clinical efficacy, including evidence-based endorsement that can assure clinicians that the quality of the algorithms that power the technology is adequate.
3. Consider how technology education may be part of the psychiatry training paradigm, so that clinicians can gain some fluency in using a future version of this technology as an adjuvant tool in their clinical work.

The Patient-Clinician Therapeutic Relationship

We noted clinicians' enthusiasm regarding how our prototype can help nurture early agenda-setting before in-person consultations. This feedback is particularly encouraging—in a busy clinical environment where time and throughput are paramount, clinicians may forego setting the stage at first based on patient feedback, to *get the work done* [50]. The clinicians in our study also thought that the information delivered through the prototype could be a helpful psychotherapeutic probe during consultations, wherein the patient's clinical presentation on social media is reconciled with what they verbally report. Hence, we conjecture that the use of this tool can facilitate that the patient's perceptions, needs, and concerns are considered appropriately by clinicians, in turn, helping to strengthen the therapeutic relationship between clinicians and patients.

However, our study also revealed potential scenarios in which the use of the tool may introduce new difficulties in managing the therapeutic alliance. According to the Agnew Relationship Measure [51], the therapeutic relationship between a patient and a clinician is defined by bond, partnership, confidence, and openness. During our study, we found that clinicians speculate on how the tool may negatively impact some of these core elements, such as patients' openness to sharing sensitive information on social media, or their partnership in care, should this tool be introduced during consultations. Our clinician participants also felt unclear about what type of patient engagement was OK under various circumstances, and if patients felt comfortable with clinicians discussing with them highly sensitive information provided by the tool, such as that relating to suicidal thoughts. They also pondered the privacy and ethical challenges they might encounter when they find themselves obliged to connect with patients in case of a potential crisis scenario flagged by the tool but when the patient's willingness to be contacted is unknown.

Even if these issues were to be mitigated in the future with deeper involvement of patients in exploring the use of the tool, a next step in this broader line of investigation, questions might arise about whether its use might undermine patients' voice and autonomy, and their power in their treatment process. Although we emphasize that the role of the tool is not to replace patients' self-reports but to augment them, it is not unusual for consented patients to feel that the tool would automatically replace the clinician's judgments and decision-making. Patients may also feel insecure and think that their clinicians may disbelieve what they say, turning their conversations confrontational, especially when patients' self-reports and social media data are not mutually consistent.

Here, we suggest the following calls to action to mitigate the challenges:

1. Consider provisions to continually negotiate patients' involvement and agency in the use and functioning of this future technology throughout the treatment process.
2. Incorporate auxiliary risk management strategies to balance protecting patients' privacy and clinicians' obligation to reach out during moments of crisis revealed by the tool. These can include involving patient collateral or family members or liaising with additional safety resources (eg, patient groups and other health service providers).
3. Identify mutually negotiated terms between the clinician and the patient so that they agree when the technology is causing more harm than benefit or when benefit is no longer present.

Institutional Infrastructures

Finally, our findings indicate the need to consider creating adequate institutional support to co-ordinate a sustainable ecosystem of stakeholders in both the deployment and maintenance of the prototype. The goal will be to assuage concerns that our study raised regarding compatibility with their existing workflows, increased burden on the clinicians, and liability and perceived lack of resources to support the use of the technology. For instance, our clinician participants noted the moral and professional quandary when they discovered alarming patterns, such as active suicidal ideation through this technology. Relatedly, they felt that the timing of when, during an ongoing consultation, to bring up the social media analyses is critical, but currently there is little counsel on it. Others expressed reservations, wondering if using the tool during an ongoing consultation was a good idea at all because it can potentially be distracting and rude, and take away the much-needed focus and eye contact desired during a conversation with a patient. Therefore, the technology will also have to be appraised continually through institutionally enforced policies so that trust and confidence in its use are maintained.

We offer the following calls to action in light of these observations:

1. Facilitate collaboration of diverse institutional stakeholders, such as management and information technology personnel, legal staff, and clinicians and patient advocacy groups, to develop institutional policies surrounding the technology.
2. Develop institutional provisions that advise clinicians on how to attend to any potential crisis discovered by the tool and standardize professional guidelines within the institution around what type of use of the technology in the context of a patient's care is acceptable.
3. Frame overarching policies governing what are the goals of care improvement when this technology is used and how efficacy and safety can be assessed throughout the period of a patient's care.
4. Suggest medical institutions to consider creating a new role to enable better assimilation of such a technology in mental health care—a "technology coach for mental health" or a "digital navigator" [52,53], similar to the notion of a patient navigator in cancer care [54] or a technology coach in web-based education [55], who can serve as an interface between the technology and the clinician, and the technology and the patient.

Ethical Implications

Although the tool we discuss in this study exclusively focuses on scenarios where a patient would have consented to have their data collected and used in the prototype for clinicians' use, we see remaining ethical concerns around the concept of using social media at the point of care.

First, we acknowledge that managing consent is a murky topic. Informed consent has been widely accepted as a legal and ethical requirement for most health care transactions; however, researchers have been reflecting on informed consent practices, especially on how much the participants should understand, how explicit their consent should be, and the delicate consideration of a patient's authenticity of choice (ie, voluntarism) [56], for instance, when a patient feels potential coercive pressures to incorporate this technology into their care, or social pressures to engage with new technology. Furthermore, patients may not fully appreciate what they are revealing when they consent, so they may share social media activities that they would otherwise choose not to share with their clinicians. The patient may also misunderstand that there might be a disadvantage if they do not participate in the sharing program. To address these problems, we need to consider a *sustained informed consent* [57] procedure in which someone will continually revisit informed consent with the patients, providing detailed information about both the sharing process and the voluntary nature of the program, as well as potential clinical and ethical harms.

Finally, we should consider the legal perspectives of future technologies. According to the Food and Drug Administration Safety and Innovation Act [58], most clinical decision support that delivers knowledge, person-specific information, and intelligently filtered information to clinicians and patients is not regulated by authorities. In addition, the source of data that will power this technology—social media—is not considered protected health information. However, because computer-aided detection or diagnosis can be considered a medical device, it raises an important question about whether such future technology should be overseen. Technology regulations also need to be considered by researchers and technology designers. We argue that even if the technology is not considered a medical device, the Food and Drug Administration Safety and Innovation Act's recommendations should be considered.

Limitations and Future Work

We note that our work suffers from some limitations, which constitute avenues for future research. First, we recruited patients from one health center and included a limited number of clinicians; therefore, the results may not be generalizable. Second, our participants explored the prototype without real interactions with their patients. By deploying the technology during actual appointments, future research can assess its ecological validity.

Finally, our study did not explore patients' opinions on the technology we proposed, although our study and the design decisions behind the prototype were situated in positive attitudes expressed by patients in sharing their social media data for diagnostic and treatment purposes [59,60]. We note here that this study is the first of a series of studies that plan to understand the potential and barriers of social media–powered technologies to support mental health treatment. We plan to explore this from a multistakeholder perspective, an important one being the patients. We believe that clinicians' feedback is a natural first step in this line of investigation. Using social media for mental health without the clinicians' guidance or support can be dangerous [24], and a lack of demonstrated clinical utility and buy-in from clinicians is likely to render subsequent studies less meaningful [61]. As argued by Baier [23], the potential harm of inappropriate social media use could be seen as a violation of this principle from indirectly encouraging boundary crossings to burdening patients with unnecessary information that could

compromise the therapeutic environment. This motivated us to consider interviewing clinicians first in the work presented in this study. As a next step, our goal is to explore patients' attitudes toward this potential technology.

Conclusions

This study presents a qualitative design study, including the design and evaluation of a prototype, to explore mental health clinicians' perspectives regarding a future technology that delivers computational insights derived from consented patients' social media. Our findings reveal the promise of the prototype beyond its compatibility with work practices. At the same time, the participants reported concerns and potential barriers to the new technology. The design of such technology should address the potential negative consequences and ethical concerns regarding credibility, liability, and institutional support. Our findings necessitate future research exploring patient perspectives on using computational insights from their social media in the context of their treatment.

Acknowledgments

The authors would like to thank Koustuv Saha, Vedant Das Swain, Qiaosi Wang, and the members of the Georgia Tech Social Dynamics and Well-Being Lab and Georgia Tech Ubicomp Group for their valuable feedback. This research was partly supported by funds from a cooperative research agreement issued by Northwell Health to Georgia Tech and a National Institute of Mental Health grant to MDC and MLB (R01MH117172).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshots of the prototype.
[PDF File (Adobe PDF File), 6110 KB - mental_v8i11e25455_app1.pdf]

Multimedia Appendix 2
The interview and focus group protocols.
[PDF File (Adobe PDF File), 81 KB - mental_v8i11e25455_app2.pdf]

Multimedia Appendix 3
The list of codes, their description, and examples.
[PDF File (Adobe PDF File), 54 KB - mental_v8i11e25455_app3.pdf]

References

1. John OP, Robins RW. Accuracy and bias in self-perception: individual differences in self-enhancement and the role of narcissism. J Pers Soc Psychol 1994 Jan;66(1):206-219. [doi: 10.1037//0022-3514.66.1.206] [Medline: 8126650]
2. Rogler LH, Malgady RG, Tryon WW. Evaluation of mental health. Issues of memory in the Diagnostic Interview Schedule. J Nerv Ment Dis 1992 Apr;180(4):215-223. [Medline: 1556561]
3. Sherman M, Trief P, Sprafkin R. Impression management in the psychiatric interview: quality, style, and individual differences. J Consult Clin Psychol 1975 Dec;43(6):867-871. [doi: 10.1037//0022-006x.43.6.867] [Medline: 1194481]
4. Arnold H, Feldman D. Social desirability response bias in self-report choice situations. Acad Manag J 1981 Jun 01;24(2):377-385 [FREE Full text] [doi: 10.2307/255848]
5. Petrik ML, Billera M, Kaplan Y, Matarazzo B, Wortzel H. Balancing patient care and confidentiality: considerations in obtaining collateral information. J Psychiatr Pract 2015 May;21(3):220-224. [doi: 10.1097/PRA.0000000000000072] [Medline: 25955265]
6. Engel GL. The need for a new medical model: a challenge for biomedicine. Science 1977 Apr 08;196(4286):129-136. [doi: 10.1126/science.847460] [Medline: 847460]
7. Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annu Rev Clin Psychol 2017 May 08;13:23-47 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
8. Chung C, Cook J, Bales E, Zia J, Munson SA. More than telemonitoring: health provider use and nonuse of life-log data in irritable bowel syndrome and weight management. J Med Internet Res 2015 Aug 21;17(8):e203 [FREE Full text] [doi: 10.2196/jmir.4364] [Medline: 26297627]
9. Kim Y, Heo E, Lee H, Ji S, Choi J, Kim J, et al. Prescribing 10,000 steps like aspirin: designing a novel interface for data-driven medical consultations. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI '17: CHI Conference on Human Factors in Computing Systems; May 6 - 11, 2017; Denver Colorado USA p. 5787-5799. [doi: 10.1145/3025453.3025570]

10. Pater JA, Farrington B, Brown A, Reining LE, Toscos T, Mynatt ED. Exploring indicators of digital self-harm with eating disorder patients. Proc ACM Hum-Comput Interact 2019 Nov 07;3(CSCW):1-26 [FREE Full text] [doi: 10.1145/3359186]
11. Pirelli G, Otto R, Estoup A. Using internet and social media data as collateral sources of information in forensic evaluations. Prof Psychol Res Pr 2016;47(1):12-17 [FREE Full text] [doi: 10.1037/pro0000061]
12. Coffey C, Batastini A, Vitacco M. Clues from the digital world: a survey of clinicians' reliance on social media as collateral data in forensic evaluations. Prof Psychol Res Pr 2018 Oct;49(5-6):345-354 [FREE Full text] [doi: 10.1037/pro0000206]
13. Golder SA, Macy MW. Diurnal and seasonal mood vary with work, sleep, and daylength across diverse cultures. Science 2011 Sep 30;333(6051):1878-1881. [doi: 10.1126/science.1202775] [Medline: 21960633]
14. Kosinski M, Stillwell D, Graepel T. Private traits and attributes are predictable from digital records of human behavior. Proc Natl Acad Sci U S A 2013 Apr 09;110(15):5802-5805 [FREE Full text] [doi: 10.1073/pnas.1218772110] [Medline: 23479631]
15. Lazer D, Pentland A, Adamic L, Aral S, Barabasi A, Brewer D, et al. Social science. Computational social science. Science 2009 Feb 06;323(5915):721-723 [FREE Full text] [doi: 10.1126/science.1167742] [Medline: 19197046]
16. 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.: Association for Computational Linguistics; 2014 Presented at: Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; June, 2014; Baltimore, Maryland, USA p. 51-60. [doi: 10.3115/v1/w14-3207]
17. De Choudhury M, Gamon M, Counts S, Horvitz E. Predicting depression via social media. Association for the Advancement of Artificial Intelligence. 2013. URL: https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/icwsm_13.pdf [accessed 2021-10-08]
18. Neimark G, Hurford MO, Digiacomo J. The internet as collateral informant. Am J Psychiatry 2006 Oct;163(10):1842. [doi: 10.1176/ajp.2006.163.10.1842]
19. Fisher CE, Appelbaum PS. Beyond googling: the ethics of using patients' electronic footprints in psychiatric practice. Harv Rev Psychiatry 2017;25(4):170-179. [doi: 10.1097/hrp.0000000000000145] [Medline: 28504978]
20. Hobbs KW, Monette PJ, Owoyemi P, Beard C, Rauch SL, Ressler KJ, et al. Incorporating information from electronic and social media into psychiatric and psychotherapeutic patient care: survey among clinicians. J Med Internet Res 2019 Jul 12;21(7):e13218 [FREE Full text] [doi: 10.2196/13218] [Medline: 31301127]
21. Deen SR, Withers A, Hellerstein DJ. Mental health practitioners' use and attitudes regarding the Internet and social media. J Psychiatr Pract 2013 Nov;19(6):454-463. [doi: 10.1097/01.pra.0000438184.74359.88] [Medline: 24241499]
22. Jent JF, Eaton CK, Merrick MT, Englebert NE, Dandes SK, Chapman AV, et al. The decision to access patient information from a social media site: what would you do? J Adolesc Health 2011 Oct;49(4):414-420 [FREE Full text] [doi: 10.1016/j.jadohealth.2011.02.004] [Medline: 21939873]
23. Baier AL. The ethical implications of social media: issues and recommendations for clinical practice. Ethics Behav 2018 Oct 11;29(5):341-351. [doi: 10.1080/10508422.2018.1516148]
24. Appelbaum PS, Kopelman A. Social media's challenges for psychiatry. World Psychiatry 2014 Feb;13(1):21-23 [FREE Full text] [doi: 10.1002/wps.20085] [Medline: 24497238]
25. Lehavot K, Barnett J, Powers D. Psychotherapy, professional relationships, and ethical considerations in the myspace generation. Prof Psychol Res Pr 2010 Apr;41(2):160-166 [FREE Full text] [doi: 10.1037/a0018709]
26. Hutton L, Henderson T. "I didn't sign up for this!": informed consent in social network research. In: Proceedings of the Ninth International AAAI Conference on Web and Social Media. 2015 Presented at: Ninth International AAAI Conference on Web and Social Media; May 26-29, 2015; Oxford, United Kingdom URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14593
27. 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. USA: ACM; 2019 Presented at: FAT* '19: Conference on Fairness, Accountability, and Transparency; January 29 - 31, 2019; Atlanta GA USA p. 79-88. [doi: 10.1145/3287560.3287587]
28. Sleeper M, Balebako R, Das S, McConahy A, Wiese J, Cranor L. The post that wasn't: exploring self-censorship on facebook. In: Proceedings of the 2013 conference on Computer supported cooperative work. USA: Association for Computing Machinery; 2013 Presented at: CSCW '13: Computer Supported Cooperative Work; February 23 - 27, 2013; San Antonio Texas USA p. 793-802. [doi: 10.1145/2441776.2441865]
29. Yoo DW, Birnbaum ML, Van Meter AR, Ali AF, Arenare E, Abowd GD, et al. Designing a clinician-facing tool for using insights from patients' social media activity: iterative co-design approach. JMIR Ment Health 2020;7(8):e16969 [FREE Full text] [doi: 10.2196/16969] [Medline: 32784180]
30. Saha K, Sugar B, Torous J, Abrahao B, Kıcıman E, De Choudhury M. A social media study on the effects of psychiatric medication use. Proc Int AAAI Conf Weblogs Soc Media 2019 Jun 07;13:440-451 [FREE Full text] [Medline: 32280562]
31. Birnbaum ML, Ernala SK, Rizvi AF, Arenare E, R Van Meter A, De Choudhury M, et al. Detecting relapse in youth with psychotic disorders utilizing patient-generated and patient-contributed digital data from Facebook. NPJ Schizophr 2019 Oct 07;5(1):17 [FREE Full text] [doi: 10.1038/s41537-019-0085-9] [Medline: 31591400]

32. Fulford H, McSwiggan L, Kroll T, MacGillivray S. Exploring the use of information and communication technology by people with mood disorder: a systematic review and metasynthesis. JMIR Ment Health 2016 Jul 01;3(3):e30 [FREE Full text] [doi: 10.2196/mental.5966] [Medline: 27370327]
33. Ernala S, Birnbaum M, Candan K, Rizvi A, Sterling W, Kane J, et al. Methodological gaps in predicting mental health states from social media: triangulating diagnostic signals. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.: Association for Computing Machinery; 2019 Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4 - 9, 2019; Glasgow Scotland, UK p. 1-16. [doi: 10.1145/3290605.3300364]
34. Gilbert E, Karahalios K. Predicting tie strength with social media. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. USA: Association for Computing Machinery; 2009 Presented at: CHI '09: CHI Conference on Human Factors in Computing Systems; April 4 - 9, 2009; Boston MA, USA p. 211-220. [doi: 10.1145/1518701.1518736]
35. Rohan KJ, Rough JN, Evans M, Ho S, Meyerhoff J, Roberts LM, et al. A protocol for the Hamilton rating scale for depression: item scoring rules, Rater training, and outcome accuracy with data on its application in a clinical trial. J Affect Disord 2016 Aug;200:111-118 [FREE Full text] [doi: 10.1016/j.jad.2016.01.051] [Medline: 27130960]
36. Carrión RE, Auther A, McLaughlin D, Olsen R, Addington J, Bearden C, et al. The global functioning: social and role scales-further validation in a large sample of adolescents and young adults at clinical high risk for psychosis. Schizophr Bull 2019 Jun 18;45(4):763-772 [FREE Full text] [doi: 10.1093/schbul/sby126] [Medline: 30351423]
37. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis. Int J Qual Methods 2017 Oct 02;16(1):160940691773384. [doi: 10.1177/1609406917733847]
38. Maizes V, Rakel D, Niemiec C. Integrative medicine and patient-centered care. Explore (NY) 2009;5(5):277-289. [doi: 10.1016/j.explore.2009.06.008] [Medline: 19733814]
39. Langhoff C, Baer T, Zubraegel D, Linden M. Therapist–patient alliance, patient–therapist alliance, mutual therapeutic alliance, therapist–patient concordance, and outcome of CBT in GAD. J Cogn Psychother 2008 Mar 01;22(1):68-79. [doi: 10.1891/0889.8391.22.1.68]
40. Capone V. Patient communication self-efficacy, self-reported illness symptoms, physician communication style and mental health and illness in hospital outpatients. J Health Psychol 2016 Jul;21(7):1271-1282. [doi: 10.1177/1359105314551622] [Medline: 25274717]
41. Di Matteo D, Fine A, Fotinos K, Rose J, Katzman M. Patient willingness to consent to mobile phone data collection for mental health apps: structured questionnaire. JMIR Ment Health 2018 Aug 29;5(3):e56 [FREE Full text] [doi: 10.2196/mental.9539] [Medline: 30158102]
42. Pina L, Sien S, Ward T, Yip J, Munson S, Fogarty J, et al. From personal informatics to family informatics: understanding family practices around health monitoring. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. USA: ACM; 2017 Presented at: CSCW '17: Computer Supported Cooperative Work and Social Computing; February 25 - March 1, 2017; Portland Oregon, USA p. 2300-2315. [doi: 10.1145/2998181.2998362]
43. Cooper AA, Kline AC, Graham B, Bedard-Gilligan M, Mello PG, Feeny NC, et al. Homework "dose," type, and helpfulness as predictors of clinical outcomes in prolonged exposure for PTSD. Behav Ther 2017 Mar;48(2):182-194. [doi: 10.1016/j.beth.2016.02.013] [Medline: 28270329]
44. McCarney R, Warner J, Iliffe S, van Haselen R, Griffin M, Fisher P. The Hawthorne effect: a randomised, controlled trial. BMC Med Res Methodol 2007;7:30 [FREE Full text] [doi: 10.1186/1471-2288-7-30] [Medline: 17608932]
45. Back MD, Stopfer JM, Vazire S, Gaddis S, Schmukle SC, Egloff B, et al. Facebook profiles reflect actual personality, not self-idealization. Psychol Sci 2010 Mar;21(3):372-374. [doi: 10.1177/0956797609360756] [Medline: 20424071]
46. Daugherty P, Wilson H. Human + Machine: Reimagining Work in the Age of AI. Massachusetts, United States: Harvard Business Press; 2018.
47. Sussskind R, Sussskind D. The Future of the Professions: How Technology Will Transform the Work of Human Experts. USA: Oxford University Press; Apr 01, 2017.
48. Ahmad M, Eckert C, Teredesai A. Interpretable machine learning in healthcare. In: Proceedings of the 2018 ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics. USA: ACM; 2018 Presented at: BCB '18: 9th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics; August 29 - September 1, 2018; Washington DC USA p. 559-560. [doi: 10.1145/3233547.3233667]
49. Ribeiro M, Singh S, Guestrin C. "Why should I trust you?": explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. USA: ACM; 2016 Presented at: KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13 - 17, 2016; San Francisco, California USA p. 1135-1144. [doi: 10.1145/2939672.2939778]
50. Torous J, Roberts LW. The ethical use of mobile health technology in clinical psychiatry. J Nerv Ment Dis 2017 Jan;205(1):4-8. [doi: 10.1097/NMD.0000000000000596] [Medline: 28005647]
51. Agnew-Davies R, Stiles WB, Hardy GE, Barkham M, Shapiro DA. Alliance structure assessed by the Agnew Relationship Measure (ARM). Br J Clin Psychol 1998 May;37(2):155-172. [doi: 10.1111/j.2044-8260.1998.tb01291.x] [Medline: 9631204]
52. Ben-Zeev D, Drake R, Marsch L. Clinical technology specialists. Br Med J 2015 Feb 19;350:h945. [doi: 10.1136/bmj.h945] [Medline: 25697164]

53. Noel VA, Carpenter-Song E, Acquilano SC, Torous J, Drake RE. The technology specialist: a 21st century support role in clinical care. NPJ Digit Med 2019;2:61 [FREE Full text] [doi: 10.1038/s41746-019-0137-6] [Medline: 31388565]
54. Braun KL, Kagawa-Singer M, Holden AE, Burhansstipanov L, Tran JH, Seals BF, et al. Cancer patient navigator tasks across the cancer care continuum. J Health Care Poor Underserved 2012 Feb;23(1):398-413 [FREE Full text] [doi: 10.1353/hpu.2012.0029] [Medline: 22423178]
55. Sugar W. Nstructional technologist as a coach: impact of a situated professional development program on teachers' technology use. J Technol Teacher Edu 2005;13:547 [FREE Full text]
56. Grady C. Enduring and emerging challenges of informed consent. N Engl J Med 2015 Feb 26;372(9):855-862. [doi: 10.1056/nejmra1411250]
57. Prentice K, Appelbaum P, Conley R, Carpenter W. Maintaining informed consent validity during lengthy research protocols. IRB Ethics Hum Res 2007;29(6):1-6. [Medline: 18237078]
58. FDASIA health IT report: strategy and recommendations for a risk-based framework. Food and Drug Administration. 2014. URL: https://www.fda.gov/about-fda/cdrh-reports/fdasia-health-it-report [accessed 2021-10-08]
59. Birnbaum ML, Rizvi AF, Confino J, Correll CU, Kane JM. Role of social media and the internet in pathways to care for adolescents and young adults with psychotic disorders and non-psychotic mood disorders. Early Interv Psychiatry 2017 Aug;11(4):290-295 [FREE Full text] [doi: 10.1111/eip.12237] [Medline: 25808317]
60. Rieger A, Gaines A, Barnett I, Baldassano CF, Gibbons MB, Crits-Christoph P. Psychiatry outpatients' willingness to share social media posts and smartphone data for research and clinical purposes: survey study. JMIR Form Res 2019 Aug 29;3(3):e14329 [FREE Full text] [doi: 10.2196/14329] [Medline: 31493326]
61. Torous J, Baker JT. Why psychiatry needs data science and data science needs psychiatry: connecting with technology. JAMA Psychiatry 2016 Jan;73(1):3-4. [doi: 10.1001/jamapsychiatry.2015.2622] [Medline: 26676879]

Abbreviations

GF:S: Global Functioning: Social
HAM-D: Hamilton Depression Rating Scale

Edited by J Torous; submitted 02.11.20; peer-reviewed by K Stawarz, R Lee; comments to author 18.12.20; revised version received 11.02.21; accepted 22.06.21; published 16.11.21.

Please cite as:
Yoo DW, Ernala SK, Saket B, Weir D, Arenare E, Ali AF, Van Meter AR, Birnbaum ML, Abowd GD, De Choudhury M
Clinician Perspectives on Using Computational Mental Health Insights From Patients' Social Media Activities: Design and Qualitative Evaluation of a Prototype
JMIR Ment Health 2021;8(11):e25455
URL: https://mental.jmir.org/2021/11/e25455
doi:10.2196/25455
PMID:34783667

Review

Machine Learning Methods for Predicting Postpartum Depression: Scoping Review

Kiran Saqib[1], MBBS, MSPH; Amber Fozia Khan[1], BSc; Zahid Ahmad Butt[1], MBBS, MSc, PhD

School of Public Health Sciences, University of Waterloo, Waterloo, ON, Canada

Corresponding Author:
Zahid Ahmad Butt, MBBS, MSc, PhD
School of Public Health Sciences
University of Waterloo
200 University Avenue West
Waterloo, ON, N2L 3G1
Canada
Phone: 1 5198884567 ext 45107
Email: zahid.butt@uwaterloo.ca

Abstract

Background: Machine learning (ML) offers vigorous statistical and probabilistic techniques that can successfully predict certain clinical conditions using large volumes of data. A review of ML and big data research analytics in maternal depression is pertinent and timely, given the rapid technological developments in recent years.

Objective: This study aims to synthesize the literature on ML and big data analytics for maternal mental health, particularly the prediction of postpartum depression (PPD).

Methods: We used a scoping review methodology using the Arksey and O'Malley framework to rapidly map research activity in ML for predicting PPD. Two independent researchers searched PsycINFO, PubMed, IEEE Xplore, and the ACM Digital Library in September 2020 to identify relevant publications in the past 12 years. Data were extracted from the articles' ML model, data type, and study results.

Results: A total of 14 studies were identified. All studies reported the use of supervised learning techniques to predict PPD. Support vector machine and random forest were the most commonly used algorithms in addition to Naive Bayes, regression, artificial neural network, decision trees, and XGBoost (Extreme Gradient Boosting). There was considerable heterogeneity in the best-performing ML algorithm across the selected studies. The area under the receiver operating characteristic curve values reported for different algorithms were support vector machine (range 0.78-0.86), random forest method (0.88), XGBoost (0.80), and logistic regression (0.93).

Conclusions: ML algorithms can analyze larger data sets and perform more advanced computations, which can significantly improve the detection of PPD at an early stage. Further clinical research collaborations are required to fine-tune ML algorithms for prediction and treatment. ML might become part of evidence-based practice in addition to clinical knowledge and existing research evidence.

(JMIR Ment Health 2021;8(11):e29838) doi:10.2196/29838

KEYWORDS

machine learning; postpartum depression; big data; mobile phone

Introduction

Background

Postpartum depression (PPD) is considered one of the most frequent maternal morbidities after delivery, with severe implications for the mother and child. According to the National Institute of Mental Health, United States, 10%-15% of women have maternal depression during and after pregnancy worldwide, whereas in low- and middle-income countries, this percentage could be as high as 18%-25% [1] and seems to depend on the cultural and traditional characteristics of the population [2]. Both the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) and the International Classification of Diseases (ICD)-10 recognize maternal depression as a mental illness with different classifications [3].

PPD, the most common complication of childbearing, is a term applied to depressive symptoms that occur within 4 weeks of giving birth and possibly as late as 30 weeks postpartum [4]. PPD is a significant public health issue that affects women as well as child's physical and mental health and cognitive and interactive development [5], thus making the child vulnerable to developing psychiatric disorders during adolescence [6]. A depressed mother may not establish a positive relationship with her infant [7], and this may continue to affect children into toddlerhood, preschool years, and beyond [8]. Infants of depressed mothers have shown poor nutrition, poor general health, and more frequent diarrheal episodes, and in extreme cases, maternal suicide and infanticide may occur [9,10]. PPD is generally an overlooked health problem that can lead to serious complications and should be addressed in a timely manner [11].

As there is no single etiology for PPD, a single prevention method or treatment will be ineffective. There is a need for a multifactorial approach combining psychological, psychosocial, and biological predictive factors of PPD to contemplate various etiological factors and individual variations [12,13]. An effective PPD prediction model can help health care providers in the early identification and effective management of at-risk patients [14], with evidence from previous studies exploring this possibility and feasibility [15].

Machine learning (ML) algorithms are broadly grouped into 3 categories: (1) supervised, (2) unsupervised, and (3) semisupervised learning. In supervised learning, data with known labels are used to train a model that can predict the label for new data [16]. ML-based predictive models are gaining popularity for combining a huge amount of data into a single model and evaluating the model's predictive value for previously unseen individuals, for example, at-risk and new patients. ML approaches rely on the use of advanced statistical and probabilistic techniques to construct systems with the ability to automatically learn from data. This enables patterns in data to be more readily and accurately identified and more accurate predictions to be made from data sources (eg, more accurate diagnosis and prognosis) [17]. ML has been used for prediction in psychiatry [18]. ML methods have been successfully used to predict major depressive disorder persistence, chronicity, severity [19], and treatment response [20]. The key to building good ML models is in the rigorous selection of appropriate features and algorithms [17]. Recently, a scoping review of ML application in mental health identified over 190 studies that applied ML in the detection and diagnosis of mental disorders and over 60 studies to predict the progression of mental health problems over time [21]. These studies reported the use of electronic health records (EHRs), mood rating scales, brain imaging data, smartphone monitoring systems, and social media platforms to predict, classify, or subgroup mental health illnesses, including depression, schizophrenia, and suicide ideation and attempts [22]. Two main ML algorithms have been commonly reported in depression prediction studies, namely, support vector machine (SVM) and random forest (RF) algorithms [21]. Depression prediction studies using these 2 methods have achieved relatively good results [23-25].

There is an opinion that ML will help mental health practitioners redefine mental illnesses more objectively than is currently done in the Diagnostic and Statistical Manual of Mental Disorders [3] and would help in the early identification of these illnesses to make interventions more effective [22]. Thus, in addition to disease-model refinement, ML may benefit psychiatry by characterizing those at risk and personalizing and discovering pharmacological therapeutics [26,27].

A literature review of ML and big data research analytics in maternal depression is pertinent and timely, given the rapid technological developments in recent years. This review aims to provide a concise snapshot of the literature on ML applications for predicting PPD. Previous reviews have demonstrated ML techniques to be robust and scalable for general depression and mental health, but no review to date has mapped ML applications within maternal mental health research and practice. Our overall aim is to examine the current state of affairs of ML applications in PPD, providing a snapshot of the methods used. Keeping in view the rapid advancements in ML and the recent use of ML in mental health research, we chose to focus specifically on exploring broadly the nature of research activity, as per the first goal of scoping reviews by Arksey and O'Malley [28].

Objective

It is hoped that this scoping review will (1) inform mental health researchers of the methods and applications of ML in the context of prediction of PPD, (2) identify the best-performing algorithm, and (3) identify the evaluation criteria for the best-performing algorithm.

Methods

Overview

The Arksey and O'Malley framework was used in addition to methodological improvements for scoping review [28-30]. Our methods also align with the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist [31]. A scoping review methodology was chosen to map the body of literature on the use of ML in predicting PPD, including a greater range of study designs and methodologies, to provide a descriptive overview of the reviewed material.

Search Strategy

The search strategy was adapted from Shatte et al [21], which is a similar review of big data applications in mental health. As ML and PPD stretch across interdisciplinary fields, the search was conducted in both health and information technology databases. First, a literature search was conducted using health-related research databases, including PsycINFO and PubMed. Next, the information technology databases IEEE Xplore and the ACM Digital Library were searched. Finally, databases that index both fields, including Scopus and Web of Science, were searched. The search period for relevant studies was conducted in September 2020. The search terms included variations in the terms for the following:

- (a) PPD (*maternal*∗, *perinatal*∗, *postpartum blues*∗, *baby blues*∗, *depression*∗, *post birth depression*∗)
- (b) ML (*machine learning**, *artificial intelligence**, *supervised learning**, *big data**)
- (c) Prediction (*predictive models*∗, *prediction*∗, *detection**)

The search was conducted on titles, keywords, and abstracts with *AND* entered into the database search to link different categories (a, b, and c) of search terms. Truncation symbols (∗) were used to search for all possible forms of a search term (Multimedia Appendix 1). Forward reference searching, that is, examining the references cited in these articles, and backward reference searching, that is, reviewing the references cited in these articles, were applied to identify further studies that met the inclusion criteria.

Study Selection

Articles were included and excluded (Textbox 1) in the review if the following criteria were met.

Textbox 1. Inclusion and exclusion criteria.

Inclusion criteria

- The article reported on a method or application of machine learning (ML) to address postpartum depression only, based on the authors' descriptions of their analyses: if they deemed it ML, the paper was included.
- The article evaluated the performance of the ML algorithm or big data technique used to predict postpartum depression.
- The article was published in a peer-reviewed publication.
- The article was available in English.
- The article was published between 2009 and 2021.

Exclusion criteria

- The article did not report ML applications in postpartum depression (eg, the paper commented on the use of ML in diagnosis, treatment, or prognosis of general depression, anxiety, and other mental health issues).
- The article did not focus on postpartum depression.
- The full text of the article was not available (eg, conference or abstracts).
- If articles were commentaries and essays. Two reviewers (KS and AFK) independently reviewed all studies and reached a consensus on all included studies after consultation with the third author (ZAB).

Data Extraction and Analysis Plan

For data extraction and analysis, we used the same framework already used in a similar scoping review [32]. For each article, data were extracted regarding (1) overall aim of research, that is, prediction and area of focus, that is, PPD; (2) input data type used; (3) type of ML algorithms used; and (4) the best-performing algorithm, that is, results.

To analyze the data, a narrative review synthesis method [32] was selected to capture the extensive range of research investigating ML and big data for PPD prediction. A meta-analysis was not deemed appropriate, given the aim of identifying research activity in the interdisciplinary field of big data and maternal mental health.

Results

Overview

The search strategies using a combination of search terms identified 1392 articles that included a search term from each category in their abstract or title (PRISMA-ScR flowchart). The range for publication year of relevant articles was 2009-2021. A total of 24 articles were duplicates. A database search was carried out by KS and AFK. Abstracts of 368 articles were read by both authors to perform an initial screening of eligibility for this scoping review. Of these, 347 were excluded because they did not focus specifically on PPD. A total of 21 articles were selected for full-text review, but 3 were conference papers and abstract only, and 4 did not use ML to predict PPD. This resulted in a total sample of 14 studies, including one preprint and one focused on predicting PPD in fathers, which met the inclusion criteria according to all authors (Figure 1). The selected 14 studies were reviewed in full by 2 authors (including KS and AFK). A mutual consensus was reached after the final approval from ZAB. In the subsequent narrative analysis, we focus on the 14 studies that reported using the ML model to predict PPD (see Tables 1 and 2 for a summary of the main study characteristics).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) procedural flowchart. ML: machine learning; PPD: postpartum depression.

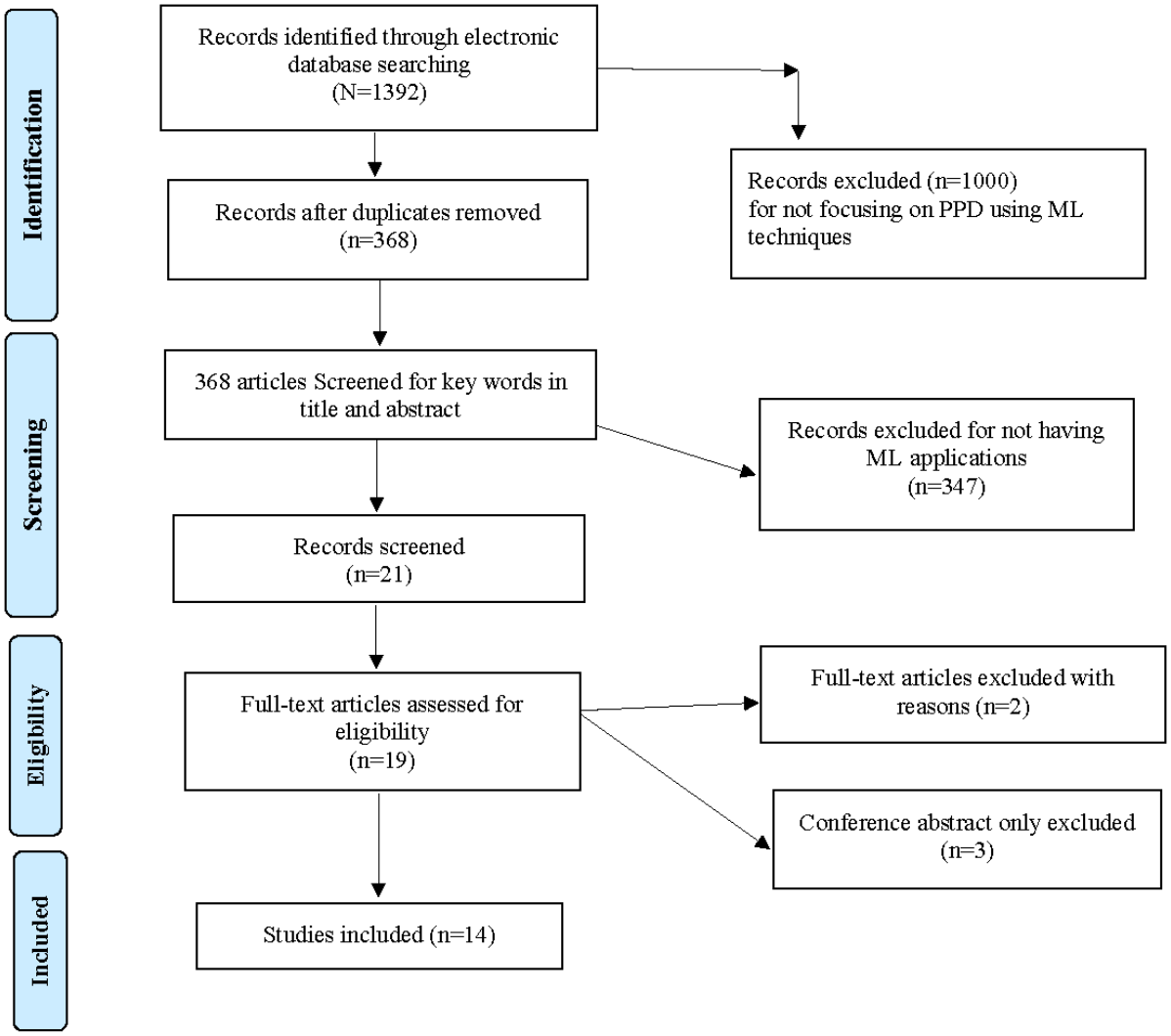


Table 1. Summary of the main study characteristics (N=14).

#	Study	Aims or objectives	Sample size; input data used	Diagnosis criteria for PPD[a]
1	Jiménez-Serrano et al [24]	Develop classification models for detecting the risk of PPD during the first week after childbirth	1880; hospital data	EPDS[b]>9; 8th or 32nd week postbirth
2	Betts et al [33]	Develop a prediction model to identify women at risk of postpartum psychiatric admission	75,054; linked administrative health data	ICD[c]-10 ($F_{20.0}$-$F_{39.9}$) or ICD-10: ($F_{53.0}$-$F_{53.1}$)
3	Tortajada et al [34]	To obtain a classification model based on feedforward multilayer perceptron to improve PPD prediction during the 32 weeks after childbirth with a high sensitivity and specificity	1397; hospital data	EPDS>9; 8th or 32nd week postbirth
4	Wang et al [35]	To develop a PPD prediction model, using EHRs[d]	179,980; EHRs	ICD-10-CM codes O99.3 and O99.34 as well as their ICD-9-CM equivalents for a diagnosis of PPD within 12 months after childbirth
5	Zhang et al [36]	To compare the effects of 4 different ML[e] models using data during pregnancy to predict PPD	508; hospital data	EPDS >9.5; within 42 days postdelivery
6	Zhang et al [37]	Propose an ML framework for PPD risk prediction	17,633 and 71,106; 2 data sets from EHRs	PPD within 1 year of childbirth
7	Hochman et al [38]	To apply ML approach to create a prediction tool for PPD to be implemented in health care systems	214,359; EHRs	PPD within first year postpartum (ICD‐9 codes: 300 and 309 or ICD-10 codes: F40-F48) or acute psychotic manic episodes (ICD‐9 codes: 296.0, 296.1, 296.4, 296.6, 296.81, 298.3, 298.4, 298.8)
8	De Choudhury et al [39]	Detect and predict PPD	165; Facebook survey using PHQ[f]-9	PHQ-9
9	Natarajan et al [23]	Propose an ML-based approach for PPD prediction and diagnosis from survey information	207; Facebook and Twitter survey data	Postpartum Depression Predictors Inventory
10	Fatima et al [40]	Use linguistic features to propose a solution for PPD that can be generalized and deployed across web-based social platforms	21; text posts from Reddit	PPD based on linguistic feature
11	Trifan et al [41]	To use social media for potential diagnosis of mothers at risk of PPD and thus the implementation of early interventions	512; Reddit text posts	Not described
12	Shatte et al [42]	To identify fathers at the risk of PPD	365; Reddit text posts	ICD-10 depression; symptom 06 months postbirth
13	Moreira et al [43]	Propose an algorithm for emotion-aware smart systems, capable for predicting the risk of PPD during pregnancy through biomedical and sociodemographic data analysis	Performance evaluation used data generated by wearable devices and sensors	Not described
14	Shin et al [44]	To develop predictive models for PPD using ML approaches	28,755; pregnancy risk assessment and monitoring system data	PHQ-2

[a]PPD: postpartum depression.

[b]EPDS: Edinburgh Postnatal Depression Scale.

[c]ICD: International Classification of Diseases.

[d]EHR: electronic health record.

[e]ML: machine learning.

[f]PHQ: Patient Health Questionnaire.

Table 2. Summary of the main study characteristics (N=14).

#	Study	Performance metric	ML[a] algorithms used	Best-performing algorithm
1	Jiménez-Serrano et al [24]	Hold-out validation	• Naive Bayes • LR[b] • SVM[c] • ANN[d]	Naive Bayes model; G function value of 0.73
2	Betts et al [33]	5-Fold cross-validation in R	• Gradient boosting • Elastic net methods	Boosted trees algorithm (AUC[e] 0.80, 95% CI 0.76-0.83)
3	Tortajada et al [34]	Hold-out validation	• ANN	Multilayer perceptrons 0.82 of G and 0.81 of accuracy (95% CI 0.76-0.86) with 0.84 of sensitivity and 0.81 of specificity
4	Wang et al [35]	10-fold cross-validation	• SVM • RF[f] • Naive Bayes • L2-regularized LR • XGBoost[g] • DT[h]	SVM with AUC (0.79)
5	Zhang et al [36]	sklearn.cross_validation package in Python	• SVM • RF	SVM and feature selection RF (sensitivity=0.69; AUC=0.78)
6	Zhang et al [37]	5-Fold cross-validation	• RF • DT • XGboost • Regularized LR • Multilayer perceptron	LR with L2 regularization; AUC (0.937, 95% CI 0.912-0.962)
7	Hochman et al [38]	Hold-out cross-validation	• XGBoost	AUC of 0.712 (95% CI 0.690-0.733), with a sensitivity of 0.349 and a specificity of 0.905)
8	De Choudhury et al [39]	Not described	• Regression models to develop a series of statistical models	Postnatal model
9	Natarajan et al [23]	Information not provided	• Functional gradient boosting • DT • SVM • NB[i]	Functional gradient boosting (Roc) 0.952
10	Fatima et al [40]	10-Fold cross-validation	• LR • SVM • Multilayer perceptron	Multilayer perceptron; 91·7% accuracy for depressive content identification and up to 869% accuracy for PPD content prediction
11	Trifan et al [41]	Hold-out validation	• SVM • Stochastic gradient descent • Passive aggressive classifiers	SVM
12	Shatte et al [42]	10-Fold cross-validation	• SVM classifiers using behavior, emotion, linguistic style, and discussion topics as features	0.67 precision, 0.68 recall, and 0.67F−measure in model including all features
13	Moreira et al [43]	10-fold cross-validation	• DT • SVM • Nearest neighbor • Ensemble classifiers	Ensemble classifiers

#	Study	Performance metric	ML[a] algorithms used	Best-performing algorithm
14	Shin et al [44]	10-Fold cross-validation	• RF • Stochastic gradient boosting • SVM • Regression trees • NB • k-nearest neighbor • LR • ANN	RF method (AUC) 0.884

[a]ML: machine learning.

[b]LR: logistic regression.

[c]SVM: support vector machine.

[d]ANN: artificial neural network.

[e]AUC: area under the curve.

[f]RF: random forest.

[g]XGBoost: Extreme Gradient Boosting.

[h]DT: decision tree.

[i]NB: Naive Bayes.

A narrative synthesis of ML activity, particularly in the context of PPD, indicated the emerging nature of this field, with most studies being published in recent years. Publication dates ranged from 2009 to 2020; however, most articles were very recent. There is a 5-year gap between the first 2009 article [34] and the next study in 2014 [39], and publications have accelerated recently with 7 papers published in 2020.

Few studies have focused on developing and testing an ML algorithm for the detection and prediction of PPD, whereas other studies focused on comparing the effects of different ML algorithms to predict PPD and explore which factors in the model are the most important for PPD prediction.

Type of Input Data

When we examined the 14 studies, we identified a subgroup of 7 studies that reported on the use of ML-based models to predict PPD using clinical or hospital data and EHRs. The other 5 studies reported on the application of ML algorithms for the prediction of PPD using data from social media platforms, including Facebook, Twitter, and Reddit. However, these studies were designed to evaluate a prediction model more broadly and did not report details on ML algorithms, training, and testing procedures. Of the remaining 2 studies, one reported on the use of population data and the other used emotion-aware system data. The outcome variable *PPD* was assessed using psychometric tools such as Patient Health Questionnaire-9, Patient Health Questionnaire-2, Edinburgh Postnatal Depression Scale, Postpartum Depression Predictors Inventory, and ICD-9 and ICD-10 codes in the case of hospital and EHR data, whereas linguistic features were used to predict PPD from text data of social networks.

Type of ML Algorithms Used

All studies reported on the use of supervised ML models, including classification and regression algorithms, to predict PPD. Most of the studies (n=7) reported using more than one algorithm, whereas one study used only regression models to develop statistical models for their data. These included SVM (n=8) logistic regression (LR; n=6), multilayer perceptron using artificial neural network (ANN; n=5), RF (n=4), Naive Bayes (n=3), decision trees (DTs; n=3), gradient boosting (n=2), XGBoost (Extreme Gradient Boosting; n=2), functional gradient boosting (n=1), elastic net methods (n=1), k-nearest neighbor (kNN; n=2), Stochastic Gradient Boosting (n=1), passive aggressive classifiers (n=1), and ensemble classifier (n=1). The data types used to develop ML algorithms included EHRs, either administrative hospital data or organizational data (n=08), mobile and wearable sensor data (n=1), and social media data (n=5).

Reported Best-Performing Algorithm

There was considerable heterogeneity in the best-performing ML algorithm across the selected studies. To report the best-performance algorithm, most studies used sensitivity, specificity, and area under the curve (AUC). Only 5 studies described the technical approaches to cross-validation using either 5-fold or 10-fold cross-validation. One study reported that of 4 ML algorithms, including Naive Bayes, LR, SVM, and ANN, Naive Bayes showed the best balance between sensitivity and specificity as a predictive model for PPD during the first week after delivery according to the G function, with a value of 0.73 [24]. Another study using 6 ML models, including SVM, RF, Naive Bayes, L2-regularized LR, XGBoost, and DT, reported that SVM had the best performance, and the difference across the performance of SVM, L2-regularized LR, RF, Naive Bayes, and XGBoost was minimal, although differences existed with respect to sensitivity and specificity [35]. In total, 9 different ML algorithms, including RF, stochastic gradient boosting, SVM, recursive partitioning and regression trees, Naive Bayes, kNN, LR, and neural network, were used to report the overall classification accuracies of the 9 models ranging from 0.650 (kNN) to 0.791 (RF). The RF method achieved the highest area under the receiver operating characteristic curve (AUROC) value of 0.884, followed by SVM, which achieved the second-best performance with an AUC value of 0.864 [44].

Using the SVM and RF algorithms, the model based on SVM and feature selection RF had the best prediction effects

(sensitivity=0.69, AUC=0.78) [36]. Five ML algorithms were trained: RF, DT, XGBoost, regularized LR, and multilayer perceptron. LR with L2 regularization was found to be the best-performing algorithm using data available up to childbirth. The AUC was 0.937 (95% CI 0.912-0.962) and 0.886 (95% CI 0.879-0.893) in hospital data sets, respectively [37]. SVM led to slightly better results in terms of F1 in the validation stage compared with stochastic gradient descent and passive aggressive classifiers [41].

Tortajada et al [34] developed 4 models for predicting PPD using a multilayer perceptron and evaluated them with the geometric mean of accuracies using a hold-out strategy. They reported that the developed models could predict PPD during the first 32 weeks after delivery with high accuracy. A similar study reported that hold-out validation showed that multilayer perceptron outperformed other techniques such as SVM and LR used in one study with 91.7% accuracy for depressive content identification and up to 86.9% accuracy for PPD content prediction [40]. Another study using gradient boosting and elastic net methods reported that the boosted trees algorithm produced the best-performing model, predicting postpartum psychiatric admission in the validation data with good discrimination (AUC 0.80, 95% CI 0.76-0.83) and achieved good calibration. This model outperformed the benchmark LR model and the elastic net model [33]. Natarajan et al [23] reported a successful functional gradient boosting algorithm that demonstrated the potential of ML in predicting PPD.

Hochman et al [38] built a model using XGBoost, an algorithm based on gradient-boosted DTs, and assessed the overall model predictive performance using the AUROC. 95% CIs were estimated using bootstrapping. The prediction model achieved an AUC of 0.712 (95% CI 0.690-0.733), with a sensitivity of 0.349 and a specificity of 0.905 at the 90th percentile risk threshold, identifying PPDs at a rate more than 3 times higher than the overall set (positive and negative predictive values were 0.074 and 0.985, respectively).

After developing a series of statistical models using regression models to predict a mother's likelihood of PPD, the postnatal model performed the best [39]. Predictive models were developed as a series of SVM classifiers using behavior, emotion, linguistic style, and discussion topics as features. The model incorporating behavior and discussion topic features alone yielded greater recall, with 0.77 and 0.82, respectively, which may be useful for screening purposes [42]. A study using hospital data showed that ensemble classifiers represent a leading solution for predicting psychological disorders related to pregnancy [43].

Many studies did not mention which statistical tools were used for analysis; however, most used a variety of software packages in R, SAS, and Python 3. Studies have reported the use of standard libraries available for data preparation (eg, missing variables), a variety of typical ML models, and natural language processing (NLP) analyses (such as topic modeling) included in their standard packages such as R.

Discussion

Principal Findings

Most of the reviewed studies used supervised classification techniques rather than other ML techniques to predict PPD. This is perhaps indicative of the extensive focus on detection and diagnosis in the literature, which is typically designed using large, retrospective, labeled data sets ideal for classification tasks [45]. All reviewed studies concluded that ML models were effective in predicting PPD, whether clinical data, EHRs, population data, and data from social media platforms. All the studies implied that the ML approach was more beneficial compared with traditional statistical approaches. However, the level of accuracy, sensitivity, or specificity that is considered acceptable varies depending on the aims of the study and the data set. None of the studies explicitly compared the ML performance with other traditional statistical analyses. In all studies, the ML approach aided researchers in answering their research questions.

The results from a cohort study for predicting PPD using hospital data reported that in the case of a small sample size, SVM can avoid overfitting while providing efficient computing time and better prediction results in depression [46,47]. The same study proposed that when the data set is small, SVM is more practical than RF in prediction research for PPD [36]. Several previous studies used the SVM algorithm to make PPD predictions, as SVM is an example of supervised learning that is most commonly used in classification problems. It focuses on minimizing the structural risks within a set of available data [36]. It has significant advantages and performs well in situations with relatively less available sample data [48]. SVM is a classifier that transforms input data into a multidimensional hyperplane using kernels to discriminate between 2 classes [49]. Jiménez-Serrano et al [24] collected data on postpartum women from 7 Spanish hospitals and used the Edinburgh Postnatal Depression Scale score as the outcome indicator to train a PPD prediction model based on SVM. Natarajan et al [23] used social media as a data source, and based on the mental health data of 173 mothers, an SVM-based PPD prediction model was established. De Choudhury [39] developed an SVM model to identify high-risk emotions and behaviors predictive of PPD using the content of Twitter posts. As these studies either target different populations or use different methods to detect the occurrence of PPD, the model prediction effects cannot be easily compared [36].

In contrast, RF models were built using a DT as the basic classifier. RF approaches have high classification accuracy, strong inductive capacity, a simple parameter adjustment process, fast calculation speed, relatively low sensitivity to missing data values, and the ability to output feature importance [50,51]. RF is an ensemble learning method that operates by constructing a multitude of DTs and outputting the class that is voted by a majority of the trees [52], and Shin et al [44] reported RF to be the best-performing algorithm for predicting PPD.

Tortajada et al [34] developed another prediction model for PPD using multilayer perceptron and pruning for pregnant Spanish women using data from 7 Spanish general hospitals

from 2003 to 2004. ANNs have a remarkable ability to characterize discriminating patterns and derive meaning from complex and noisy data sets. They have been widely applied in general medicine for the differential diagnosis, classification, prediction of disease, and condition prognosis. For instance, ANNs have been applied to the diagnosis of dementia using clinical data [53] and more recently for predicting Alzheimer disease using mixed effects neural networks [54].

There is a great deal of debate about which ML model evaluation metric is best [55]. Making sense of reported ML evaluation metrics is made even more difficult because different performance parameters often provide conflicting results and the optimal ML algorithm also depends significantly on the composition of the data set [56]. Some reviewed studies reported varying degrees of accuracy and were not always explicitly clear regarding the meaning of the resulting performance metrics. Owing to the negative effects of PPD on mothers and infants [57,58], such as the negative effects on the physical and mental health of mothers, the closeness of the mother-infant bond, and infant development, it is important to have a model with high sensitivity while maintaining a high AUROC value. The selection of indicators for evaluating depression prediction models varies across studies. For example, Natarajan et al [23] and De Choudhury [39] emphasized the accuracy of the model's prediction of PPD. Jiménez-Serrano et al [24] emphasized the sensitivity and specificity of the model. The balance between the two is the geometric mean. The AUROC is also widely used to evaluate the comprehensive performance of a model [23,25].

PPD is a highly prevalent problem but frequently goes undetected, leading to substantial treatment delays [59]. EHRs collect a large number of biometric markers and patient characteristics that could foster the detection of PPD in primary care settings. NLP and ML have the potential to complement clinical practice by categorizing and analyzing data from clinical notes [60]. NLP is a computerized process that analyzes and codes human language into text [61] that ML algorithms can analyze and use to predict outcomes [62]. Advances in technology, such as social media, smartphones, wearables, and neuroimaging, have allowed mental health researchers and clinicians to collect a vast range of data at a rapidly growing rate [63]. ML is a vigorous technique with the ability to analyze these data. A data-driven primary intervention approach using ML and EHR data may be leveraged to reduce the burden of health care providers in identifying PPD risk [37].

In the studies included in our review, individuals experiencing PPD were identified through screening surveys, their public sharing of a diagnosis on social media, Twitter, Facebook, or Reddit, and were distinguishable from control users by patterns in their language and web-based activity [23,40,42]. Automated detection methods may help identify depressed or otherwise at-risk individuals through the large-scale passive monitoring of social media and, in the future, may complement existing screening procedures [64]. Social media data and EHRs both hold the promise of innovating in the maternal mental health domain, particularly when leveraged by ML techniques [21].

Finally, there are some challenges to consider when using ML techniques in mental health applications. ML models are inevitably limited by the quality of the data used to develop the model. As such, ML does not replace other research or analytic approaches; rather, it has the potential to add value to mental health research. Many ML techniques require access to training data sets, which calls for collaboration between researchers and clinicians to maximize the usefulness of the models developed. It is important to highlight that ML might become part of evidence-based practice, in addition to clinical knowledge and existing research evidence. Greater collaboration between mental health researchers and clinicians (eg, for the provision of training data sets and for feedback on the clinical usefulness of ML algorithms) will be needed to continue to advance the applications of ML in mental health. Analyzing *big data* on clinical outcomes, in addition to genetic, biomedical, behavioral, environmental, and demographic patient characteristics, could help predict maternal depression. EHR databases can provide valuable, real-world, practice-based evidence to support better prediction models for at-risk patients [65]. In this way, ML offers a solution for analyzing idiographic research questions in big data [66].

Limitations

This study has a few limitations. The aim of this scoping review was to provide a snapshot of the research activity in a summarized format while using a systematic search method. In line with the aims of a scoping review, we did not identify specific study designs in advance and did not assess the quality of the included studies [28]. Moreover, because of restrictions in the search methodology, there may be a chance to have missed some relevant articles, for example, broad search terms and the exclusion of nonpeer reviewed literature. This is a common limitation reported in scoping review studies attributed to maintaining a balance between breadth and depth of analysis within a rapid timeframe [67]. This review successfully mapped a cross-section of the literature on the use of ML for PPD prediction and provides a useful synthesis for researchers and clinicians to understand the potential of ML in this field. This study did not examine the effectiveness of individual ML models for predicting PPD. Such research questions would be suitable for future systematic reviews, guided by the framework outlined in our results tables, that is, the effectiveness of specific ML techniques within specific data types for specific clinical applications.

Conclusions

To conclude, the use of ML to predict PPD has revealed exciting advances, particularly in recent years. Compared with traditional statistical methods, ML algorithms are capable of analyzing larger data sets and performing more advanced computations. Overall, it is clear that ML can significantly improve the detection of PPD at an early stage. Research into the applications of ML to identify potential PPD predictors has demonstrated positive results. However, this work is currently limited, and further research is required to identify additional benefits of ML on maternal mental health. ML techniques and the performance of ML models may differ depending on the type, content, and accuracy of the original data; thus, it may be challenging to evaluate the performance of a single model. With ML tools becoming more accessible to researchers and

clinicians, it is expected that the field will continue to grow and that novel applications for mental health will follow. Further clinical research collaborations are required to fine-tune ML algorithms for prediction and treatment. As ML algorithms continue to be refined and improved, it might be possible to help clinicians identify maternal mental illnesses at an earlier stage when interventions may be more effective and personalized treatments based on an individual's unique characteristics. Moreover, the current lack of procedural evaluation guidelines leaves many clinicians and researchers in the field with no means to systematically evaluate the claims, maturity, and clinical readiness of an ML study [68].

Acknowledgments

We are thankful to Jackie Stapleton for her continuous support and help with the review.

Authors' Contributions

KS conceived the study, participated in its design and coordination, performed the search and data extraction, interpreted the data, and drafted the manuscript. AFK assisted with the search and data extraction and helped revise the manuscript. ZAB conceived the study, participated in its design and coordination, contributed to the interpretation of the data, and helped to draft and revise the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Databases and search strings used for this review.
[DOCX File , 16 KB - mental_v8i11e29838_app1.docx]

References

1. Fisher J, Cabral de Mello M, Patel V, Rahman A, Tran T, Holton S, et al. Prevalence and determinants of common perinatal mental disorders in women in low- and lower-middle-income countries: a systematic review. Bull World Health Organ 2011 Nov 24;90(2):139-149 [FREE Full text] [doi: 10.2471/blt.11.091850] [Medline: 22423165]
2. Gaynes BN, Gavin N, Meltzer-Brody S, Lohr KN, Swinson T, Gartlehner G, et al. Perinatal depression: prevalence, screening accuracy, and screening outcomes. Evid Rep Technol Assess (Summ) 2005 Feb(119):1-8. [doi: 10.1037/e439372005-001] [Medline: 15760246]
3. Wisner KL, Moses-Kolko EL, Sit DK. Postpartum depression: a disorder in search of a definition. Arch Womens Ment Health 2010 Feb;13(1):37-40 [FREE Full text] [doi: 10.1007/s00737-009-0119-9] [Medline: 20127453]
4. Andrews-Fike C. A review of postpartum depression. Prim Care Companion J Clin Psychiatry 1999 Feb 01;1(1):9-14 [FREE Full text] [doi: 10.4088/pcc.v01n0103] [Medline: 15014700]
5. Murray L, Halligan S, Cooper P. Effects of postnatal depression on mother-infant interactions and child development. In: Wachs T, Bremner G, editors. Handbook of Infant Development (2nd ed.). Oxford, United Kingdom: Wiley-Blackwell; 2010:192-220.
6. Pawlby S, Sharp D, Hay D, O'Keane V. Postnatal depression and child outcome at 11 years: the importance of accurate diagnosis. J Affect Disord 2008 Apr;107(1-3):241-245 [FREE Full text] [doi: 10.1016/j.jad.2007.08.002] [Medline: 17854906]
7. Wisner KL, Sit DK, McShea MC, Rizzo DM, Zoretich RA, Hughes CL, et al. Onset timing, thoughts of self-harm, and diagnoses in postpartum women with screen-positive depression findings. JAMA Psychiatry 2013 May;70(5):490-498 [FREE Full text] [doi: 10.1001/jamapsychiatry.2013.87] [Medline: 23487258]
8. Goodman S, Brand S. Parental psychopathology and its relation to child psychopathology. In: Hersen M, Gross A, editors. Handbook of clinical psychology, vol 2: Children and adolescents. Hoboken, New Jersey, United States: John Wiley & Sons Inc; 2008:937-965.
9. Muzik M, Borovska S. Perinatal depression: implications for child mental health. Ment Health Fam Med 2010 Dec;7(4):239-247 [FREE Full text] [Medline: 22477948]
10. Rahman A, Malik A, Sikander S, Roberts C, Creed F. Cognitive behaviour therapy-based intervention by community health workers for mothers with depression and their infants in rural Pakistan: a cluster-randomised controlled trial. Lancet 2008 Sep;372(9642):902-909. [doi: 10.1016/s0140-6736(08)61400-2]
11. Aliani R, Khuwaja B. Epidemiology of postpartum depression in Pakistan: a review of literature. Nat J Health Sci 2017 Feb 01;2(1):24-30. [doi: 10.21089/njhs.21.0024]
12. Johnstone S, Boyce P, Hickey A, Morris-Yatees AD, Harris M. Obstetric risk factors for postnatal depression in urban and rural community samples. Aust N Z J Psychiatry 2001 Feb;35(1):69-74 [FREE Full text] [doi: 10.1046/j.1440-1614.2001.00862.x] [Medline: 11270460]

13. Beck CT. Predictors of postpartum depression: an update. Nurs Res 2001;50(5):275-285 [FREE Full text] [doi: 10.1097/00006199-200109000-00004] [Medline: 11570712]
14. Robertson E, Grace S, Wallington T, Stewart DE. Antenatal risk factors for postpartum depression: a synthesis of recent literature. Gen Hosp Psychiatry 2004;26(4):289-295 [FREE Full text] [doi: 10.1016/j.genhosppsych.2004.02.006] [Medline: 15234824]
15. Righetti-Veltema M, Conne-Perréard E, Bousquet A, Manzano J. Risk factors and predictive signs of postpartum depression. J Affect Disord 1998 Jun;49(3):167-180 [FREE Full text] [doi: 10.1016/s0165-0327(97)00110-9]
16. Naqa IE, Murphy MJ. What is machine learning? In: Naqa IE, Li R, Murphy M, editors. Machine Learning in Radiation Oncology. Cham: Springer; 2015:3-11.
17. 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]
18. Gillan CM, Whelan R. What big data can do for treatment in psychiatry. Curr Opin Behav Sci 2017 Dec;18:34-42. [doi: 10.1016/j.cobeha.2017.07.003]
19. Kessler RC, van Loo HM, Wardenaar KJ, Bossarte RM, Brenner LA, Cai T, et al. Testing a machine-learning algorithm to predict the persistence and severity of major depressive disorder from baseline self-reports. Mol Psychiatry 2016 Oct;21(10):1366-1371 [FREE Full text] [doi: 10.1038/mp.2015.198] [Medline: 26728563]
20. Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, Trivedi MH, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. Lancet Psychiatry 2016 Mar;3(3):243-250. [doi: 10.1016/S2215-0366(15)00471-X] [Medline: 26803397]
21. Shatte AB, 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]
22. Graham S, Depp C, Lee EE, Nebeker C, Tu X, Kim HC, et al. Artificial intelligence for mental health and mental illnesses: an overview. Curr Psychiatry Rep 2019 Nov 07;21(11):116 [FREE Full text] [doi: 10.1007/s11920-019-1094-0] [Medline: 31701320]
23. Natarajan S, Prabhakar A, Ramanan N, Bagilone A, Siek K, Connelly K. Boosting for postpartum depression prediction. In: Proceedings of the IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE). 2017 Presented at: IEEE/ACM International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE); July 17-19, 2017; Philadelphia, PA, USA. [doi: 10.1109/chase.2017.82]
24. Jiménez-Serrano S, Tortajada S, García-Gómez JM. A mobile health application to predict postpartum depression based on machine learning. Telemed J E Health 2015 Jul;21(7):567-574 [FREE Full text] [doi: 10.1089/tmj.2014.0113] [Medline: 25734829]
25. Jin H, Wu S, Di Capua P. Development of a clinical forecasting model to predict comorbid depression among diabetes patients and an application in depression screening policy making. Prev Chronic Dis 2015 Sep 03;12:E142 [FREE Full text] [doi: 10.5888/pcd12.150047] [Medline: 26334714]
26. Ćosić K, Popović S, Šarlija M, Kesedžić I, Jovanovic T. Artificial intelligence in prediction of mental health disorders induced by the COVID-19 pandemic among health care workers. Croat Med J 2020 Jul 05;61(3):279-288 [FREE Full text] [Medline: 32643346]
27. Bickman L. Improving mental health services: a 50-year journey from randomized experiments to artificial intelligence and precision mental health. Adm Policy Ment Health 2020 Sep 26;47(5):795-843 [FREE Full text] [doi: 10.1007/s10488-020-01065-8] [Medline: 32715427]
28. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32 [FREE Full text] [doi: 10.1080/1364557032000119616]
29. Daudt HM, van Mossel C, Scott SJ. Enhancing the scoping study methodology: a large, inter-professional team's experience with Arksey and O'Malley's framework. BMC Med Res Methodol 2013 Mar 23;13:48 [FREE Full text] [doi: 10.1186/1471-2288-13-48] [Medline: 23522333]
30. O'Brien KK, Colquhoun H, Levac D, Baxter L, Tricco A, Straus S, et al. Advancing scoping study methodology: a web-based survey and consultation of perceptions on terminology, definition and methodological steps. BMC Health Serv Res 2016 Jul 26;16:305 [FREE Full text] [doi: 10.1186/s12913-016-1579-z] [Medline: 27461419]
31. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
32. van Doorn KA, Kamsteeg C, Bate J, Aafjes M. A scoping review of machine learning in psychotherapy research. Psychother Res 2021 Jan 29;31(1):92-116 [FREE Full text] [doi: 10.1080/10503307.2020.1808729] [Medline: 32862761]
33. Betts KS, Kisely S, Alati R. Predicting postpartum psychiatric admission using a machine learning approach. J Psychiatr Res 2020 Nov;130:35-40 [FREE Full text] [doi: 10.1016/j.jpsychires.2020.07.002] [Medline: 32771679]
34. Tortajada S, García-Gomez JM, Vicente J, Sanjuán J, de Frutos R, Martín-Santos R, et al. Prediction of postpartum depression using multilayer perceptrons and pruning. Methods Inf Med 2009;48(3):291-298 [FREE Full text] [doi: 10.3414/ME0562] [Medline: 19387507]

35. Wang S, Pathak J, Zhang Y. Using electronic health records and machine learning to predict postpartum depression. Stud Health Technol Inform 2019 Aug 21;264:888-892. [doi: 10.3233/SHTI190351] [Medline: 31438052]
36. Zhang W, Liu H, Silenzio VM, Qiu P, Gong W. Machine learning models for the prediction of postpartum depression: application and comparison based on a cohort study. JMIR Med Inform 2020 Apr 30;8(4):e15516 [FREE Full text] [doi: 10.2196/15516] [Medline: 32352387]
37. Zhang Y, Wang S, Hermann A, Joly R, Pathak J. Development and validation of a machine learning algorithm for predicting the risk of postpartum depression among pregnant women. J Affect Disord 2021 Jan 15;279:1-8 [FREE Full text] [doi: 10.1016/j.jad.2020.09.113] [Medline: 33035748]
38. Hochman E, Feldman B, Weizman A, Krivoy A, Gur S, Barzilay E, et al. Development and validation of a machine learning-based postpartum depression prediction model: a nationwide cohort study. Depress Anxiety 2021 Apr 07;38(4):400-411. [doi: 10.1002/da.23123] [Medline: 33615617]
39. De Choudhury M, Counts S, Horvitz EJ, 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. USA: Association for Computing Machinery; 2014 Presented at: CSCW'14: Computer Supported Cooperative Work; February 15 - 19, 2014; Baltimore Maryland USA p. 626-638. [doi: 10.1145/2531602.2531675]
40. Fatima I, Abbasi BU, Khan S, Al-Saeed M, Ahmad HF, Mumtaz R. Prediction of postpartum depression using machine learning techniques from social media text. Expert Syst 2019 Apr 26;36(4):e12409. [doi: 10.1111/exsy.12409]
41. Trifan A, Semeraro D, Drake J, Bukowski R, Oliveira JL. Social media mining for postpartum depression prediction. Stud Health Technol Inform 2020 Jun 16;270:1391-1392. [doi: 10.3233/SHTI200457] [Medline: 32570674]
42. Shatte AB, 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 [FREE Full text] [doi: 10.1089/cyber.2019.0746] [Medline: 32915660]
43. Moreira MW, Rodrigues JJ, Kumar N, Saleem K, Illin IV. Postpartum depression prediction through pregnancy data analysis for emotion-aware smart systems. Inf Fusion 2019 May;47:23-31. [doi: 10.1016/j.inffus.2018.07.001]
44. Shin D, Lee KJ, Adeluwa T, Hur J. Machine learning-based predictive modeling of postpartum depression. J Clin Med 2020 Sep 08;9(9):2899 [FREE Full text] [doi: 10.3390/jcm9092899] [Medline: 32911726]
45. Ghassemi M, Naumann T, Schulam P, Beam AL, Chen IY, Ranganath R. A review of challenges and opportunities in machine learning for health. AMIA Jt Summits Transl Sci Proc 2020;2020:191-200 [FREE Full text] [Medline: 32477638]
46. Malki K, Koritskaya E, Harris F, Bryson K, Herbster M, Tosto MG. Epigenetic differences in monozygotic twins discordant for major depressive disorder. Transl Psychiatry 2016 Jun 14;6(6):e839 [FREE Full text] [doi: 10.1038/tp.2016.101] [Medline: 27300265]
47. Patel MJ, Khalaf A, Aizenstein HJ. Studying depression using imaging and machine learning methods. Neuroimage Clin 2016;10:115-123 [FREE Full text] [doi: 10.1016/j.nicl.2015.11.003] [Medline: 26759786]
48. Boser B, Guyon I, Vapnik V. A training algorithm for optimal margin classifiers. In: Proceedings of the Fifth Annual Workshop on Computational Learning Theory. New York, New York, USA: ACM Press; 1992 Presented at: COLT92: 5th Annual Workshop on Computational Learning Theory; July 27 - 29, 1992; Pittsburgh Pennsylvania USA p. 144-152. [doi: 10.1145/130385.130401]
49. Gultepe E, Green J, Nguyen H, Adams J, Albertson T, Tagkopoulos I. From vital signs to clinical outcomes for patients with sepsis: a machine learning basis for a clinical decision support system. J Am Med Inform Assoc 2014;21(2):315-325 [FREE Full text] [doi: 10.1136/amiajnl-2013-001815] [Medline: 23959843]
50. Yao D, Yang J. Research on feature selection and classification method based on random forest for medical data. Dissertation - Harbin Engineering University. 2017. URL: http://cdmd.cnki.com.cn/Article/CDMD-10217-1017245762.htm [accessed 2021-11-09]
51. Hapfelmeier A, Hothorn T, Ulm K, Strobl C. A new variable importance measure for random forests with missing data. Stat Comput 2012 Aug 28;24(1):21-34. [doi: 10.1007/s11222-012-9349-1]
52. Breiman L. Random forests. Mach Learn 2001;45:5-32. [doi: 10.1023/A:1010933404324]
53. Mulsant B, Servan-Schreiber E. A connectionist approach to the diagnosis of dementia. In: Proceedings of the Annual Symposium on Computer Application in Medical Care. 1988 Presented at: Annual Symposium on Computer Application in Medical Care; November 7-11, 1988; Orlando, FL, USA URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2245294/pdf/procascamc00018-0260.pdf
54. Tandon R, Adak S, Kaye JA. Neural networks for longitudinal studies in Alzheimer's disease. Artif Intell Med 2006 Mar;36(3):245-255 [FREE Full text] [doi: 10.1016/j.artmed.2005.10.007] [Medline: 16427257]
55. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: machine learning and the future of medicine. J Intern Med 2018 Dec 03;284(6):603-619 [FREE Full text] [doi: 10.1111/joim.12822] [Medline: 30102808]
56. Rácz A, Bajusz D, Héberger K. Multi-level comparison of machine learning classifiers and their performance metrics. Molecules 2019 Aug 01;24(15):2811 [FREE Full text] [doi: 10.3390/molecules24152811] [Medline: 31374986]
57. Martini J, Petzoldt J, Einsle F, Beesdo-Baum K, Höfler M, Wittchen HU. Risk factors and course patterns of anxiety and depressive disorders during pregnancy and after delivery: a prospective-longitudinal study. J Affect Disord 2015 Apr 01;175:385-395. [doi: 10.1016/j.jad.2015.01.012] [Medline: 25678171]

58. Glasheen C, Richardson GA, Fabio A. A systematic review of the effects of postnatal maternal anxiety on children. Arch Womens Ment Health 2010 Feb;13(1):61-74 [FREE Full text] [doi: 10.1007/s00737-009-0109-y] [Medline: 19789953]
59. Hansotte E, Payne SI, Babich SM. Positive postpartum depression screening practices and subsequent mental health treatment for low-income women in Western countries: a systematic literature review. Public Health Rev 2017 Jan 31;38:3 [FREE Full text] [doi: 10.1186/s40985-017-0050-y] [Medline: 29450075]
60. Longhurst CA, Harrington RA, Shah NH. A 'green button' for using aggregate patient data at the point of care. Health Aff (Millwood) 2014 Jul;33(7):1229-1235 [FREE Full text] [doi: 10.1377/hlthaff.2014.0099] [Medline: 25006150]
61. Lacson R, Khorasani R. Natural language processing: the basics (part 1). J Am Coll Radiol 2011 Jun;8(6):436-437 [FREE Full text] [doi: 10.1016/j.jacr.2011.04.020] [Medline: 21636059]
62. Monuteaux MC, Stamoulis C. Machine learning: a primer for child psychiatrists. J Am Acad Child Adolesc Psychiatry 2016 Oct;55(10):835-836 [FREE Full text] [doi: 10.1016/j.jaac.2016.07.766] [Medline: 27663936]
63. Chen M, Mao S, Liu Y. Big data: a survey. Mobile Netw Appl 2014 Jan 22;19(2):171-209 [FREE Full text] [doi: 10.1007/s11036-013-0489-0]
64. Guntuku S, Yaden D, Kern M, Ungar L, Eichstaedt J. Detecting depression and mental illness on social media: an integrative review. Curr Opin Behav Sci 2017 Dec;18:43-49. [doi: 10.1016/j.cobeha.2017.07.005]
65. Lutz W, Rubel JA, Schwartz B, Schilling V, Deisenhofer AK. Towards integrating personalized feedback research into clinical practice: development of the Trier Treatment Navigator (TTN). Behav Res Ther 2019 Sep;120:103438 [FREE Full text] [doi: 10.1016/j.brat.2019.103438] [Medline: 31301550]
66. Silberschatz G. Improving the yield of psychotherapy research. Psychother Res 2017 Jan 11;27(1):1-13. [doi: 10.1080/10503307.2015.1076202] [Medline: 26360343]
67. Pham MT, Rajić A, Greig JD, Sargeant JM, Papadopoulos A, McEwen SA. A scoping review of scoping reviews: advancing the approach and enhancing the consistency. Res Synth Methods 2014 Dec;5(4):371-385 [FREE Full text] [doi: 10.1002/jrsm.1123] [Medline: 26052958]
68. Cearns M, Hahn T, Baune BT. Recommendations and future directions for supervised machine learning in psychiatry. Transl Psychiatry 2019 Oct 22;9(1):271 [FREE Full text] [doi: 10.1038/s41398-019-0607-2] [Medline: 31641106]

Abbreviations

ANN: artificial neural network
AUC: area under the curve
AUROC: area under the receiver operating characteristic curve
DT: decision tree
EHR: electronic health record
ICD: International Classification of Diseases
kNN: k-nearest neighbor
LR: logistic regression
ML: machine learning
NLP: natural language processing
PPD: postpartum depression
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RF: random forest
SVM: support vector machine
XGBoost: Extreme Gradient Boosting

Edited by J Torous; submitted 22.04.21; peer-reviewed by H Oyama, C Gorrostieta; comments to author 15.08.21; revised version received 26.08.21; accepted 30.08.21; published 24.11.21.

Please cite as:
Saqib K, Khan AF, Butt ZA
Machine Learning Methods for Predicting Postpartum Depression: Scoping Review
JMIR Ment Health 2021;8(11):e29838
URL: https://mental.jmir.org/2021/11/e29838
doi:10.2196/29838
PMID:34822337

Original Paper

Momentary Manifestations of Negative Symptoms as Predictors of Clinical Outcomes in People at High Risk for Psychosis: Experience Sampling Study

Isabell Paetzold[1], MSc; Karlijn S F M Hermans[2], PhD; Anita Schick[1], PhD; Barnaby Nelson[3,4], PhD; Eva Velthorst[5], PhD; Frederike Schirmbeck[6,7], PhD; EU-GEI High Risk Study[8]; Jim van Os[9,10,11], PhD, MD; Craig Morgan[12,13], PhD; Mark van der Gaag[14,15], PhD; Lieuwe de Haan[16], PhD; Lucia Valmaggia[17], PhD; Philip McGuire[10,18], PhD; Matthew Kempton[10], PhD; Inez Myin-Germeys[2], PhD; Ulrich Reininghaus[1,12,13], PhD

[1]Department of Public Mental Health, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

[2]Department of Neuroscience, Center for Contextual Psychiatry, KU Leuven, Leuven, Belgium

[3]Orygen, Parkville, Victoria, Australia

[4]Centre for Youth Mental Health, The University of Melbourne, Parkville, Victoria, Australia

[5]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, NY, United States

[6]Department of Psychiatry, Amsterdam UMC, Location AMC, University of Amsterdam, Amsterdam, Netherlands

[7]Arkin, Institute for Mental Health, Amsterdam, Netherlands

[8]See Acknowledgments, Maastricht, Netherlands

[9]Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, Netherlands

[10]Department of Psychosis Studies, Institute of Psychiatry, King's College London, London, United Kingdom

[11]Department of Psychiatry, Brain Center Rudolf Magnus, Utrecht University Medical Centre, Utrecht, Netherlands

[12]ESRC Centre for Society and Mental Health, King's College London, London, United Kingdom

[13]Department of Health Service and Population Research, Centre for Epidemiology and Public Health, Institute of Psychiatry, Psychology & Neuroscience, School of Mental Health & Psychological Sciences, King's College London, London, United Kingdom

[14]Department of Clinical, Neuro and Developmental Psychology, Vrije Universiteit, Amsterdam, Netherlands

[15]Department of Psychosis Research, Parnassia Psychiatric Institute, The Hague, Netherlands

[16]Department of Early Psychosis, Amsterdam UMC, Location AMC, University of Amsterdam, Amsterdam, Netherlands

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

[18]NIHR Biomedical Research Centre, South London and Maudsley NHS Foundation Trust, London, London, United Kingdom

Corresponding Author:
Ulrich Reininghaus, PhD
Department of Public Mental Health, Central Institute of Mental Health
Medical Faculty Mannheim
Heidelberg University
J 5 1
Mannheim, 68159
Germany
Phone: 49 62117031931
Email: ulrich.reininghaus@zi-mannheim.de

Abstract

Background: Negative symptoms occur in individuals at ultrahigh risk (UHR) for psychosis. Although there is evidence that observer ratings of negative symptoms are associated with level of functioning, the predictive value of subjective experience in daily life for individuals at UHR has not been studied yet.

Objective: This study therefore aims to investigate the predictive value of momentary manifestations of negative symptoms for clinical outcomes in individuals at UHR.

Methods: Experience sampling methodology was used to measure momentary manifestations of negative symptoms (blunted affective experience, lack of social drive, anhedonia, and social anhedonia) in the daily lives of 79 individuals at UHR. Clinical

outcomes (level of functioning, illness severity, UHR status, and transition status) were assessed at baseline and at 1- and 2-year follow-ups.

Results: Lack of social drive, operationalized as greater experienced pleasantness of being alone, was associated with poorer functioning at the 2-year follow-up (b=−4.62, P=.01). Higher levels of anhedonia were associated with poorer functioning at the 1-year follow-up (b=5.61, P=.02). Higher levels of social anhedonia were associated with poorer functioning (eg, disability subscale: b=6.36, P=.006) and greater illness severity (b=−0.38, P=.045) at the 1-year follow-up. In exploratory analyses, there was evidence that individuals with greater variability of positive affect (used as a measure of blunted affective experience) experienced a shorter time to remission from UHR status at follow-up (hazard ratio=4.93, P=.005).

Conclusions: Targeting negative symptoms in individuals at UHR may help to predict clinical outcomes and may be a promising target for interventions in the early stages of psychosis.

(JMIR Ment Health 2021;8(11):e30309) doi:10.2196/30309

KEYWORDS

ecological momentary assessment; psychotic disorder; psychopathology

Introduction

Background

Negative symptoms occur in individuals at ultrahigh risk (UHR, also known as clinical high risk) for psychosis and have been reported to be associated with reduced quality of life and impaired functioning in cross-sectional and longitudinal studies [1-4]. Recently, several studies have demonstrated the predictive value of negative symptoms for social aspects of functioning [5,6]. Furthermore, negative symptoms have been found to be predictive of transition to psychotic disorder in UHR samples [7-12].

To date, clinical outcomes in UHR studies have primarily focused on transition to psychosis. Given that most individuals at UHR do not develop psychosis (71%-76%) as indicated in meta-analyses and systematic reviews [13-16], investigating other outcomes has received increasing attention in recent years [17,18]. Meta-analyses have found that most individuals at UHR who do not transition to psychosis do not remit from UHR status within 2 years either [19]. In addition, individuals at UHR—regardless of whether they transition to psychosis—show other clinical symptoms and marked impairments in functioning that are comparable with those reported in patients with social phobia and major depressive disorder [18-23]. The level of functioning in individuals at UHR is more similar to that which is observed in patients with psychotic disorders than in controls [20]. Hence, level of functioning and persistence of clinical symptoms are important outcomes other than transition to psychosis.

Standard measures used to assess negative symptoms (eg, the Positive and Negative Syndrome Scale) [24,25], though valid in their own right, have been criticized for being overly reliant on behavioral observation and third-party anamnesis [26-28]. In addition, standardized self-report questionnaires and laboratory measures of negative symptoms in patients with psychosis do not seem to converge with real-time and real-world reports generated using the experience sampling methodology (ESM) and hence may capture different constructs [29,30]. ESM is a semistructured diary method that captures daily behavior and experience of company with high ecological validity [31]. A recent systematic review of experience sampling studies investigating everyday social experiences of individuals with schizophrenia [32] concluded that, compared with other methods, experience sampling allows a more granular assessment of social experience. This underscores the importance of examining the perspective of individuals' experience of negative symptoms in daily life (ie, momentary manifestations of negative symptoms), as this is when psychiatric symptoms naturally emerge. Experience sampling studies have made important contributions to our understanding of psychosis, but until now, studies of momentary experience of social context and manifestations of negative symptoms have mainly focused on individuals with a psychotic disorder [26,33].

Previous experience sampling studies have investigated blunted affective experience, lack of social drive, anhedonia, and social anhedonia as momentary manifestations of negative symptoms in daily life. Blunted affective experience has been operationalized as intensity (ie, mean level), instability (ie, differences in affect from one moment to the following), and variability (ie, differences between affect in the moment and the average individual affect) of positive and negative affect [26,33-35]. Lack of social drive has been assessed using the amount of time spent alone, the preference to be alone when in company, and the experienced pleasantness of being alone [35,36]. Anhedonia has been operationalized as a smaller increase of positive affect in moments of pleasant events [26,35]. Similarly, social anhedonia has been operationalized as a smaller increase in positive affect associated with being in pleasant company [26,35,36].

To our knowledge, only 2 experience sampling studies have, to date, investigated momentary manifestations of negative symptoms in individuals at UHR [35,37]. Although differing in focus and operationalization of constructs, both studies compared momentary manifestations of negative symptoms across individuals at UHR, patients with first-episode psychosis, and controls. In line with findings in enduring psychosis [26,38], both studies concluded that there may be a mismatch between what individuals at UHR experience and how they express this in their behavior, that may be interpreted as 2 distinct dimensions of negative symptoms (ie, experience vs expression). Hence, assessing individuals' subjective experience of negative symptoms is important to gain a more comprehensive understanding of internal, experiential aspects [27]. However,

both studies used a cross-sectional design. No experience sampling study to date has used momentary manifestations of negative symptoms for predicting clinical outcomes in individuals at UHR in a longitudinal design. This is an important gap that needs to be addressed, as a shift in research toward subjective experience of momentary symptoms may offer new insights into the social nature and development of negative symptoms in UHR and its outcomes.

Objectives

This study aims to investigate whether momentary manifestations of negative symptoms predict clinical outcomes (ie, illness severity, level of functioning, and remission from UHR status and transition to psychosis) in individuals at UHR for psychosis at the 1- and 2-year follow-ups. We tested the following hypotheses:

Momentary manifestations of negative symptoms in daily life predict clinical outcomes in individuals at UHR at 1- and 2-year follow-up such that higher levels of (1) blunted affective experience (ie, lower intensity, variability and instability of positive and negative affect; H1, hypothesis 1); (2) lack of social drive (ie, amount of time spent alone, pleasantness of being alone, and preference to be alone when in company; H2); (3) anhedonia (ie, no or low increase of positive affect in moments of pleasant events; H3); and (4) social anhedonia (ie, no or low increase of positive affect in moments of pleasant company; H4) are associated with greater illness severity and poorer functioning at follow-up.

In exploratory analyses, we further aimed to examine whether momentary manifestations of negative symptoms are associated with time to transition to psychosis or remission from UHR status.

Methods

Sample

We recruited a sample of individuals at UHR aged 15-35 years, who were assessed at baseline and 1- and 2-year follow-up. Participants were recruited in London (United Kingdom), Melbourne (Australia), and Amsterdam and The Hague (the Netherlands) as a part of the high-risk study of the European Network of National Schizophrenia Networks Studying Gene-Environment Interactions (EU-GEI [39]). EU-GEI is a naturalistic prospective multicenter study that aims to identify the interactive genetic, clinical, and environmental determinants of schizophrenia.

To be eligible to participate, individuals had to meet at least one of the UHR criteria as defined by the Comprehensive Assessment of At Risk Mental States [40]: (1) attenuated psychotic symptoms: the presence of subthreshold positive psychotic symptoms for at least 1 month during the past year; (2) brief limited intermittent psychotic symptoms: an episode of frank psychotic symptoms that lasted no longer than 1 week, which abated spontaneously; or (3) vulnerability: a first-degree relative with a psychotic disorder or schizotypal personality disorder in combination with a significant drop in functioning during at least 1 month in the previous year or enduring low functioning. Exclusion criteria were (1) presence of a current or past psychotic disorder; (2) symptoms for inclusion explained by a medical disorder, drugs or alcohol dependency; or (3) intelligence quotient<60.

Data Collection

Experience Sampling Measures

Data on momentary manifestations of negative symptoms were collected using ESM [31,41]. Participants were asked to report their thoughts, feelings, and symptoms as well as the context (eg, location, company, activity) and the appraisal of the context in their normal daily lives [41-44]. For data collection, participants used a dedicated digital device (the Psymate), which prompted participants with a *beep* to complete a brief questionnaire 10 times a day on 6 consecutive days at random moments within set blocks of time.

A detailed description of ESM items and compliance procedure is provided in Table 1. Momentary manifestations of negative symptoms were operationalized as follows: for blunted affective experiences, we computed mean levels of intensity, variability, and instability of positive and negative affect across beeps within participants. We used 3 operationalizations for lack of social drive: the amount of time spent alone as the percentage of total time, the preference for being alone when in company, and the pleasantness of being alone. To represent anhedonia, we obtained fitted values of positive affect predicted by event pleasantness. As anhedonia is by definition related to pleasant events, only ratings of 1 to 3 were used to test associations with positive affect [26,35]. Observations that indicated unpleasant events (−3 to −1) were excluded from analysis, and neutral events (0) were set as the reference category [26]. We fitted a 2-level, linear mixed model with pleasantness of being in company as the independent and positive affect as the outcome variable and obtained fitted values for representing social anhedonia.

Consistent with previous research, psychometric properties for measures of momentary manifestations of negative symptoms were assessed by evaluating their convergent validity. Therefore, we examined the association between momentary manifestations and observer-rated measures of negative symptoms at baseline (assessed with the expanded Brief Psychiatric Rating Scale [BPRS] [45]; Multimedia Appendix 1). We found small to moderate correlations between the BPRS total score and intensity of negative (r=0.28, P=.02) and positive affect (r=−0.34, P=.004), variability of negative affect (r=0.26, P=.03), anhedonia (r=−0.34, P=.003), and social anhedonia (r=−0.31, P=.008). We found no evidence that the BPRS negative symptom subscale was associated with momentary manifestations of negative symptoms. In addition, we used observer-rated measures of negative symptoms to predict momentary manifestations of negative symptoms in a multilevel model (Multimedia Appendix 1). BPRS total score predicted intensity of positive (b=0.04, P=.01) and negative affect (b=−0.04, P<.001), instability (b=0.04, P=.03) and variability (b=0.03, P=.003) of negative affect, anhedonia (b=−0.04, P=.001), and social anhedonia (b=−0.04, P=.001). The BPRS negative symptoms scale did not predict momentary manifestations of negative symptoms in the multilevel model.

Clinical Outcomes

Clinical outcomes were assessed at baseline, and approximately 1 and 2 years after the baseline assessment. As participants were not seen at exactly 1 and 2 years from their baseline appointment, the exact time points for follow-up assessments varied. Hence, the data closest to 1 and 2 years after baseline were selected as follow-up data. Transition to psychosis and UHR status were assessed using the Comprehensive Assessment of At Risk Mental States [40]. If participants could not be reinterviewed for the follow-up assessments, clinical notes were used to determine transition status. Participants' level of functioning was assessed using the symptoms and the functioning subscales of the Global Assessment of Functioning (GAF [46]) scale. Illness severity was assessed using the severity of illness subscale of the Clinical Global Impression [47] scale. A detailed description of the outcome measures is provided in Table 1.

Table 1. Overview of experience sampling and clinical outcome measures.

Domain	Measure
Experience sampling[a]	
Positive affect	• Positive affect was measured by asking participants to rate how cheerful, relaxed, satisfied, and enthusiastic they felt on a Likert scale ranging from 1 (not at all) to 7 (very much). We found satisfying internal consistency, Cronbach α=.73. In line with previous studies [34,48], we used high and low physiological arousal items.
Negative affect	• Negative affect was measured by asking participants to rate the extent to which they felt insecure, down, lonely, anxious and irritated on a Likert scale ranging from 1 (not at all) to 7 (very much). We found satisfying internal consistency, Cronbach α=.73.
Blunted affect	• Intensity was operationalized as the mean levels of positive and negative affect. • Instability was computed as the squared difference between beep-level positive and negative affect intensity at beep t and beep-level positive affect intensity at beep t-1 (previous beep), within days, within persons (mean of the squared successive differences), and only calculated if there was a maximum of 2 observations missing between 2 consecutive observations. Difference scores between 2 observations overnight were excluded [33]. • Variability was computed as the squared difference between beep-level intensity of positive and negative affect at each observation and individual mean positive and negative affect over observations, over days within persons [33].
Social drive	• Lack of social drive was conceptualized as the amount of time spent alone in percentage of total time, the experienced pleasantness of being alone, and the preference of being alone when in company. Pleasantness of being alone and preference to be alone when in company were rated on a Likert scale ranging from 1 (not at all) to 7 (very much). If participants were alone: "I find it pleasant to be alone" and "I would prefer to have company." If participants were in company: "I find being with these people pleasant." and "I would prefer to be alone."
Anhedonia	• Anhedonia was conceptualized as the relationship between positive affect and the occurrence of pleasant events. Participants were asked to think about the most important event that happened since the last beep. The pleasantness of this event was rated on a bipolar scale ranging from −3 (very unpleasant) to 3 (very pleasant). We only used ratings of 1 to 3 to test associations with positive affect, as anhedonia is per definition related to pleasant events. Observations indicating unpleasant events (−3 to −1) were excluded, and neutral events (0) were used as a reference category [26].
Social anhedonia	• Social anhedonia was defined as the association between positive affect and pleasantness of being in company [26]. Participants were asked whether they were alone or in company. If participants indicated to be in company, they were asked to rate "I find being with these people pleasant." on a Likert scale ranging from 1 (not at all) to 7 (very much).
Clinical outcome measures	
CAARMS	• Transition to psychosis and UHR status were assessed using the Comprehensive Assessment of At Risk Mental State (CAARMS [40]), a semistructured interview to assess attenuated psychotic symptoms in individuals at high risk for psychosis. The CAARMS comprises 27 items clustered in 7 subscales: positive symptoms, cognitive change (attention and concentration), emotional disturbance, negative symptoms, behavioral change, motor or physical changes, and general psychopathology. Scores on each item range from 0 (absent) to 6 (extreme).
GAF	• The Global Assessment of Functioning (GAF [46]) obtains ratings of burdening symptoms and disabilities in the last month on a scale from 100 (no symptoms or superior functioning in a wide range of activities) to 1 (persistent danger of severely hurting self or others or serious suicidal act with clear expectation of death or persistent inability to maintain minimal personal hygiene).
CGI	• The Clinical Global Impression Scale (CGI [47]) symptoms severity subscale is an expert rating of average illness severity during the last week ranging from 1 (normal, not at all ill) to 7 (among the most extremely ill patients).

[a]Experience sampling procedure: During an initial briefing, the study team ensured that the week of data collection was a typical week for the participant. Each time the device emitted the beep signal, participants were asked to stop their activity and answer the questions. The experience sampling questionnaire was available to participants for the duration of 10 minutes after emission of the beep signal. Participants were contacted at least once during the assessment period to assess their adherence to instructions, identify any potential distress associated with the method, and maximize the number of observations per participant. At the end of the assessment period, participants' reactivity to, and compliance with, the method was examined in a debriefing session. Participants were required to provide valid responses to at least one-third (ie, 20 valid answers) of the emitted beeps to be included in the analysis [49]. Procedures to ensure data quality are reported in Multimedia Appendix 2.

Statistical Analysis

Momentary manifestations of negative symptoms—operationalized and computed as detailed above—were used as independent variables to predict clinical outcomes at 1- and 2-year follow-up using Stata 15. We fitted linear regression models using the command *regress* with level of functioning and illness severity as outcome variables and momentary manifestations of negative symptoms as independent variables. In exploratory analyses, we examined the predictive value of momentary manifestations of negative symptoms for transition to psychosis and remission from UHR status as

outcomes. Survival analyses using the Stata commands *stset* and *streg* were performed to account for the time to event structure of the data. We used time to follow-up as a proxy for time to remission. In both survival analyses, a Weibull distribution was assumed.

Analyses were adjusted for a priori confounders (ie, age, gender, ethnicity, center, time to follow-up; unadjusted results are provided in Multimedia Appendix 3). In a sensitivity analysis, we included current depressive episode (Multimedia Appendix 4) and comorbid disorders (Multimedia Appendix 5) as additional independent variables to control for potential confounding. We corrected for multiple testing to reduce the probability of type I errors because of the number of tests performed. We corrected within domains of momentary manifestations of negative symptoms and clinical outcomes. As in previous experience sampling studies [50,51], Simes correction method was used to account for multiple tests of significance [52]. Simes correction is a modified version of the more conservative Bonferroni correction in case of dependent hypotheses given significance tests in the current analyses were not independent [52]. With the Simes correction, the most significant *P* value is tested against α=.05/n (total number of tests), the second most significant *P* value is tested against α=.05/(n−1), the third *P* value against α=.05/(n−2), and so on. Simes-corrected significant results are highlighted in a footnote in tables.

Results

Basic Sample and Clinical Characteristics

The ESM sample comprised 79 individuals at UHR, of whom 9 transitioned to psychosis during the study period. Data on clinical outcomes were obtained for 48 individuals at 1-year follow-up and 36 individuals at 2-year follow-up. Participants were on average aged 23 (SD 4.93) years and 56% (44/79) were women. The majority (53/79, 67%) of the sample was White, followed by 15% (12/79) with Black ethnicity. In addition to their UHR status, 76% (60/79) of the participants were diagnosed with a comorbid axis I disorder (further details are provided in Multimedia Appendix 6). Compared with the sample of individuals included in the EU-GEI High Risk Study for whom experience sampling data were not collected (no ESM sample, N=266), samples in this study showed no differences in demographic characteristics (age: $t_{343}=-1.33$, *P*=.19; gender: $\chi^2_1=3.6$, *P*=.06; ethnicity: $\chi^2_5=6.5$, *P*=.26) or prevalence of comorbid disorders ($\chi^2_1=1.8$, *P*=.18). However, the current sample showed poorer functioning (symptoms: $t_{315}=2.29$, *P*=.02) and lower levels of observer-rated negative symptoms ($t_{320}=2.27$, *P*=.02) at baseline. Comparing participants who completed follow-up assessments, the sample with no ESM data collected (N=134, 1-year follow-up; N=89, 2-year follow-up) showed a lower BPRS total score at 1-year follow-up ($t_{159}=-2.07$, *P*=.04). There were no significant differences in demographic or clinical characteristics at 2-year follow-up. Table 2 gives an overview of relevant sample characteristics.

Table 2. Basic sample and clinical characteristics.[a]

Characteristics	ESM[b] sample			No ESM sample			ESM vs no ESM		
	Baseline	1-year follow-up	2-year follow-up	Baseline	1-year follow-up	2-year follow-up	Baseline	1-year follow-up	2-year follow-up
Sample size	79	48	36	266	134	89	N/A[c]	N/A	N/A
Age at baseline (years), mean (SD)	23.0 (4.93)	23.6 (5.24)	23.8 (5.18)	22.2 (4.89)	22.5 (4.82)	23.3 (5.14)	t_{343}=−1.33, *P*=.19	t_{180}=−1.30, *P*=.19	t_{123}=−.45, *P*=.65
Gender, n (%)							χ^2_1=3.58, *P*=.06	χ^2_1=3.76, *P*=.05	χ^2_1=2.74, *P*=.10
Male	35 (44)	22 (46)	16 (44)	150 (56)	83 (62)	54 (61)			
Female	44 (56)	26 (54)	20 (56)	116 (44)	51 (38)	35 (39)			
Ethnicity, n (%)							χ^2_5=6.5, *P*=.26	χ^2_5=6.7, *P*=.24	χ^2_5=6.25, *P*=.28
White	53 (67)	33 (69)	27 (75)	193 (73)	99 (74)	63 (71)			
Black	12 (15)	9 (19)	5 (14)	22 (8)	9 (7)	6 (7)			
Other	14 (18)	6 (13)	4 (11)	50 (19)	26 (19)	20 (22)			
Comorbidity at baseline, n(%)	60 (76)	37 (77)	28 (78)	220 (83)	111 (83)	79 (89)	χ^2_1=1.8, *P*=.18	χ^2_1=0.77, *P*=.38	χ^2_1=2.5, *P*=.11
BPRS[d]									
Total score, mean (SD)	44.00 (9.46)	39.69 (11.63)	37.31 (12.53)	43.37 (10.57)	36.01 (9.70)	33.45 (10.85)	t_{314}=−0.46, *P*=.65	t_{159}=−2.07, *P*=.04	t_{111}=−1.97, *P*=.05
Negative symptom score, mean (SD)	4.49 (1.86)	4.04 (1.74)	3.75 (1.78)	5.21(2.51)	4.42 (2.05)	4.12(1.87)	t_{320}=2.27, *P*=.02	t_{160}=1.22, *P*=.26	t_{112}=−.98, *P*=.33
GAF[e]									
Symptoms, mean (SD)	52.88 (9.85)	56.96 (10.76)	61.00 (11.73)	55.92 (10.23)	59.49 (13.08)	61.25 (15.02)	t_{315}=2.29, *P*=.02	t_{180}=1.20, *P*=.23	t_{123}=0.09, *P*=.93
Disability, mean (SD)	56.27 (13.00)	58.92 (13.41)	63.78 (13.62)	55.36 (12.20)	60.40 (13.77)	61.81 (16.09)	t_{330}=−0.57, *P*=.57	t_{196}=0.65, *P*=.51	t_{132}=−0.65, *P*=.51
CGI[f]									
Illness severity, mean (SD)	3.57 (1.21)	3.15 (1.32)	2.89 (1.25)	3.60 (1.09)	3.33 (1.37)	3.22 (1.51)	t_{319}=0.21, *P*=.83	t_{203}=0.83, *P*=.41	t_{148}=1.21, *P*=.23
UHR[g] criteria met, n (%)	N/A	36 (73)	23 (62)	N/A	107 (73)	71 (66)	N/A	χ^2_1=.00, *P*=.97	χ^2_1=.15, *P*=.69

[a]Follow-up values for age, gender, ethnicity, and comorbidity based on individuals with valid Global Assessment of Functioning Scale at follow-up.

[b]ESM: Experience Sampling Methodology.

[c]N/A: not applicable.

[d]BPRS: Brief Psychiatric Rating Scale.

[e]GAF: Global Assessment of Functioning Scale.

[f]CGI: Clinical Global Impression Scale.

[g]UHR: ultrahigh risk.

Blunted Affective Experience and Clinical Outcomes

Tables 3 and 4 show the results on clinical outcomes at follow-up predicted by blunted affective experience at baseline. We found no evidence that blunted affective experience predicted illness severity or level of functioning. In exploratory analyses, time to remission from UHR status was predicted by variability of positive affect (hazard ratio [HR]=4.93, 95% CI 1.61-15.11, *P*=.005, statistically significant after Simes correction). Participants with greater variability were more likely to experience a shorter time to remission from the UHR status. We found no evidence that blunted affective experience predicted transition to psychosis.

Table 3. Level of functioning at 1- and 2-year follow-up predicted by blunted affective experience at baseline (ie, intensity, instability, and variability of negative and positive affect) and clinical outcome at baseline.[a]

Outcomes	Level of functioning							
	Symptoms[b]				Disability[b]			
	1-year follow-up (n=48)		2-year follow-up (n=36)		1-year follow-up (n=48)		2-year follow-up (n=36)	
	b (CI)	*P* value	*b* (CI)	*P* value	*b* (CI)	*P* value	*b* (CI)	*P* value
Predictor: Intensity NA[c]								
Outcome at baseline	0.18 (−0.16 to 0.52)	.29	0.00 (−0.61 to 0.61)	.99	0.34 (−0.01 to 0.70)	.06	0.55 (0.06 to 1.04)	.03
Intensity NA	−2.51 (−6.54 to 1.52)	.22	−1.36 (−6.89 to 4.18)	.62	−3.17 (−7.83 to 1.48)	.18	1.26 (−4.90 to 7.42)	.68
Predictor: Intensity PA[d]								
Outcome at baseline	0.15 (−0.18 to 0.48)	.36	−0.01 (−0.62 to 0.60)	.97	0.34 (0.00 to 0.68)	.05	0.55 (0.06 to 1.05)	.03
Intensity PA	3.84 (−0.09 to 7.77)	.06	1.15 (−4.58 to 6.88)	.68	5.04 (0.59 to 9.49)	.03	0.07 (−6.34 to 6.48)	.98
Predictor: Instability NA								
Outcome at baseline	0.24 (−0.10 to 0.58)	.17	−0.02 (−0.62 to 0.58)	.94	0.36 (0.00 to 0.73)	.05	0.55 (0.08 to 1.01)	.02
Instability NA	0.80 (−1.43 to 3.04)	.47	−1.72 (−5.94 to 2.50)	.41	−0.36 (−2.98 to 2.27)	.79	−3.66 (−8.21 to 0.88)	.11
Predictor: Instability PA								
Outcome at baseline	0.22 (−0.12 to 0.56)	.21	−0.06 (−0.64 to 0.52)	.83	0.37 (0.01 to 0.73)	.046	0.55 (0.11 to 0.99)	.02
Instability PA	−0.24 (−3.75 to 3.28)	.89	−4.68 (−10.43 to 1.07)	.11	−0.46 (−4.57 to 3.64)	.82	−7.52 (−13.54 to −1.51)	.02
Predictor: Variability NA								
Outcome at baseline	0.23 (−0.11 to 0.56)	.18	0.00 (−0.59 to 0.60)	.99	0.37 (0.01 to 0.73)	.046	0.52 (0.07 to 0.98)	.03
Variability NA	1.60 (−3.30 to 6.50)	.51	−3.80 (−11.52 to 3.92)	.32	0.03 (−5.76 to 5.82)	.99	−7.96 (−16.15 to 0.22)	.06
Predictor: Variability PA								
Outcome at baseline	0.21 (−0.13 to 0.55)	.22	0.02 (−0.56 to 0.60)	.93	0.37 (0.01 to 0.73)	.04	0.47 (−0.01 to 0.95)	.05
Variability PA	1.12 (−4.54 to 6.77)	.6.92	−5.55 (−12.51 to 1.42)	.11	2.22 (−4.34 to 8.78)	.50	−6.30 (−14.23 to 1.63)	.11

[a]Results adjusted for age, gender, ethnicity, center, and time to follow-up.

[b]Level of functioning assessed with the Global Assessment of Functioning Scale.

[c]NA: negative affect.

[d]PA: positive affect.

Table 4. Illness severity, remission from ultrahigh risk (UHR) status and transition status 1- and 2-year follow-up predicted by blunted affective experience at baseline (ie, intensity, instability, and variability of negative and positive affect) and clinical outcome at baseline.[a]

Outcomes	Illness severity[b]				Remission from UHR status		Transition status	
	1-year follow-up (n=47)		2-year follow-up (n=37)					
	b (CI)	*P* value	*b* (CI)	*P* value	HR[c] (CI)	*P* value	HR (CI)	*P* value
Predictor: Intensity NA[d]								
Outcome at baseline	0.43 (0.13 to 0.73)	.006	0.28 (−0.19 to 0.75)	.24	N/A[e]	N/A	N/A	N/A
Intensity NA	0.32 (−0.07 to 0.71)	.11	−0.03 (−0.59 to 0.53)	.91	0.34 (0.12 to 0.98)	.045	1.44 (0.66 to 3.13)	.36
Predictor: Intensity PA[f]								
Outcome at baseline	0.44 (0.15 to 0.74)	.004	0.17 (−0.30 to 0.64)	.46	N/A	N/A	N/A	N/A
Intensity PA	−0.31 (−0.69 to 0.07)	.11	−0.35 (−0.98 to 0.28)	.26	2.08 (0.88 to 4.93)	.10	0.62 (0.23 to 1.65)	.34
Predictor: Instability NA								
Outcome at baseline	0.52 (0.22 to 0.81)	.001	0.27 (−0.18 to 0.72)	.23	N/A	N/A	N/A	N/A
Instability NA	−0.02 (−0.23 to 0.18)	.81	−0.02 (−0.46 to 0.42)	.94	1.19 (0.57 to 2.48)	.64	1.02 (0.67 to 1.54)	.92
Predictor: Instability PA								
Outcome at baseline	0.51 (0.22 to 0.81)	.001	0.28 (−0.16 to 0.73)	.20	N/A	N/A	N/A	N/A
Instability PA	−0.06 (−0.38 to 0.26)	.71	0.24 (−0.38 to 0.86)	.43	1.75 (1.69 to 4.44)	.24	0.99 (0.50 to 1.94)	.97
Predictor: Variability NA								
Outcome at baseline	0.51 (0.22 to 0.81)	.001	0.26 (−0.20 to 0.72)	.25	N/A	N/A	N/A	N/A
Variability NA	−0.10 (−0.53 to 0.33)	.64	−0.08 (−0.89 to 0.74)	.85	1.24 (0.30 to 5.14)	.77	1.21 (0.55 to 2.63)	.64
Predictor: Variability PA								
Outcome at baseline	0.51 (0.21 to 0.81)	.001	0.37 (−0.09 to 0.83)	.11	N/A	N/A	N/A	N/A
Variability PA	0.04 (−0.47 to 0.54)	.88	0.49 (−0.29 to 1.28)	.21	4.93 (1.61 to 15.11)	.005[g]	1.49 (0.52 to 4.23)	.46

[a]Results adjusted for age, gender, ethnicity, center, and time to follow-up.

[b]Illness severity assessed with the Clinical Global Impression Scale.

[c]HR: hazard ratio.

[d]NA: negative affect.

[e]N/A: not applicable.

[f]PA: positive affect.

[g]Statistically significant after Simes correction.

Lack of Social Drive and Clinical Outcomes

Tables 5 and 6 show findings on clinical outcomes predicted by lack of social drive. We found no evidence that the amount of time spent alone and the preference to be alone when in company predicted level of functioning or illness severity. Experienced pleasantness of being alone predicted the GAF disability subscale at 2-year follow-up (b=−4.62; 95% CI −8.19 to −1.04, *P*=.01 [statistically significant after Simes correction]), such that individuals who experienced greater pleasantness of being alone showed poorer functioning. However, there was no evidence that pleasantness of being alone predicted illness

severity and the GAF symptoms score. In exploratory analyses, there was no evidence that lack of social drive predicted time to transition or remission from UHR status.

Table 5. Level of functioning at 1- and 2-year follow-up predicted by lack of social drive (ie, amount of time spent alone, preference to be alone when in company and experienced pleasantness of being alone) and clinical outcome at baseline.[a]

Outcomes	Level of functioning: symptoms[b]				Level of functioning: disability[b]			
	1-year follow-up (N=48)		2-year follow-up (N=36)		1-year follow-up (N=48)		2-year follow-up (N=36)	
	b (CI)	*P* value	*b* (CI)	*P* value	*b* (CI)	*P* value	*b* (CI)	*P* value
Predictor: Amount of time spent alone								
Outcome at baseline	0.22 (−0.12 To 0.56)	.20	−0.10 (−0.69 to 0.48)	.72	0.37 (0.01 to 0.73)	.045	0.49 (0.03 to 0.96)	.04
Amount of time spent alone	1.71 (−12.18 to 15.60)	.80	13.63 (−3.99 to 31.24)	.12	4.50 (−11.75 to 20.76)	.58	17.77 (−1.38 to 36.93)	.07
Predictor: Preference to be alone when in company								
Outcome at baseline	0.20 (−0.13 to 0.54)	.23	0.04 (−0.58 to 0.67)	.89	0.37 (0.01 to 0.72)	.04	0.54 (0.05 to 1.03)	.03
Preference to be alone	−1.61 (−4.20 to 0.97)	.21	−1.38 (−5.41 to 2.65)	.49	−1.88 (−4.90 to 1.15)	.22	−1.12 (−5.46 to 3.23)	.60
Predictor: Pleasantness of being alone								
Outcome at baseline	0.22 (−0.12 to 0.57)	.21	0.08 (−0.51 to 0.67)	.77	0.41 (0.04 to 0.79)	.03	0.51 (0.07 to 0.95)	.02
Pleasantness of being alone	0.04 (−2.88 to 2.96)	.98	−2.87 (−6.36 to 0.62)	.10	−1.38 (−4.86 to 2.10)	.43	−4.62 (−8.19 to −1.04)	.01[c]

[a]Results adjusted for age, gender, ethnicity, center, and time to follow-up.

[b]Level of functioning assessed with the Global Assessment of Functioning Scale.

[c]Statistically significant after Simes correction.

Table 6. Illness severity, remission from ultrahigh risk (UHR) status and transition status at 1- and 2-year follow-up predicted by lack of social drive (ie, amount of time spent alone, preference to be alone when in company and experienced pleasantness of being alone) and clinical outcome at baseline.[a]

Outcomes	Illness severity[b]				Remission from UHR status		Transition status	
	1-year follow-up (N=47)		2-year follow-up (N=37)		(N=54)		(N=57)	
	b (CI)	*P* value	*b* (CI)	*P* value	HR[c] (CI)	*P*	HR (CI)	*P* value
Predictor: Amount of time spent alone								
Outcome at baseline	0.49 (0.19 to 0.80)	.002	0.23 (−0.22 to 0.67)	.30	N/A[d]	N/A	N/A	N/A
Amount of time spent alone	−0.29 (−1.59 to 1.02)	.66	−1.17 (−3.00 to 0.66)	.20	3.91 (0.25 to 60.64)	.33	0.07 (0.00 to 2.07)	.13
Predictor: Preference to be alone when in company								
Outcome at baseline	0.48 (0.18 to 0.78)	.002	0.24 (−0.20 to 0.68)	.27	N/A	N/A	N/A	N/A
Preference to be alone	0.11 (−0.14 to 0.36)	.37	0.23 (−0.17 to 0.63)	.24	0.97 (0.51 to 1.84)	.92	1.20 (0.65 to 2.22)	.56
Predictor: Pleasantness of being alone								
Outcome at baseline	0.51 (0.21 to 0.81)	.002	0.32 (−0.13 to 0.77)	.15	N/A	N/A	N/A	N/A
Pleasantness of being alone	0.05 (−0.19 to 0.30)	.68	0.19 (−0.16 to 0.54)	.28	0.82 (0.44 to 1.54)	.54	1.39 (0.75 to 2.56)	.30

[a]Results adjusted for age, gender, ethnicity, center, and time to follow-up.

[b]Illness severity assessed with the Clinical Global Impression Scale.

[c]HR: hazard ratio.

[d]N/A: not applicable.

Anhedonia and Clinical Outcomes

Tables 7 and 8 show findings on clinical outcomes at 1- and 2-year follow-up predicted by anhedonia. Anhedonia predicted the GAF disability subscale at 1-year follow-up (b=5.61, 95% CI 1.08-10.15; *P*=.02 [statistically significant after Simes correction]). Lower positive affect in moments of pleasant events or, in other words, higher levels of anhedonia, were associated with poorer functioning. However, we found no evidence that anhedonia predicted functioning at 2-year follow-up. In addition, anhedonia did not predict illness severity at 1- and 2-year follow-up. In exploratory analyses, we found no evidence that anhedonia predicted time to remission or transition to psychosis.

Table 7. Level of functioning at 1- and 2-year follow-up predicted by Anhedonia, Social Anhedonia and clinical outcome at baseline.[a]

Outcomes	Level of functioning: symptoms[b]				Level of functioning: disability[b]			
	1-year follow-up (N=48)		2-year follow-up (N=36)		1-year follow-up (N=48)		2-year follow-up (N=36)	
	b (CI)	*P* value	*b* (CI)	*P* value	*b* (CI)	*P* value	*b* (CI)	*P* value
Predictor: anhedonia								
Outcome at baseline	0.16 (−0.17 to 0.49)	.32	−0.02 (−0.62 to 0.59)	.96	0.34 (0.00 to 0.68)	.048	0.56 (0.06 to 1.05)	.03
Anhedonia events	3.73 (−0.32 to 7.78)	.07	0.25 (−5.37 to 5.88)	.93	5.61 (1.08 to 10.15)	.02[c]	−0.43 (−6.70 to 5.85)	.89
Predictor: social anhedonia								
Outcome at baseline	0.17 (−0.15 to 0.49)	.28	0.01 (−0.59 to 0.61)	.97	0.33 (0.01 to 0.66)	.046	0.53 (0.04 to 1.01)	.04
Social anhedonia	4.61 0.74 to 8.48)	.02[c]	2.29 (−3.65 to 8.23)	.44	6.36 (1.97 to 10.74)	.006[c]	3.09 (−3.51 to 9.70)	.35

[a]Results adjusted for age, gender, ethnicity, center and time to follow-up.

[b]Level of functioning assessed with the Global Assessment of Functioning Scale.

[c]Statistically significant after Simes correction.

Table 8. Illness severity, remission from ultrahigh risk (UHR) status and transition status at 1- and 2-year follow-up predicted by Anhedonia, Social Anhedonia and clinical outcome at baseline.[a]

Outcomes	Illness severity[b]				Remission from UHR status		Transition status	
	1-year follow-up (N=47)		2-year follow-up (N=37)		(N=54)		(N=57)	
	b (CI)	*P* value	*b* (CI)	*P* value	HR[c] (CI)	*P* value	HR (CI)	*P* value
Predictor: anhedonia								
Outcome at baseline	0.45 (0.15 to 0.75)	.004	0.19 (−0.29 to 0.66)	.42	N/A[d]	N/A	N/A	N/A
Anhedonia	−0.30 (−0.69 to 0.09)	.13	−0.30 (−0.91 to 0.32)	.34	2.02 (0.82 to 4.96)	.13	0.66 (0.23 to 1.88)	.44
Predictor: social anhedonia								
Outcome at baseline	0.44 (0.15 to 0.73)	.004	0.14 (−0.30 to 0.59)	.52	N/A	N/A	N/A	N/A
Social anhedonia	−0.38 (−0.74 to −0.01)	.045[e]	−0.57 (−1.18 to 0.05)	.07	2.22 (0.85 to 5.81)	.10	0.69 (0.26 to 1.80)	.45

[a]Results adjusted for age, gender, ethnicity, center and time to follow-up.

[b]Illness severity assessed with the Clinical Global Impression Scale.

[c]HR: hazard ratio.

[d]N/A: not applicable.

[e]Statistically significant after Simes correction.

Social Anhedonia and Clinical Outcomes

As displayed in Tables 7 and 8, reduced positive affect in moments of pleasant company or, in other words, higher levels of social anhedonia at baseline were associated with higher levels of illness severity (b=−0.38; 95% CI −0.74 to 0.01; *P*=.045 [statistically significant after Simes correction]) and lower scores on both GAF subscales (symptoms: b=4.61; 95% CI 0.74 to 8.48; *P*=.02 [statistically significant after Simes correction]; disability: b=6.36; 95% CI 1.97 to 10.74; *P*=.006 [statistically significant after Simes correction]) at 1-year follow-up. However, we found no evidence that social anhedonia predicted clinical outcomes at 2-year follow-up. In exploratory analyses, we found no evidence that social anhedonia predicted time to remission or transition to psychosis.

Discussion

Principal Findings

Using an experience sampling design, this study found no evidence that blunted affective experience predicted functioning

or illness severity at follow-up (H1). However, there was some evidence that higher experienced pleasantness of being alone was associated with poorer functioning at 2-year follow-up (H2). In addition, our results tentatively suggest that higher levels of anhedonia were associated with poorer functioning at 1-year follow-up (H3). Finally, we found robust evidence that higher levels of social anhedonia were associated with higher levels of illness severity and poorer functioning at 1-year follow-up (H4). In our exploratory analysis, we found no evidence that momentary manifestations of negative symptoms in daily life predicted transition status. However, our results tentatively suggest that blunted affective experience predicted time to remission from UHR status.

Methodological Considerations

Our findings should be interpreted in light of several methodological considerations. First, the sample selection should be critically evaluated: ESM is a burdensome research method, which may lead to selection bias, such that individuals with more intense symptoms might be underrepresented in the sample. However, compared with the no ESM sample of the EU-GEI High Risk Study, the participants in this showed comparable levels of illness severity and lower scores on the GAF symptoms subscale at baseline. In addition, the sample showed high comorbidity rates of nonpsychotic disorders, which replicates findings from previous studies and systematic reviews [53,54]. High rates of comorbidity, especially comorbid depressive disorders, may have attenuated the observed effects. However, when controlling for current depressive episodes or comorbid disorders in our sensitivity analysis, we found a similar pattern in terms of magnitude of associations but slightly wider 95% CIs and some differences in statistical significance. In addition, it is important to consider the small-to-moderate sample size and the small absolute number of 9 individuals (11% of the sample) who transitioned to psychosis within the follow-up period, although this transition rate is rather common in the field [14,55]. Second, measuring social isolation and affect repeatedly over longer periods might provide a better prediction of outcomes. However, given burden on participants, this would require a less intense longitudinal data collection method, as is the case for ESM. Third, it is important to consider some limitations regarding data collection at follow-up: Although this was planned for 1- and 2-year follow-up, follow-up intervals varied in some individuals. Yet, analyses were controlled for time to follow-up and sensitivity analyses conducted with the subsample of individuals assessed ±6 months to the ideal follow-up time point showed a similar pattern of findings though varying statistical significance due to reduced sample size (Multimedia Appendix 7). Moreover, experience sampling data was not collected at follow-up. Nonetheless, using the Clinical Global Impression scale and the GAF scale, we obtained ratings of several widely used outcome measures at follow-up. In addition, the follow-up period of 2 years was, arguably, rather short in this study. However, previous research has demonstrated that the highest risk for transition in UHR samples is over the first 2 years after ascertainment [56]. Fourth, one should consider some statistical issues: For anhedonia and social anhedonia, we used fitted values of positive affect predicted by event pleasantness or pleasantness of social contact, to predict, in turn, clinical outcomes at follow-up. For blunted affective experience and lack of social drive, we aggregated data on the person-level. Aggregation of momentary manifestations of negative symptoms on the person-level led to a loss of information in comparison with the beep-level, as the variance of beeps is not reflected in the aggregated scores. Nonetheless, compared with a single questionnaire assessment, the aggregated experience sampling measures used in this study still provide higher levels of precision in measurement. The number of statistical analyses performed may have resulted in multiple testing problems. However, in order to control for type I error, results were corrected using the Simes method [52] by momentary manifestation of negative symptom and outcome domain. In addition, time to follow-up was used as a crude proxy to impute for time to remission from UHR status (eg, for participants who remitted at any time between baseline and 1-year follow-up, the date of the 1-year follow-up assessment was used as proxy), which might lead to imprecision in these exploratory survival analyses. Future research should attempt to establish a more precise data collection for time to remission.

Comparison With Previous Research

To our knowledge, this is the first study using an experience sampling design to investigate the predictive value of momentary manifestations of negative symptoms measured in individuals UHR. In accordance with our hypotheses, we found evidence for more intense momentary manifestations of negative symptoms to be associated with poorer functioning and higher illness severity at follow-up. In addition, we found evidence that individuals with greater variability of positive affect (as a measure of blunted affective experience), experienced a shorter time to remission from UHR status. This is in line with findings from previous studies using other operationalizations of negative symptoms [2-5]. Given that ESM measures of momentary manifestations of negative symptoms are intended to capture subjective experience of social context, our findings primarily pertain to the experiential level rather than to the level of expression [27].

Our findings tentatively suggest that blunted affect, lack of social drive, and anhedonia are associated with some clinical outcomes, but findings on social anhedonia were most robust. We may speculate that changes in affective response to social contact (ie, social anhedonia) in daily life may be most relevant in individuals at UHR, whereas other types of momentary manifestations of negative symptoms (eg, lack of social drive) may be more relevant in later stages of psychosis. Social anhedonia may contribute to a loss of reinforcement of social contact, which might encourage a progressive decrease of social interaction and social functioning more downstream, closer to, or directly at, onset of psychotic disorder [35,57-59].

The findings have important implications for clinicians and researchers aiming to improve functional outcomes of individuals at UHR. Recent meta-analyses found no evidence for psychosocial treatment to improve functioning in individuals at UHR [60], with poor functioning at baseline being, in turn, a predictor for later psychopathology [61]. Taken together, this may contribute to a vicious cycle of symptom burden and poor functioning amplifying each other in this group at risk.

Therefore, new intervention approaches are urgently required and the experience of momentary manifestations of negative symptoms, especially social anhedonia, in daily life may be a promising target. Possibly, improving social anhedonia may diminish social isolation, and thereby improve outcomes.

In addition, we found only weak correlations between momentary manifestations of negative symptoms and the BPRS scores, highlighting the relevance of participants' subjective experience. These discrepancies may be interpreted in different ways. First, discrepancies may evolve due to varying modes of assessment and, hence, precision of measurement. Gerritsen, Bagby [37] claim that some negative symptoms may be associated with no or very limited subjective distress and, hence, difficult to measure via self-report. However, one may argue that aggregating multiple momentary measurements across several days may provide a more precise measure of affective and motivational processes than cross-sectional clinical interviews [27]. Second, the discrepancies may, in fact, reflect 2 distinct dimensions of negative symptoms (ie, experience vs expression), and therefore, relying on purely behavioral indicators in assessing negative symptoms may result in a more limited understanding of internal, experiential aspects [27]. Both interpretations highlight the potential of ESM as a diagnostic tool over and above traditional clinical measures of symptoms [62].

Conclusions

We found evidence for momentary manifestations of negative symptoms, especially social anhedonia, to predict clinical outcomes at follow-up. These findings emphasize that the assessment of momentary manifestations of negative symptoms in individuals at UHR is of considerable potential value for both diagnostic assessment and early intervention. The assessment of momentary manifestations of negative symptoms may provide a more comprehensive picture of patients' symptoms in the context of their daily life for clinicians and researchers and contribute to a better understanding of individuals' subjective experience. In addition, the experience of momentary manifestations of negative symptoms, especially social anhedonia, in daily life may be a promising target for interventions aiming to improve clinical outcomes in the early stages of psychosis.

Acknowledgments

The authors sincerely thank the participants and all researchers involved in the data collection. This work was supported by a Postdoctoral Research Fellowship of the UK National Institute for Health Research (UR, NIHR-PDF-201104065); a Heisenberg professorship from the German Research Foundation (UR, grant no. 389624707); financial support from the National Institute for Health Research (NIHR) Biomedical Research Centre for Mental Health at South London and Maudsley NHS Foundation Trust and King's College London; a FWO (Flemish Science Foundation) Odysseus grant (IM-G, G0F8416N); a NHMRC Senior Research Fellowship (BN, 1137687); the Beatrice and Samuel A Seaver Foundation (EV); and the Wellcome Trust (CM, WT087417). This work is an approved add-on study of the "The European Network of National Schizophrenia Networks studying Gene-Environment Interactions " (EU-GEI), which is supported by funding from the European Union (European Community's Seventh Framework Program [HEALTH-F2-2009–241909; Project EU-GEI]).

The members of the EU-GEI High Risk Study are as follows: Philip McGuire[1], Lucia R. Valmaggia[2], Emily Hedges[1], Maria Calem[1], Stefania Tognin[1], Gemma Modinos[1], Lieuwe de Haan[3], Mark van der Gaag[4,5], Eva Velthorst[3,6], Tamar C Kraan[3], Nadine Burger[5], Daniella S van Dam[3], Neus Barrantes-Vidal[7,8,9,10], Tecelli Domínguez-Martínez[7], Paula Cristóbal-Narváez[7], Thomas R Kwapil[8] , Manel Monsonet-Bardají[7], Lídia Hinojosa[7], Anita Riecher-Rössler[11], Stefan Borgwardt[11], Charlotte Rapp[11], Sarah Ittig[11], Erich Studerus[11], Renata Smieskova[11], Rodrigo Bressan[12], Ary Gadelha[12], Elisa Brietzke[13], Graccielle Asevedo[12], Elson Asevedo[12], Andre Zugman[12], Stephan Ruhrmann[14], Dominika Gebhard[14], Julia Arnhold[15], Joachim Klosterkötter[14], Dorte Nordholm[16], Lasse Randers[16], Kristine Krakauer[16], Tanya Louise Naumann[16], Louise Birkedal Glenthøj[16], Merete Nordentoft[16], Marc De Hert[17], Ruud van Winkel[17], Barnaby Nelson[18], Patrick McGorry[18], Paul Amminger[18], Christos Pantelis[18], Athena Politis[19], Joanne Goodall[19], Gabriele Sachs[20], Iris Lasser[20], Bernadette Winklbaur[20], Mathilde Kazes[21], Claire Daban[21], Julie Bourgin[21], Olivier Gay[21], Célia Mam-Lam-Fook[21], Marie-Odile Krebs[21], Bart P Rutten[22], Jim van Os1[1,23]

Affiliations

[1]Department of Psychosis Studies, Institute of Psychiatry, King's College London, London, United Kingdom

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

[3]Department of Psychiatry, Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands

[4]Department of Clinical Psychology, VU University and Amsterdam; Public Mental Health research institute, Amsterdam, The Netherlands

[5]Department of Psychosis Research, Parnassia Psychiatric Institute, The Hague, The Netherlands

[6]Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, United States of America

[7]Departament de Psicologia, Clínica i de la Salut, Universitat Autònoma de Barcelona, Barcelona, Spain

[8]Departament de Salut Mental, Sant Pere Claver-Fundació Sanitària, Barcelona, Spain

[9]Spanish Mental Health Research Network, CIBERSAM, Spain

[10]Department of Psychology, University of North Carolina at Greensboro, Greensboro, United States of America

[11]Center for Gender Research and Early Detection, Psychiatric University Clinics Basel, Basel, Switzerland

[12]LiNC - Lab Interdisciplinar Neurociências Clínicas, Depto Psiquiatria, Escola Paulista de Medicina, Universidade Federal de São Paulo – UNIFESP, São Paulo, Brazil

[13]Pogram for cognition and Intervention in Individuals in At-Risk Mental States (PRISMA), Department of Psychiatry, Universidade Federal de São Paulo, São Paulo, Brazil

[14]Department of Psychiatry and Psychotherapy, University of Cologne, Cologne, Germany

[15]Psyberlin, Berlin, Germany

[16]Mental Health Center Copenhagen and Center for Clinical Intervention and Neuropsychiatric Schizophrenia Research, CINS, Mental Health Center Glostrup, Mental Health Services in the Capital Region of Copenhagen, University of Copenhagen, Copenhagen, Denmark

[17]Department of Neuroscience, University Psychiatric Centre, Catholic University Leuven, Leuven, Belgium

[18]Melbourne Neuropsychiatry Centre, The University of Melbourne, Melbourne, Australia

[19]Centre for Youth Mental Health, University of Melbourne, Melbourne, Australia

[20]Department of Psychiatry and Psychotherapy, Medical University of Vienna, Vienna, Austria

[21]University Paris Descartes, Hôpital Sainte-Anne, C'JAAD, Service Hospitalo-Universitaire, Inserm U894, Institut de Psychiatrie, Paris, France

[22]Department of Psychiatry and Neuropsychology, School for Mental Health and Neuroscience, Maastricht University Medical Centre, Maastricht, The Netherlands

[23]Department of Psychiatry and Psychology, Maastricht University Medical Center, Maastricht, The Netherlands

Conflicts of Interest

None declared.

Multimedia Appendix 1
Convergent validity of momentary manifestations of negative symptoms and interviewer-rated measures of negative symptoms.
[DOCX File , 17 KB - mental_v8i11e30309_app1.docx]

Multimedia Appendix 2
Data quality of clinical outcome measures.
[DOCX File , 13 KB - mental_v8i11e30309_app2.docx]

Multimedia Appendix 3
Unadjusted analyses.
[DOCX File , 28 KB - mental_v8i11e30309_app3.docx]

Multimedia Appendix 4
Sensitivity analysis with current depressive episode as an additional independent variable to control for potential confounding.
[DOCX File , 31 KB - mental_v8i11e30309_app4.docx]

Multimedia Appendix 5

Sensitivity analysis with comorbid Axis-I disorder as an additional independent variable to control for potential confounding.
[DOCX File , 29 KB - mental_v8i11e30309_app5.docx]

Multimedia Appendix 6
Comorbid Axis-I diagnoses at baseline.
[DOCX File , 15 KB - mental_v8i11e30309_app6.docx]

Multimedia Appendix 7
Restricted analyses.
[DOCX File , 23 KB - mental_v8i11e30309_app7.docx]

References

1. Azar M, Pruessner M, Baer LH, Iyer S, Malla AK, Lepage M. A study on negative and depressive symptom prevalence in individuals at ultra-high risk for psychosis. Early Interv Psychiatry 2018 Oct;12(5):900-906. [doi: 10.1111/eip.12386] [Medline: 27653624]
2. Fulford D, Niendam TA, Floyd EG, Carter CS, Mathalon DH, Vinogradov S, et al. Symptom dimensions and functional impairment in early psychosis: more to the story than just negative symptoms. Schizophr Res 2013 Jun;147(1):125-131 [FREE Full text] [doi: 10.1016/j.schres.2013.03.024] [Medline: 23587696]
3. Schlosser DA, Campellone TR, Biagianti B, Delucchi KL, Gard DE, Fulford D, et al. Modeling the role of negative symptoms in determining social functioning in individuals at clinical high risk of psychosis. Schizophr Res 2015 Dec;169(1-3):204-208 [FREE Full text] [doi: 10.1016/j.schres.2015.10.036] [Medline: 26530628]
4. Svirskis T, Korkeila J, Heinimaa M, Huttunen J, Ilonen T, Ristkari T, et al. Quality of life and functioning ability in subjects vulnerable to psychosis. Compr Psychiatry 2007;48(2):155-160. [doi: 10.1016/j.comppsych.2006.10.008] [Medline: 17292706]
5. Burton CZ, Tso IF, Carrión RE, Niendam T, Adelsheim S, Auther AM, et al. Baseline psychopathology and relationship to longitudinal functional outcome in attenuated and early first episode psychosis. Schizophr Res 2019 Oct;212:157-162 [FREE Full text] [doi: 10.1016/j.schres.2019.07.048] [Medline: 31395490]
6. Glenthøj LB, Kristensen TD, Wenneberg C, Hjorthøj C, Nordentoft M. Investigating cognitive and clinical predictors of real-life functioning, functional capacity, and quality of life in individuals at ultra-high risk for psychosis. Schizophr Bull Open 2020 Jun 25;1(1). [doi: 10.1093/schizbullopen/sgaa027]
7. Demjaha A, Valmaggia L, Stahl D, Byrne M, McGuire P. Disorganization/cognitive and negative symptom dimensions in the at-risk mental state predict subsequent transition to psychosis. Schizophr Bull 2012 Mar;38(2):351-359 [FREE Full text] [doi: 10.1093/schbul/sbq088] [Medline: 20705805]
8. Piskulic D, Addington J, Cadenhead KS, Cannon TD, Cornblatt BA, Heinssen R, et al. Negative symptoms in individuals at clinical high risk of psychosis. Psychiatry Res 2012 Apr 30;196(2-3):220-224 [FREE Full text] [doi: 10.1016/j.psychres.2012.02.018] [Medline: 22445704]
9. Oliver D, Reilly T, Baccaredda Boy O, Petros N, Davies C, Borgwardt S, et al. What causes the onset of psychosis in individuals at clinical high risk? A meta-analysis of risk and protective factors. Schizophr Bull 2020 Jan 04;46(1):110-120 [FREE Full text] [doi: 10.1093/schbul/sbz039] [Medline: 31219164]
10. Valmaggia LR, Stahl D, Yung AR, Nelson B, Fusar-Poli P, McGorry PD, et al. Negative psychotic symptoms and impaired role functioning predict transition outcomes in the at-risk mental state: a latent class cluster analysis study. Psychol Med 2013 Nov;43(11):2311-2325. [doi: 10.1017/S0033291713000251] [Medline: 23442767]
11. Zimmermann R, Gschwandtner U, Wilhelm FH, Pflueger MO, Riecher-Rössler A, Fuhr P. EEG spectral power and negative symptoms in at-risk individuals predict transition to psychosis. Schizophr Res 2010 Nov;123(2-3):208-216. [doi: 10.1016/j.schres.2010.08.031] [Medline: 20850950]
12. Alderman T, Addington J, Bearden C, Cannon TD, Cornblatt BA, McGlashan TH, et al. Negative symptoms and impaired social functioning predict later psychosis in Latino youth at clinical high risk in the North American prodromal longitudinal studies consortium. Early Interv Psychiatry 2015 Dec;9(6):467-475 [FREE Full text] [doi: 10.1111/eip.12128] [Medline: 24576057]
13. Ferrarelli F, Mathalon D. The prodromal phase: time to broaden the scope beyond transition to psychosis? Schizophr Res 2020 Feb;216:5-6 [FREE Full text] [doi: 10.1016/j.schres.2019.12.035] [Medline: 31924373]
14. Simon AE, Velthorst E, Nieman DH, Linszen D, Umbricht D, de Haan L. Ultra high-risk state for psychosis and non-transition: a systematic review. Schizophr Res 2011 Oct;132(1):8-17. [doi: 10.1016/j.schres.2011.07.002] [Medline: 21784618]
15. Fusar-Poli P, Bonoldi I, Yung AR, Borgwardt S, Kempton MJ, Valmaggia L, et al. Predicting psychosis: meta-analysis of transition outcomes in individuals at high clinical risk. Arch Gen Psychiatry 2012 Mar;69(3):220-229. [doi: 10.1001/archgenpsychiatry.2011.1472] [Medline: 22393215]

16. Addington J, Liu L, Buchy L, Cadenhead K, Cannon T, Cornblatt B, et al. North American Prodrome Longitudinal Study (NAPLS 2): the prodromal symptoms. J Nerv Ment Dis 2015 May;203(5):328-335 [FREE Full text] [doi: 10.1097/NMD.0000000000000290] [Medline: 25919383]
17. Polari A, Lavoie S, Yuen H, Amminger P, Berger G, Chen E, et al. Clinical trajectories in the ultra-high risk for psychosis population. Schizophr Res 2018 Jul;197:550-556. [doi: 10.1016/j.schres.2018.01.022] [Medline: 29463457]
18. Addington J, Cornblatt BA, Cadenhead KS, Cannon TD, McGlashan TH, Perkins DO, et al. At clinical high risk for psychosis: outcome for nonconverters. Am J Psychiatry 2011 Aug;168(8):800-805 [FREE Full text] [doi: 10.1176/appi.ajp.2011.10081191] [Medline: 21498462]
19. Simon AE, Borgwardt S, Riecher-Rössler A, Velthorst E, de Haan L, Fusar-Poli P. Moving beyond transition outcomes: meta-analysis of remission rates in individuals at high clinical risk for psychosis. Psychiatry Res 2013 Oct 30;209(3):266-272. [doi: 10.1016/j.psychres.2013.03.004] [Medline: 23871169]
20. Fusar-Poli P, Rocchetti M, Sardella A, Avila A, Brandizzi M, Caverzasi E, et al. Disorder, not just state of risk: meta-analysis of functioning and quality of life in people at high risk of psychosis. Br J Psychiatry 2015 Sep;207(3):198-206. [doi: 10.1192/bjp.bp.114.157115] [Medline: 26329563]
21. Lin A, Wood SJ, Nelson B, Beavan A, McGorry P, Yung AR. Outcomes of nontransitioned cases in a sample at ultra-high risk for psychosis. Am J Psychiatry 2015 Mar 01;172(3):249-258. [doi: 10.1176/appi.ajp.2014.13030418] [Medline: 25727537]
22. Beck K, Andreou C, Studerus E, Heitz U, Ittig S, Leanza L, et al. Clinical and functional long-term outcome of patients at clinical high risk (CHR) for psychosis without transition to psychosis: a systematic review. Schizophr Res 2019 Aug;210:39-47. [doi: 10.1016/j.schres.2018.12.047] [Medline: 30651204]
23. Michel C, Ruhrmann S, Schimmelmann BG, Klosterkötter J, Schultze-Lutter F. Course of clinical high-risk states for psychosis beyond conversion. Eur Arch Psychiatry Clin Neurosci 2018 Feb;268(1):39-48. [doi: 10.1007/s00406-016-0764-8] [Medline: 28054132]
24. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr Bull 1987;13(2):261-276. [doi: 10.1093/schbul/13.2.261] [Medline: 3616518]
25. Kay SR, Opler LA, Lindenmayer J. The Positive and Negative Syndrome Scale (PANSS): rationale and standardisation. Br J Psychiatry Suppl 1989 Nov(7):59-67. [Medline: 2619982]
26. Oorschot M, Lataster T, Thewissen V, Lardinois M, Wichers M, van Os J, et al. Emotional experience in negative symptoms of schizophrenia--no evidence for a generalized hedonic deficit. Schizophr Bull 2013 Jan;39(1):217-225 [FREE Full text] [doi: 10.1093/schbul/sbr137] [Medline: 22021660]
27. Blanchard JJ, Shan L, Andrea A, Savage C, Kring AM, Weittenhiller L. Negative symptoms and their assessment in schizophrenia and related disorders. In: Badcock JC, Paulik-White G, editors. A Clinical Introduction to Psychosis. Cambridge, MA: Academic Press,; 2020:153-175.
28. Galderisi S, Mucci A, Buchanan RW, Arango C. Negative symptoms of schizophrenia: new developments and unanswered research questions. Lancet Psychiatry 2018 Aug;5(8):664-677. [doi: 10.1016/S2215-0366(18)30050-6] [Medline: 29602739]
29. Kring AM, Caponigro JM. Emotion in schizophrenia: where feeling meets thinking. Curr Dir Psychol Sci 2010 Aug;19(4):255-259 [FREE Full text] [doi: 10.1177/0963721410377599] [Medline: 22557707]
30. Cohen AS, Najolia GM, Brown LA, Minor KS. The state-trait disjunction of anhedonia in schizophrenia: potential affective, cognitive and social-based mechanisms. Clin Psychol Rev 2011 Apr;31(3):440-448. [doi: 10.1016/j.cpr.2010.11.001] [Medline: 21130553]
31. Myin-Germeys I, Kasanova Z, Vaessen T, Vachon H, Kirtley O, Viechtbauer W, et al. Experience sampling methodology in mental health research: new insights and technical developments. World Psychiatry 2018 Jun;17(2):123-132 [FREE Full text] [doi: 10.1002/wps.20513] [Medline: 29856567]
32. Mote J, Fulford D. Ecological momentary assessment of everyday social experiences of people with schizophrenia: a systematic review. Schizophr Res 2020 Feb;216:56-68. [doi: 10.1016/j.schres.2019.10.021] [Medline: 31874743]
33. Myin-Germeys I, Delespaul PA, deVries MW. Schizophrenia patients are more emotionally active than is assumed based on their behavior. Schizophr Bull 2000;26(4):847-854. [doi: 10.1093/oxfordjournals.schbul.a033499] [Medline: 11087017]
34. Dejonckheere E, Mestdagh M, Houben M, Rutten I, Sels L, Kuppens P, et al. Complex affect dynamics add limited information to the prediction of psychological well-being. Nat Hum Behav 2019 May;3(5):478-491. [doi: 10.1038/s41562-019-0555-0] [Medline: 30988484]
35. Hermans KS, Myin-Germeys I, Gayer-Anderson C, Kempton MJ, Valmaggia L, McGuire P, et al. Elucidating negative symptoms in the daily life of individuals in the early stages of psychosis. Psychol Med 2020 May 22:1-11. [doi: 10.1017/S0033291720001154] [Medline: 32438944]
36. Kwapil TR, Silvia PJ, Myin-Germeys I, Anderson A, Coates SA, Brown LH. The social world of the socially anhedonic: exploring the daily ecology of asociality. J Res Pers 2009 Feb;43(1):103-106. [doi: 10.1016/j.jrp.2008.10.008]
37. Gerritsen C, Bagby RM, Sanches M, Kiang M, Maheandiran M, Prce I, et al. Stress precedes negative symptom exacerbations in clinical high risk and early psychosis: a time-lagged experience sampling study. Schizophr Res 2019 Aug;210:52-58. [doi: 10.1016/j.schres.2019.06.015] [Medline: 31248749]

38. Aghevli MA, Blanchard JJ, Horan WP. The expression and experience of emotion in schizophrenia: a study of social interactions. Psychiatry Res 2003 Aug 01;119(3):261-270. [doi: 10.1016/s0165-1781(03)00133-1] [Medline: 12914897]
39. European Network of National Networks studying Gene-Environment Interactions in Schizophrenia (EU-GEI), van Os J, Rutten BP, Myin-Germeys I, Delespaul P, Viechtbauer W, et al. Identifying gene-environment interactions in schizophrenia: contemporary challenges for integrated, large-scale investigations. Schizophr Bull 2014 Jul;40(4):729-736 [FREE Full text] [doi: 10.1093/schbul/sbu069] [Medline: 24860087]
40. Yung AR, Yuen HP, McGorry PD, Phillips LJ, Kelly D, Dell'Olio M, et al. Mapping the onset of psychosis: the Comprehensive Assessment of At-Risk Mental States. Aust N Z J Psychiatry 2005;39(11-12):964-971. [doi: 10.1080/j.1440-1614.2005.01714.x] [Medline: 16343296]
41. Myin-Germeys I, Oorschot M, Collip D, Lataster J, Delespaul P, van Os J. Experience sampling research in psychopathology: opening the black box of daily life. Psychol Med 2009 Sep;39(9):1533-1547. [doi: 10.1017/S0033291708004947] [Medline: 19215626]
42. Palmier-Claus JE, Myin-Germeys I, Barkus E, Bentley L, Udachina A, Delespaul PA, et al. Experience sampling research in individuals with mental illness: reflections and guidance. Acta Psychiatr Scand 2011 Jan;123(1):12-20. [doi: 10.1111/j.1600-0447.2010.01596.x] [Medline: 20712828]
43. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
44. Palmier-Claus JE, Dunn G, Lewis SW. Emotional and symptomatic reactivity to stress in individuals at ultra-high risk of developing psychosis. Psychol Med 2012 May;42(5):1003-1012. [doi: 10.1017/S0033291711001929] [Medline: 22067414]
45. Overall JE, Gorham DR. The brief psychiatric rating scale. Psychol Rep 1962 Jun 1;10(3):799-812. [doi: 10.2466/pr0.1962.10.3.799]
46. Diagnostic and Statistical Manual of Mental Disorders: DSM-IV [Internet]. 4th edition. Washington, DC: American Psychiatric Assoication; 1994.
47. Guy W. ECDEU assessment manual for psychopharmacology. US Dept Health, Education, and Welfare, Public Health Service, Alcohol, Drug Abuse, and Mental Health Administration, National Institute of Mental Health, Psychopharmacology Research Branch, Division of Extramural Research Programs. 1976. URL: https://openlibrary.org/books/OL24341821M/ECDEU_assessment_manual_for_psychopharmacology [accessed 2021-10-14]
48. McManus MD, Siegel JT, Nakamura J. The predictive power of low-arousal positive affect. Motiv Emot 2019;43(1):130-144. [doi: 10.1007/s11031-018-9719-x]
49. Delespaul P, deVries M, van Os J. Determinants of occurrence and recovery from hallucinations in daily life. Soc Psychiatry Psychiatr Epidemiol 2002 Mar;37(3):97-104. [doi: 10.1007/s001270200000] [Medline: 11990012]
50. Janssens M, Lataster T, Simons C, Oorschot M, Lardinois M, van Os J, GROUP. Emotion recognition in psychosis: no evidence for an association with real world social functioning. Schizophr Res 2012 Dec;142(1-3):116-121 [FREE Full text] [doi: 10.1016/j.schres.2012.10.003] [Medline: 23122740]
51. van Winkel R, Henquet C, Rosa A, Papiol S, Fananás L, De Hert M, et al. Evidence that the COMT(Val158Met) polymorphism moderates sensitivity to stress in psychosis: an experience-sampling study. Am J Med Genet B Neuropsychiatr Genet 2008 Jan 05;147B(1):10-17. [doi: 10.1002/ajmg.b.30559] [Medline: 17525974]
52. Simes RJ. An improved Bonferroni procedure for multiple tests of significance. Biometrika 1986 Dec;73(3):751-754. [doi: 10.1093/biomet/73.3.751]
53. Lim J, Rekhi G, Rapisarda A, Lam M, Kraus M, Keefe RS, et al. Impact of psychiatric comorbidity in individuals at Ultra High Risk of psychosis - findings from the Longitudinal Youth at Risk Study (LYRIKS). Schizophr Res 2015 May;164(1-3):8-14. [doi: 10.1016/j.schres.2015.03.007] [Medline: 25818728]
54. Albert U, Tomassi S, Maina G, Tosato S. Prevalence of non-psychotic disorders in ultra-high risk individuals and transition to psychosis: a systematic review. Psychiatry Res 2018 Dec;270:1-12. [doi: 10.1016/j.psychres.2018.09.028] [Medline: 30243126]
55. Malda A, Boonstra N, Barf H, de Jong S, Aleman A, Addington J, et al. Individualized prediction of transition to psychosis in 1,676 individuals at clinical high risk: development and validation of a multivariable prediction model based on individual patient data meta-analysis. Front Psychiatry 2019;10:345 [FREE Full text] [doi: 10.3389/fpsyt.2019.00345] [Medline: 31178767]
56. Nelson B, Yuen HP, Wood SJ, Lin A, Spiliotacopoulos D, Bruxner A, et al. Long-term follow-up of a group at ultra high risk ("prodromal") for psychosis: the PACE 400 study. JAMA Psychiatry 2013 Aug;70(8):793-802. [doi: 10.1001/jamapsychiatry.2013.1270] [Medline: 23739772]
57. Corcoran CM, Kimhy D, Parrilla-Escobar MA, Cressman VL, Stanford AD, Thompson J, et al. The relationship of social function to depressive and negative symptoms in individuals at clinical high risk for psychosis. Psychol Med 2011 Feb;41(2):251-261 [FREE Full text] [doi: 10.1017/S0033291710000802] [Medline: 20444306]
58. Reininghaus U, Kempton MJ, Valmaggia L, Craig TK, Garety P, Onyejiaka A, et al. Stress sensitivity, aberrant salience, and threat anticipation in early psychosis: an experience sampling study. Schizophr Bull 2016 May;42(3):712-722 [FREE Full text] [doi: 10.1093/schbul/sbv190] [Medline: 26834027]

59. Edwards CJ, Cella M, Emsley R, Tarrier N, Wykes TH. Exploring the relationship between the anticipation and experience of pleasure in people with schizophrenia: an experience sampling study. Schizophr Res 2018 Dec;202:72-79 [FREE Full text] [doi: 10.1016/j.schres.2018.06.040] [Medline: 30007868]
60. Devoe DJ, Farris MS, Townes P, Addington J. Interventions and social functioning in youth at risk of psychosis: a systematic review and meta-analysis. Early Interv Psychiatry 2019 Apr;13(2):169-180. [doi: 10.1111/eip.12689] [Medline: 29938910]
61. Velthorst E, Nelson B, Wiltink S, de Haan L, Wood SJ, Lin A, et al. Transition to first episode psychosis in ultra high risk populations: does baseline functioning hold the key? Schizophr Res 2013 Jan;143(1):132-137. [doi: 10.1016/j.schres.2012.10.025] [Medline: 23182438]
62. van Os J, Lataster T, Delespaul P, Wichers M, Myin-Germeys I. Evidence that a psychopathology interactome has diagnostic value, predicting clinical needs: an experience sampling study. PLoS One 2014;9(1):e86652 [FREE Full text] [doi: 10.1371/journal.pone.0086652] [Medline: 24466189]

Abbreviations

BPRS: Brief Psychiatric Rating Scale
ESM: Experience Sampling Methodology
GAF: Global Assessment of Functioning
UHR: ultrahigh risk

Edited by J Torous; submitted 10.05.21; peer-reviewed by S Steingrimsson; comments to author 19.07.21; accepted 02.08.21; published 19.11.21.

Please cite as:
Paetzold I, Hermans KSFM, Schick A, Nelson B, Velthorst E, Schirmbeck F, EU-GEI High Risk Study, van Os J, Morgan C, van der Gaag M, de Haan L, Valmaggia L, McGuire P, Kempton M, Myin-Germeys I, Reininghaus U
Momentary Manifestations of Negative Symptoms as Predictors of Clinical Outcomes in People at High Risk for Psychosis: Experience Sampling Study
JMIR Ment Health 2021;8(11):e30309
URL: https://mental.jmir.org/2021/11/e30309
doi:10.2196/30309
PMID:34807831

Review

Mobile Apps That Promote Emotion Regulation, Positive Mental Health, and Well-being in the General Population: Systematic Review and Meta-analysis

Mia Eisenstadt[1,2*], PhD; Shaun Liverpool[1,2,3*], PhD; Elisa Infanti[2,4,5], PhD; Roberta Maria Ciuvat[2,4], BSc; Courtney Carlsson[2], MBA

[1]Evidence Based Practice Unit, Anna Freud National Centre for Children and Families and University College London, London, United Kingdom
[2]Paradym Ltd, Bloomsbury, London, United Kingdom
[3]Faculty of Health, Social Care and Medicine, Edge Hill University, Ormskirk, United Kingdom
[4]Division of Psychology and Language Sciences, University College London, London, United Kingdom
[5]Birkbeck, Department of Psychological Sciences, University of London, London, United Kingdom
[*]these authors contributed equally

Corresponding Author:
Mia Eisenstadt, PhD
Evidence Based Practice Unit
Anna Freud National Centre for Children and Families and University College London
4-8 Rodney Street
London, N1 9JH
United Kingdom
Phone: 44 7989165986
Email: mia.eisenstadt@annafreud.org

Abstract

Background: Among the general public, there appears to be a growing need and interest in receiving digital mental health and well-being support. In response to this, mental health apps (MHapps) are becoming available for monitoring, managing, and promoting positive mental health and well-being. Thus far, evidence supports favorable outcomes when users engage with MHapps, yet there is a relative paucity of reviews on apps that support positive mental health and well-being.

Objective: We aimed to systematically review the available research on MHapps that promote emotion regulation, positive mental health, and well-being in the general population aged 18-45 years. More specifically, the review aimed at providing a systematic description of the theoretical background and features of MHapps while evaluating any potential effectiveness.

Methods: A comprehensive literature search of key databases, including MEDLINE (via Ovid), EMBASE (via Ovid), PsycINFO (via Ovid), Web of Science, and the Cochrane Register of Controlled Trials (CENTRAL), was performed until January 2021. Studies were included if they described standalone mental health and well-being apps for adults without a formal mental health diagnosis. The quality of all studies was assessed against the Mixed Methods Appraisal Tool. In addition, the Cochrane Risk-of-Bias tool (RoB-2) was used to assess randomized control trials (RCTs). Data were extracted using a modified extraction form from the Cochrane Handbook of Systematic Reviews. A narrative synthesis and meta-analysis were then undertaken to address the review aims.

Results: In total, 3156 abstracts were identified. Of these, 52 publications describing 48 MHapps met the inclusion criteria. Together, the studies evaluated interventions across 15 countries. Thirty-nine RCTs were identified suggesting some support for the role of individual MHapps in improving and promoting mental health and well-being. Regarding the pooled effect, MHapps, when compared to controls, showed a small effect for reducing mental health symptoms (k=19, Hedges g=–0.24, 95% CI –0.34 to –0.14; P<.001) and improving well-being (k=13, g=0.17, 95% CI 0.05-0.29, P=.004), and a medium effect for emotion regulation (k=6, g=0.49, 95% CI 0.23-0.74, P<.001). There is also a wide knowledge base of creative and innovative ways to engage users in techniques such as mood monitoring and guided exercises. Studies were generally assessed to contribute unclear or a high risk of bias, or to be of medium to low methodological quality.

Conclusions: The emerging evidence for MHapps that promote positive mental health and well-being suggests promising outcomes. Despite a wide range of MHapps, few apps specifically promote emotion regulation. However, our findings may

position emotion regulation as an important mechanism for inclusion in future MHapps. A fair proportion of the included studies were pilot or feasibility trials (*k*=17, 33%), and full-scale RCTs reported high attrition rates and nondiverse samples. Given the number and pace at which MHapps are being released, further robust research is warranted to inform the development and testing of evidence-based programs.

(JMIR Ment Health 2021;8(11):e31170) doi:10.2196/31170

KEYWORDS

systematic review; MHapp, mHealth; mental health; well-being; emotion regulation; mobile apps; effectiveness; monitoring; management; mental health app

Introduction

Background

Globally, the prevalence of mental health disorders has been increasing [1]. Statistics from the United Kingdom indicate that between 16% and 21% of working adults experience mental health difficulties [2]. In 2019, a study found that there were 51.5 million North American adults with mental illness, with the highest prevalence among young adults aged 18-25 years (29.4%) compared to adults aged 26-49 (25%) or 50 years and older (14.1%) [3]. In light of these statistics, neuropsychiatric conditions continue to be among the main causes of disability globally [1]. In recent times, unfortunate events such as the COVID-19 pandemic have further exacerbated the problem, contributing to rising rates of anxiety, loneliness, and depression [1,4,5], along with decreased access to in-person support [6]. In this context, there is an increased demand for accessible and scalable mental health services (eg, mobile health and electronic health) for both the promotion of well-being and the prevention of mental disorders [7].

Potential of Mental Health Apps to Fill the Need

Delivering mental health care online has become more feasible with the rapid increase in smartphone usage. Smartphone ownership is estimated to reach over 6 billion users globally in 2021 and these numbers are expected to increase by several hundred million in the coming few years [8]. Moreover, research suggests that in August 2017, smartphone owners in the United Kingdom spent on average 62 hours per month using the internet, as compared with 75 hours in the United States and 58 hours in Germany [9]. In general, digital apps may offer users opportunities to manage their mental and physical health, and support behavior change efforts. An estimate suggested that nearly 325,000 health apps were available for users to download in 2017, with mental health apps (MHapps) constituting about one-third of disease-specific apps [10,11]. In addition, a report from Statista found that “health and lifestyle” was one of the most popular categories of apps in the App Store [12]. However, the vast majority of MHapps have not been scientifically tested [13].

The developing evidence base for MHapps suggests that apps accessed via smart devices are increasingly able to play an important role in mental health care provision [13,14]. MHapps have been researched in terms of their effectiveness for the treatment and management of mental health disorders, but they are also increasingly understood to have potential in the prevention of mental health disorders and in the promotion of positive mental health [13-15]. In particular, the dynamic multifeatured nature of MHapps provides a platform for monitoring [16], preventing [14], and reducing mental health symptoms [17]. In addition, MHapps are able to facilitate emotion regulation and enhance mental well-being [14]. Moreover, MHapps are appealing to users due to advantages of cost-effectiveness, privacy, personalization features, and scope for use at any time in any location and setting [18]. MHapps also have the potential to overcome barriers to seeking help, such as stigma, as well as to promote positive habits for improved long-term well-being and mental health outcomes [14,18]. However, scientific research has not been able to keep up with the pace of the new developments in MHapps. Consequently, a large number of apps are available without any published scientific evidence base or peer-reviewed acceptability studies [19,20].

Mental Well-being, Positive Mental Health, and Emotion Regulation

Although enhanced psychological well-being has been consistently linked to positive health and mental health outcomes, it is increasingly understood that mental health and mental well-being are separate entities with separate determinants [21,22]. The concept of mental well-being goes beyond the absence of mental health disorders and symptoms, and can address psychological parameters such as subjective well-being, autonomy, positive relationships, and personal growth [23,24]. In the same vein, positive mental health refers to “a positive emotion (affect) such as feelings of happiness, a personality trait inclusive of the psychological resources of self-esteem and mastery, and as resilience, which is the capacity to cope with adversity” [25]. Thus, both concepts overlap in highlighting that well-being is a broader concept that goes beyond the absence of mental health disorders [21]. Further, emotion regulation has been considered to be a focal point to address psychological disorders [26] and to enhance well-being [27]. Emotion regulation refers to the experience and expression of both positive and negative emotions [28,29]. Difficulties with emotion regulation are linked with increased stress [28,29], and represent an established risk factor for a range of mental health disorders such as depression [26] and bipolar disorder [30]. Emotion regulation strategies can be used in both adaptive and maladaptive ways depending on the context and the purpose. However, frequent use of maladaptive emotion regulation strategies is linked to mood disorders [30]. Therefore, being mindful or having emotional awareness is considered to facilitate emotion regulation [31], and may be considered an underlying

influencing factor to achieve positive mental health and well-being [32].

Previous Research and Systematic Reviews

Interestingly, despite the large number of apps available, the evidence of their effectiveness is not yet widely accepted. In a review of 52 commercially available anxiety apps, the authors reported that 67.3% did not include health care professionals in their creation and only 3.8% were supported by robust research [19]. Nonetheless, the growing evidence base suggests the potential efficacy of MHapps [33-35]. For example, Firth et al [34] found app-delivered interventions to be effective in decreasing anxiety (Hedges g=0.32, 95% CI 0.17-0.48) and depression (g=0.38, 95% CI 0.24-0.52) [34]. Similarly, other studies reported small to moderate effect sizes, specifically for mindfulness apps that reduced perceived stress (g=0.46, 95% CI 0.24-0.68), anxiety (g=0.28, 95% CI 0.16-0.40), and depression (g=0.33, 95% CI 0.24-0.43), and increased psychological well-being (g=0.29, 95% CI 0.14-0.45) [35]. Other reviewers corroborated this research, reporting positive findings for mindfulness apps in improving overall mental health (g=0.23, 95% CI 0.09-0.38) [36] and reducing perceived stress (g=–0.43, 95% CI –0.20 to –0.66) [37]. Other findings indicated some support for MHapps targeting alcohol disorder, sleep disorder, depression, suicidal behaviors, self-injurious thoughts/behaviors, and posttraumatic stress disorder (PTSD) [16,17]. Although this knowledge base points in a positive direction, there has been a stronger emphasis on the effectiveness of MHapps for monitoring and managing mental health disorders [16,17,33], cognitive behavioral therapy (CBT)-based MHapps, and MHapps tested within specific clinical populations [16,34,38]. With respect to evidence of the effectiveness of MHapps for the general public, McKay and colleagues [39] reviewed commercially available healthy lifestyle apps, and found that behavior change strategies mainly focused on rehearsal or practice (of new habits) and self-monitoring. A recent meta-analysis found a small effect of mindfulness MHapps for psychological well-being but found no significant effects for general well-being [35]. Building on these findings could provide support for MHapps that are underpinned by other psychological theories and highlight benefits for a broader sample of users.

Rationale for This Review

Based on the above evidence, prior reviews examining the evidence for the effectiveness of MHapps did not include emotion regulation and rarely focused on mental well-being apps [14,20,40]. However, it is increasingly recognized that MHapps can support emotion regulation and may offer an advantage for users to manage their emotional states [14]. Moreover, the effectiveness of an intervention is usually associated with the level of user engagement [41], and therefore more research highlighting the components and features of the interface and design of MHapps would be beneficial. Although there are studies emerging that provide some recommendations of features that could be included in MHapps [14], it is still unclear how these features are being incorporated and the dominant theoretical approaches applied to the design. In this review, we were particularly interested in adults (18-45 years) owing to the concerning prevalence data indicating that among US adults aged 18 years or older, less than half of the population with a mental health disorder accessed mental health services (44.8% of adults >18 years old in 2019) [3]. With smartphone ownership being the greatest among young to middle-aged adults (91%-100% aged 18-44 years), and the evidence that this age group is the most likely to access and engage with smartphone apps [8], any important findings from this review may be readily transferable.

Thus, the overall objective of this review was to provide an overview of the available evidence on MHapps that promote emotion regulation, positive mental health, and well-being in the general adult population. This review will complement and expand upon the existing systematic reviews that have focused on apps for mental health disorders by focusing on MHapps for mental well-being and positive mental health. We aimed to identify, evaluate, and summarize the findings of relevant individual studies, thereby making the available evidence more accessible to both researchers and commercially based developers. More specifically, we describe and assess the characteristics and theoretical background of the apps themselves and the studies undertaken to evaluate them. We then highlight any gaps in the current knowledge base that may require further investigation. In doing so, we aimed to address the following research questions: (1) What are the characteristics and theoretical background of MHapps designed to improve (a) mental well-being (eg, psychological, subjective, and emotional), (b) emotion regulation (eg, emotion awareness), and (c) positive mental health (eg, reduce early mental health symptoms)? (2) Is there potential for MHapps to be effective in improving emotion regulation, positive mental health, and well-being in the general population (18-45 years)?

For the purpose of this study, an MHapp was defined as a digital psychological intervention or program that can be directly downloaded onto a mobile device. MHapps aim to promote positive mental health and well-being, including a reduction in mental health symptoms such as stress, and anxiety and depression symptomology. These apps are expected to be standalone interventions serving as a form of psychological intervention by assisting the user to draw on their own capacities to facilitate behavior change, and increase psychoeducation and self-help provisions [42,43].

Methods

Design

The review was performed in accordance with the Cochrane Handbook for Systematic Reviews of Interventions [44] and is reported according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) guidelines [45]. This study protocol was registered in the International Prospective Register of Systematic Reviews of the National Institute for Health Research (PROSPERO) [46] (ID CRD42020213051).

Changes to the Review Protocol

Notably, changes were made after publication of the study protocol. We initially planned to only focus on studies that

examined emotion regulation and well-being; however, we also ultimately included studies that measured positive mental health–related outcomes. This decision was based on the fact that many papers tend to interchange terms related to mental well-being and positive mental health [35]. Owing to the limited qualitative data found in the review, we did not summarize the results of qualitative studies as per the review protocol. As the review retrieved a large number of eligible studies that included outcomes of interest, we did not examine secondary outcomes as stated previously (ie, physical health and behavioral outcomes such as improved sleep). Such outcomes would be appropriately studied in a separate review and meta-analysis. Other changes included the use of the Mixed Methods Appraisal Tool (MMAT) instead of ROBINS-I to perform the quality assessment of the identified studies. MMAT was selected as it provided a single tool to assess methodological quality criteria for different study designs [47]. We initially intended to perform subgroup analysis based on age, gender, and ethnicity of the samples or the underpinning theory of the MHapp; however, owing to the high levels of heterogeneity, we were unable to perform these analyses. Therefore, we instead captured some key differences in the narrative synthesis (eg, the overrepresentation of groups with specific demographic characteristics).

Search Strategy

A systematic literature search was completed using the following 5 electronic databases: MEDLINE (via Ovid), EMBASE (via Ovid), PsycINFO (via Ovid), Web of Science, and Cochrane Register of Controlled Trials (CENTRAL). The key terms for the intervention type (mobile apps) and the outcome themes (well-being OR emotion regulation OR mental health) were searched in the title, abstract, keywords, and, when available, subject headings. In addition, study type terms (including randomized controlled trials [RCTs] and before-and-after studies) were searched in the full text. Queries for each key area were combined with a logical “AND” operator, and were adapted to the different syntax and technical support of each individual database (see Multimedia Appendix 1). We aimed to identify records matching our inclusion criteria that were published between January 2008 and January 2021. We selected 2008 as the starting point in correspondence with the year when apps were first available to download on smart devices [33]. In addition, the bibliography of the relevant reviews and included studies were manually searched to identify additional publications for inclusion.

Eligibility Criteria

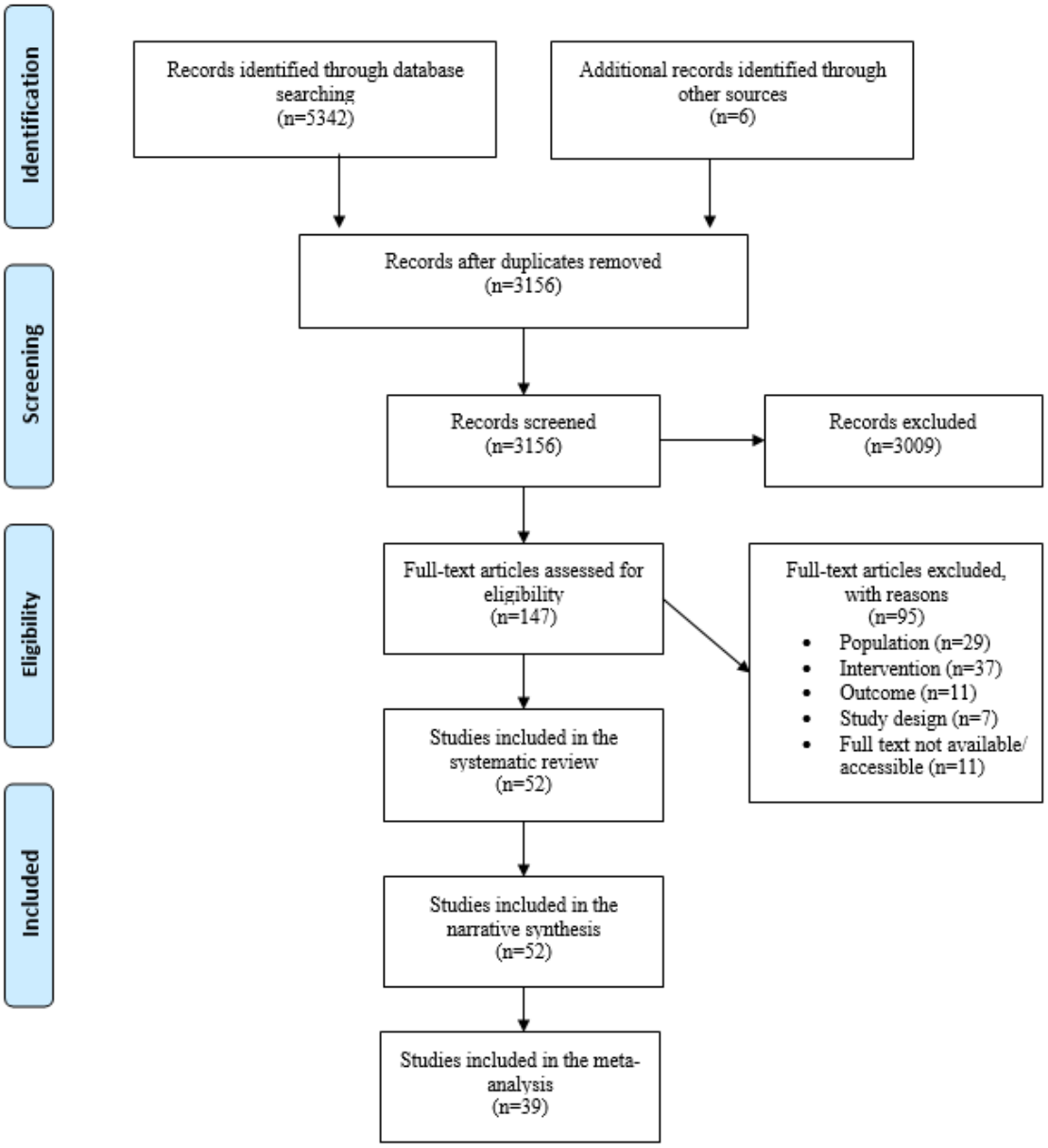
We included (a) both qualitative and quantitative experimental (eg, any type of RCT, controlled before-after) or quasiexperimental (eg, one-group pretest-posttest design, time-series) studies; (b) studies investigating the effects of standalone psychological interventions focused on promoting the outcomes of psychological, mental, or emotional well-being, promoting emotion regulation and positive mental health; (c) studies in which interventions were delivered via a digital app accessed via smartphones or tablets, or other portable devices. In addition, studies were included if they were (d) targeting adults in the age range of 18-45 years or interventions that partially overlapped with the target population where the mean age of the participants fell between 18 and 45 years. Moreover, only (e) peer-reviewed studies and (f) those published in English were considered.

We excluded records focusing on a physical characteristic (eg, weight loss, physical activity, tracking alcohol consumption) as a primary outcome or those focusing on diagnosis or assessment only. Studies reporting digital interventions delivered in conjunction with in-person interventions or that focused on the evaluation of in-person therapies including a digital component or online services for scheduling/booking appointments were also excluded. Similarly, telehealth interventions such as therapy delivered by phone, text message, video platform, or personal computer were excluded. Owing to the focus of this study on the general population, we also excluded records focusing on diagnosed health disorders (eg, PTSD, schizophrenia, major depression) and neurodiverse conditions (eg, dyslexia and autism spectrum disorders).

Study Selection

The results of the searches were downloaded and imported into the online tool CADIMA [48] for duplicates removal and study selection. All studies were independently screened by three reviewers in two stages. In the first instance, records were screened based on the inclusion/exclusion criteria, using the title and abstract. Subsequently, the full text was retrieved for the eligible articles and a full-text screening was performed. Disagreements arose for <8% of the records in both stages of the screening process. Disagreements centered around whether to include interventions that incorporated an in-person training or in-person therapy component, or what cutoff on various anxiety and stress levels was accepted for the general population. It was agreed that mild to moderate symptoms of anxiety and depression could be included but not severe symptoms. Consequently, studies including participants with scores above a clinical threshold were excluded. These disagreements were resolved through consultations with independent experts and through discussions at weekly meetings convened for the purpose of the review. Figure 1 presents the PRISMA diagram displaying the flow of records throughout the selection process.

Figure 1. Prisma diagram of studies selected for inclusion in the systematic review and meta-analysis.



Quality Assessment

First, MMAT (v2018) [47] was used to assess the methodological quality of each selected study. This tool was selected based on its capacity to assess both mixed methods and quantitative studies. The tool was recently updated in 2018 and has shown evidence of good interrater reliability, usability, and content validity [47]. Studies were rated on a categorical scale as "no," "can't tell," or "yes" to indicate whether they met the methodological quality criteria assessed. The number of items rated "yes" was counted to provide an overall score out of a possible 5 [47], with a higher number corresponding to stronger methodological quality. Second, the Cochrane Collaboration Risk of Bias tool (RoB-2) was applied to the identified controlled trials [49]. Each RCT was assessed for bias against six domains (ie, random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, and selective outcome reporting). Each domain was ranked as low risk, high risk, or unclear risk of bias. This two-fold approach was seen as a strength to present and describe the current state of the available evidence. Publication bias was assessed via visual inspection of the funnel plot asymmetry. At least two reviewers independently performed the quality assessments. The rate of agreement between the two reviewers was 56.8%. The two reviewers discussed all discrepancies and, when necessary, consulted a third team member to reach a final decision.

Data Extraction

Data were extracted using an adaptation of the data extraction form from the Cochrane Handbook of Systematic Reviews [50]. We extracted information relating to study characteristics (eg, title, author, year of publication, aim of study, study design, type of data); participant demographic details (eg, population, setting, sample size, age, gender, ethnicity); intervention and comparator details (eg, theoretical basis, features, duration);

and overall findings of the outcome measures to address the aims and objectives of this study. Two reviewers independently extracted the data and where discrepancies arose, a consensus was reached through discussions.

Synthesis and Aggregation of Data

Narrative Synthesis

First, we adopted techniques from a content analytic approach [51] looking across studies to evaluate the heterogeneity of the data, similarities and differences within and between the literature and the corresponding apps, and to identify patterns or gaps within the literature. A narrative synthesis, as a textual drawing together of the findings of the studies' and apps' characteristics, was then performed following guidance from Popay et al [51]. Where applicable, a descriptive numerical summary is also presented to group similar articles and provide an overview of the available evidence.

Statistical Data Analysis

RevMan [52] was used to calculate pooled effect sizes and their 95% CIs, and to generate the corresponding figures. Guided by the Cochrane handbook [50], between-group standardized mean differences were calculated based on posttest means, SDs, and sample sizes for each of the outcomes of interest (ie, emotion regulation, well-being, and mental health). We used Hedges *g* as an index of the effect size, allowing us to include sufficiently similar outcome measures and provide adjustment for studies with small sample sizes. In agreement with the registered protocol, we included postintervention period data but excluded follow-up period data, as not all studies included a follow-up assessment, and when this was performed, it varied extensively between studies. The effect sizes were conventionally considered as small (0.2), medium (0.5), or large (0.8). Since considerable heterogeneity was expected, we used random-effects models for all analyses. Higgins I^2 was used as a measurement for heterogeneity [53], which was categorized as low (0-40%), moderate (30%-60%), substantial (50%-90%), or considerable (75%-100%). Sensitivity analysis was performed to test the influence of outliers and the inclusion of similar-type MHapps (eg, similar features or purpose). When necessary, we contacted primary authors for further information or used the RevMan calculator to convert the relevant data. R software using the "meta" packages were employed to estimate and account for publication bias. To estimate the risk of publication bias, funnel plots were generated and Egger statistical tests were performed for analyses with an adequate number (k>10) of studies [50]. When appropriate, the "trim and fill" procedures were applied to impute potential missing data and provide an adjusted estimate effect [54].

Results

Included Studies

A total of 52 articles [42,43,55-104] (see Multimedia Appendix 2) describing 48 interventions, published between 2008 and 2020, met the inclusion criteria. Together, the studies evaluated interventions across 15 distinct countries with a total of 22,090 research participants. Studies were mainly performed in the United States (13/52, 25%), United Kingdom (6/52, 12%), and Australia (5/52, 9.6%), with fewer studies performed in low- and middle-income countries such as Brazil (1/52, 2%) or Iran (1/52, 2%). Several studies (10/52, 19%) drew participants from a range of countries through online recruitment. Demographic profiles of the participants varied across study type and settings with the mode number of studies including participants who were younger (18-25 years; 18/52, 35%), female (44/50, 88%; gender was not reported in two studies), and of White ethnicity (18/21, 86% of the studies reporting ethnicity data). Several studies adopted an RCT study design (39/52, 75%) or a nonrandomized study design (11/52, 21%), and only a few mixed method studies were found (4/52, 8%). Study participants were mainly recruited from universities (17/52, 33%), online (15/52, 29%), or workplaces (7/52, 13%).

The rest of the Results section is organized according to the emerging evidence on mental well-being (k=5), mental health (k=10), emotion regulation (k=1), or any combination of these outcomes (k=36). This is followed by a summary of the theoretical underpinnings of the identified MHapps and an overview of the range of technological features deployed within the MHapps. Lastly, we present the results of meta-analyses that integrate the available research on the effectiveness of the identified MHapps.

Summary of the Emerging Research Evidence

Most of the studies adopted a randomized study design (k=39) and incorporated a combination or subsample of outcome measures to test effectiveness of the MHapp. Taken together, the methodological quality varied across studies (see Multimedia Appendix 2) and study samples were mainly recruited from educational settings or online. Eligibility to participate in MHapp research generally required ownership of a smart device and access to the internet. Intervention periods also varied vastly between studies (eg, 12 days versus 12 weeks). Studies captured app usage/engagement data in a range of ways, with objective app usage data only provided in less than half of the total studies (22/52, 42%). Lack of generalizability of the findings and low participant adherence to MHapp usage or the study protocol were commonly reported as limitations in the individual studies. Nonetheless, the individual studies generally reported significant positive findings for reducing mental health symptoms and/or promoting well-being or emotion regulation. A detailed overview of the emerging evidence is provided in Multimedia Appendix 3.

Summary of the Theoretical Underpinnings of MHapps

The majority of MHapps covered in the reviewed studies were based on the theoretical principles of mindfulness, CBT, acceptance and commitment therapy (ACT), or a combination of any of these approaches, such as mindfulness-based resilience training that draws on mindfulness and ACT. In addition, positive psychology principles such as encouraging users to practice gratitude, recognize strengths, and engage in positive activities informed the design of several MHapps with the wider aim of promoting positive mental health [56]. A psychological theory specific to the treatment of particular disorders was applied in specific apps, such as providing CBT relevant to the treatment of depression or prevention of depressive thinking

styles, or addressing beliefs linked with low self-esteem [42]. A detailed textual overview of the theories that are applied within the range of MHapps is provided in Multimedia Appendix 4.

Technological Features

Mood Monitoring

The majority of MHapps included a mood monitoring feature where the app collects data on the user's mood (*HeadGear, Mood Prism, MoodKit, Catch it, Pacifica, MoodHacker, Wildflowers*). Mood monitoring involved users either selecting their mood on a scale, choosing from a menu of emotions, manually inputting their mood into the app (eg, *MoodPrism*), or identifying their mood on a map (*MoodMap*) [99]. Some apps also recorded the situation where the mood was felt, the time of day, and the strength of the mood on a scale. Several MHapps also provided opportunities for users to journal or record diary entries, either responding to prompts or in free form (eg, *Oiva* and *MoodKit*) [99].

Assessments

Psychological assessments were built into several MHapps to enable the collection of outcome data, to inform the prescription of specific activities and exercises, and to provide data so that the user could see changes in their mood over time (*MoodPrism,MoodKit*) [99]. Some MHapps tracked mood and then provided an exercise or meditation (eg, *Wildflowers*). In addition, some MHapps included a risk calculator (eg, *HeadGear*) that would collect data and give users a risk score for the risk of developing a mental health condition [82].

Guided Meditations and Breathing Exercises

A range of studies examined MHapps that contained prerecorded mindfulness meditations, provided via an audio recording or a video clip ranging between 10 and 30 minutes, that aimed to increase the user's capacity to practice meditation through tutorials or practice (eg, *Calm,Headspace, VGZ Mindfulness coach, It's Time to Relax!, Wildflowers*). In the absence of explicit guided meditations, apps incorporated exercises to encourage users to focus their attention on their breathing alongside other features (eg, *Tactical Breather, the wellbeing mobile app*) [65].

Psychoeducation

Psychoeducation was delivered through in-app video, audio, and written content. Following watching or listening to psychoeducational content, users were provided with an activity, quiz, or challenge. For example, *HeadGear* set a challenge that involved planning a values-based activity or a positive activity. Many MHapps prescribed specific strategies in response to a mood that was provided via text (eg, *Jibun kiroku*) or through a suggested mindfulness exercise (eg, *Stop, Breathe, Think*). Other apps set specific social challenges to encourage users to build social connections to expand their social network or practice acts of kindness (eg, *Nod*) [64].

Narrative Storytelling and Gamification

Some MHapps (eg, *Equoo)* used creative approaches such as storytelling. In some cases, a fictional character would convey psychoeducation or key app concepts [100]. Several apps used the metaphor of a journey and provided users with challenges or tasks (eg, *MoodMission*) [99]. One app included real-life stories of hope selected by researchers to provide examples centered around overcoming adversity [58]. Gamification was applied to promote engagement and provide rewards such as "virtual coins" for "level completion." Gamification refers to the use of mechanisms and game-based thinking to engage users and encourage action and problem-solving [105]. In one app, gamification was used to rate the advice the user received through the app, with the user assigning points called "sprouts" (*Spring*) [101]. Several apps included quizzes to test a user's knowledge following completing a psychoeducation module (eg, *GG Self Esteem*) [61]. Additional gamified features included earning stickers (eg, *Living with Heart*) or imaginative aspects such as placing positive messages into a virtual bottle within the app (eg, *Feel Stress Free*) [42].

In-App Notifications and Other Features

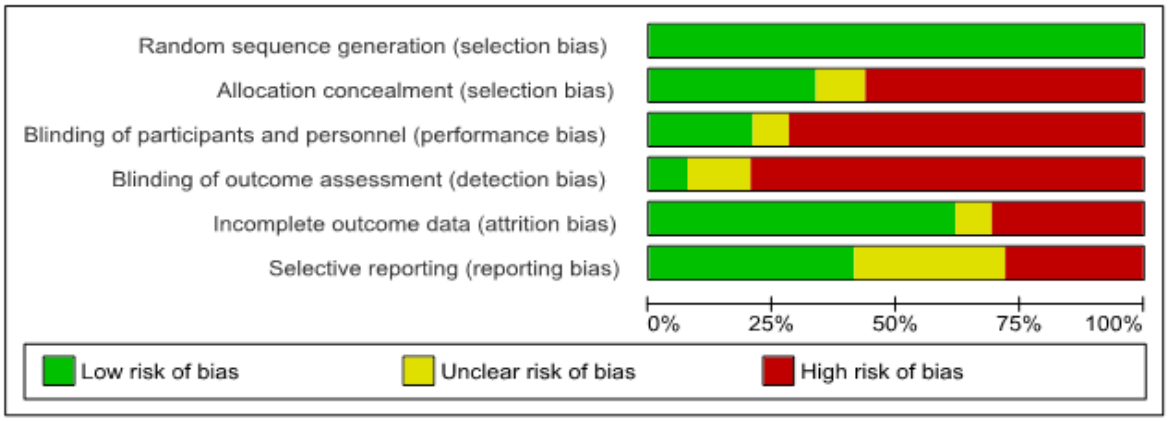
In-app notifications were applied in many MHapps at scheduled times and at random (eg, *Act Daily)*, which included reminders to complete psychological assessments [68]. In one app, notifications were sent containing messages of hope to promote well-being [58]. In other apps (eg, *OL@-OR@)*, tips for reducing stress and eating healthily were given via notifications [59]. Some MHapps had a virtual assistant to guide users around the app and make activity recommendations (eg, *Feel Stress Free*). Others featured a conversational agent, also described as a chatbot, to provide support to the user to adopt CBT strategies (eg, *Shim*) [56]. Although the focus of this review was on MHapps that aimed to improve emotion regulation, mental well-being, and mental health, there were some MHapps that also used sensors and trackers to detect sleep patterns (eg, *Kelaa Mental Resilience*). Other features included the analysis of the moods based on choice of music [102].

Effectiveness of MHapps

Risk of Bias

According to the Cochrane Risk of Bias assessment RoB-2 [49], among the 39 RCTs (including 2 mixed methods studies), 8 (21%) studies were scored as low risk in one out of six domains, 9 (23%) were scored as low risk in two out of six domains, 12 studies (31%) were rated as low risk in three out of six domains, and 5 studies (13%) were rated as low risk in four out of six domains; 4 (10%) studies were low risk in five out of six domains and 0 studies were assessed as low risk in all six domains. Studies generally were rated as high risk for allocation concealment (22/39, 56%) and blinding domains (25/39, 64%), and provided little information when reporting findings and were therefore rated unclear (19/39, 49%). All studies used an adequate randomization strategy and were thus rated as low risk. Eight studies explicitly mentioned the blinding procedures of participants and personnel. The most common method applied was random number generation via computer software; 23 studies explicitly mentioned allocation concealment. Figure 2 provides a summary of the general risk of bias of the sample of the RCTs considered for the meta-analyses. See Multimedia Appendix 5 for a summary of each study's detailed risk of bias evaluation.

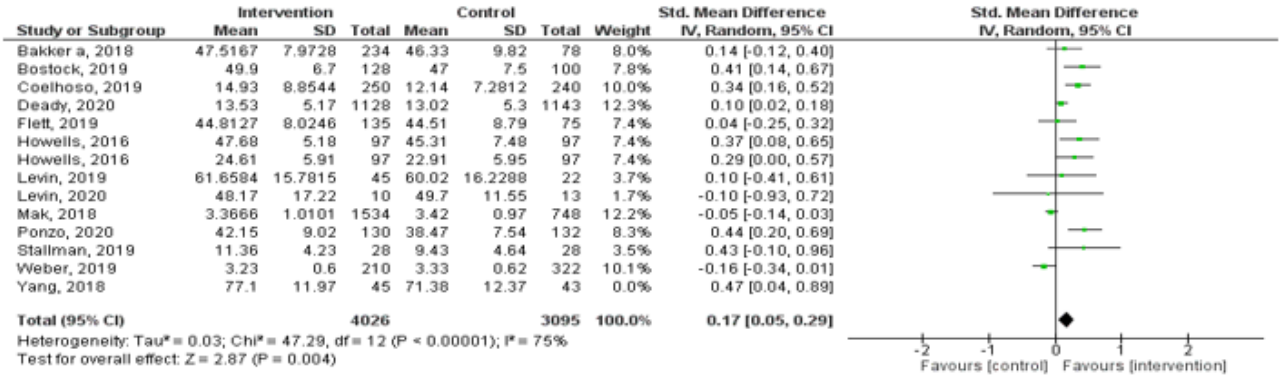
Figure 2. Cochrane Risk of Bias summary of randomized controlled trials included in the meta-analysis.



Well-being

Using data from 13 studies (Figure 3), we compared the effects of MHapps with any control group (eg, assessment only, waitlist, treatment as usual, or active control groups). The meta-analysis revealed a very small pooled effect size (g=0.17, 95% CI 0.05-0.29, P=.004) in favor of MHapps having the potential to improve well-being. However, a considerable amount of heterogeneity was present (I^2=75%). Five studies were excluded based on our planned sensitivity analysis. Two studies were judged as outliers [71,88]. Three studies were judged to be dissimilar in purpose and/or features of the MHapp or outcome measures, or had insufficient data available [58,74,89]. The Egger test was not significant and therefore trim-and-fill procedures were not applied [44].

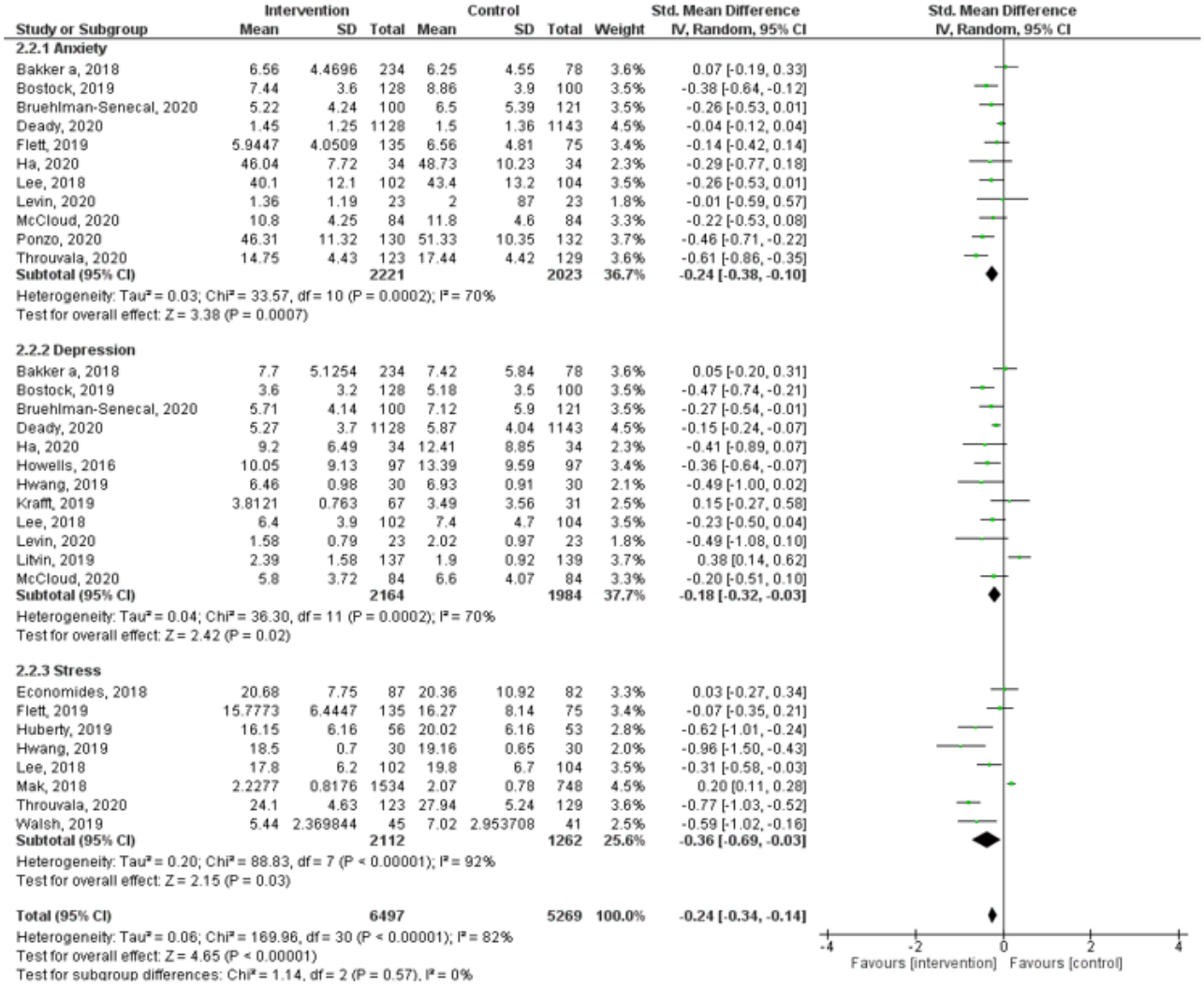
Figure 3. Forest plot showing the effect of MHapps vs control conditions for mental well-being.



Mental Health

Using data from 19 studies (Figure 4) revealed an overall small pooled effect (g=–0.24, 95% CI –0.34 to –0.14, I^2=82%, P<.001) with a considerable amount of heterogeneity, indicating the potential of MHapps to lower any of the mental health symptoms (ie, anxiety, depression, stress) when compared with controls. Subgroup analyses indicated that MHapps had a small pooled effect on reducing stress (g=–0.36, 95% CI –0.69 to –0.03, I^2=92%). For reducing anxiety symptoms, MHapps had a small pooled effect (g=–0.24, 95% CI –0.38 to –0.10, I^2=70%), and for reducing depression symptoms, MHapps had a very small pooled effect (g=–0.18, 95% CI –0.32 to 0.03, I^2=70%). However, substantial or considerable heterogeneity was observed among the studies. Eight studies were excluded based on our planned sensitivity analysis. Four studies were judged as outliers [72,80,81,88]. Four studies were judged to be dissimilar in purpose and/or features of the MHapp or outcome measures used, or had insufficient data available [67,71,74,98]. An examination of the funnel plot and a significant Egger test indicated a high level of potential publication bias. After trim-and-fill procedures were applied, adding 17 potential studies (see Appendix 6, Figure S3), the overall pooled effect of MHapps for reducing mental health symptoms was no longer significant (g=0.08, 95% CI –0.17 to 0.34, P=.50).

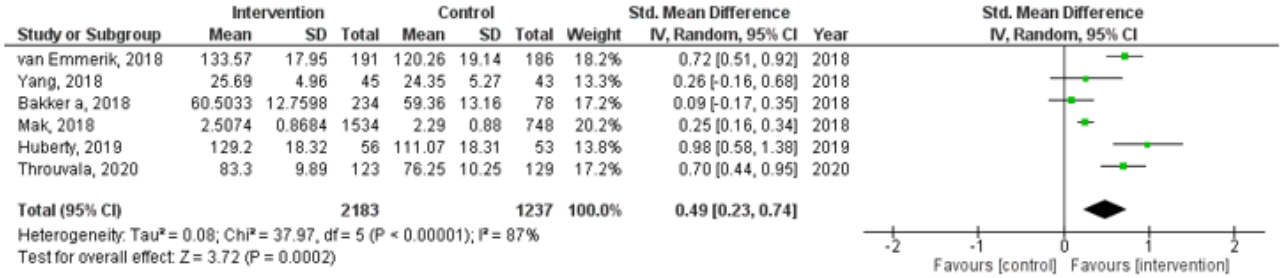
Figure 4. Forest plot of effect of MHapps versus a control condition on anxiety, depression, and stress.



Emotion Regulation

Using data from 6 studies (Figure 5) revealed a medium effect (g=0.49, 95% CI 0.23-0.74) in favor of MHapps having the potential to improve emotion regulation over control conditions. A considerable amount of heterogeneity was also present (I^2=87%, P<.001). This analysis included less than 10 studies (k=6) and therefore the Egger test was deemed inappropriate [50]. Owing to methodological weaknesses and unreliability of their estimates, trim and fill was also not attempted [44].

Figure 5. Forest plot showing the effect of MHapps versus a control condition on the outcome of emotion regulation.



Other Important Findings

None of the reviewed studies reported well-being and mental health apps to be harmful or to have any kind of adverse effect on users. Similarly, we observed that none of the reviewed studies reported on cost-effectiveness.

Discussion

Summary of Findings

This review documents the emerging evidence on available MHapps that promote emotion regulation, well-being, and mental health in the general population. Fifty-two articles

describing and evaluating 48 MHapps met the inclusion criteria, with a total of 39 RCTs identified. Overall, the review found that there is growing evidence and support for the role of MHapps in improving and promoting positive mental health, emotion regulation, and well-being in the general population. Results from the meta-analyses of RCTs indicated significant small effects of MHapps compared to control conditions for well-being, symptoms of stress, depression, and anxiety. The meta-analysis also indicated a significant medium effect of MHapps compared to control conditions for emotion regulation. There is also a foundation of theoretically based empirical studies documenting creative and innovative ways to engage users; mindfulness and cognitive behavioral approaches appear to be the most common among app developers.

Based on the reviewed studies, the evidence on mobile apps to promote emotion regulation, well-being, and mental health in the general population is still in its infancy. A fair proportion of studies were pilot or feasibility trials (17/52, 33%), and full-scale RCTs reported high attrition rates and nondiverse samples, limiting the extent to which the findings could be generalized. The evidence is also limited, with few studies on MHapps specifically for emotion regulation in the target population, and on safety and cost-effectiveness related to mobile apps, which are important components of digital interventions. Moreover, heterogeneity was generally substantial and large confidence intervals were observed in many studies. In the same vein, the lack of qualitative data limits our understanding of in-depth user experience. Lastly, techniques and methods used in MHapp research varied vastly in terms of intervention period, adherence measurements, and recruitment strategies, with implications for pooling the findings to assess overall effectiveness.

Comparison to Other Reviews

This study adds to the review performed by Wang et al [16] suggesting that MHapps have potential to be effective in monitoring or improving symptoms of certain mental disorders. The findings also extend recent findings by Gál et al [35] that suggested the potential for mindfulness MHapps to improve well-being. Specifically, our findings add that the potential may extend beyond mindfulness-focused MHapps, while also highlighting the potential for success within a specific target audience (18-45 years old). Our findings also align with previous evidence suggesting that apps are important for mood management, improving mental well-being, and life satisfaction through better management of emotions [14]. In the absence of imputed data, the current findings also suggested small to medium pooled effect sizes for mental health, including anxiety, depression [17], and stress [17,106]. Similarly, small pooled effects were repeated for well-being [35]. Among MHapp research, heterogeneity appears to be consistently high [39], as was repeated in this study.

As per the systematic review of distant mood monitoring apps by van der Watt et al [107], participants included in the present review were mostly female. In the previous review, the authors suggested that this gender bias is related to the higher rates of depressive and bipolar disorders in women [107]. However, an in-depth understanding of why samples tend to have a majority of female participants would be important future steps to meet the needs of men and other genders who may be currently underserved [107]. In accordance with other reviews, our findings also highlighted that none of the identified MHapps was harmful to users [108].

Interpretation of the Findings

Given that the extant literature shows that the effectiveness of the vast majority of available MHapps is not well supported by evidence-based research [33], this review has provided evidence of a range of mental well-being, mental health, and emotion regulation apps that are supported by the literature. There is substantial empirical support for MHapps that apply theoretical insights from CBT and mindfulness approaches. Mood monitoring in various forms is also well-supported empirically, as was ecological momentary assessment and the value of users recording moods as part of their daily lives. Other approaches are much less supported from a small number of empirical studies and warrant further research.

This review found a lack of empirical studies investigating the effectiveness of apps for emotion regulation for adults. Although there is substantial literature addressing the impact of virtual reality emotion regulation interventions for well-being and emotion regulation in adults, and video games in children [109,110], the effectiveness of MHapps that aim specifically to support emotion regulation in adults has not been systematically investigated. Notwithstanding the substantial amount of heterogeneity, this review found a medium effect of MHapps on emotion regulation, suggesting a promising area for further exploration.

Interestingly, within the reviewed MHapps, some strategies employed for well-being and mental health outcomes are also relevant to emotion regulation. For example, cognitive reappraisal is a common emotion regulation strategy [111]. Cognitive reappraisal was taught via psychoeducation in several apps, but emotion regulation was rarely a primary outcome in the studies or among the aims of the included MHapps. Similarly, mood monitoring has been found to increase emotional self-awareness, and there is some evidence to suggest that a low level of self-awareness is a risk factor for anxiety, depression, and stress [26,112]. Thus, increasing emotion regulation has been explored as an approach to improve mental health and reduce the risk of mental health disorders [112].

From the current evidence, beyond mood monitoring and in-app meditations, it is difficult to say which specific features are the most effective in MHapps. Additional features such as gamification, use of virtual assistants, rewards, and chatbots were also observed in some of the studies, albeit these were assessed in a relatively less systematic manner. Moreover, some apps adopted innovative approaches such as “crowdsourcing” therapeutic advice, in-app rating of the advice received, and use of music to assist users with moods. These less-studied approaches would also benefit from further research. In particular, the use of mixed methods designs could prove particularly helpful for illustrating the limitations of a particular feature or the limitations of technology in general as a source of provision of psychological support.

Implications for Research

It is clear that app content, design, and features, including the underpinning theory of change and the mode of delivery, will influence engagement with the app and, by extension, its impact on an individual. We agree with Hollis et al [113] that this can make it difficult to judge whether outcomes of a study are associated with the intervention content and theory of the change, the digital delivery platform, or an interaction between the two. It may be recommended that implementation science account for these dynamic and sometimes complex designs, and include components that are capable of evaluating individual differences. Equally, it is important that studies seek to understand the mechanisms such as particular features or theories that contribute to changes in user outcomes in further research.

Adherence and retention continue to be challenges to the quality of research, with little or no information about reasons for dropouts given across studies. Research designs may also need to adapt to capture such information that might be needed to inform future trials of similar apps. For example, follow-ups on dropouts and inclusion of insights from qualitative data might provide vital information. Although some studies reported objective app use data (eg, [68]), other studies relied on user reports of app usage. It is anticipated that app research would benefit from reporting more objective app use data. Another important point to consider is the high heterogeneity in the measure of retention/adherence across different MHapps. In the absence of standardized measures, researchers should aim to justify the choice of specific adherence/retention measures and how they can best be adapted to the app designs. For example, adhering to an intervention for a certain period (eg, 2 weeks or 1 month) may be different from completing a set number of modules or tasks. Similarly, usage of 10 minutes per day compared to "logging on" at least once during the intervention period could impact how results are interpreted. These variations may further implicate how findings can or should be pooled together to inform future reviews to assess overall effectiveness.

Alongside a general need for more RCTs and qualitative or mixed methods studies, this review identified a dearth of evidence on apps aiming to improve emotion regulation (despite mood management being integrated within many apps). Another finding of this review was an absence of diverse samples in included studies; thus, researchers should aim to address the lack of variance in ethnicity and socioeconomic status of the populations included in the current literature. In particular, studies should focus on minority populations, and should be performed in both low- and middle-income countries to aid with generalizability and to identify any differences in MHapp effectiveness or engagement for users from different demographic backgrounds.

Implications for App Development

The increased use of digital devices to support mental health suggests that MHapps are likely to become a relevant aspect of a proactive mental health and well-being model in the next few years. There is substantial positive support for the model of an app that captures the user's moods or emotional states and then provides information, and there is support for apps that provide a short, regular meditation session (eg, *Calm*). It is recommended that developers continue to implement these strategies with a focus on engagement. We agree with Huberty et al [114] that it is helpful to take the different stages of user engagement into account as part of app development (eg, mood check-ins). Nonetheless, some features continue to be underexplored and underused (eg, chatbots, diaries, games, storytelling, rewards, crowdsourcing, avatars, personalization); such features may have the potential to be effective in other digital interventions. In particular, there is potential for the development of features (eg, notifications and gamification) that can improve retention in both app use and studies.

Moreover, despite the proliferation of MHapps, a wider range of psychological theories (eg, attachment theory) could be explored and incorporated to broaden our understanding of applying technology to achieve improved outcomes and therapeutic change via MHapps. It may be just as important to include researchers and other relevant stakeholders in the early stages of the design process to ensure that the prototypes are useful and usable for a wider audience and will be recommended or endorsed by experts. Equally, it is also important to continue to include end users as part of the human-computer interaction approach to app development.

Strengths and Limitations

This review followed established guidelines for performing systematic reviews [50]. The protocol for the review was published on PROSPERO and search terms were reviewed with an experienced university librarian at University College London. As recommended, the screening, quality assessment, and data extraction processes were undertaken by at least two independent reviewers and checked by a senior researcher to reduce the chance of selection bias or omitting any relevant information. Although several measures were in place to ensure that a rigorous systematic review process was undertaken, this review is not devoid of limitations. This study may have been limited by the search terms used. There is a wide range of terminology used to describe apps; multiple definitions of well-being, mental health, and emotion regulation; and many well-being mediators. There were also challenges in separating standalone apps versus human-mediated interventions, as this was not always clearly stated or we were not aware of the extent to which phone calls, emails, and text messages were related to the study design. Similarly, although we excluded studies and MHapps targeting formally diagnosed mental health clients, an objective definition of "general population" is not straightforward. As a result, we could have unknowingly failed to identify and include relevant articles.

Lastly, studies were judged to be of varying quality across the different categories and of high risk of bias. Based on the findings of this review, several other sources of bias may also need to be considered in future studies. For example, the differences in sample demographics should be incorporated, which may not always be discernible from published results, but may contribute to differences in response rates. As a result, our findings should be interpreted in light of these shortcomings.

Conclusions

This systematic review focused on the relatively less studied but important domain of MHapps for mental well-being, positive mental health, and emotion regulation. The review found 52 publications describing 48 MHapps. When pooled, the MHapps demonstrated a small effect for reducing mental health symptoms and improving well-being and a medium effect for increasing emotion regulation. The findings of the meta-analysis suggest that MHapps have potential to assist users to manage mental health symptoms, boost well-being, and foster emotion regulation. Therefore, MHapps may be an important source of mental health care in the current climate of increased rates of mental health disorders and poor well-being.

In terms of the current state of the evidence for MHapps, existing research indicates some benefits to app usage, and several creative and innovative ways to engage and reward users over time. Such features enable users to learn and apply psychoeducational content, learn mindfulness or other techniques, as well as to input changes in moods and emotions. However, there remain some areas for further development within the evidence base. The body of knowledge could benefit from more large-scale RCTs and qualitative research with diverse research samples, consistent and standardized approaches to measurements (eg, reporting objective app use data), further granularity on which features are effective and how, and a specific focus on apps that support users with emotion regulation. More evidence-based apps incorporating multiple psychological theories and other innovative modes of delivery are also welcomed. Researchers, app developers, end users, and other relevant stakeholders should continue to work together to ensure that the apps are not only effective but also useful, usable, safe, cost-efficient, and sustainable over time.

Acknowledgments

The authors thank Athina-Marina Metaxa, Narissa Hanim, and Eugenia Cavarretta (all undergraduate BSc Psychology volunteers) for their significant contribution to the data extraction and quality appraisal processes. We thank Dr Bettina Moltrecht for discussions in the early stage of the research. We also would like to thank Chloe Austerberry for statistical guidance in the later stages of the study. This review was funded by Paradym Ltd.

Conflicts of Interest

The authors declare no direct financial gains from carrying out this study. However, it must be noted that all five authors have varying degrees of association with Paradym Ltd, the funder of this work and an organization that developed Paradym, which is an app for supporting well-being (not included in the current review). CC is the founder and CEO of Paradym Ltd and ME is employed as the Research Lead. Apart from commissioning a call for evidence in this area, CC had no role in the study design, data collection, data analysis, and data interpretation. However, CC was responsible for the oversight of the project and was involved in editing the drafts of the manuscript. During the period of the study, EI and RC were commissioned as researchers at Paradym, and SL was commissioned as an independent research consultant. The remaining authors have no other conflicts of interest.

Multimedia Appendix 1
Search strategy.
[DOCX File , 21 KB - mental_v8i11e31170_app1.docx]

Multimedia Appendix 2
Characteristics of included studies.
[DOCX File , 45 KB - mental_v8i11e31170_app2.docx]

Multimedia Appendix 3
An overview of the emerging evidence for MHapps.
[PDF File (Adobe PDF File), 134 KB - mental_v8i11e31170_app3.pdf]

Multimedia Appendix 4
An overview of the theoretical underpinning of the reviewed MHapps.
[PDF File (Adobe PDF File), 128 KB - mental_v8i11e31170_app4.pdf]

Multimedia Appendix 5
Risk of bias summary.
[DOCX File , 435 KB - mental_v8i11e31170_app5.docx]

Multimedia Appendix 6

Funnel plots for mental health, mental well-being and emotion regulation outcomes.
[DOCX File , 88 KB - mental_v8i11e31170_app6.docx]

References

1. Shah SMA, Mohammad D, Qureshi MFH, Abbas MZ, Aleem S. Prevalence, psychological responses and associated correlates of depression, anxiety and stress in a global population, during the coronavirus disease (COVID-19) pandemic. Community Ment Health J 2021 Jan 27;57(1):101-110 [FREE Full text] [doi: 10.1007/s10597-020-00728-y] [Medline: 33108569]
2. Stansfeld SA, Rasul FR, Head J, Singleton N. Occupation and mental health in a national UK survey. Soc Psychiatry Psychiatr Epidemiol 2011 Mar 24;46(2):101-110 [FREE Full text] [doi: 10.1007/s00127-009-0173-7] [Medline: 20033130]
3. 2019 NSDUH Annual National Report. National Survey on Drug Use and Health. Substance Abuse and Mental Health Services Administration. 2020 Sep 11. URL: https://www.samhsa.gov/data/report/2019-nsduh-annual-national-report [accessed 2021-09-29]
4. Ozamiz-Etxebarria N, Idoiaga Mondragon N, Dosil Santamaría M, Picaza Gorrotxategi M. Psychological symptoms during the two stages of lockdown in response to the COVID-19 outbreak: an investigation in a sample of citizens in northern Spain. Front Psychol 2020 Jun 18;11:2116. [doi: 10.3389/fpsyg.2020.01491]
5. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry 2020 Jun;7(6):547-560 [FREE Full text] [doi: 10.1016/S2215-0366(20)30168-1] [Medline: 32304649]
6. Moreno C, Wykes T, Galderisi S, Nordentoft M, Crossley N, Jones N, et al. How mental health care should change as a consequence of the COVID-19 pandemic. Lancet Psychiatry 2020 Sep;7(9):813-824 [FREE Full text] [doi: 10.1016/S2215-0366(20)30307-2] [Medline: 32682460]
7. Peek N, Sujan M, Scott P. Digital health and care in pandemic times: impact of COVID-19. BMJ Health Care Inform 2020 Jun 21;27(1):e100166 [FREE Full text] [doi: 10.1136/bmjhci-2020-100166] [Medline: 32565418]
8. Smartphone Ownership and Usage Smartphones are the most important personal item – almost everywhere. Statista. 2021. URL: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/ [accessed 2021-09-20]
9. The International Communications Market 2017: Key findings: the UK in context. 2017 Dec 18. URL: https://www.ofcom.org.uk/__data/assets/pdf_file/0019/108910/icmr-2017-uk-context.pdf [accessed 2021-09-29]
10. Aitkin M, Clancy B, Nass D. The Growing Value of Digital Health.: IQVIA Institute for Human Data Sciences; 2017. URL: https://regresearchnetwork.org/wp-content/uploads/the-growing-value-of-digital-health.pdf [accessed 2021-09-29]
11. IMS Institute for Healthcare Informatics. Patient adoption of mHealth: use, evidence and remaining barriers to mainstream acceptance. IQVIA. 2015 Sep 01. URL: https://www.iqvia.com/-/media/iqvia/pdfs/institute-reports/patient-adoption-of-mhealth.pdf [accessed 2021-09-29]
12. Antezana G, Venning A, Blake V, Smith D, Winsall M, Orlowski S, et al. An evaluation of behaviour change techniques in health and lifestyle mobile applications. Health Informatics J 2020 Mar 30;26(1):104-113 [FREE Full text] [doi: 10.1177/1460458218813726] [Medline: 30501364]
13. Neary M, Schueller SM. State of the field of mental health apps. Cogn Behav Pract 2018 Nov;25(4):531-537 [FREE Full text] [doi: 10.1016/j.cbpra.2018.01.002] [Medline: 33100810]
14. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016 Mar 01;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
15. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile apps for bipolar disorder: a systematic review of features and content quality. J Med Internet Res 2015 Aug 17;17(8):e198 [FREE Full text] [doi: 10.2196/jmir.4581] [Medline: 26283290]
16. 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]
17. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019 Dec 02;2(1):118. [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
18. Musiat P, Goldstone P, Tarrier N. Understanding the acceptability of e-mental health - attitudes and expectations towards computerised self-help treatments for mental health problems. BMC Psychiatry 2014 Apr 11;14(1):109. [doi: 10.1186/1471-244x-14-109]
19. Sucala M, Cuijpers P, Muench F, Cardo R, Soflau R, Dobrean A, et al. Anxiety: There is an app for that. A systematic review of anxiety apps. Depress Anxiety 2017 Jun;34(6):518-525. [doi: 10.1002/da.22654] [Medline: 28504859]
20. 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]

21. Huppert FA. The state of wellbeing science: concepts, measures, interventions, and policies. In: Huppert FA, Cooper CL, editors. Interventions and Policies to Enhance Wellbeing: Wellbeing: A Complete Reference Guide, Volume VI. Hoboken, NJ: John Wiley & Sons; Sep 27, 2013:1-49.
22. McAneney H, Tully MA, Hunter RF, Kouvonen A, Veal P, Stevenson M, et al. Individual factors and perceived community characteristics in relation to mental health and mental well-being. BMC Public Health 2015 Dec 12;15(1):1237 [FREE Full text] [doi: 10.1186/s12889-015-2590-8] [Medline: 26655388]
23. Keyes CLM. The mental health continuum: from languishing to flourishing in life. J Health Soc Behav 2002 Jun;43(2):207-222. [Medline: 12096700]
24. Deci EL, Ryan RM. Hedonia, eudaimonia, and well-being: an introduction. J Happiness Stud 2006 Nov 18;9(1):1-11. [doi: 10.1007/s10902-006-9018-1]
25. World Health Organization, Department of Mental Health and Substance Abuse in collaboration with the Victorian Health Promotion Foundation and The University of Melbourne. Promoting mental health: concepts, emerging evidence, practice, summary report. World Health Organization. 2004. URL: https://www.who.int/mental_health/evidence/en/promoting_mhh.pdf [accessed 2021-09-29]
26. Visted E, Vøllestad J, Nielsen M, Schanche E. Emotion regulation in current and remitted depression: a systematic review and meta-analysis. Front Psychol 2018 May 18;9:756-720. [doi: 10.3389/fpsyg.2018.00756] [Medline: 29867700]
27. Kraiss JT, Ten Klooster PM, Moskowitz JT, Bohlmeijer ET. The relationship between emotion regulation and well-being in patients with mental disorders: A meta-analysis. Compr Psychiatry 2020 Oct;102:152189 [FREE Full text] [doi: 10.1016/j.comppsych.2020.152189] [Medline: 32629064]
28. Côté S, Gyurak A, Levenson RW. The ability to regulate emotion is associated with greater well-being, income, and socioeconomic status. Emotion 2010 Dec;10(6):923-933 [FREE Full text] [doi: 10.1037/a0021156] [Medline: 21171762]
29. English T, Eldesouky L, Gross JJ. Emotion regulation: personality processes individual differences. In: John OP, Robins RW, Ackerman RA, Allen TA, Omotilewa Aro-Lambo M, editors. Handbook of personality: theory and research. New York: Routledge; 2008:701-722.
30. Townsend J, Altshuler L. Emotion processing and regulation in bipolar disorder: a review. Bipolar Disord 2012 Jun;14(4):326-339. [doi: 10.1111/j.1399-5618.2012.01021.x] [Medline: 22631618]
31. Szczygieł D, Buczny J, Bazińska R. Emotion regulation and emotional information processing: The moderating effect of emotional awareness. Person Ind Diff 2012 Feb;52(3):433-437. [doi: 10.1016/j.paid.2011.11.005]
32. Houben M, Van Den Noortgate W, Kuppens P. The relation between short-term emotion dynamics and psychological well-being: A meta-analysis. Psychol Bull 2015 Jul;141(4):901-930. [doi: 10.1037/a0038822] [Medline: 25822133]
33. 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]
34. Firth J, Torous J, Nicholas J, Carney R, Pratap A, Rosenbaum S, et al. The efficacy of smartphone-based mental health interventions for depressive symptoms: a meta-analysis of randomized controlled trials. World Psychiatry 2017 Oct;16(3):287-298. [doi: 10.1002/wps.20472] [Medline: 28941113]
35. Gál É, Stefan 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]
36. Spijkerman M, Pots W, Bohlmeijer E. 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]
37. Jayewardene W, Lohrmann D, Erbe R, Torabi M. Effects of preventive online mindfulness interventions on stress and mindfulness: A meta-analysis of randomized controlled trials. Prev Med Rep 2017 Mar 14;5:150-159 [FREE Full text] [doi: 10.1016/j.pmedr.2016.11.013] [Medline: 28050336]
38. Garrido S, Millington C, Cheers D, Boydell K, Schubert E, Meade T, et al. What works and what doesn't work? A systematic review of digital mental health interventions for depression and anxiety in young people. Front Psychiatry 2019 Nov 13;10:759. [doi: 10.3389/fpsyt.2019.00759] [Medline: 31798468]
39. McKay FH, Wright A, Shill J, Stephens H, Uccellini M. Using health and well-being apps for behavior change: a systematic search and rating of apps. JMIR Mhealth Uhealth 2019 Jul 04;7(7):e11926 [FREE Full text] [doi: 10.2196/11926] [Medline: 31274112]
40. Marshall JM, Dunstan DA, Bartik W. The role of digital mental health resources to treat trauma symptoms in Australia during COVID-19. Psychol Trauma 2020 Aug;12(S1):S269-S271. [doi: 10.1037/tra0000627] [Medline: 32496103]
41. 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]
42. McCloud T, Jones R, Lewis G, Bell V, Tsakanikos E. Effectiveness of a mobile app intervention for anxiety and depression symptoms in university students: randomized controlled trial. JMIR Mhealth Uhealth 2020 Jul 31;8(7):e15418 [FREE Full text] [doi: 10.2196/15418] [Medline: 32735221]

43. Meinlschmidt G, Lee J, Stalujanis E, Belardi A, Oh M, Jung EK, et al. Smartphone-based psychotherapeutic micro-interventions to improve mood in a real-world setting. Front Psychol 2016 Jul 28;7:1112. [doi: 10.3389/fpsyg.2016.01112] [Medline: 27516747]
44. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions version 6.2. 2021 Feb 01. URL: https://training.cochrane.org/handbook/current [accessed 2021-09-29]
45. Moher D, Liberati A, Tetzlaff J, Altman D, PRISMA Group. 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]
46. Eisenstadt M, Infanti E, Liverpool S. Characteristics and effectiveness of mobile apps that promote emotion regulation and well-being in the general population: a systematic review protocol. National Institute for Health Research. PROSPERO. 2020 Nov 11. URL: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020213051 [accessed 2021-09-29]
47. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Educ Inf 2018 Dec 18;34(4):285-291. [doi: 10.3233/efi-180221]
48. Kohl C, McIntosh E, Unger S, Haddaway N, Kecke S, Schiemann J, et al. Online tools supporting the conduct and reporting of systematic reviews and systematic maps: a case study on CADIMA and review of existing tools. Environ Evid 2018 Feb 1;7(1):2-17. [doi: 10.1186/s13750-018-0115-5]
49. Sterne JAC, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
50. Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions version 6. In: Higgins J, Green S, editors. Cochrane Book Series. Hoboken, NJ: John Wiley & Sons; 2019.
51. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews: A product from the ESRC Methods Programme. 2006 Apr. URL: https://www.lancaster.ac.uk/media/lancaster-university/content-assets/documents/fhm/dhr/chir/NSsynthesisguidanceVersion1-April2006.pdf [accessed 2021-09-29]
52. RevMan for non-Cochrane reviews. The Cochrane Collaboration. 2020. URL: https://training.cochrane.org/online-learning/core-software-cochrane-reviews/revman/revman-non-cochrane-reviews [accessed 2021-09-22]
53. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002 Jun 15;21(11):1539-1558. [doi: 10.1002/sim.1186] [Medline: 12111919]
54. Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 2000 Jun;56(2):455-463. [doi: 10.1111/j.0006-341x.2000.00455.x] [Medline: 10877304]
55. Ahtinen A, Mattila E, Välkkynen P, Kaipainen K, Vanhala T, Ermes M, et al. Mobile mental wellness training for stress management: feasibility and design implications based on a one-month field study. JMIR Mhealth Uhealth 2013 Jul 10;1(2):e11 [FREE Full text] [doi: 10.2196/mhealth.2596] [Medline: 25100683]
56. Ly KH, Ly AM, Andersson G. A fully automated conversational agent for promoting mental well-being: A pilot RCT using mixed methods. Internet Interv 2017 Dec 10;10:39-46 [FREE Full text] [doi: 10.1016/j.invent.2017.10.002] [Medline: 30135751]
57. Weber S, Lorenz C, Hemmings N. Improving stress and positive mental health at work via an app-based intervention: a large-scale multi-center randomized control trial. Front Psychol 2019 Dec 06;10:2745-2714. [doi: 10.3389/fpsyg.2019.02745] [Medline: 31866915]
58. Daugherty DA, Runyan JD, Steenbergh TA, Fratzke BJ, Fry BN, Westra E. Smartphone delivery of a hope intervention: Another way to flourish. PLoS One 2018;13(6):e0197930 [FREE Full text] [doi: 10.1371/journal.pone.0197930] [Medline: 29856805]
59. Firestone R, Cheng S, Dalhousie S, Hughes E, Funaki T, Henry A, et al. Exploring Pasifika wellbeing: findings from a large cluster randomised controlled trial of a mobile health intervention programme. N Z Med J 2020 Oct 30;133(1524):82-101. [Medline: 33119572]
60. van Emmerik AAP, Berings F, Lancee J. Efficacy of a mindfulness-based mobile application: a randomized waiting-list controlled trial. Mindfulness 2018;9(1):187-198 [FREE Full text] [doi: 10.1007/s12671-017-0761-7] [Medline: 29387266]
61. Giraldo‐O'Meara M, Doron G. Can self‐esteem be improved using short daily training on mobile applications? Examining real world data of GG self‐esteem users. Clin Psychol 2020 May 13:online ahead of print. [doi: 10.1111/cp.12219]
62. Carissoli C, Villani D, Riva G. Does a meditation protocol supported by a mobile application help people reduce stress? Suggestions from a controlled pragmatic trial. Cyberpsychol Behav Soc Netw 2015 Jan;18(1):46-53. [doi: 10.1089/cyber.2014.0062] [Medline: 25584730]
63. Birney A, Gunn R, Russell J, Ary DV. MoodHacker mobile web app with email for adults to self-manage mild-to-moderate depression: randomized controlled trial. JMIR Mhealth Uhealth 2016 Jan 26;4(1):e8-18 [FREE Full text] [doi: 10.2196/mhealth.4231] [Medline: 26813737]
64. Bruehlman-Senecal E, Hook C, Pfeifer J, FitzGerald C, Davis B, Delucchi K, et al. Smartphone app to address loneliness among college students: pilot randomized controlled trial. JMIR Ment Health 2020 Oct 20;7(10):e21496 [FREE Full text] [doi: 10.2196/21496] [Medline: 33079071]

65. Roy MJ, Costanzo ME, Highland KB, Olsen C, Clayborne D, Law W. An app a day keeps the doctor away: guided education and training via smartphones in subthreshold post traumatic stress disorder. Cyberpsychol Behav Soc Netw 2017 Aug;20(8):470-478. [doi: 10.1089/cyber.2017.0221] [Medline: 28737954]
66. Krafft J, Potts S, Schoendorff B, Levin ME. A randomized controlled trial of multiple versions of an acceptance and commitment therapy matrix app for well-being. Behav Modif 2019 Mar 20;43(2):246-272. [doi: 10.1177/0145445517748561] [Medline: 29262693]
67. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The use and effectiveness of mobile apps for depression: results from a fully remote clinical trial. J Med Internet Res 2016 Dec 20;18(12):e330 [FREE Full text] [doi: 10.2196/jmir.6482] [Medline: 27998876]
68. Levin M, Haeger J, Cruz R. Tailoring acceptance and commitment therapy skill coaching in the moment through smartphones: results from a randomized controlled trial. Mindfulness 2018 Jul 21;10(4):689-699. [doi: 10.1007/s12671-018-1004-2]
69. Morris ME, Kathawala Q, Leen TK, Gorenstein EE, Guilak F, Labhard M, et al. Mobile therapy: case study evaluations of a cell phone application for emotional self-awareness. J Med Internet Res 2010 Apr 30;12(2):e10 [FREE Full text] [doi: 10.2196/jmir.1371] [Medline: 20439251]
70. Bostock S, Crosswell AD, Prather AA, Steptoe A. Mindfulness on-the-go: Effects of a mindfulness meditation app on work stress and well-being. J Occup Health Psychol 2019 Mar;24(1):127-138 [FREE Full text] [doi: 10.1037/ocp0000118] [Medline: 29723001]
71. Mak W, Wong C, Chan A, Lau J. Mobile self-compassion programme for promotion of public mental health: a randomised controlled trial. Hong Kong Med J 2019 Feb;25(Suppl 2):15-17 [FREE Full text] [Medline: 30674702]
72. Moberg C, Niles A, Beermann D. Guided self-help works: randomized waitlist controlled trial of Pacifica, a mobile app integrating cognitive behavioral therapy and mindfulness for stress, anxiety, and depression. J Med Internet Res 2019 Jun 08;21(6):e12556 [FREE Full text] [doi: 10.2196/12556] [Medline: 31199319]
73. Ponzo S, Morelli D, Kawadler JM, Hemmings NR, Bird G, Plans D. Efficacy of the digital therapeutic mobile app BioBase to reduce stress and improve mental well-being among university students: randomized controlled trial. JMIR Mhealth Uhealth 2020 Apr 06;8(4):e17767 [FREE Full text] [doi: 10.2196/17767] [Medline: 31926063]
74. Yang E, Schamber E, Meyer RML, Gold JI. Happier healers: randomized controlled trial of mobile mindfulness for stress management. J Altern Complement Med 2018 May;24(5):505-513. [doi: 10.1089/acm.2015.0301] [Medline: 29420050]
75. Stallman H. Efficacy of the My Coping Plan mobile application in reducing distress: a randomised controlled trial. Clin Psychol 2020 Nov 09;23(3):206-212. [doi: 10.1111/cp.12185]
76. Throuvala MA, Griffiths MD, Rennoldson M, Kuss DJ. Mind over matter: testing the efficacy of an online randomized controlled trial to reduce distraction from smartphone use. Int J Environ Res Public Health 2020 Jul 05;17(13):4842 [FREE Full text] [doi: 10.3390/ijerph17134842] [Medline: 32635650]
77. Walsh K, Saab B, Farb N. Effects of a mindfulness meditation app on subjective well-being: active randomized controlled trial and experience sampling study. JMIR Ment Health 2019 Jan 08;6(1):e10844 [FREE Full text] [doi: 10.2196/10844] [Medline: 30622094]
78. McEwan K, Richardson M, Sheffield D, Ferguson FJ, Brindley P. A smartphone app for improving mental health through connecting with urban nature. Int J Environ Res Public Health 2019 Sep 12;16(18):3373 [FREE Full text] [doi: 10.3390/ijerph16183373] [Medline: 31547286]
79. Bakker D, Rickard N. Engagement with a cognitive behavioural therapy mobile phone app predicts changes in mental health and wellbeing: MoodMission. Austral Psychol 2020 Nov 12;54(4):245-260. [doi: 10.1111/ap.12383]
80. Champion L, Economides M, Chandler C. The efficacy of a brief app-based mindfulness intervention on psychosocial outcomes in healthy adults: A pilot randomised controlled trial. PLoS One 2018;13(12):e0209482 [FREE Full text] [doi: 10.1371/journal.pone.0209482] [Medline: 30596696]
81. Coelhoso CC, Tobo PR, Lacerda SS, Lima AH, Barrichello CRC, Amaro E, et al. A new mental health mobile app for well-being and stress reduction in working women: randomized controlled trial. J Med Internet Res 2019 Nov 07;21(11):e14269 [FREE Full text] [doi: 10.2196/14269] [Medline: 31697244]
82. Deady M, Glozier N, Calvo R, Johnston D, Mackinnon A, Milne D, et al. Preventing depression using a smartphone app: a randomized controlled trial. Psychol Med 2020 Jul 06:1-10. [doi: 10.1017/S0033291720002081] [Medline: 32624013]
83. Economides M, Martman J, Bell M, Sanderson B. Improvements in stress, affect, and irritability following brief use of a mindfulness-based smartphone app: a randomized controlled trial. Mindfulness 2018 Mar 01;9(5):1584-1593 [FREE Full text] [doi: 10.1007/s12671-018-0905-4] [Medline: 30294390]
84. Flett J, Hayne H, Riordan B, Thompson L, Conner T. Mobile mindfulness meditation: a randomised controlled trial of the effect of two popular apps on mental health. Mindfulness 2018 Oct 31;10(5):863-876. [doi: 10.1007/s12671-018-1050-9]
85. Lee R, Jung M. Evaluation of an mHealth app (DeStressify) on university students' mental health: pilot trial. JMIR Ment Health 2018 Jan 23;5(1):e2 [FREE Full text] [doi: 10.2196/mental.8324] [Medline: 29362209]
86. Levin ME, Hicks ET, Krafft J. J Am Coll Health 2020 Mar 09:online ahead of print. [doi: 10.1080/07448481.2020.1728281] [Medline: 32150519]

87. Huberty J, Green J, Glissmann C, Larkey L, Puzia M, Lee C. Efficacy of the mindfulness meditation mobile app "Calm" to reduce stress among college students: randomized controlled trial. JMIR Mhealth Uhealth 2019 Jun 25;7(6):e14273 [FREE Full text] [doi: 10.2196/14273] [Medline: 31237569]
88. Hwang WJ, Jo HH. Evaluation of the effectiveness of mobile app-based stress-management program: a randomized controlled trial. Int J Environ Res Public Health 2019 Nov 03;16(21):4270 [FREE Full text] [doi: 10.3390/ijerph16214270] [Medline: 31684169]
89. Litvin S, Maier M. How mHealth programmes can treat depression: A randomised controlled trial. Psychreg J Psychol 2019;3(2):66-78 [FREE Full text]
90. Bakker D, Rickard N. Engagement in mobile phone app for self-monitoring of emotional wellbeing predicts changes in mental health: MoodPrism. J Affect Disord 2018 Feb;227:432-442. [doi: 10.1016/j.jad.2017.11.016] [Medline: 29154165]
91. Deady M, Johnston D, Milne D, Glozier N, Peters D, Calvo R, et al. Preliminary effectiveness of a smartphone app to reduce depressive symptoms in the workplace: feasibility and acceptability study. JMIR Mhealth Uhealth 2018 Dec 04;6(12):e11661 [FREE Full text] [doi: 10.2196/11661] [Medline: 30514694]
92. Kawadler JM, Hemmings NR, Ponzo S, Morelli D, Bird G, Plans D. Effectiveness of a smartphone app (BioBase) for reducing anxiety and increasing mental well-being: pilot feasibility and acceptability study. JMIR Form Res 2020 Nov 10;4(11):e18067 [FREE Full text] [doi: 10.2196/18067] [Medline: 32969341]
93. Morrison Wylde C, Mahrer NE, Meyer RM, Gold JI. Mindfulness for novice pediatric nurses: smartphone application versus traditional intervention. J Pediatr Nurs 2017 Sep;36:205-212. [doi: 10.1016/j.pedn.2017.06.008] [Medline: 28888505]
94. Haeger J, Davis C, Levin M. Utilizing ACT daily as a self-guided app for clients waiting for services at a college counseling center: A pilot study. J Am Coll Health 2020 Jun 12:1-8. [doi: 10.1080/07448481.2020.1763366] [Medline: 32529933]
95. Hamamura T, Suganuma S, Ueda M, Mearns J, Shimoyama H. Standalone effects of a cognitive behavioral intervention using a mobile phone app on psychological distress and alcohol consumption among Japanese workers: pilot nonrandomized controlled trial. JMIR Ment Health 2018 Mar 22;5(1):e24 [FREE Full text] [doi: 10.2196/mental.8984] [Medline: 29567634]
96. Wen L, Sweeney TE, Welton L, Trockel M, Katznelson L. Encouraging mindfulness in medical house staff via smartphone app: a pilot study. Acad Psychiatry 2017 Oct 9;41(5):646-650. [doi: 10.1007/s40596-017-0768-3] [Medline: 28795335]
97. Ly KH, Dahl J, Carlbring P, Andersson G. Development and initial evaluation of a smartphone application based on acceptance and commitment therapy. Springerplus 2012 Jul 31;1(1):11 [FREE Full text] [doi: 10.1186/2193-1801-1-11] [Medline: 23961342]
98. Borjalilu S, Mazaheri MA, Talebpour A. Effectiveness of mindfulness-based stress management in the mental health of Iranian university students: a comparison of blended therapy, face-to-face sessions, and mHealth app (Aramgar). Iran J Psychiatry Behav Sci 2019 May 12;13(2):1-9. [doi: 10.5812/ijpbs.84726]
99. Bakker D, Kazantzis N, Rickwood D, Rickard N. A randomized controlled trial of three smartphone apps for enhancing public mental health. Behav Res Ther 2018 Oct;109:75-83. [doi: 10.1016/j.brat.2018.08.003] [Medline: 30125790]
100. Litvin S, Saunders R, Maier MA, Lüttke S. Gamification as an approach to improve resilience and reduce attrition in mobile mental health interventions: A randomized controlled trial. PLoS One 2020;15(9):e0237220 [FREE Full text] [doi: 10.1371/journal.pone.0237220] [Medline: 32877425]
101. Ha SW, Kim J. Designing a scalable, accessible, and effective mobile app based solution for common mental health problems. Int J Hum Comput Interact 2020 Apr 26;36(14):1354-1367. [doi: 10.1080/10447318.2020.1750792]
102. Hides L, Dingle G, Quinn C, Stoyanov SR, Zelenko O, Tjondronegoro D, et al. Efficacy and outcomes of a music-based emotion regulation mobile app in distressed young people: randomized controlled trial. JMIR Mhealth Uhealth 2019 Jan 16;7(1):e11482 [FREE Full text] [doi: 10.2196/11482] [Medline: 30664457]
103. Howells A, Ivtzan I, Eiroa-Orosa F. Putting the 'app' in happiness: a randomised controlled trial of a smartphone-based mindfulness intervention to enhance wellbeing. J Happiness Stud 2014 Oct 29;17(1):163-185. [doi: 10.1007/s10902-014-9589-1]
104. Mak W, Tong A, Yip S, Lui W, Chio F, Chan A, et al. Efficacy and moderation of mobile app-based programs for mindfulness-based training, self-compassion training, and cognitive behavioral psychoeducation on mental health: randomized controlled noninferiority trial. JMIR Ment Health 2018 Oct 11;5(4):e60 [FREE Full text] [doi: 10.2196/mental.8597] [Medline: 30309837]
105. Kapp KM. The Gamification of Learning and Instruction: Game-based Methods and Strategies for Training and Education. San Francisco, CA: John Wiley & Sons; 2012.
106. Linardon J, Cuijpers P, Carlbring P, Messer M, Fuller-Tyszkiewicz M. The efficacy of app-supported smartphone interventions for mental health problems: a meta-analysis of randomized controlled trials. World Psychiatry 2019 Oct;18(3):325-336. [doi: 10.1002/wps.20673] [Medline: 31496095]
107. van der Watt ASJ, Odendaal W, Louw K, Seedat S. Distant mood monitoring for depressive and bipolar disorders: a systematic review. BMC Psychiatry 2020 Jul 22;20(1):383 [FREE Full text] [doi: 10.1186/s12888-020-02782-y] [Medline: 32698802]
108. Abd-Alrazaq AA, Rababeh A, Alajlani M, Bewick BM, Househ M. Effectiveness and safety of using chatbots to improve mental health: systematic review and meta-analysis. J Med Internet Res 2020 Jul 13;22(7):e16021 [FREE Full text] [doi: 10.2196/16021] [Medline: 32673216]

109. Montana J, Matamala-Gomez M, Maisto M, Mavrodiev P, Cavalera C, Diana B, et al. The benefits of emotion regulation interventions in virtual reality for the improvement of wellbeing in adults and older adults: a systematic review. J Clin Med 2020 Mar 12;9(2):500 [FREE Full text] [doi: 10.3390/jcm9020500] [Medline: 32059514]
110. Villani D, Carissoli C, Triberti S, Marchetti A, Gilli G, Riva G. Videogames for emotion regulation: a systematic review. Games Health J 2018 Apr;7(2):85-99. [doi: 10.1089/g4h.2017.0108] [Medline: 29424555]
111. Shapero BG, Stange JP, McArthur BA, Abramson LY, Alloy LB. Cognitive reappraisal attenuates the association between depressive symptoms and emotional response to stress during adolescence. Cogn Emot 2019 May 11;33(3):524-535 [FREE Full text] [doi: 10.1080/02699931.2018.1462148] [Medline: 29637806]
112. Subic-Wrana C, Beutel ME, Brähler E, Stöbel-Richter Y, Knebel A, Lane RD, et al. How is emotional awareness related to emotion regulation strategies and self-reported negative affect in the general population? PLoS One 2014 Mar 17;9(3):e91846 [FREE Full text] [doi: 10.1371/journal.pone.0091846] [Medline: 24637792]
113. 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]
114. Huberty J, Green J, Puzia M, Stecher C. Evaluation of mood check-in feature for participation in meditation mobile app users: retrospective longitudinal analysis. JMIR Mhealth Uhealth 2021 Apr 23;9(4):e27106 [FREE Full text] [doi: 10.2196/27106] [Medline: 33890865]

Abbreviations

ACT: acceptance and commitment therapy
CBT: cognitive behavior therapy
MHapp: mental health app
MMAT: Mixed Methods Appraisal Tool
PRISMA: Preferred Items for Reporting a Systematic Review and Meta-analysis
PTSD: posttraumatic stress disorder
RCT: randomized controlled trial
RoB: Risk of Bias

Edited by J Torous; submitted 15.06.21; peer-reviewed by M Salimi, J Huberty; comments to author 17.07.21; revised version received 21.07.21; accepted 14.08.21; published 08.11.21.

Please cite as:
Eisenstadt M, Liverpool S, Infanti E, Ciuvat RM, Carlsson C
Mobile Apps That Promote Emotion Regulation, Positive Mental Health, and Well-being in the General Population: Systematic Review and Meta-analysis
JMIR Ment Health 2021;8(11):e31170
URL: https://mental.jmir.org/2021/11/e31170
doi:10.2196/31170
PMID:34747713

Review

Implementing the Routine Use of Electronic Mental Health Screening for Youth in Primary Care: Systematic Review

Rhiannon Martel[1], MN, PhD; Matthew Shepherd[2], DClinPsy, PhD; Felicity Goodyear-Smith[1], MBChB, MD

[1]Department of General Practice & Primary Health Care, Faculty of Medical & Health Science, University of Auckland, Auckland, New Zealand
[2]School of Psychology, Massey University, Auckland, New Zealand

Corresponding Author:
Felicity Goodyear-Smith, MBChB, MD
Department of General Practice & Primary Health Care
Faculty of Medical & Health Science
University of Auckland
22 Park Road
Grafton
Auckland, 1142
New Zealand
Phone: 64 99232357
Email: f.goodyear-smith@auckland.ac.nz

Abstract

Background: Adolescents often present at primary care clinics with nonspecific physical symptoms when, in fact, they have at least 1 mental health or risk behavior (psychosocial) issue with which they would like help but do not disclose to their care provider. Despite global recommendations, over 50% of youths are not screened for mental health and risk behavior issues in primary care.

Objective: This review aimed to examine the implementation, acceptability, feasibility, benefits, and barriers of e-screening tools for mental health and risk behaviors among youth in primary care settings.

Methods: Electronic databases—MEDLINE, CINAHL, Scopus, and the Cochrane Database of Systematic Reviews—were searched for studies on the routine screening of youth in primary care settings. Screening tools needed to be electronic and screen for at least 1 mental health or risk behavior issue. A total of 11 studies that were reported in 12 articles, of which all were from high-income countries, were reviewed.

Results: e-Screening was largely proven to be feasible and acceptable to youth and their primary care providers. Preconsultation e-screening facilitated discussions about sensitive issues and increased disclosure by youth. However, barriers such as the lack of time, training, and discomfort in raising sensitive issues with youth continued to be reported.

Conclusions: To implement e-screening, clinicians need to change their behaviors, and e-screening processes must become normalized into their workflows. Co-designing and tailoring screening implementation frameworks to meet the needs of specific contexts may be required to ensure that clinicians overcome initial resistances and perceived barriers and adopt the required processes in their work.

(JMIR Ment Health 2021;8(11):e30479) doi:10.2196/30479

KEYWORDS

adolescent; mental health; risk behavior; screening; primary care

Introduction

More than 90% of New Zealand secondary school students visit a primary care provider, such as a family physician or primary care nurse, at least once per year [1]. Adolescents often present at primary care clinics with nonspecific physical symptoms when, in fact, they have at least 1 mental health or risk behavior (psychosocial) issue with which they would like help but do not disclose to their care provider [2-4]. Incidence rates of youth psychosocial issues are higher for New Zealand's indigenous Māori population, whose access to appropriate care is less than that of the general New Zealand population [5]. Mental health issues generally include anxiety and depression but may also include more general distress resulting from a variety of stressors

and difficulty with controlling anger. Risk behaviors include substance misuse (nicotine, alcohol, and recreational drugs), eating and conduct distress, sexual health, physical inactivity, and exposure to abuse and problem gambling or gaming. A full psychosocial assessment can help with identifying these concerns, but the young person must be willing to discuss personal and delicate issues, sometimes with someone they do not know [6]. Screening can reveal issues that could otherwise be overlooked and can facilitate discussions about psychosocial concerns between care providers and youth [7-9].

Currently, year 9 students (aged 13-14 years) in New Zealand decile 1 to decile 3 secondary schools undergo a routine psychosocial assessment that uses the Home, Education/Employment, Eating, Activities, Drugs and Alcohol, Sexuality, Suicide/Depression, and Safety (HEEADSSS) assessment tool [10]. This is a multi-item, interview-based assessment tool that is also used by some clinicians in primary health care during consultations with adolescents. Although the HEEADSSS tool is used nationally and internationally, it is not validated, it can be time-consuming to use (sometimes taking up to 2 hours to complete), and the results are variable [11,12].

A number of screening tools are available for youth psychosocial issues, but most cover a single domain [13], and administering and interpreting these tools can be time-consuming [14]. Primary care clinicians may be uncertain about which screening tools are suitable for use in certain clinical contexts. Many tools rely on care providers having the skills, knowledge, expertise, and experience to initiate the screen, interpret the results, and provide appropriate interventions [7]. Care providers often describe being underresourced in terms of time, the availability of appropriate tools, training, and their experience in youth health [15]. Care providers have also cited a lack of awareness of appropriate agencies and available support services as a further barrier to screening [3,7,15,16].

Underpinned by national and international policies and strategies, global recommendations state that young people who seek help from their care providers should be routinely screened for psychosocial issues [17]. Despite this, such screening occurs in less than 50% of primary care consultations with youth, meaning that over half of adolescent mental health concerns go undetected [7,8].

The aim of this literature review was to examine the implementation of e-screening tools for psychosocial issues among youth in primary care settings. Specifically, we aimed to determine whether e-screening has been performed opportunistically or systematically, whether such screening has targeted those who were deemed at risk for mental health or risk behavior (psychosocial) issues, whether e-screening has been conducted in the waiting room prior to consultation or at another time, and whether e-screening has been initiated by an administrator (either a research assistant or a clinic staff member). The objectives were to explore different conditions and settings where routine e-screening for youth psychosocial issues is undertaken and to identify the perceived acceptability and benefits of, barriers to, and feasibility of the implementation of such screening.

Methods

Search Strategy

The search strategy was devised through discussion with a specialist librarian and all review authors. The electronic databases MEDLINE, CINAHL, Scopus, and the Cochrane Database of Systematic Reviews were searched for studies on screening for mental health issues and risky behaviors among youth. The search was conducted by using search strings that incorporated wildcard symbols (Multimedia Appendix 1). Search results were exported to bibliography software and recorded in a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram.

Inclusion and Exclusion Criteria

All research studies published in English that conducted the e-screening of psychosocial issues among young people up to May 2020 were included. There was no publication date range for excluding studies. The inclusion criteria included studies involving the e-screening of youth in primary care. e-Screening involved the use of web-based screening tools that were delivered by a mobile device, an e-tablet, a computer, or another digital device. Youth were defined as young people aged between 12 and 25 years. Primary care settings were community-based health settings that catered to either all patients (general practice or family health services) or youth specifically (school-based clinics or youth clinics). The inclusion criteria included studies that addressed facilitators and barriers to and the process, implementation, and feasibility of using e-screening tools in primary care. The exclusion criteria were study protocols (no data available) and studies in which screening was not conducted on young people, screening was not for psychosocial issues, or screening was not the focus of the research. Studies were also excluded if the screen was not electronic or was not conducted in a primary care setting. Non-English papers were excluded.

Screening

Titles were screened for initial eligibility, and duplicates were removed by using bibliography software. Abstracts were independently screened by 2 authors and cross-checked for agreement. The included abstracts were reviewed and further excluded if they did not meet the eligibility criteria. Afterward, the full papers of included studies and further studies identified through hand searching were reviewed, and those that did not meet the eligibility criteria were excluded. A second researcher checked that the full-text papers were eligible for inclusion.

Analyses

The items to be coded from the included papers were decided upon via discussion among the research team members. The studies were classified based on the country of origin, study design, type of data, clinical setting, people who were selected as participants (eg, age range), and people who had recruited them (eg, the research assistant of a clinical staff member). The lead author tabulated the specific tools and screening domains, along with any additional tools that were used, the time and duration of screening, and the place in which screening occurred. Data on the types of measures used (eg, utility, acceptability,

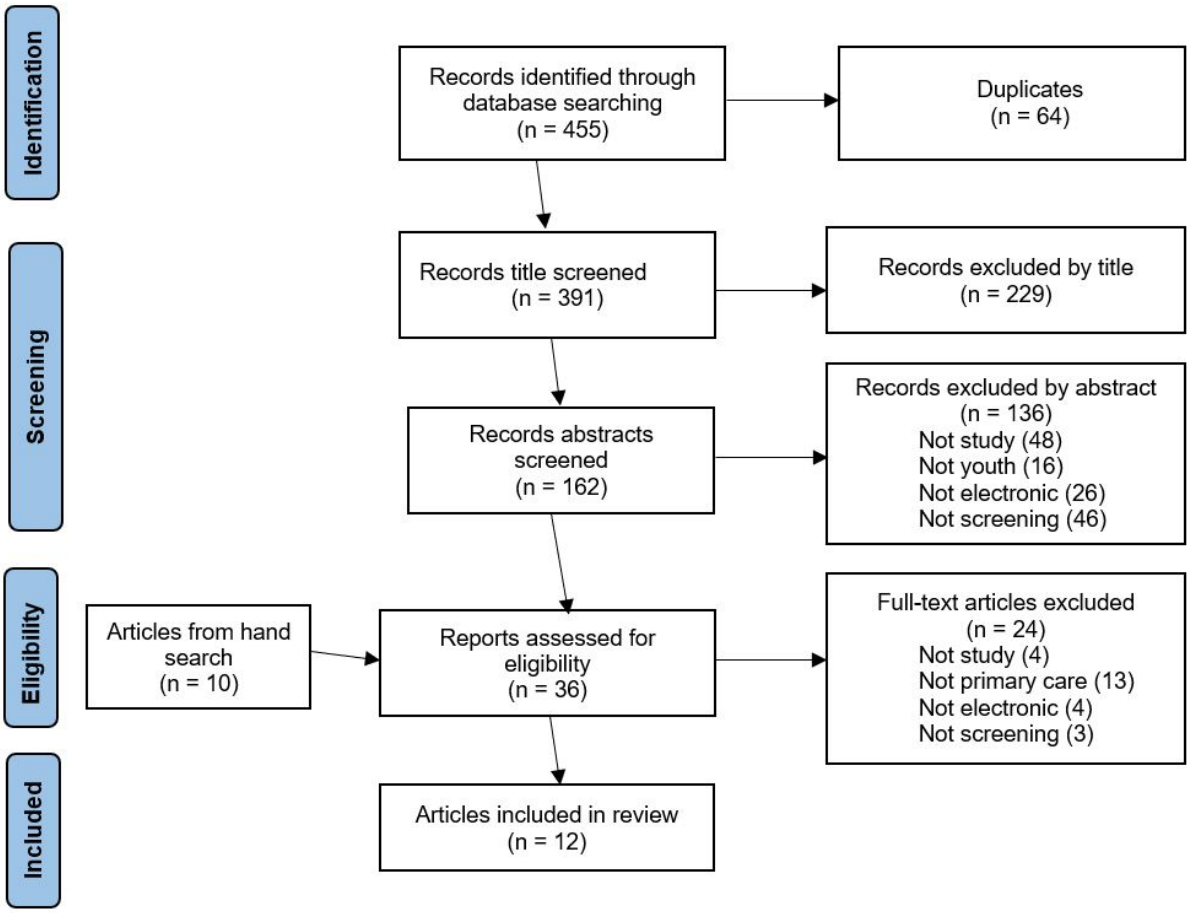
feasibility) and the analyses undertaken were extracted and synthesized from the studies. The study quality was assessed by identifying potential biases, limitations, and strengths. FGS reviewed the process at various stages, as well as the included papers and tables, and provided feedback. Due to the heterogeneity of the studies, a meta-analysis was not possible.

Results

Identification and Screening of Studies

A total of 455 articles were identified, and after the screening and hand-searching processes, 12 articles reporting 11 studies were included in the review (Figure 1).

Figure 1. PRISMA flow diagram.



Study Characteristics

The included papers described 11 studies that were conducted between 2009 and 2018 [6,16,18-27]. The designs used in the reviewed studies were a case study [16], co-design and descriptive studies [19,22], a translational study [20], quasi-experimental studies [18,21,23,24], and randomized trials [25-27]. All studies included quantitative data, and 4 were mixed methods studies [16,18,23]. All studies were carried out in high-income countries, and nearly half (5/11, 45%) were conducted in family health clinics, general practice clinics, or primary care clinics. Study sites also included pediatric primary care clinics, an integrated health clinic, school clinics, and a colocated youth clinic. Most of the studies (8/11, 73%) recruited both young people and care providers as participants. Youth were recruited from clinic waiting rooms when they attended their routine medical reviews, while care providers were recruited from participating clinics. In one study conducted in New Zealand, the youth participants were mostly indigenous Māori [19].

All sixth- to 12th-grade students at a public school were eligible to participate in 1 project, and one study did not recruit young people per se but used deidentified data from electronic medical records. Youth participants' ages ranged from 11 to 25 years across all of the studies (Table 1).

Table 1. Study designs, participants, and settings.

Study authors	Country	Study design	Data type	Setting	Participant selection criteria
Bilardi et al [18]	Australia	Quasi-experimental	Mixed	Family health clinic	All 16- to 24-year-olds attending their annual reviews (N=871) and primary care providers
Bradford and Rickwood [6]	Australia	Quasi-experimental	Quantitative	Youth clinic	12- to 25-year-olds (n=339) and 13 clinicians
Curtis et al [20]	United States of America	Translational	Quantitative	School clinic	All sixth- to 12th-grade pupils from 1 school (N=248)
Diamond et al [22]	United States of America	Descriptive	Quantitative	Family health clinic	12- to 21-year-olds in primary health care waiting rooms (N=415)
Gadomski et al [23]	United States of America	Quasi-experimental	Mixed	Urban and rural clinics	Consecutive patients aged <18 years attending their annual reviews (N=72) and primary care providers
Goodyear-Smith et al [19]	New Zealand	Co-design	Mixed	Youth clinic	Consecutive patients aged 12-24 years (N=30) and care providers
Harris and Knight [21]	United States of American and Czech Republic	Quasi-experimental	Quantitative	Family health clinic	All patients aged 12-18 years undergoing routine care (United States of America: n=2106; Czech Republic: n=589)
Olson et al [24]	United States of America	Quasi-experimental	Quantitative	Primary care clinic	Consecutive patients aged 11-19 years (N=1052) and primary care providers
Riese et al [25]	United States of America	Randomized controlled trial	Quantitative	Pediatric primary care clinic	13- to 19-year-olds (n=120) and primary care providers (n=14)
Sterling et al [27]	United States of America	Randomized controlled trial	Quantitative	Integrated health clinic	Primary care providers caring for ≥50 eligible youths (N=52; EMR[a] data on 1871 youths were analyzed)
Webb et al [16]	Australia	Case study	Mixed	General practice clinic	14- to 25-year-olds (n=87), general practitioners (n=4), and support staff (n=10)

[a]EMR: electronic medical record.

Initiation and Completion of Screening

In a majority of studies (7/11, 64%), e-screening was initiated by a research assistant before a young person's consultation with their clinician (Table 2). In one study, young people were given the details of a web-based tool at the end of their consultation by either a clinician or a clinic administration staff member. Youth participants were invited to access and complete the e-screen either before leaving the clinic or later at home, but completion rates were low [18]. Clinic administration staff initiated the screen in 2 of the studies, and in a school-based project, a guidance counselor initiated it. Young people completed the screen on a mobile device; most did so in the waiting room preconsultation. Once the screen was completed, the results were immediately available to the care provider.

A variety of screening tools were used in the studies reviewed, of which some (4/11, 36%) were validated. The majority of the tools were multi-item tools, and all but one study [18] included screening for alcohol and drugs. Screens that only covered 1 domain were used in 3 studies—2 studied substance abuse screening and 1 studied sexual health risk assessment (Table 2).

Table 2. Screening tools, domains, screen validation, the location and duration of screens, and screen initiators.

Study authors	Tool	Domains screened	Links	Screening time (location)	Screening duration	Screen initiator
Bilardi et al [18]	Check Your Risk	Sexual health	—[a]	Postconsultation (clinic or home)	—	Youth
Bradford and Rickwood [6]	My Assessment	Home, education, eating, activities, alcohol or drug use, tobacco, sexual health, emotions, and safety	—	Preconsultation (clinic)	10-15 minutes	Research assistant
Curtis et al [20]	CRAFFT[b] instrument (validated)	Alcohol and drugs	Alcohol and drug information	Preconsultation (school clinic)	15 minutes	School counselor
Diamond et al [22]	BHS[c] (validated)	Medical, family, school, safety, sexuality, abuse, nutrition, eating, anxiety, trauma, depression, alcohol or drug use, suicidality, and psychosis	BDI-II[d], MSSI[e], and TSC[f]	Preconsultation (waiting room)	8-12 minutes	Research assistant
Gadomski et al [23]	DartScreen	Nutrition, exercise, alcohol or drug use, school, mental health, depression and anxiety, and sexual health	PHQ[g], GAD-2[h], and SBQ[i]	Preconsultation (waiting room)	9.5 minutes	Research assistant
Goodyear-Smith et al [19]	YouthCHAT (validated)	Smoking, alcohol or drug use, gambling, eating disorder, depression, anxiety, stress, sexual health, abuse, conduct, anger, and inactivity	PHQ-A[j], GAD-7[k], SACS[l], and ASSIST[m]	Preconsultation (waiting room)	—	Research assistant
Harris and Knight [21]	CRAFFT instrument (validated)	Alcohol or drug use	CRAFFT instrument	Preconsultation	5 minutes	Research assistant
Olson et al [24]	Based on GAPSQ[n]	Family, medical, safety, smoking, sexuality, activity, mental health, body image, school, relationships, nutrition, conduct	Alcohol and drug information	Preconsultation (clinic)	9-11 minutes	Admin staff
Riese et al [25]	TickiT (with and without the YRBS[o])	Home, education, eating, activities, alcohol or drug use tobacco, sexual health, emotions, safety	Selected YRBS	Preconsultation (waiting room)	8.4 minutes	Research assistant
Sterling et al [26,27]	TWCQ[p]	Alcohol or drug use, mood, and suicidality	CRAFFT instrument	Preconsultation (clinic)	—	Admin staff
Webb et al [16]	Check Up general practitioner app	Home, education, eating, activities, alcohol or drug use, tobacco, sexual health, emotions, and safety	—	Preconsultation (general practice clinic)	10-14 minutes	Research assistant

[a]Not applicable.

[b]CRAFFT: Car, Relax, Alone, Forget, Friends, Trouble.

[c]BHS: Behavioral Health Screen.

[d]BDI-II: Beck Depression Inventory-II.

[e]MSSI: Modified Scale for Suicidal Ideation.

[f]TSC: Trauma Symptom Checklist.

[g]PHQ: Patient Health Questionnaire.

[h]GAD-2: Generalized Anxiety Disorder 2-item.

[i]SBQ: Suicide Behavior Questionnaire.

[j]PHQ-A: Patient Health Questionnaire-Adolescent Version.

[k]GAD-7: Generalized Anxiety Disorder 7-item.

[l]SACS: Substances and Choices Scale.

[m]ASSIST: Alcohol, Smoking and Substance Involvement Screening Test.

[n]GAPSQ: Guidelines for Adolescent Preventive Services Questionnaire.

[o]YRBS: Youth Risk Behavior Survey.

[p]TWCQ: Teen Well Check Questionnaire.

Implementation Factors Included in the Studies

The acceptability and utility of e-screening tools for both care providers and young people were outcomes that were measured in 5 of the studies, and 8 studies described the impact that reviewing the results of a screen had on discussions and engagement during the postscreen consultation (Table 3). Two studies evaluated whether training care providers, providing them with resources, and obtaining support from other clinicians had any influence on rates of the psychosocial e-screening of youth. Another analyzed screening rates after the implementation of a computer-based, self-reported, previsit screen for youth psychosocial issues.

Table 3. Sources of data, study measures, potential biases, limitations, and strengths.

Study authors	Data sources	Measures	Analysis	Bias	Limitations	Strengths
Bilardi et al [18]	EMR[a] data, and interviews	Number of tests at 6 months pre- and postintervention, youth feedback, and barriers to use	2-sided *P* values, descriptive statistics, and thematic analysis	Training increases screening awareness	Small sample and no feedback	Real clinical situation
Bradford and Rickwood [6]	My Assessment data and questionnaires	Acceptability, feasibility, utility, reported behaviors, and barriers to use	Descriptive statistics and the comparison of control and intervention psychometrics	Missing data	Single center	Large sample size, a response rate of 87%, and a quasi-experimental design
Curtis et al [20]	EMR data	Utility in school, screening and detection rates, counseling acceptability, sustainability barriers, and barriers to use	Formative evaluation	Bias toward financially stable families	No usage data	Tested in school
Diamond et al [22]	Survey	Utility and acceptability, screen understandability, honest disclosure, and barriers to use	Descriptive statistics and odds ratios	Researcher-created tool	Nonrandom sample	Identifies barriers
Gadomski et al [23]	Interviews, audio recordings, and a youth survey	Information provided, question types, brief intervention delivery rates, engagement, and issues addressed	Inductive thematic approach	Effect of recording	Nonrandom sample	Real clinical situation
Goodyear-Smith et al [19]	Surveys, focus groups, and interviews	Assessment utility, youth and care provider acceptability, and barriers to use	Descriptive statistics and thematic analysis	Nonrepresentative sample	Small sample and no control	Real clinical situation
Harris and Knight [21]	Postvisit survey and EMR data	Advice-to-quit rates, likelihood of following advice, youth satisfaction, responses to the 3- and 6-month postscreen survey, and barriers to use	Chi-square tests (categorical data), *t* tests (continuous data), and longitudinal data	Self-reported data (potential recall error and the social desirability effect)	Nonrandomized study and small sample	Consistent with previous study
Olson et al [24]	Exit surveys	Youth satisfaction, youths' perceptions of care provider attention and discussions, and barriers to use	Chi-square and Fisher exact tests	Sample mostly consisting of White, middle-class participants	Small study	—[b]
Riese et al [25]	Exit survey	Care providers' impressions of the utility of disclosures and discussions and barriers to use	Descriptive statistics	Specific setting and population	Small sample	Cluster-randomized study
Sterling et al [27]	EMR data	Effect on screening rates, effect of adding BHCs[c] (initiation and engagement with and without a BHC), and barriers to use	Descriptive statistics and bivariate and logistic models	Integrated clinics	Established EMR	Diverse population
Webb et al [16]	Focus groups, interviews, and utility measures	Rates of use, barriers and facilitators, and the feasibility of use	Descriptive statistics and thematic analysis	Socioeconomically advantaged population	Single case study	—

[a]EMR: electronic medical record.

[b]Not available.

[c]BHC: behavioral health clinician.

Data were gathered by using a range of methods. Acceptability and feasibility data were gathered via questionnaires, focus groups with young people, interviews with clinic staff, and exit surveys. Transcripts of audio recordings of consultations, focus

groups, and free text in surveys were used to obtain qualitative data on the effect that screens had on consultations. Deidentified pre- and postintervention data from electronic health records, Likert-style survey questions, and yes-no survey questions provided quantitative data. Further information about the study measures, potential biases, limitations, and strengths of each study are summarized in Table 3.

Findings of Studies

Summary of Studies' Findings

None of the reviewed studies had changes in screening rates as the main focus. Nonetheless, offering access to a web-based screening tool increased screening rates in all of the studies except one, in which access to the screening tool was provided at the end of the consult. In this study, care providers often forgot to give the link or only gave it to youth who they perceived to be at high risk [18]. When the screen covered several domains, multiple risk behaviors were disclosed by over one-third of young people.

Preintervention Training

Preintervention education was available to participating care providers in 7 studies and was not discussed in the remainder of the studies. Care providers in one study received no formal training, although they were supplied with printed instructional materials on the guidelines for the screening, management, and treatment of chlamydia. In other studies, the research team trained staff to use the screening tool and offered support and resources to guide the delivery of brief interventions. When care providers attended 2 or more of these education sessions, the likelihood of e-screening for psychosocial issues taking place and brief interventions being delivered increased.

Barriers to Using Web-Based Screening Tools

Despite the considerable heterogeneity of these studies, commentaries on barriers to use were successfully extracted from all but one study. Barriers, which were identified by young people, to using web-based screening tools were only mentioned in 1 study. In this study, youth perceived a lack of privacy when completing the screen in the waiting room [25]. However, all but one study [23] identified barriers preventing care providers from routinely e-screening youth for psychosocial issues. The cited barriers included a lack of time, knowledge, training, and awareness of referral options [18,24,25,28]. Some care providers were uncomfortable with raising sensitive issues with young people, as they were concerned that youth might be too embarrassed or worried about confidentiality to discuss psychosocial issues with them [18,20]. Additionally, a lack of staff and high staff turnover [24,28] resulted in a barrier to screening, and in one study, staff were worried that technology could impair face-to-face engagement with young people [19].

Effect on Consultation

In two studies, care providers found that they were able to include e-screening and brief counseling into the time allocated for standard consultations [25,26]. Following the completion of a preconsultation e-screen, there was a nonsignificant increase or no increase in consultation length; care providers felt that a slightly longer appointment was acceptable, given the increased disclosure of psychosocial issues [23-26,28]. Reviewing the results of e-screens helped care providers to plan consultations, set priorities, and engage with youth in useful discussions [19,22-24]. Adolescents believed that completing a screen by using a computer or mobile device afforded them increased privacy and confidentiality, which increased the likelihood of them disclosing psychosocial issues. In consultations, young people felt listened to, felt encouraged to talk, and felt that all of the issues that they wished to discuss had been addressed. Young people reported that the delivery and quality of brief interventions improved, and their satisfaction with care increased.

Acceptability and Feasibility

e-Screening for psychosocial concerns was found to be acceptable in 7 studies and was generally feasible to implement. However, all studies concluded that more research is needed into making e-screening for youth psychosocial issues feasible in primary care.

Discussion

Principal Results

More than one-third of adolescents engage in multiple risk behaviors [19,24]; therefore, the ability to conduct screening across several domains quickly and effectively in primary care might help with detecting issues that are not typically screened for by care providers. Multi-item e-screening tools for youth psychosocial issues have the potential to facilitate increased disclosure and, hence, early intervention in primary care settings. This review found 12 papers describing 11 studies that were carried out in a variety of settings in high-income countries. A range of study designs were used to evaluate the acceptability and feasibility of implementing e-screening tools for youth psychosocial issues in primary care settings. A lack of time is the most common barrier to screening among care providers; yet, when this was measured preconsultation, e-screening and subsequent discussion made little to no difference in consultation length [19,23-26,28]. The review of an e-screening report during ensuing discussions allows care providers to raise subjects that they may otherwise have found difficult to discuss. Despite concerns that young people may not want to address psychosocial issues in their consultations, youth participants reported increased satisfaction and felt more involved with their care when such discussions were initiated by their care providers. Additionally, reviewing e-screening results with young people directs discussions toward psychosocial issues and better meets the unique health and well-being needs of youth.

The reviewed studies found that while e-screening in primary care is effective in detecting youth psychosocial issues and enabling timely brief interventions, common barriers (a lack of time, training, tools, and staff and discomfort in raising sensitive issues) to their use persist [3,13,15,18,24,25,28,29]. The initiation of screens by a research assistant creates an artificial environment, and the initiation of screening does not become a part of daily workflows. Further, because some staff believe that e-screening requires extra resources, they may resist its integration into daily practice. The use of e-screening tools

under research conditions does not represent their use within real, complex clinical environments, where context-specific barriers and challenges can inhibit the assimilation of e-screening tools into routine practice [30-39]. To overcome barriers to use, e-screening tools must be acceptable to intended users and involve usage processes that are feasible and can be easily assimilated into routine use [30,40-44]. Further, to reduce inequities among indigenous youth, screening tools also need to be culturally appropriate.

Comparison With Prior Work

Existing evidence suggests that youth would prefer to complete an initial self-assessment electronically (e-screening) rather than undergo a face-to-face interview [21,45-47]. e-Screening not only saves staff time but also provides reliable and consistent results. Young people who believe that computers provide more privacy may disclose more sensitive issues in e-screens than they would in face-to-face interviews. Adolescents perceive e-screening as an appropriate method of collecting information in clinical settings [23,48-50]. Youth prefer to complete an e-screen in the waiting room prior to consultations with their clinicians [51]. This augments engagement with care providers, increases disclosure, and facilitates shared goal setting in ensuing consultations [52]. e-Screening for youth psychosocial issues in primary health care can improve health outcomes and help to reduce the incidence rates of youth suicide, self-harm, accidental death, and mental health issues [53].

For e-screening tools to be effective in improving patient outcomes, their use must become established in routine clinical practice. This challenges all clinic staff, individually and collectively, to make some degree of change in their ways of working and interacting with colleagues and patients.

The implementation of complex web-based interventions, such as e-screening, in particular clinical settings is influenced by how well these interventions are accepted, users' perceptions of the benefits and barriers of these interventions' uptake, and the impact that using these interventions has on the workflows of potential users and existing systems of practice [39].

As an implementation theory of action, the Normalization Process Theory [54] consists of 4 constructs (coherence, cognitive participation, collective action, and reflexive monitoring) that outline what intended users need to do to make sense of, commit to, engage with, and evaluate complex interventions [41,55-57]. The successful implementation of e-screening needs to begin by working in collaboration with stakeholders, community and cultural leaders, and end users, so that interventions are tailored to be acceptable and feasible for use in each specific setting. This co-design approach gives researchers a unique insight into the challenges faced by users in any given setting. Further, in a co-design approach, the experience, knowledge, and skills of users are used to inform the development of implementation processes and overcome context-specific barriers.

Strengths and Limitations

This study's strengths include the searching of 4 databases and a hand search, which were conducted to find studies for inclusion in this review. Explicit inclusion and exclusion criteria were used to identify the 12 studies that were finally reviewed, and to ensure that a comprehensive search was conducted, expansive search strings were developed. However, there are limitations to this review. There is a paucity of literature in this area, and most studies had considerable limitations to their methodologies and generalizability. Further, most of our findings pertained to only a subset of the reviewed studies. Finally, the heterogeneity of the studies included in this review precluded the ability to conduct a meta-analysis.

Conclusion

The efficacy and acceptability of using e-screening tools are not in doubt. Nonetheless, their use in practice is sporadic and is often limited to youth who are considered to be at high risk [18,22]. The feasibility of implementing e-screening does not only rely on the availability of appropriate technological infrastructures; the effect that e-screening tools have on those who use them is also crucial to their efficacy [40]. When those who conduct screening recognize that there are clear benefits, such as improving the health outcomes of patients while reducing workloads, then the routine use of e-screening becomes viable [40,54]. To be truly effective, screening tools must be implemented in clinical settings, and their use must become a part of routine practice [30,40]. Co-designing and tailoring e-screening tools and processes to meet the needs of specific clinical contexts may be required to enable clinicians to overcome perceived barriers and integrate the use of e-screening processes into their practices.

Acknowledgments

This review was undertaken as part of the doctoral study of RM at the University of Auckland, New Zealand, and is supported by the University of Auckland Health Advances through Behavioural Intervention Technologies Digital Mental Health Doctoral Scholarship in Psychological Medicine, the University of Auckland Scholarship in General Practice, and a New Zealand Nursing Education and Research Foundation McCutchan Trust Scholarship.

Conflicts of Interest

FGS and RM are involved in the refinement and implementation of the psychosocial screening tool YouthCHAT. MS has no competing or potential conflicts of interest to declare.

Multimedia Appendix 1
Search strings.

[DOCX File , 15 KB - mental_v8i11e30479_app1.docx]

References

1. Coker T, Sareen H, Chung P, Kennedy DP, Weidmer BA, Schuster MA. Improving access to and utilization of adolescent preventive health care: the perspectives of adolescents and parents. J Adolesc Health 2010 Aug;47(2):133-142. [doi: 10.1016/j.jadohealth.2010.01.005] [Medline: 20638005]
2. Mariu KR, Merry SN, Robinson EM, Watson PD. Seeking professional help for mental health problems, among New Zealand secondary school students. Clin Child Psychol Psychiatry 2012 Apr;17(2):284-297. [doi: 10.1177/1359104511404176] [Medline: 21852315]
3. Sanci L, Grabsch B, Chondros P, Shiell A, Pirkis J, Sawyer S, et al. The prevention access and risk taking in young people (PARTY) project protocol: a cluster randomised controlled trial of health risk screening and motivational interviewing for young people presenting to general practice. BMC Public Health 2012 Jun 06;12:400 [FREE Full text] [doi: 10.1186/1471-2458-12-400] [Medline: 22672481]
4. Taliaferro LA, Borowsky IW. Beyond prevention: promoting healthy youth development in primary care. Am J Prev Med 2012 Jun;42(6 Suppl 2):S117-S121. [doi: 10.1016/j.amepre.2012.03.017] [Medline: 22704429]
5. Crengle S, Clark T, Robinson E. The health and wellbeing of Maori New Zealand secondary school students in 2012. In: Te Ara Whakapiki Taitamariki. Youth 2000 series. Auckland, New Zealand: The University of Auckland; 2013.
6. Bradford S, Rickwood D. Acceptability and utility of an electronic psychosocial assessment (myAssessment) to increase self-disclosure in youth mental healthcare: a quasi-experimental study. BMC Psychiatry 2015 Dec 01;15:305 [FREE Full text] [doi: 10.1186/s12888-015-0694-4] [Medline: 26627041]
7. Wissow LS, Brown J, Fothergill KE, Gadomski A, Hacker K, Salmon P, et al. Universal mental health screening in pediatric primary care: a systematic review. J Am Acad Child Adolesc Psychiatry 2013 Nov;52(11):1134-1147.e23 [FREE Full text] [doi: 10.1016/j.jaac.2013.08.013] [Medline: 24157388]
8. Brown JD, Wissow LS. Screening to identify mental health problems in pediatric primary care: considerations for practice. Int J Psychiatry Med 2010;40(1):1-19. [doi: 10.2190/PM.40.1.a] [Medline: 20565041]
9. Hacker K, Arsenault L, Franco I, Shaligram D, Sidor M, Olfson M, et al. Referral and follow-up after mental health screening in commercially insured adolescents. J Adolesc Health 2014 Jul;55(1):17-23. [doi: 10.1016/j.jadohealth.2013.12.012] [Medline: 24525226]
10. Thabrew H, Corter A, Goodyear-Smith F, Goldfinch M. Randomized trial comparing the electronic composite psychosocial screener YouthCHAT with a clinician-interview assessment for young people: A study protocol. JMIR Res Protoc 2017 Jul 31;6(7):e135 [FREE Full text] [doi: 10.2196/resprot.7995] [Medline: 28760727]
11. Buckley S, McDonald J, Mason D, Gerring Z, Churchward M, Cumming J. Nursing services in New Zealand secondary schools. Victoria University of Wellington. 2009. URL: https://www.wgtn.ac.nz/health/centres/health-services-research-centre/docs/reports/downloads/Nursing-Services-in-Schools-Edited-19-Feb.pdf [accessed 2021-10-24]
12. Evaluation of Healthy Community Schools Initiative in AIMHI schools. Ministry of Health. 2009. URL: https://www.health.govt.nz/system/files/documents/publications/aimhi-08-10-09.pdf [accessed 2021-10-21]
13. Webb MJ, Kauer SD, Ozer EM, Haller DM, Sanci LA. Does screening for and intervening with multiple health compromising behaviours and mental health disorders amongst young people attending primary care improve health outcomes? A systematic review. BMC Fam Pract 2016 Aug 04;17:104 [FREE Full text] [doi: 10.1186/s12875-016-0504-1] [Medline: 27488823]
14. Connors EH, Arora P, Blizzard AM, Bower K, Coble K, Harrison J, et al. When behavioral health concerns present in pediatric primary care: Factors influencing provider decision-making. J Behav Health Serv Res 2018 Jul;45(3):340-355. [doi: 10.1007/s11414-017-9580-9] [Medline: 29209899]
15. Ham P, Allen C. Adolescent health screening and counseling. Am Fam Physician 2012 Dec 15;86(12):1109-1116. [Medline: 23316983]
16. Webb MJ, Wadley G, Sanci LA. Experiences of general practitioners and practice support staff using a health and lifestyle screening app in primary health care: Implementation case study. JMIR Mhealth Uhealth 2018 Apr 24;6(4):e105 [FREE Full text] [doi: 10.2196/mhealth.8778] [Medline: 29691209]
17. Global Accelerated Action for the Health of Adolescents (AA-HA!): guidance to support country implementation. World Health Organization. 2017. URL: https://www.who.int/publications/i/item/9789241512343 [accessed 2021-10-24]
18. Bilardi JE, Sanci LA, Fairley CK, Hocking JS, Mazza D, Henning DJ, et al. The experience of providing young people attending general practice with an online risk assessment tool to assess their own sexual health risk. BMC Infect Dis 2009 Mar 12;9:29 [FREE Full text] [doi: 10.1186/1471-2334-9-29] [Medline: 19284635]
19. Goodyear-Smith F, Corter A, Suh H. Electronic screening for lifestyle issues and mental health in youth: a community-based participatory research approach. BMC Med Inform Decis Mak 2016 Nov 08;16(1):140 [FREE Full text] [doi: 10.1186/s12911-016-0379-z] [Medline: 27821128]
20. Curtis BL, McLellan AT, Gabellini BN. Translating SBIRT to public school settings: an initial test of feasibility. J Subst Abuse Treat 2014 Jan;46(1):15-21. [doi: 10.1016/j.jsat.2013.08.001] [Medline: 24029623]
21. Harris SK, Knight JR. Putting the screen in screening: Technology-based alcohol screening and brief interventions in medical settings. Alcohol Res 2014;36(1):63-79 [FREE Full text] [Medline: 26259001]

22. Diamond G, Levy S, Bevans KB, Fein JA, Wintersteen MB, Tien A, et al. Development, validation, and utility of internet-based, behavioral health screen for adolescents. Pediatrics 2010 Jul;126(1):e163-e170. [doi: 10.1542/peds.2009-3272] [Medline: 20566613]
23. Gadomski AM, Fothergill KE, Larson S, Wissow LS, Winegrad H, Nagykaldi ZJ, et al. Integrating mental health into adolescent annual visits: impact of previsit comprehensive screening on within-visit processes. J Adolesc Health 2015 Mar;56(3):267-273 [FREE Full text] [doi: 10.1016/j.jadohealth.2014.11.011] [Medline: 25586231]
24. Olson AL, Gaffney CA, Hedberg VA, Gladstone GR. Use of inexpensive technology to enhance adolescent health screening and counseling. Arch Pediatr Adolesc Med 2009 Feb;163(2):172-177. [doi: 10.1001/archpediatrics.2008.533] [Medline: 19188650]
25. Riese A, Mello MJ, Baird J, Steele DW, Ranney ML. Prompting discussions of youth violence using electronic previsit questionnaires in primary care: a cluster randomized trial. Acad Pediatr 2015;15(3):345-352. [doi: 10.1016/j.acap.2015.01.005] [Medline: 25906703]
26. Sterling S, Kline-Simon AH, Jones A, Satre DD, Parthasarathy S, Weisner C. Specialty addiction and psychiatry treatment initiation and engagement: Results from an SBIRT randomized trial in pediatrics. J Subst Abuse Treat 2017 Nov;82:48-54 [FREE Full text] [doi: 10.1016/j.jsat.2017.09.005] [Medline: 29021115]
27. Sterling S, Kline-Simon AH, Satre DD, Jones A, Mertens J, Wong A, et al. Implementation of screening, brief intervention, and referral to treatment for adolescents in pediatric primary care: A cluster randomized trial. JAMA Pediatr 2015 Nov;169(11):e153145 [FREE Full text] [doi: 10.1001/jamapediatrics.2015.3145] [Medline: 26523821]
28. Harris SK, Csémy L, Sherritt L, Starostova O, Van Hook S, Johnson J, et al. Computer-facilitated substance use screening and brief advice for teens in primary care: an international trial. Pediatrics 2012 Jun;129(6):1072-1082 [FREE Full text] [doi: 10.1542/peds.2011-1624] [Medline: 22566420]
29. Population of Northland DHB. Ministry of Health NZ. URL: https://www.health.govt.nz/new-zealand-health-system/my-dhb/northland-dhb/population-northland-dhb [accessed 2021-10-24]
30. Murray E, Treweek S, Pope C, MacFarlane A, Ballini L, Dowrick C, et al. Normalisation process theory: a framework for developing, evaluating and implementing complex interventions. BMC Med 2010 Oct 20;8:63 [FREE Full text] [doi: 10.1186/1741-7015-8-63] [Medline: 20961442]
31. Smith MJ, Liehr P. Understanding middle range theory by moving up and down the ladder of abstraction. In: Middle Range Theory for Nursing. New York: Springer Publishing Company; 2014:21.
32. Risjord M. Middle-range theories as models: New criteria for analysis and evaluation. Nurs Philos 2019 Jan;20(1):e12225. [doi: 10.1111/nup.12225] [Medline: 30264549]
33. Kaidesoja T. A dynamic and multifunctional account of middle-range theories. Br J Sociol 2019 Sep;70(4):1469-1489. [doi: 10.1111/1468-4446.12621] [Medline: 30460990]
34. Case D. Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior. Bingley, UK: Emerald Group Publishing Ltd; 2016.
35. Worum H, Lillekroken D, Ahlsen B, Roaldsen KS, Bergland A. Bridging the gap between research-based knowledge and clinical practice: a qualitative examination of patients and physiotherapists' views on the Otago exercise Programme. BMC Geriatr 2019 Oct 21;19(1):278 [FREE Full text] [doi: 10.1186/s12877-019-1309-6] [Medline: 31638912]
36. Lenz ER, Suppe F, Gift AG, Pugh LC, Milligan RA. Collaborative development of middle-range nursing theories: toward a theory of unpleasant symptoms. ANS Adv Nurs Sci 1995 Mar;17(3):1-13. [doi: 10.1097/00012272-199503000-00003] [Medline: 7778887]
37. Ridde V, Pérez D, Robert E. Using implementation science theories and frameworks in global health. BMJ Glob Health 2020 Apr 16;5(4):e002269 [FREE Full text] [doi: 10.1136/bmjgh-2019-002269] [Medline: 32377405]
38. Hassan NR, Lowry PB. Seeking middle-range theories in information systems research. 2015 Presented at: International Conference on Information Systems (ICIS 2015); December 13-18, 2015; Fort Worth, Texas.
39. May C, Sibley A, Hunt K. The nursing work of hospital-based clinical practice guideline implementation: an explanatory systematic review using Normalisation Process Theory. Int J Nurs Stud 2014 Feb;51(2):289-299 [FREE Full text] [doi: 10.1016/j.ijnurstu.2013.06.019] [Medline: 23910398]
40. May CR, Mair F, Finch T, MacFarlane A, Dowrick C, Treweek S, et al. Development of a theory of implementation and integration: Normalization Process Theory. Implement Sci 2009 May 21;4:29 [FREE Full text] [doi: 10.1186/1748-5908-4-29] [Medline: 19460163]
41. May C, Finch T. Implementing, embedding, and integrating practices: An outline of normalization process theory. Sociology 2009 Jun 15;43(3):535-554. [doi: 10.1177/0038038509103208]
42. May C, Finch T, Mair F, Ballini L, Dowrick C, Eccles M, et al. Understanding the implementation of complex interventions in health care: the normalization process model. BMC Health Serv Res 2007 Sep 19;7:148 [FREE Full text] [doi: 10.1186/1472-6963-7-148] [Medline: 17880693]
43. May CR, Cummings A, Girling M, Bracher M, Mair FS, May CM, et al. Using Normalization Process Theory in feasibility studies and process evaluations of complex healthcare interventions: a systematic review. Implement Sci 2018 Jun 07;13(1):80 [FREE Full text] [doi: 10.1186/s13012-018-0758-1] [Medline: 29879986]

44. May CR, Johnson M, Finch T. Implementation, context and complexity. Implement Sci 2016 Oct 19;11(1):141 [FREE Full text] [doi: 10.1186/s13012-016-0506-3] [Medline: 27756414]
45. Bradford S, Rickwood D. Young people's views on electronic mental health assessment: Prefer to type than talk? J Child Fam Stud 2015;24(5):1213-1221 [FREE Full text] [doi: 10.1007/s10826-014-9929-0] [Medline: 25960628]
46. Watson PD, Denny SJ, Adair V, Ameratunga SN, Clark TC, Crengle SM, et al. Adolescents' perceptions of a health survey using multimedia computer-assisted self-administered interview. Aust N Z J Public Health 2001 Dec;25(6):520-524. [doi: 10.1111/j.1467-842x.2001.tb00316.x] [Medline: 11824987]
47. Chisolm DJ, Gardner W, Julian T, Kelleher KJ. Adolescent satisfaction with computer-assisted behavioural risk screening in primary care. Child Adolesc Ment Health 2008 Nov 01;13(4):163-168 [FREE Full text] [doi: 10.1111/j.1475-3588.2007.00474.x] [Medline: 19779573]
48. Renker PR. Breaking the barriers: the promise of computer-assisted screening for intimate partner violence. J Midwifery Womens Health 2008;53(6):496-503. [doi: 10.1016/j.jmwh.2008.07.017] [Medline: 18984505]
49. Harris F, Macfarlane S, Macfarlane A, Jolly M. Māori values in the workplace: Investing in diversity. MAI Journal 2016 Jun 13;5(1):48-62. [doi: 10.20507/maijournal.2016.5.1.4]
50. Kadivar H, Thompson L, Wegman M, Chisholm T, Khan M, Eddleton K, et al. Adolescent views on comprehensive health risk assessment and counseling: assessing gender differences. J Adolesc Health 2014 Jul;55(1):24-32 [FREE Full text] [doi: 10.1016/j.jadohealth.2013.12.002] [Medline: 24613096]
51. Jasik CB, Berna M, Martin M, Ozer EM. Teen preferences for clinic-based behavior screens: Who, where, when, and how? J Adolesc Health 2016 Dec;59(6):722-724. [doi: 10.1016/j.jadohealth.2016.08.009] [Medline: 27884300]
52. Amaro H, Reed E, Rowe E, Picci J, Mantella P, Prado G. Brief screening and intervention for alcohol and drug use in a college student health clinic: feasibility, implementation, and outcomes. J Am Coll Health 2010;58(4):357-364. [doi: 10.1080/07448480903501764] [Medline: 20159759]
53. Goodyear-Smith F, Martel R, Darragh M, Warren J, Thabrew H, Clark TC. Screening for risky behaviour and mental health in young people: the YouthCHAT programme. Public Health Rev 2017 Oct 13;38:20 [FREE Full text] [doi: 10.1186/s40985-017-0068-1] [Medline: 29450092]
54. May C. A rational model for assessing and evaluating complex interventions in health care. BMC Health Serv Res 2006 Jul 07;6:86 [FREE Full text] [doi: 10.1186/1472-6963-6-86] [Medline: 16827928]
55. McNaughton RJ, Steven A, Shucksmith J. Using Normalization Process Theory as a practical tool across the life course of a qualitative research project. Qual Health Res 2020 Jan;30(2):217-227. [doi: 10.1177/1049732319863420] [Medline: 31347440]
56. Ristea I. Reflections on mechanisms influencing human behavior. Procedia Soc Behav Sci 2013 Oct 10;92:799-805. [doi: 10.1016/j.sbspro.2013.08.757]
57. May CR, Finch T, Ballini L, MacFarlane A, Mair F, Murray E, et al. Evaluating complex interventions and health technologies using normalization process theory: development of a simplified approach and web-enabled toolkit. BMC Health Serv Res 2011 Sep 30;11:245 [FREE Full text] [doi: 10.1186/1472-6963-11-245] [Medline: 21961827]

Abbreviations

HEEADSSS: Home, Education/Employment, Eating, Activities, Drugs and Alcohol, Sexuality, Suicide/Depression, and Safety
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by J Torous; submitted 17.05.21; peer-reviewed by E Sturgiss, J Piqueras; comments to author 25.06.21; revised version received 09.07.21; accepted 12.07.21; published 19.11.21.

Please cite as:
Martel R, Shepherd M, Goodyear-Smith F
Implementing the Routine Use of Electronic Mental Health Screening for Youth in Primary Care: Systematic Review
JMIR Ment Health 2021;8(11):e30479
URL: https://mental.jmir.org/2021/11/e30479
doi:10.2196/30479
PMID:34807833

Review

Self-directed Technology-Based Therapeutic Methods for Adult Patients Receiving Mental Health Services: Systematic Review

Anthony Saad[1], BSc, MPH, MD; Deanna Bruno[1,2], MD; Bettina Camara[3], BSc; Josephine D'Agostino[1], RSSWK; Blanca Bolea-Alamanac[1,2], MSc, MD, PhD

[1]Department of Psychiatry, Women's College Hospital, Toronto, ON, Canada

[2]Department of Psychiatry, University of Toronto, Toronto, ON, Canada

[3]Department of Immunology, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Blanca Bolea-Alamanac, MSc, MD, PhD
Department of Psychiatry
Women's College Hospital
76 Grenville St
Toronto, ON, M5S 1B2
Canada
Phone: 1 6473312707
Email: Blanca.BoleaAlamanac@wchospital.ca

Abstract

Background: Technological interventions used to treat illnesses and promote health are grouped under the umbrella term of digital therapeutics. The use of digital therapeutics is becoming increasingly common in mental health. Although many technologies are currently being implemented, research supporting their usability, efficacy, and risk requires further examination, especially for those interventions that can be used without support.

Objective: This review aims to identify the evidence-based, self-directed, technology-based methods of care that can be used in adult patients after they are discharged from mental health services. The interventions reviewed are automated with no human input required (either at the patient's or at the technology's end), so the patients can implement them without any support.

Methods: A systematic review was conducted according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) and PROSPERO (International Prospective Register of Systematic Reviews) guidelines in 3 databases: PubMed, Web of Science, and OVID. The inclusion criteria were self-directed, automated, and technology-based interventions related to mental health, primarily for adults, having a solid evaluation process. The interventions had to be *self-directed*, in that the participants could use the technology without any external guidance.

Results: We identified 36 papers that met the inclusion criteria: 26 randomized controlled trials, 9 nonrandomized controlled trial quantitative studies, and 1 qualitative study. The technologies used included websites, automated text messaging, phone apps, videos, computer software, and integrated voice response. There were 22 studies focused on internet-based cognitive behavioral therapies as a therapeutic paradigm compared with the waitlist, web-based human-delivered therapy, and other interventions. Among these studies, 14 used paradigms other than the internet-based cognitive behavioral therapy. Of the 8 studies comparing guided and unguided digital care, 3 showed no differences, 3 favored guided interventions, and 2 favored unguided interventions. The research also showed that dropout rates were as high as 80%, citing potential problems with the acceptability of the suggested technologies.

Conclusions: There is limited research on the efficacy and suitability of self-directed technology-based care options for mental health. Digital technologies have the potential to bridge the gap between ambulatory care and independent living. However, these interventions may need to be developed collaboratively with the users to encourage their acceptability and to avoid high dropout rates.

(JMIR Ment Health 2021;8(11):e27404) doi:10.2196/27404

KEYWORDS

digital therapeutics; self-directed; mental health; telehealth; technology; mobile applications; telemedicine; internet; mobile phone

Introduction

Background

Health care systems have changed dramatically over the last 50 years. The COVID-19 pandemic has specifically disrupted the traditional health care delivery model. New methods of care have been developed that can be delivered safely and that complement and improve the way treatment is provided both in and outside the physician's office. The technological interventions used to treat illnesses and to promote health are grouped under the umbrella term of digital therapeutics [1]. There is a growing interest in digital therapeutics and their applications in the field of mental health. Digital forms of treatment have been investigated in various domains of mental health treatment, including psychotherapy, treatment of addictive behavior, medication adherence, e-therapy, obsessive-compulsive disorder, and posttraumatic stress disorder [2-6]. Maintenance of health and prevention of relapse are key concerns in mental health. For example, it is estimated that as many as one-third of patients with depression relapse during the 18 months following their recovery [7]. Mental health practitioners require not only tools that can treat their patients in the short term, but also *postdischarge* tools that will maintain health and prevent relapse. Digital therapeutics, if designed and evaluated appropriately, can be used independent of the health care providers and after having left the care of mental health services [8]. Accessibility is a key advantage of digital therapeutics. Patients who do not have access to traditional care or those who may face stigma in their communities for accessing mental health services can use digital therapeutics to obtain mental health care and avoid these problems [9]. This allows a distinctive approach to mental health practice that may improve the health of not only the individual, but also the entire population, through a better allocation of resources.

It is, therefore, essential to evaluate digital interventions regarding their usability, efficacy, and risk before they are recommended to the public [8]. Patients discharged from mental health services have access to many digital therapeutic options in the free market. They often ask physicians about these technologies and expect their technical appraisals [8]. Physicians are also understandably reluctant to endorse products that may not have been evaluated scientifically.

Objective

Digital therapeutic methods raise issues of privacy, confidentiality, and the possible weakening of the clinician-patient relationship. Therefore, such technologies may not be accepted by potential users. It has also been suggested that the discord between the systematic nature of new technologies and the psychiatrists' professional culture may lead to a *disruption* in mental health practice [10]. Therefore, there is a need for evidence-based research into digital therapeutics. Although other systematic reviews have examined the evidence for self-guided interventions in the past, those reviews differ in some respects to this review. Many focused on only 1 mental health condition (eg, depression), studied only 1 digital modality (eg, internet-based cognitive behavioral therapy [iCBT]), or examined interventions that were not truly independent or self-directed [11-15]. The aim of this review is to identify the self-directed digital technologies (eg, apps and websites) used to treat mental health conditions in adults with published evidence of evaluation at any level (qualitative or quantitative). The motivation was to find evidence-based *digital therapeutics* that could be used by patients after their discharge from mental health services. Once the patients are discharged, they may not remain under the guidance of mental health care professionals. Therefore, we sought the interventions that were suitable for independent use by the patients.

Methods

The research question for this study can be summarized as follows: *What self-directed digital therapeutic options can be used by adult patients receiving psychiatric care and what is the evidence supporting their effectiveness?*

Inclusion and Exclusion Criteria

The inclusion criteria were: (1) the studies evaluated a technology that was an internet-based or remote communication-based intervention for mental health, (2) the studies had at least 1 part or group that was self-directed (ie, the patient could perform the intervention on their own), (3) the study participants were at least 18 years of age, and (4) the studies had an evaluation component (ie, the effect, acceptability, usability, or feasibility of the technology-based intervention was studied). The studies were excluded if: (1) their primary outcome was not related to mental health or to participants with a mental health diagnosis; (2) the intervention was not completely automated (ie, required other human input for the treatment to be administered in full); (3) they had a group therapy or group forum component, as this was not deemed truly independent because group work often requires mediation and moderation by a specialist. However, the studies were not excluded if the assistance provided was carefully documented as entirely technical in nature (ie, not considered part of a therapeutic treatment).

When including the studies in this review, we enforced a strict *self-directed* criterion. Studies were only included if a digital intervention was given to at least 1 study group without any notable human support. We defined human support as any interaction between the patient and the health care team, which can be interpreted as a treatment that is psychologically beneficial. This was done to simulate the conditions of real-life practice in which the patients would use these technologies independently, without any support.

Database Review

Three primary databases were used in this review: PubMed, Web of Science and OVID. The primary purpose of using OVID was to identify papers not captured by PubMed and Web of Science, using the *National Library of Medicine's MEDLINE and former HealthSTAR databases* (per the OVID description page). However, as all the articles that were found in OVID either overlapped with PubMed or were ultimately excluded by our criteria, we felt assured that we had thoroughly assessed the current literature on the aforementioned topic. Appropriate keywords, including MeSH (Medical Subject Headings) terms,

were used in searching the databases. The search was conducted on November 8, 2019 and included articles from the respective databases' inception. The earliest study dated back to 1995. However, only the articles published in English were included in the study. The review broadly followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) and PROSPERO (International Prospective Register of Systematic Reviews) guidelines [16]. Conventional systematic review methods were applied to this paper, including screening by title and abstract, as well as full text review. We applied a double-coding systematic review procedure, with 2 separate reviewers assessing each article. We also followed the PRISMA guidelines and completed the checklist [16]. Automated tools, beyond conventional bibliographical methods, were not used in this study. Database software was used to organize and review the studies [17].

Levels of Evidence

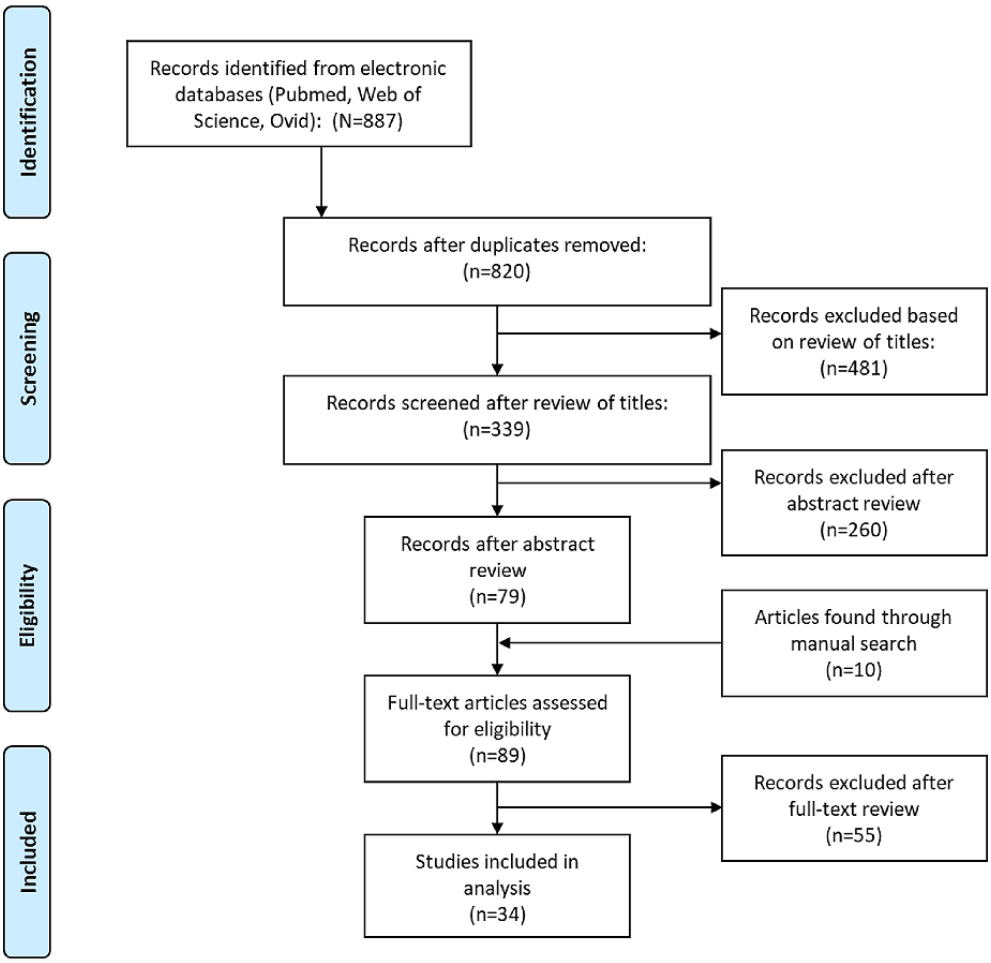
This review uses the Oxford Centre for Evidence-Based Medicine—levels of evidence (LOE) [18]. Oxford Centre for Evidence-Based Medicine has set out a methodology for systematizing the process of evaluating evidence. A number and letter grading system is used, with a designation of *1a* being the highest level (for systematic reviews with homogeneity) and a rating of *5* being the lowest (expert opinion and qualitative only studies).

Results

Overview

A total of 889 articles were identified on searching the databases. Using the PRISMA screening process, 36 studies were included in this review: 26 (72%) were randomized controlled trials (RCTs), 9 (25%) were non-RCT quantitative studies, and 1 (3%) was a qualitative study. This process flow has been illustrated in Figure 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram for each step of the screening process.



Many studies were identified as iCBTs; therefore, these studies were examined as a group. Tables 1 and 2 describe the interventions studied, whereas further information about the studies has been summarized in Multimedia Appendix 1 [2,19-39] and Multimedia Appendix 2 [40-53].

Table 1. Description of interventions found in the 22 internet-based cognitive behavioral therapy (iCBT) studies examined.

Author (year)	Type of study	Intervention	Number and type of sessions	Description of study design
Batterham et al [19] (2018)	RCT[a]	FindMindKit: email delivered, CBT[b]- and web-based modules	18 modules in total, with 2 symptom-specific modules: for symptoms of fear disorders, distress or mood disorders, suicidal ideation, and substance disorders. Modules have scenarios and a fictional character serving as a role model and as an expert narrator. Modules were followed by a worksheet for practice.	Participants were divided into 3 groups: one receiving the personalized FindMindKit modules, one receiving the generalized modules, and an attention control group that received access to a control mental health program.
Berger et al [20] (2017)	RCT	Velibra: CBT-based web program for anxiety	6 sessions: transdiagnostic measures of treating anxiety, a form of treatment that applies similar principles across mental disorders without tailoring to specific diagnoses (eg, same treatments for GAD[c] and social phobia). These could be tailored automatically following the user's responses.	The intervention group had access to Velibra; the control group received access after the study was completed.
Berger et al [21] (2011)	RCT	Deprexis: self-help iCBT website	10 modules and a summary session: the content is mainly text-based, with illustrations, exercises, and user response feedback. Subsequent content is automatically tailored by the program.	All the participants had access to Deprexis. The unguided group received access without support. The guided group also received a scheduled weekly email feedback with a therapist and the freedom to contact that therapist at will.
Botella et al [22] (2016)	RCT	Smiling is fun: web-delivered, CBT-based self-help program for the treatment of depression [38].	No sessions; the website contains general multimedia, images, and an interactive platform.	Both the intervention groups had access to Smiling is fun. One of the intervention groups also had access to EEG,[d] EKG,[e] and ACT[f] sensors to monitor the users' cognitive, physiological, and physical states, as well as provide feedback. The control group did not have access to iCBT or the sensors.
Brettschneider et al [23] (2015)	Cross-sectional	Web-based text and video program for social anxiety	8 sessions: has scenarios and a fictional character serving as a role model, who also provides automatically generated feedback	All participants used the program; there was no control condition.
Christensen et al [24] (2014)	RCT	Website with CBT-based program for anxiety	10 sessions, having 1 session per week: CBT education and CBT techniques (weeks 1-7), relaxation (weeks 8-9), and physical activity promotion (week 10)	The study had 4 parts. The active condition only used the website. The control condition was a website that provided only general information on anxiety and general health. The *call* condition had a weekly telephone call, with a progress check and a reminder to use the program. The email condition had a weekly reminder via email, with similar content as the call condition.
Donker et al [25] (2013)	Randomized controlled noninferiority trial	MoodGYM: focus on dysfunctional thinking and self-esteem training [26]; CBT e-couch: deals with negative thoughts and behavioral activation; IPT[g] e-couch: focusing on roles and interpersonal deficits	Each group used 1 of the 3 programs for a 4-week period.	This was a 3-part study that compared 2 new iCBT programs to MoodGYM (as a control) for 4 weeks.

Author (year)	Type of study	Intervention	Number and type of sessions	Description of study design
Ebert et al [54] (2015)	RCT	Internet-based recovery training, focusing on psyhoeducation and mindfulness for the treatment of insomnia, with automated, adaptive, and tailored feedback based on user response	6 sessions	The intervention group used the program; the control group was a waitlist condition.
Gilbody et al [26] (2017)	RCT	MoodGYM: iCBT focused on dysfunctional thinking and self-esteem training	6 modules released sequentially, lasting approximately 30-45 minutes each. The participants were asked to complete 1 session per week.	The intervention group also received 8 phone calls from a graduate-level support worker, which consisted of introducing the participant to MoodGYM (first call), provide motivation and help identify the barriers to engagement (second to seventh call), and then consolidate the information and discuss the next steps (eighth call). Control group received MoodGYM without phone calls (no guidance).
Gosling et al [27] (2018)	RCT	SHUTi[h]: web- and CBT-based treatment for insomnia with modules and a sleep diary; HealthWatch: interactive lifestyle website having general health information (eg, nutrition)	N/A[i]	Two-part study comparing *SHUTi* with HealthWatch.
Hagatun et al [28] (2018)	RCT	SHUTi: see Gosling et al [27]; education website: emulates the information presented by general practitioners on insomnia	N/A	Two-part study comparing SHUTi with an education website (control group).
Hagatun et al [29] (2019)	RCT	SHUTi: see Gosling et al [27]; website: information on sleep hygiene and insomnia education	N/A	Two-part study comparing SHUTi with an informational website (control group).
Lien et al [30] (2019)	Posthoc analysis of RCT	SHUTi: see Gosling et al [27]; website: information on sleep hygiene and insomnia education	N/A	Posthoc analysis of Hagatun (2019) comparing morning versus evening persons (ie, persons with either diurnal or nocturnal sleeping habits) in the same treatment groups as the study of comparison.
Lintvedt et al [31] (2013)	RCT	MoodGYM: see [26]; BluePages: website with over 400 pages of evidence-based information on depression	N/A	Two-part study: the intervention group had access to both MoodGYM and BluePages. The control group was a waitlist condition with no intervention.
Lokman et al [32] (2017)	RCT	CDMIs,[j] based on CBT techniques	3-4 web-based, unguided self-help modules	Two-part study comparing CDMIs to a waitlist control condition.
Loughnan et al [33] (2019)	RCT	MUMentum: pregnancy-focused, CBT-based program for antenatal depression and anxiety (illustrated, story-based exercises)	4-week unguided programs	Two-part study comparing a TAU[k] control group with an intervention group that was provided access to MUMentum.
Mewton et al [34] (2013)	Cross-sectional	This Way Up: fully automated, unassisted web-based CBT program	N/A	All the participants were offered access to This Way Up.
Moloney et al [35] (2019)	Cross-sectional	SHUTi [27]	9 weeks	All the participants had access to SHUTi.
Noguchi et al [36] (2017)	RCT	Simplified iCBT: 5-minute exercise; sEFM[l]: based on taking time to feel negative thoughts and emotions without judgment	N/A	Three-part study comparing iCBT, sEFM, and a waiting list control group.

Author (year)	Type of study	Intervention	Number and type of sessions	Description of study design
Proudfoot et al [37] (2013)	RCT	myCompass: fully automated self-help monitoring system, completed via mobile phone or computer	12 modules, 10 minutes each in length, comprised of skill-building activities.	Participants were randomly placed in 3 groups: the myCompass intervention group; the attention control group, which received a control mental health program; and the waitlist group that did not receive access to the intervention until after the study period.
Romero-Sanchiz et al [38] (2017)	RCT	Smiling is fun: CBT-based, self-help program for depression [22]	N/A	Participants were either TSG,[m] or were provided with LITG,[n] which involved emails sent from a therapist offering support with the program.
Van Kessel et al [39] (2016)	RCT	MSInvigor8: CBT-based internet program	8 sessions for each group	Two-part study: MSInvigor8-Plus received regular email support from a trained clinical psychologist, while MSInvigor8-Only did not receive any support except the iCBT program.

[a]RCT: randomized controlled trial.

[b]CBT: cognitive behavioral therapy.

[c]GAD: generalized anxiety disorder.

[d]EEG: electroencephalogram.

[e]EKG: electrocardiogram.

[f]ACT: actigraphy.

[g]IPT: interpersonal therapy.

[h]SHUTi: sleep healthy using the internet.

[i]N/A: not applicable.

[j]CDMI: complaint-directed mini-intervention.

[k]TAU: treatment as usual.

[l]sEFM: simple mindfulness exercise.

[m]TSG: totally self-guided.

[n]LITG: low-intensity therapist guidance.

Table 2. Description of the interventions found in the non–internet-based cognitive behavioral therapy studies.

Study designs and author (year)	Treatment paradigm	Intervention technology	Description
RCTs[a]			
Aardoom et al [40] (2016)	Psychoeducation	Website	Featback was a website that offered psychoeducation and general information on eating disorders, along with monitoring and tailored feedback (automatically by the program) on progress. Examined 4 dimensions: (1) body dissatisfaction, (2) concern with body weight or shape, (3) unbalanced nutrition and dieting, and (4) binge eating and compensatory behaviors. Therapist support was by email, teleconferencing, or chat.
Bernstein et al [41] (2016)	CBT[b]	Automated text messages (and phone)	All participants received a brochure on the benefits of quitting smoking and a phone number for a smokers' quitline. Intervention participants also received 4 weeks of nicotine patches and gum, a referral faxed to a quitline, and enrollment in SmokefreeTXT, an automatic texting library of 128 texts. Five random messages were sent per day. The evaluation used EMA,[c] allowing users to send feedback to the automated system about mood, craving, use, or health care contact.
Constant et al [42] (2014)	Informational only	Automated text messages	Intervention involved automated text messages starting on the first day. 13 timed text messages were sent with reminders to take medication and to provide information on bleeding, cramping, and side effects. This was compared with SOC,[d] which was abortion counseling (eg, information on mifepristone side effects), administration of mifepristone on site, self-administration at home (1-2 days), and follow-up clinical assessment (2-3 weeks). Intervention group received both the intervention and the SOC.
Kannisto et al [43] (2017)	Informational only	Automated text messages	Intervention was Mobile.Net, a tailored SMS text message system designed for medication adherence and outpatient care in adult patients with psychosis. Participants received semiautomatic texts for 12 months (approximately 10/month, 2-25 text messages) based on preferences. They could decide the amount, timing, frequency, and the content of the messages.
Kleiboer et al [44] (2015)	Problem-solving therapy	Website	Five-part study that looked at varying levels of support with an internet-based, PST[e] for depression and anxiety called Allesondercontrole, which had 5 weekly lessons with exercises guiding on problem-solving in a structured format. Condition 1 received no support, condition 2 received support upon request and condition 3 received weekly support from a coach. Condition 4 did not receive the internet-based treatment but did receive nonspecific support via chat or email. Condition 5 was a waitlist condition with access to a website containing psychoeducation about depression and anxiety.
Mason et al [45] (2012)	Social cognitive theory	Website	Tailored advice consisted of an advice report based on several variables (eg, sex, previous quit attempts, current health, etc). Participants reported a quit date (past or future) and received a progress report 4 weeks later, which included baseline variables, quit date reminders, slip-ups, and changes in variables. Standard reports were generated using similar algorithms but with default content and modal responses and were all identical. Advice reports could be accessed and filled out at the iQUIT website.

Study designs and author (year)	Treatment paradigm	Intervention technology	Description
Pictet et al [46] (2016)	CBM[f]	Website	Three-part study that compared 2 types of cognitive bias modification programs, as well as comparing them to a waitlist condition. Both intervention groups received access to a website that introduced photographic illustrations and audio recordings depicting everyday situations, and then the patients were instructed to imagine the situations. In the Imagery CBM group, the situations always ended positively. In the Control CBM group, the situations ended positively half the time and negatively in the other half.
Sherman et al [47] (2012)	Psychoeducation, crisis intervention model	Videos; Phone support	4-part RCT receiving either: usual care, usual care and videos, telephone counseling, or telephone counseling and videos. Usual care involved office or inpatient visits, offering education, support group access and options for referral. Psychoeducational videos were offered in the institution or in the home, with 4 phase-specific videos on coping with breast cancer diagnoses. Telephone counseling consisted of 4 phase-specific telephone calls conducted by a nurse interventionist trained in telephone counseling approaches. These were also on coping with breast cancer.
Non-RCT quantitative studies			
Ahmedani et al [48] (2015)	Motivational interviewing and CBT	Computer tablet with app	Intervention was a handheld tablet in which an animated narrator interacts with the participants by user input. Responses by the participants on the tablet would lead to varying responses by the program, allowing for branching down unique pathways and feedback tailored specifically to the user. This system combined motivational interviewing and CBT models. Intervention was delivered via a handheld computer tablet with headphones.
Kipping et al [49] (2016)	Informational only	Website	HealthCheck was a patient portal that allowed access of patients to their health care. It included access to EMR,[g] the ability to request medication renewals on the web, view upcoming appointments and educational materials, and access to communication with the providers.
Piette et al [50] (2013)	Informational only (medication adherence)	IVR[h]	CarePartner program (Depression Version) was an IVR system that monitored the patients' depression symptoms using PHQ[i]-9 and provided advice to improve medication adherence and prompt clinical follow-up. Suicidal ideation led to an alert to the clinical team, instructions to call 911 or the provider, or a suicide hotline. Faxes were sent to the providers when there was a sharp rise in PHQ-9 or medication adherence problems.
Pratap et al [51] (2018)	Cognitive control, problem-solving therapy, informational	App	Three-part study that compared 3 different self-guided phone apps for the treatment of depression. The first group used a video-game inspired app called Project EVO, a cognitive-based program designed to modulate cognitive control abilities. The second app was an iPST[j] program. The third was daily health tips (HTips), a program designed to provide information control to overcome depressed mood through self-care and physical activity. Each app had daily reminders. All programs were self-guided

Study designs and author (year)	Treatment paradigm	Intervention technology	Description
Stein et al [52] (2012)	Informational only (medication adherence)	Computer software	CommonGround was a computerized support system that the participants could use before a medication visit. It included an introductory video about recovery from mental illness and brief videos of patients discussing their recovery. It was followed by a customized survey of the patient's concerns, decisional balance, and trade-off exercises.
Qualitative study			
Bauer et al [53] (2018)	Informational only	App	Ginger.io was a smartphone app with a web-based dashboard with notifications to complete regular clinical surveys, occasional satisfaction surveys, and with health tips (eg, self-care activities) related to depression and anxiety 3-4 times a week. The dashboard allowed for the monitoring of patient app use. Participants used this app while continuing collaborative care treatment, which was care with a general practitioner, a care manager, and a psychiatric consultant.

[a]RCT: randomized controlled trial.

[b]CBT: cognitive behavioral therapy.

[c]EMA: ecological momentary assessment.

[d]SOC: standard of care.

[e]PST: problem-solving treatment.

[f]CBM: cognitive bias modification.

[g]EMR: electronic medical record.

[h]IVR: integrated voice response.

[i]PHQ: patient health questionnaire.

[j]iPST: internet-based problem-solving therapy.

Studies Using iCBTs

Overview

We identified 22 studies that used iCBT (summarized in Table 1 and Multimedia Appendix 1). All these studies incorporated an internet-based program (either via a website or via a program downloaded from a website) that followed cognitive behavioral principles for the treatment of various psychological conditions. Although the websites and programs varied in their content, they all provided access to cognitive behavioral therapy–based modules. In most programs, the users could provide feedback.

The sample sizes varied from 39 to 2413. The targeted populations had diagnoses that varied from insomnia to anxiety and depression. Most studies were RCTs in design (18/22, 82%); 14% (3/22) were cross-sectional, and 5% (1/22) were a posthoc analysis of 1 RCT. While most of the studies were rated 1b for LOE, several were rated as 2b or 2c because they had small sample sizes, had a single part with no comparison, or because they did not report the *P* values or CIs [23,34,35].

iCBT studies could be further subcategorized based on the type of comparison that was made. Of the 18 RCT studies in this category, 6 (33%) compared iCBT against a waitlist condition [2,20,31-33,37]; 5 (28%) studies compared unguided intervention with guided controls [21,24,26,38,39]; 8 (44%) studies compared iCBT with other types of interventions [19,25,27-30,36,37]; 1 (6%) study used a sensor-based approach and compared it to unguided iCBT without a sensor [22]. The 3 non-RCT studies were cross-sectional studies that used a single group to assess the feasibility, accessibility, and preliminary effectiveness of iCBT programs [23,34,35]. The iCBT studies were categorized and reviewed in more detail based on their study design.

Studies Using iCBT: RCTs With a Waitlist Condition Group

A total of 6 RCT studies used a single comparison: participants with access to an iCBT program against participants who either did not have access to any intervention or those who received access to the intervention after the study was completed (ie, waitlist) [20,31-33,37,54]. These studies have been reviewed in Table 1 and Multimedia Appendix 1. They were categorized separately from the other RCTs because of concerns regarding the use of waiting lists for a comparison group, as waiting lists are not comparable with placebo interventions [55]. Berger et al [20] demonstrated significant decreases in depression, anxiety, and other mental health measures when compared with a waitlist condition, with many of the participants no longer warranting the diagnoses of anxiety disorders after 6 sessions. Ebert et al [54] showed greater improvement in insomnia measures than the waitlist control group, along with more participants achieving a symptom-free state and improving on secondary measures such as depression and sleep quality. Lokman et al [32], who compared mini-cognitive behavioral therapy–based interventions to a waitlist, found a significant decrease in depression, anxiety, and sleep-related problems and a higher well-being in the intervention group. Loughnan et al [33] found that iCBT produced moderate to large effect reductions in anxiety and psychological distress compared with a waitlist

condition group. Finally, Lintvedt et al [31] demonstrated lower levels of depressive symptoms, negative thoughts, and improved depression literacy compared with a waitlist control group.

Studies Using iCBT: RCTs Compared With Guided Interventions

A valuable approach is to compare an unguided technological intervention to a similar intervention completed under the guidance of a trained professional. Five studies in this review used this strategy and have been summarized in Table 1 and Multimedia Appendix 1 [21,24,26,38,39]. In 2011, Berger et al [21] showed an improvement in depression symptoms when compared with a waitlist condition, with no significant difference seen whether the iCBT intervention was guided by a psychotherapist or not. Christensen et al [24] did not observe improved anxiety outcomes on generalized anxiety disorder at 6- or 12-month periods on any measures (guided or unguided, iCBT, or non-iCBT treatment), but did find higher completion rates in the 3 study arms that used phone or email guidance. Gilbody and colleagues showed an improvement in depression (by PHQ-9) in the guided group over the unguided group at 4 months but not at 12 months [26]. Romero-Sanchiz et al [38] were able to show cost-effectiveness per point improvement on Beck's Depression Inventory (BDI-II) and quality-adjusted life years in the self-directed and therapist-supported groups when compared with care as usual, although it was more pronounced in the self-directed group than in the therapist-intervention group. Van Kessel et al [39] found greater reductions in fatigue in the guided group than in the unguided group but observed no significant differences in anxiety or depression.

Studies Using iCBT: RCTs Compared With Other Interventions

Another effective strategy to demonstrate the utility of unguided iCBT is to compare it with other psychological interventions. These studies have been reviewed in Table 1 and Multimedia Appendix 1 [19,25,27-30,36,37]. Donker et al [25] compared a specific unguided iCBT program against other unguided iCBT programs. They found that although there were no differences between the 3 groups at baseline or follow-up, their dropout rates varied. Gosling et al [27] demonstrated that the insomnia-based iCBT program sleep healthy using the internet lead to greater improvements on measures of anxiety (at posttest and at 6-month follow-up) than a website with general health tips. In 2018, Hagatun et al [28] again showed sleep healthy using the internet's superiority over a patient education website on measures of anxiety. They also showed improvements in the measures of insomnia [29]. A posthoc analysis of this study team's research in 2019 demonstrated that this effect was not mediated by whether a person was a morning or an evening person (ie, persons with either diurnal or nocturnal sleeping habits) [30]. Noguchi et al [36] did not find any differences between iCBT and mindfulness-based training on depression measures.

Studies Using iCBT: RCT Comparing Self-guided Intervention With or Without Sensors

One study used a novel intervention strategy added to iCBT, which is reviewed in Table 1 and Multimedia Appendix 1 [22]. Botella et al [22] compared 2 intervention groups. Although both groups had access to an iCBT program for depression (*Smiling is fun*), one group also had access to electroencephalogram, electrocardiogram, and actigraphy sensors to monitor the physiological states and to provide feedback to the users. There was also a comparison with the waitlist control. This study found that the most effective treatment for depression was the sensor group, followed by the nonsensor intervention group [22].

Studies Using iCBT With a Cross-sectional Study Design

Three studies used a single-part, cross-sectional study approach, and have been summarized in Table 1 and Multimedia Appendix 1 [23,34,35]. Brettschneider et al [23] observed less social anxiety and depressive symptoms over an 8-week iCBT program, with a dropout rate of 26% (10/39). Mewton et al [34] found lower scores on the measures of psychological distress and disability after a 6-course lesson, with greater adherence in older adults (>60 years old) than in younger adults. Moloney et al [35] were able to show positive and significant improvements in US women on measures of insomnia, sleep quality, depression, and the likelihood of using medication after a 6-week intervention.

Non-iCBT Digital Therapeutic Studies

The other 14 studies in this review, which did not use iCBT, were categorized into RCTs, non-RCT quantitative studies, and qualitative studies.

RCTs With Non-iCBT Interventions

We identified 8 RCTs, which have been summarized in Table 2 and Multimedia Appendix 2. The RCTs were heterogeneous in nature. They encompassed several types of interventions, including websites, automated text message systems, and videos. The websites varied in content, although many of them provided access to psychoeducation modules, with some allowing users to provide their feedback. One website allowed the patients to create tailored advice reports that were generated based on user responses to preset questions [45]. Automated text messaging services allowed the participants to receive programmed text messages in the form of reminders, education, and questions about mood, craving, or use [20,41-43]. In one study, the participants could respond to text messages, allowing for ecological momentary assessment, or the immediate reporting of participants' behaviors in real time [41]. One study provided videos for the participants to watch at home [47].

The sample size varied from n=60 to n=1758. The targeted populations included those with mental health diagnoses, as well as healthy participants who were measured using a mental health–related outcome (ie, adjustment). Although most of the studies were rated 1b for LOE, both the Bernstein and Sherman studies were given a 2b LOE rating because they had small sample sizes and the results did not report CIs [41,47].

Of the 8 RCTs, 4 (50%) had a waitlist group. Of these 4 studies, 2 (50%) had no other comparison [42,43], whereas the other 2 (50%) used at least one other comparison group [40,47]. There were 38% (3/8) of studies that compared unguided interventions with guided interventions [40,44,47]. In addition, 38% (3/8) of

studies compared a novel technological approach to usual care (ie, psychoeducational websites, brochures, or usual care) [41,45,46].

Aardoom et al [40] showed that *Featback* (a website using psychoeducation principles) was superior to a waitlist condition with regard to bulimic-related psychopathology. Bernstein et al [41] demonstrated that 47% (14/30) of the intervention group showed a 7-day smoking abstinence at 1-month compared with 10% (3/30) in the control group, but this effect was less significant at 3 months (9/30, 30% vs 4/30, 13%). Constant et al [42] reported lower anxiety using the Hospital Anxiety and Depression Scale score in women in the intervention group (ie, the group receiving automated text messaging for medical abortion self-management) and that these women were better prepared for the side effects of their medication. Kannisto et al [43] looked at recruitment and attrition, finding that one-third of those screened were eligible, but two-thirds of the eligible patients refused. Many were involved in the data retrieval stage, but very few were followed up at the postal survey stage. Participants mentioned a lack of interest, lack of mobile use adherence, or the lack of ability to use a mobile device as the main influences on their adherence. Kleiboer et al [44], who examined the effect of an internet-based problem-solving therapy, found that weekly, scheduled guidance by a trained professional had a small but significant effect on depressive symptoms compared with an internet-based problem-solving therapy–only intervention. All the groups showed improvement posttreatment. Mason and colleagues, who used a website-based advice report to quit smoking, did not find a difference in prolonged smoking abstinence between a tailored advice report group and a standardized advice report control group, regardless of the socioeconomic status and whether the participants were smoking at baseline or had recently quit [45]. Pictet et al [46] showed that positive scenarios had a considerable effect on whether a treatment (here, a website-based cognitive bias modification program) was effective. Finally, Sherman et al [47] compared 4 groups that received psychoeducational videos with varying levels of support and showed that although there were improvements in all groups in adjustment to illness, there were no significant differences among the groups in the adjustment scores.

Non-RCT Quantitative Studies With Non-iCBT Interventions

We identified 5 non-RCT quantitative studies, as summarized in Table 2 and Multimedia Appendix 2. Study designs included feasibility, cohort, and case-control studies. The intervention types included an application program, a website, phone apps, computer software, and integrated voice response (IVR), which is a technology that allows a computer to interact with humans through the use of voice and dual tone, multi-frequency tone input via a keypad.

The application program allowed for an interactive experience between the user and the program. Notifications and surveys were also used [48]. One study used a website design that allowed the portal access to patients receiving mental health services, looking at their use, appointment keeping, and mental health recovery measures [49]. This portal included psychoeducational materials that the patients could access, as well as information about their appointments. One study used a computer software program that the participants could download at home [52]. Another study used IVR, which allowed the participants to receive automated phone calls where they could provide feedback to the system on their depression symptoms [50].

The sample size varied from n=75 to n=3158. The populations included those with mental health diagnoses, such as depression, anxiety, or psychosis. Each study examined a different outcome. Ahmedani et al [48] used scales to evaluate the interventions (eg, patient health questionnaire; PHQ-9). They found that there was a statistically significant reduction in depression scores, along with a one-third decrease in the number of patients having moderate to mild depression scores in the study cohort. Kipping et al [49] examined the use, recovery measures, and surveys for interventions. Their study showed an increased activation of service users and caregivers, with improved recovery scores (based on mental health recovery measures domains). The users were more likely to attend scheduled appointments than the nonusers. Piette et al [50] used IVR to reach the patients and measured the call completion rates between 4 different disease groups, showing that depression had the lowest call completion rates among the 4 disease groups (314/442, 71%), and the call completion rates decreased over time with the increased severity of mental health. Pratap et al [51] compared phone apps and found that they could decrease the depressive symptoms in participants, with no significant differences between the types of apps used. Stein et al [52] focused on medication adherence and found that the users of their program did not have higher medication adherence than the nonusers. Although 3 studies were rated at a 2b LOE (individual data and cohort studies), the study by Stein et al [52] was given a 3b rating because it was a case-control design that did not control the treatment allocation.

Qualitative Study With Non-iCBT Intervention

We identified one qualitative study, as summarized in Table 2 and Multimedia Appendix 2. Bauer et al [53] reported on a pilot feasibility and acceptability study (N=17) of Ginger.io, a smartphone app with a web-based dashboard designed to offer support and activities related to anxiety and depression in adults diagnosed with these conditions. The primary outcome was the participants' use of the app and their survey completion rates. As a qualitative study, it was given level 5 on the LOE.

Although all 17 participants used it at first, only 6 (35%) used it for 8 weeks. Many reported feeling satisfied with the app (11/17, 67%) and found it easy to use (13/17, 77%), but few reported concerns (2/17, 13%). Despite this, 88% (15/17) of the participant completed all the weekly symptom measures before discontinuing the use of the app.

Discussion

Principal Findings

This review highlights the potential of digital interventions to improve mental health, as well as the areas where new research is required. The main challenges include the heterogeneity of interventions and the low-quality comparators. Patient-related

issues that were identified include high dropout rates, variable efficacy, and a lack of safety evaluations.

We identified 36 studies that examined various types of digital therapeutics. Six studies (17%) had a single group and 30 (83%) used between-group comparisons. Of the 9 studies that compared a digital treatment against a waitlist, only one did not find a beneficial effect from the use of a digital therapeutic (ie, medication adherence [52]), whereas all others showed a positive effect on primary outcomes. However, these findings must be interpreted with caution; although waitlists have often been used as control conditions when assessing psychotherapy, they are not equivalent to the placebo group in a pharmacological study and may not be a suitable comparison to show effectiveness in this context [55]. Patients know that they are not receiving an intervention, that they may not receive any alternative support, or they may be frustrated by being on a waitlist. In addition, many psychiatric disorders worsen with time if left untreated. To demonstrate that digital therapeutics are a viable alternative to other treatments, future research into these programs should focus on using groups that are comparable with the intervention, instead of using the lack of any intervention as a control group.

Eight studies compared an unguided intervention to varying levels of support from a trained professional. Of these, 5 studies found a difference between the guided and unguided groups (with 3 favoring the guided interventions), and 2 did not find any difference. The treatment effects varied across studies in terms of their quality, size, and duration. One study found effects on depressive symptoms at 4 months that were not sustained at 12 months [54]. Another study found an effect on fatigue but not on depression [39]. These findings warrant further study to ascertain the specific factors that influence the effectiveness of such interventions.

There are many potential explanations for the variable effects of treatment. Perhaps the most salient point and indeed the reason why a meta-analysis could not be done is that studies are too heterogeneous. As the tables show, they differ in their target populations (eg, external population vs clinical setting), severity of disease, nature of the interventions, length and structure of the assessments used, reminders used, the cultural and ethnic backgrounds of study participants, and the social support structures that ultimately may help explain why some interventions seemed to work better than others.

Most studies compared self-guided interventions against each other or to other treatment methods, which included educational websites and traditional treatment with a mental health team. In these studies, digital therapeutic interventions were comparable with psychoeducational websites in mental health outcomes. When interventions were compared with *standard of care*, this term was usually not well-defined, preventing any conclusions to be extended outside the context of the specific study.

Of the 36 studies examined, 22 were identified as iCBT, showing the popularity of this modality of web-based psychotherapy compared with other psychological paradigms. This therapeutic approach appears to be a preferred treatment method, with many randomized studies having large sample sizes. However, many studies have compared these interventions against other digital technologies or waitlist conditions, which may not be comparable. Digital technologies are relatively new, and this fact may limit the body of research available. While iCBT is driving most of the research available, there are also other types of psychotherapy delivered digitally that warrant further study.

Strengths and Limitations

This review focused specifically on self-directed automated interventions that patients could implement without a therapist. Independent technology-based care options can be implemented at minimal cost by the organizations and patients and can be done at home, without having to access hospital or clinic resources. The immediate availability of these technologies has important advantages regarding the access and universality of care. Their potential accessibility is far broader than other methods of care delivery, contributing to equality in health care. They can be adapted to monitor compliance and side effects of medications, and to consolidate the gains obtained through individual psychotherapy, group psychotherapy, or psychoeducation after the discharge from health care services, thus liberating the time and resources used in follow-up and potentially preventing relapse. The study participants often reported satisfaction with technology-based care and attributed benefits to the intervention [53]. The universal accessibility of these types of interventions can help reach patients who are unable to receive traditional care because of the lack of local resources or the stigma attached to mental health, thus providing a low-barrier alternative to their care [56].

The methods used to evaluate technology-based care may differ from those for traditional RCT methods. This is partly due to the way in which technology is constantly being updated. A study by Desveaux et al [57] notes that the rigor by which we evaluate health care systems is usually applied to a static, fixed intervention, which is in direct conflict with the dynamic and ever-changing nature of technology. RCTs, by definition, require blinding, and this is often not possible in psychotherapy and digital therapeutics. RCTs with blinding are the benchmark in interventions such as medication because of the lack of contextual factors affecting their use, and the context of the intervention is vital in the development of the intervention itself [57]. This includes factors related to the interaction between the technology and the user, the environmental factors, and the access to technology. Therefore, the user must be a crucial part of not only the evaluation of the intervention, but also its design. The evaluation of new digital therapeutics requires a combination of traditional RCT methodology and novel methods of evaluation that consider the adaptive nature of sociotechnological systems of technology-based care.

Research and advice on how these novel methods of evaluation should be like can currently be found in the literature, with some going as far as having designed models such as the multiphase optimization strategy and sequential multiple assignment randomized trial evaluation system [58]. These 2 approaches apply various strategies (such as the use of screening and refining phases or time-varying adaptive interventions) that account for the changing needs of digital interventions and of

their target population [58]. Other researchers have described several criteria that would help evaluate digital health interventions, including the application of a multidisciplinary approach (ie, clinical and behavioral intervention, as well as computer and engineering science), or the notion of adopting the iterative approach (ie, several cycles of development and optimization), such as the accelerated creation-to-sustainability model [59,60]. Advocates of digital interventions should also consider aspects such as safety, data security, and engagement [61].

Although the aim of this study was to evaluate self-directed interventions, several studies had a part with some human-assisted support. The effect of this therapist support suggests that human interaction may play a role in the acceptability of these programs. For example, although Aardoom et al [40] found no significant differences in the improvement of eating disorder symptoms between participants in the 3 intervention groups, qualitative data suggest that participants who received therapist support showed more satisfaction with the intervention. A similar trend in which satisfaction and engagement increased with human support was found in other studies [47,53,54].

Guided treatments can have different qualitative effects than unguided treatments, but these differences are not always detected and require further study. It may be that both unguided and guided treatments are effective but in different ways and for different groups of patients. However, a notable limitation of some of the unguided interventions is that many have *technical support*, raising the question of whether a simple call from a nontrained professional provides some therapeutic benefit. In this regard, it may be that some of the study procedures that examined *unguided* interventions were not truly unguided. Therefore, these interventions might not have as much of a therapeutic effect as the studies suggested. A similar trend in which satisfaction and engagement increased with human support was found in other studies [47,53,54].

The attrition rates varied between the studies. Previous studies have shown dropout rates of up to 80% [62]. In our review, Kannisto et al [43] found that despite having only 4.8% dropout at baseline, more than half of the intervention participants (52.45%) did not complete the final study surveys. Bauer et al [53] found a similar trend: the use was 100% of the participants in the first 4 weeks but dropped off to 35% of the participants by 8 weeks (a loss of 65% in 4 weeks). Piette et al [50] found that call completion rates were lower in depression when compared with other medical conditions such as diabetes or cancer, suggesting higher dropout rates in mental health interventions. Batterham et al [19] found that only 34% of the participants completed most intervention models.

Although many studies have shown high dropout rates, there are a couple of important points to note regarding adherence to these types of interventions. First, though the dropout rates in automated community-based interventions are likely to be high, the resources needed to reach those individuals who would otherwise not have any other form of treatment are relatively low. This suggests that there is merit in delivering self-guided, low-intensity technological interventions in this subgroup. It is also worth noting that there are many reasons why the dropout rates may be high, and although it is likely that many are negative, it is entirely possible that some of these reasons could be positive. For example, if someone feels that they have benefited from the program and stops early, or if they recover before the program has concluded, they may have dropped out because of this improvement. Given the heterogeneity of the studies, we could not identify a particular patient who would benefit more from these interventions. However, as more information becomes available and more RCTs are published, the profile of an ideal patient who responds well to digital interventions can be defined [63,64].

Perhaps the greatest limitation of this review is that technology changes at a rapid pace and despite the authors' attempts to consider a broad range of interventions, new technology-based care methods are constantly being developed and evaluated. Some of these evaluations may not have been published yet or may even remain unpublished if the results are not positive. This is an expanding field, and it is likely that more research will be published in the future.

As digital therapeutics become more available, there is a need to establish acceptable guidelines and evidence-based approaches to determine the efficiency and suitability of technology-based treatments. This need has already been recognized. The American Psychiatric Association has established the *App Evaluation Model*, which is a set of guidelines that help health care providers evaluate the safety, benefits, and potential harms of phone apps [8]. Safety issues would include: implementing safeguards on patient data and potential data sharing, scientific review of the content, and continuous evaluation of the potential harms via a user or provider feedback system [65]. Lagan et al [66] developed a framework to translate these qualitative guidelines into objective metrics using a set of standardized questions, facilitating access to critically evaluated apps for providers as well as general audiences. The development of guidelines is crucial to not only orientate clinician advice on digital therapeutics, but also to direct research to those areas that require it while ensuring safe practices for the patients.

Conclusions

The use of technology-based interventions in health care is increasing, but there needs to be more specific outcomes to assess their efficacy over time and the maintenance of those gains. In addition, although there are many papers that examine the use of technology-based interventions, reducing the list of research articles to those that only have *fully self-guided interventions* shows that considerably less research is addressing the issues mentioned above. To be effective, the interventions should be developed in collaboration with the users. This is evidenced by the fact that dropout rates were high in most of the studies evaluated in this review. Studies on culturally and linguistically diverse communities have found that co-design of mental health services can help recognize and account for the issues related to trust, power differential, communication, and confidentiality regarding the relationships between the researchers and the communities and users of their interventions [67]. Other research has also found the benefit of co-design in

children and young people, as well as in specific mental health program designs [68-70].

Current research suggests that the effectiveness of technology-based care interventions is superior to that of waitlist controls and other interventions. However, to show their effectiveness over traditional psychiatric care, the studies should use comparison groups that are comparable with the intervention studied, thus avoiding waiting lists or other nonintervention parts.

Self-directed interventions may lead to lower costs and fewer hours spent by health care providers in supporting a treatment. These interventions will also become accessible to people lacking access to health care, such as those who live far from health care centers, those who cannot travel because of disability or family commitments, or those who cannot afford traditional care. In times of crisis or quarantine, these methods of care can become crucial instruments to deliver treatment. For many people, technology-based care methods are their first point of access to care. Thus, self-directed digital therapeutics can contribute to health care equality.

Acknowledgments

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sector.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of the characteristics of the 22 internet-based cognitive behavioral therapy studies.
[PDF File (Adobe PDF File), 370 KB - mental_v8i11e27404_app1.pdf]

Multimedia Appendix 2
Summary of data extracted from non–internet-based cognitive behavioral therapy studies.
[PDF File (Adobe PDF File), 367 KB - mental_v8i11e27404_app2.pdf]

References

1. Cho CH, Lee HJ. Could digital therapeutics be a game changer in psychiatry? Psychiatry Investig 2019 Feb;16(2):97-98 [FREE Full text] [doi: 10.30773/pi.2019.01.20] [Medline: 30808114]
2. Ferreri F, Bourla A, Mouchabac S, Karila L. e-Addictology: an overview of new technologies for assessing and intervening in addictive behaviors. Front Psychiatry 2018 Mar 01;9:51 [FREE Full text] [doi: 10.3389/fpsyt.2018.00051] [Medline: 29545756]
3. Steinkamp JM, Goldblatt N, Borodovsky JT, LaVertu A, Kronish IM, Marsch LA, et al. Technological interventions for medication adherence in adult mental health and substance use disorders: a systematic review. JMIR Ment Health 2019 Mar 12;6(3):e12493 [FREE Full text] [doi: 10.2196/12493] [Medline: 30860493]
4. Bourla A, Mouchabac S, El Hage W, Ferreri F. e-PTSD: an overview on how new technologies can improve prediction and assessment of Posttraumatic Stress Disorder (PTSD). Eur J Psychotraumatol 2018 Feb 06;9(sup1):1424448 [FREE Full text] [doi: 10.1080/20008198.2018.1424448] [Medline: 29441154]
5. Ferreri F, Bourla A, Peretti C, Segawa T, Jaafari N, Mouchabac S. How new technologies can improve prediction, assessment, and intervention in Obsessive-Compulsive Disorder (e-OCD): review. JMIR Ment Health 2019 Dec 10;6(12):e11643 [FREE Full text] [doi: 10.2196/11643] [Medline: 31821153]
6. Sucala M, Schnur JB, Constantino MJ, Miller SJ, Brackman EH, Montgomery GH. The therapeutic relationship in e-therapy for mental health: a systematic review. J Med Internet Res 2012 Aug 02;14(4):e110 [FREE Full text] [doi: 10.2196/jmir.2084] [Medline: 22858538]
7. Simon GE, Von Korff M, Ludman EJ, Katon WJ, Rutter C, Unützer J, et al. Cost-effectiveness of a program to prevent depression relapse in primary care. Med Care 2002 Oct;40(10):941-950. [doi: 10.1097/00005650-200210000-00011] [Medline: 12395027]
8. App evaluation model. American Psychiatric Association. 2018. URL: https://www.psychiatry.org/psychiatrists/practice/mental-health-apps/app-evaluation-model [accessed 2021-11-06]
9. Leung R, Hastings JF, Keefe RH, Brownstein-Evans C, Chan KT, Mullick R. Building mobile apps for underrepresented mental health care consumers: a grounded theory approach. Soc Work Ment Health 2016;14(6):625-636 [FREE Full text] [doi: 10.1080/15332985.2015.1130010] [Medline: 29056878]
10. Bourla A, Ferreri F, Ogorzelec L, Peretti C, Guinchard C, Mouchabac S. Psychiatrists' attitudes toward disruptive new technologies: mixed-methods study. JMIR Ment Health 2018 Dec 14;5(4):e10240 [FREE Full text] [doi: 10.2196/10240] [Medline: 30552086]

11. Cuijpers P, Donker T, Johansson R, Mohr DC, van Straten A, Andersson G. Self-guided psychological treatment for depressive symptoms: a meta-analysis. PLoS One 2011;6(6):e21274 [FREE Full text] [doi: 10.1371/journal.pone.0021274] [Medline: 21712998]
12. Karyotaki E, Riper H, Twisk J, Hoogendoorn A, Kleiboer A, Mira A, et al. Efficacy of self-guided internet-based cognitive behavioral therapy in the treatment of depressive symptoms: a meta-analysis of individual participant data. JAMA Psychiatry 2017 Apr 01;74(4):351-359. [doi: 10.1001/jamapsychiatry.2017.0044] [Medline: 28241179]
13. 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]
14. Fischer R, Bortolini T, Karl JA, Zilberberg M, Robinson K, Rabelo A, et al. Rapid review and meta-meta-analysis of self-guided interventions to address anxiety, depression, and stress during COVID-19 social distancing. Front Psychol 2020 Oct 28;11:563876 [FREE Full text] [doi: 10.3389/fpsyg.2020.563876] [Medline: 33192837]
15. 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]
16. Welcome to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) website!. PRISMA. URL: http://www.prisma-statement.org/ [accessed 2021-11-06]
17. Mendeley [computer software] v 1.19.3. Mendeley. URL: https://www.mendeley.com/?interaction_required=true [accessed 2021-11-06]
18. Oxford Centre for Evidence-Based Medicine: Levels of Evidence. Centre for Evidence-Based Medicine. 2009. URL: https://www.cebm.net/2009/06/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/ [accessed 2021-11-06]
19. Batterham PJ, Calear AL, Farrer L, McCallum SM, Sze Cheng VW. FitMindKit: randomised controlled trial of an automatically tailored online program for mood, anxiety, substance use and suicidality. Internet Interv 2018 Jun;12:91-99 [FREE Full text] [doi: 10.1016/j.invent.2017.08.002] [Medline: 30135773]
20. Berger T, Urech A, Krieger T, Stolz T, Schulz A, Vincent A, et al. Effects of a transdiagnostic unguided Internet intervention ('velibra') for anxiety disorders in primary care: results of a randomized controlled trial. Psychol Med 2017 Jan;47(1):67-80. [doi: 10.1017/S0033291716002270] [Medline: 27655039]
21. Berger T, Hämmerli K, Gubser N, Andersson G, Caspar F. Internet-based treatment of depression: a randomized controlled trial comparing guided with unguided self-help. Cogn Behav Ther 2011;40(4):251-266. [doi: 10.1080/16506073.2011.616531] [Medline: 22060248]
22. Botella C, Mira A, Moragrega I, García-Palacios A, Bretón-López J, Castilla D, et al. An internet-based program for depression using activity and physiological sensors: efficacy, expectations, satisfaction, and ease of use. Neuropsychiatr Dis Treat 2016 Feb 23;12:393-406 [FREE Full text] [doi: 10.2147/NDT.S93315] [Medline: 27042067]
23. Brettschneider M, Neumann P, Berger T, Renneberg B, Boettcher J. Internet-based interpretation bias modification for social anxiety: a pilot study. J Behav Ther Exp Psychiatry 2015 Dec;49(Pt A):21-29. [doi: 10.1016/j.jbtep.2015.04.008] [Medline: 26005201]
24. Christensen H, Batterham P, Mackinnon A, Griffiths K, Kalia Hehir K, Kenardy J, et al. Prevention of generalized anxiety disorder using a web intervention, iChill: randomized controlled trial. J Med Internet Res 2014 Sep 02;16(9):e199 [FREE Full text] [doi: 10.2196/jmir.3507] [Medline: 25270886]
25. Donker T, Bennett K, Bennett A, Mackinnon A, van Straten A, Cuijpers P, et al. Internet-delivered interpersonal psychotherapy versus internet-delivered cognitive behavioral therapy for adults with depressive symptoms: randomized controlled noninferiority trial. J Med Internet Res 2013 May 13;15(5):e82 [FREE Full text] [doi: 10.2196/jmir.2307] [Medline: 23669884]
26. Gilbody S, Brabyn S, Lovell K, Kessler D, Devlin T, Smith L, REEACT Collaborative. Telephone-supported computerised cognitive-behavioural therapy: REEACT-2 large-scale pragmatic randomised controlled trial. Br J Psychiatry 2017 May;210(5):362-367. [doi: 10.1192/bjp.bp.116.192435] [Medline: 28254959]
27. Gosling JA, Batterham P, Ritterband L, Glozier N, Thorndike F, Griffiths KM, et al. Online insomnia treatment and the reduction of anxiety symptoms as a secondary outcome in a randomised controlled trial: the role of cognitive-behavioural factors. Aust N Z J Psychiatry 2018 Dec;52(12):1183-1193. [doi: 10.1177/0004867418772338] [Medline: 29717621]
28. Hagatun S, Vedaa Ø, Harvey AG, Nordgreen T, Smith OR, Pallesen S, et al. Internet-delivered cognitive-behavioral therapy for insomnia and comorbid symptoms. Internet Interv 2018 Jun;12:11-15 [FREE Full text] [doi: 10.1016/j.invent.2018.02.003] [Medline: 30135764]
29. Hagatun S, Vedaa Ø, Nordgreen T, Smith OR, Pallesen S, Havik OE, et al. The short-term efficacy of an unguided internet-based cognitive-behavioral therapy for insomnia: a randomized controlled trial with a six-month nonrandomized follow-up. Behav Sleep Med 2019;17(2):137-155. [doi: 10.1080/15402002.2017.1301941] [Medline: 28345961]
30. Lien M, Bredeli E, Sivertsen B, Kallestad H, Pallesen S, Smith OR, et al. Short and long-term effects of unguided internet-based cognitive behavioral therapy for chronic insomnia in morning and evening persons: a post-hoc analysis. Chronobiol Int 2019 Oct;36(10):1384-1398. [doi: 10.1080/07420528.2019.1647435] [Medline: 31368382]

31. Lintvedt OK, Griffiths KM, Sørensen K, Østvik AR, Wang CE, Eisemann M, et al. Evaluating the effectiveness and efficacy of unguided internet-based self-help intervention for the prevention of depression: a randomized controlled trial. Clin Psychol Psychother 2013;20(1):10-27. [doi: 10.1002/cpp.770] [Medline: 21887811]
32. Lokman S, Leone SS, Sommers-Spijkerman M, van der Poel A, Smit F, Boon B. Complaint-directed mini-interventions for depressive complaints: a randomized controlled trial of unguided web-based self-help interventions. J Med Internet Res 2017 Jan 04;19(1):e4 [FREE Full text] [doi: 10.2196/jmir.6581] [Medline: 28052840]
33. Loughnan S, Sie A, Hobbs M, Joubert A, Smith J, Haskelberg H, et al. A randomized controlled trial of 'MUMentum Pregnancy': internet-delivered cognitive behavioral therapy program for antenatal anxiety and depression. J Affect Disord 2019 Jan 15;243:381-390. [doi: 10.1016/j.jad.2018.09.057] [Medline: 30266030]
34. Mewton L, Sachdev PS, Andrews G. A naturalistic study of the acceptability and effectiveness of internet-delivered cognitive behavioural therapy for psychiatric disorders in older australians. PLoS One 2013 Aug 12;8(8):e71825 [FREE Full text] [doi: 10.1371/journal.pone.0071825] [Medline: 23951253]
35. Moloney ME, Martinez AI, Badour CL, Moga DC. Internet-based cognitive behavioral therapy for insomnia in appalachian women: a pilot study. Behav Sleep Med 2020;18(5):680-689 [FREE Full text] [doi: 10.1080/15402002.2019.1661249] [Medline: 31470745]
36. Noguchi R, Sekizawa Y, So M, Yamaguchi S, Shimizu E. Effects of five-minute internet-based cognitive behavioral therapy and simplified emotion-focused mindfulness on depressive symptoms: a randomized controlled trial. BMC Psychiatry 2017 Mar 04;17(1):85 [FREE Full text] [doi: 10.1186/s12888-017-1248-8] [Medline: 28259151]
37. Proudfoot J, Clarke J, Birch M, Whitton AE, Parker G, Manicavasagar V, et al. Impact of a mobile phone and web program on symptom and functional outcomes for people with mild-to-moderate depression, anxiety and stress: a randomised controlled trial. BMC Psychiatry 2013 Nov 18;13:312 [FREE Full text] [doi: 10.1186/1471-244X-13-312] [Medline: 24237617]
38. Romero-Sanchiz P, Nogueira-Arjona R, García-Ruiz A, Luciano JV, García Campayo J, Gili M, et al. Economic evaluation of a guided and unguided internet-based CBT intervention for major depression: results from a multi-center, three-armed randomized controlled trial conducted in primary care. PLoS One 2017 Feb 27;12(2):e0172741 [FREE Full text] [doi: 10.1371/journal.pone.0172741] [Medline: 28241025]
39. van Kessel K, Wouldes T, Moss-Morris R. A New Zealand pilot randomized controlled trial of a web-based interactive self-management programme (MSInvigor8) with and without email support for the treatment of multiple sclerosis fatigue. Clin Rehabil 2016 May;30(5):454-462. [doi: 10.1177/0269215515584800] [Medline: 25952587]
40. Aardoom J, Dingemans A, Spinhoven P, van Ginkel JR, de Rooij M, van Furth EF. Web-based fully automated self-help with different levels of therapist support for individuals with eating disorder symptoms: a randomized controlled trial. J Med Internet Res 2016 Jun 17;18(6):e159 [FREE Full text] [doi: 10.2196/jmir.5709] [Medline: 27317358]
41. Bernstein S, Rosner J, Toll B. A multicomponent intervention including texting to promote tobacco abstinence in emergency department smokers: a pilot study. Acad Emerg Med 2016 Jul;23(7):803-808 [FREE Full text] [doi: 10.1111/acem.12990] [Medline: 27146116]
42. Constant D, de Tolly K, Harries J, Myer L. Mobile phone messages to provide support to women during the home phase of medical abortion in South Africa: a randomised controlled trial. Contraception 2014 Sep;90(3):226-233. [doi: 10.1016/j.contraception.2014.04.009] [Medline: 24850188]
43. Kannisto K, Korhonen J, Adams C, Koivunen M, Vahlberg T, Välimäki MA. Factors associated with dropout during recruitment and follow-up periods of a mHealth-based randomized controlled trial for mobile.net to encourage treatment adherence for people with serious mental health problems. J Med Internet Res 2017 Feb 21;19(2):e46 [FREE Full text] [doi: 10.2196/jmir.6417] [Medline: 28223262]
44. Kleiboer A, Donker T, Seekles W, van Straten A, Riper H, Cuijpers P. A randomized controlled trial on the role of support in internet-based problem solving therapy for depression and anxiety. Behav Res Ther 2015 Sep;72:63-71. [doi: 10.1016/j.brat.2015.06.013] [Medline: 26188373]
45. Mason D, Gilbert H, Sutton S. Effectiveness of web-based tailored smoking cessation advice reports (iQuit): a randomized trial. Addiction 2012 Dec;107(12):2183-2190. [doi: 10.1111/j.1360-0443.2012.03972.x] [Medline: 22690882]
46. Pictet A, Jermann F, Ceschi G. When less could be more: investigating the effects of a brief internet-based imagery cognitive bias modification intervention in depression. Behav Res Ther 2016 Sep;84:45-51. [doi: 10.1016/j.brat.2016.07.008] [Medline: 27485252]
47. Sherman DW, Haber J, Hoskins CN, Budin WC, Maislin G, Shukla S, et al. The effects of psychoeducation and telephone counseling on the adjustment of women with early-stage breast cancer. Appl Nurs Res 2012 Feb;25(1):3-16. [doi: 10.1016/j.apnr.2009.10.003] [Medline: 20974079]
48. Ahmedani BK, Crotty N, Abdulhak MM, Ondersma SJ. Pilot feasibility study of a brief, tailored mobile health intervention for depression among patients with chronic pain. Behav Med 2015;41(1):25-32. [doi: 10.1080/08964289.2013.867827] [Medline: 24313728]
49. Kipping S, Stuckey MI, Hernandez A, Nguyen T, Riahi S. A web-based patient portal for mental health care: benefits evaluation. J Med Internet Res 2016 Nov 16;18(11):e294 [FREE Full text] [doi: 10.2196/jmir.6483] [Medline: 27852556]

50. Piette J, Rosland A, Marinec N, Striplin D, Bernstein S, Silveira M. Engagement with automated patient monitoring and self-management support calls: experience with a thousand chronically ill patients. Med Care 2013 Mar;51(3):216-223 [FREE Full text] [doi: 10.1097/MLR.0b013e318277ebf8] [Medline: 23222527]
51. Pratap A, Renn BN, Volponi J, Mooney SD, Gazzaley A, Arean PA, et al. Using mobile apps to assess and treat depression in Hispanic and Latino populations: fully remote randomized clinical trial. J Med Internet Res 2018 Aug 09;20(8):e10130 [FREE Full text] [doi: 10.2196/10130] [Medline: 30093372]
52. Stein BD, Kogan JN, Mihalyo MJ, Schuster J, Deegan PE, Sorbero MJ, et al. Use of a computerized medication shared decision making tool in community mental health settings: impact on psychotropic medication adherence. Community Ment Health J 2013 Apr;49(2):185-192 [FREE Full text] [doi: 10.1007/s10597-012-9528-8] [Medline: 22837104]
53. Bauer AM, Iles-Shih M, Ghomi RH, Rue T, Grover T, Kincler N, et al. Acceptability of mHealth augmentation of collaborative care: a mixed methods pilot study. Gen Hosp Psychiatry 2018;51:22-29 [FREE Full text] [doi: 10.1016/j.genhosppsych.2017.11.010] [Medline: 29272712]
54. Ebert DD, Berking M, Thiart H, Riper H, Laferton JA, Cuijpers P, et al. Restoring depleted resources: efficacy and mechanisms of change of an internet-based unguided recovery training for better sleep and psychological detachment from work. Health Psychol 2015 Dec;34S:1240-1251. [doi: 10.1037/hea0000277] [Medline: 26651465]
55. Bandelow B, Reitt M, Röver C, Michaelis S, Görlich Y, Wedekind D. Efficacy of treatments for anxiety disorders: a meta-analysis. Int Clin Psychopharmacol 2015 Jul;30(4):183-192. [doi: 10.1097/YIC.0000000000000078] [Medline: 25932596]
56. Donker T, Petrie K, Proudfoot J, Clarke J, Birch MR, 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]
57. Desveaux L, Shaw J, Wallace R, Bhattacharyya O, Bhatia RS, Jamieson T. Examining tensions that affect the evaluation of technology in health care: considerations for system decision makers from the perspective of industry and evaluators. JMIR Med Inform 2017 Dec 08;5(4):e50 [FREE Full text] [doi: 10.2196/medinform.8207] [Medline: 29222075]
58. Collins LM, Murphy SA, Strecher V. The Multiphase Optimization Strategy (MOST) and the Sequential Multiple Assignment Randomized Trial (SMART): new methods for more potent eHealth interventions. Am J Prev Med 2007 May;32(5 Suppl):S112-S118 [FREE Full text] [doi: 10.1016/j.amepre.2007.01.022] [Medline: 17466815]
59. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851 [FREE Full text] [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
60. Mohr D, Lyon A, Lattie E, Reddy M, Schueller S. Accelerating digital mental health research from early design and creation to successful implementation and sustainment. J Med Internet Res 2017 May 10;19(5):e153 [FREE Full text] [doi: 10.2196/jmir.7725] [Medline: 28490417]
61. Batterham PJ, Calear AL, O'Dea B, Larsen ME, J Kavanagh D, Titov N, et al. Stakeholder perspectives on evidence for digital mental health interventions: implications for accreditation systems. Digit Health 2019;5:2055207619878069 [FREE Full text] [doi: 10.1177/2055207619878069] [Medline: 31565238]
62. Twomey C, O'Reilly G, Byrne M, Bury M, White A, Kissane S, et al. A randomized controlled trial of the computerized CBT programme, MoodGYM, for public mental health service users waiting for interventions. Br J Clin Psychol 2014 Nov;53(4):433-450. [doi: 10.1111/bjc.12055] [Medline: 24831119]
63. Marshall JM, Dunstan DA, Bartik W. Effectiveness of using mental health mobile apps as digital antidepressants for reducing anxiety and depression: protocol for a multiple baseline across-individuals design. JMIR Res Protoc 2020 Jul 05;9(7):e17159 [FREE Full text] [doi: 10.2196/17159] [Medline: 32623368]
64. Connolly SL, Miller CJ, Koenig CJ, Zamora KA, Wright PB, Stanley RL, et al. Veterans' attitudes toward smartphone app use for mental health care: qualitative study of rurality and age differences. JMIR Mhealth Uhealth 2018 Aug 22;6(8):e10748 [FREE Full text] [doi: 10.2196/10748] [Medline: 30135050]
65. Huckvale K, Nicholas J, Torous J, Larsen M. Smartphone apps for the treatment of mental health conditions: status and considerations. Curr Opin Psychol 2020 Dec;36:65-70 [FREE Full text] [doi: 10.1016/j.copsyc.2020.04.008] [Medline: 32553848]
66. Lagan S, Aquino P, Emerson MR, Fortuna K, Walker R, Torous J. Actionable health app evaluation: translating expert frameworks into objective metrics. NPJ Digit Med 2020;3:100 [FREE Full text] [doi: 10.1038/s41746-020-00312-4] [Medline: 32821855]
67. O'Brien J, Fossey E, Palmer VJ. A scoping review of the use of co-design methods with culturally and linguistically diverse communities to improve or adapt mental health services. Health Soc Care Community 2021 Jan;29(1):1-17. [doi: 10.1111/hsc.13105] [Medline: 32686881]
68. Thabrew H, Fleming T, Hetrick S, Merry S. Co-design of eHealth interventions with children and young people. Front Psychiatry 2018 Oct 18;9:481 [FREE Full text] [doi: 10.3389/fpsyt.2018.00481] [Medline: 30405450]
69. Bevan Jones R, Stallard P, Agha SS, Rice S, Werner-Seidler A, Stasiak K, et al. Practitioner review: co-design of digital mental health technologies with children and young people. J Child Psychol Psychiatry 2020 Aug;61(8):928-940. [doi: 10.1111/jcpp.13258] [Medline: 32572961]

70. 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-S39. [doi: 10.5694/mja2.50349] [Medline: 31587276]

Abbreviations

iCBT: internet-based cognitive behavioral therapy
IVR: integrated voice response
LOE: levels of evidence
MeSH: Medical Subject Headings
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: International Prospective Register of Systematic Reviews
RCT: randomized controlled trial

Edited by J Torous; submitted 23.01.21; peer-reviewed by A Bourla, P Batterham, T Saheb; comments to author 07.03.21; revised version received 26.07.21; accepted 12.08.21; published 26.11.21.

Please cite as:
Saad A, Bruno D, Camara B, D'Agostino J, Bolea-Alamanac B
Self-directed Technology-Based Therapeutic Methods for Adult Patients Receiving Mental Health Services: Systematic Review
JMIR Ment Health 2021;8(11):e27404
URL: https://mental.jmir.org/2021/11/e27404
doi:10.2196/27404
PMID:34842556

Review

Examining the Effectiveness of Gamification in Mental Health Apps for Depression: Systematic Review and Meta-analysis

Stephanie G Six[1], BA; Kaileigh A Byrne[1], PhD; Thomas P Tibbett[2], PhD; Irene Pericot-Valverde[3], PhD

[1]Department of Psychology, Clemson University, Clemson, SC, United States

[2]SAP National Security Services, Inc, Newtown Square, PA, United States

[3]Clemson University School of Health Research, Greenville, SC, United States

Corresponding Author:
Kaileigh A Byrne, PhD
Department of Psychology
Clemson University
418 Brackett Hall
Clemson, SC, 29634
United States
Phone: 1 864 656 3935
Email: kaileib@clemson.edu

Abstract

Background: Previous research showed that computerized cognitive behavioral therapy can effectively reduce depressive symptoms. Some mental health apps incorporate gamification into their app design, yet it is unclear whether features differ in their effectiveness to reduce depressive symptoms over and above mental health apps without gamification.

Objective: The aim of this study was to determine whether mental health apps with gamification elements differ in their effectiveness to reduce depressive symptoms when compared to those that lack these elements.

Methods: A meta-analysis of studies that examined the effect of app-based therapy, including cognitive behavioral therapy, acceptance and commitment therapy, and mindfulness, on depressive symptoms was performed. A total of 5597 articles were identified via five databases. After screening, 38 studies (n=8110 participants) remained for data extraction. From these studies, 50 total comparisons between postintervention mental health app intervention groups and control groups were included in the meta-analysis.

Results: A random effects model was performed to examine the effect of mental health apps on depressive symptoms compared to controls. The number of gamification elements within the apps was included as a moderator. Results indicated a small to moderate effect size across all mental health apps in which the mental health app intervention effectively reduced depressive symptoms compared to controls (Hedges g=–0.27, 95% CI –0.36 to –0.17; *P*<.001). The gamification moderator was not a significant predictor of depressive symptoms (β=–0.03, SE=0.03; *P*=.38), demonstrating no significant difference in effectiveness between mental health apps with and without gamification features. A separate meta-regression also did not show an effect of gamification elements on intervention adherence (β=–1.93, SE=2.28; *P*=.40).

Conclusions: The results show that both mental health apps with and without gamification elements were effective in reducing depressive symptoms. There was no significant difference in the effectiveness of mental health apps with gamification elements on depressive symptoms or adherence. This research has important clinical implications for understanding how gamification elements influence the effectiveness of mental health apps on depressive symptoms.

(JMIR Ment Health 2021;8(11):e32199) doi:10.2196/32199

KEYWORDS

depression; reward; gamification; mental health apps; apps

Introduction

Depression is a highly prevalent mental disorder in the United States that affects 17.3 million people [1]. Effective treatments are available to treat depression, including pharmacotherapy and psychological treatment [2,3]. However, widespread barriers to treatment exist, such as problems of consistent adherence, access to mental health care resources, and cost [4-7]. The

United States alone spends approximately US $71 billion annually on depression treatment [8], which underscores the substantial financial burden that depression can incur. Furthermore, of all adult Americans who experienced a major depressive episode in 2017, 35% did not receive treatment [1].

To mitigate these challenges, companies have created technology-based mental health apps, with the goal of alleviating symptoms of various mental disorders [8,9]. As of 2019, 81% of all Americans owned a smartphone, which highlights the potential impact of technology-based mental health apps: providing a platform, consuming less time, requiring less commitment, and allowing users to move at their own pace [10]. This ubiquity suggests that app technology could make treatment more accessible by providing individuals with cost-efficient tools and apps to aid them between sessions. However, development of these apps is recent and additional attention is needed to identify what is most effective and rewarding about these digital tool kits.

People with depression frequently experience anhedonia, which may result in blunted sensitivity to reward [11-15]. Depressed individuals view rewards, like money or social encouragement, as less motivating than individuals without depressive symptoms [16-18]. Effective therapeutic approaches may benefit from improving this reward-processing deficit, potentially through the use of gamification elements. Gamification is defined as the use of game-design elements and incentives combined with desired behaviors in order to positively influence user motivation, behavior of users, and adherence [19-22]. Previous research suggests that different gamification elements represent motivational affordances that can influence psychological outcomes [23]. These elements include leaderboards, achievements, badges, levels, challenges, and points [23]. The Unified Gamification and Motivation (UGM) model lends a framework for understanding how including gamification elements in therapeutic intervention can enhance treatment engagement [24]. Based on this model, the inclusion of game-like elements would make the intervention more salient, which could increase motivation to use the intervention, thereby resulting in greater treatment usage.

Much of the work demonstrating the effects of gamification elements on reward motivation stems from video game research. For example, in a recent study, young adults were randomized to either a video game training or a control task [25]. After 2 months playing the video game, participants randomized to the video game intervention exhibited higher activation in the ventral striatum (ie, increase in reward activation) during a nongamified task than those assigned to the control task at 2 months posttest. This finding suggests that (1) the effect of video games on reward motivation can transfer to other tasks and (2) video games can enhance individuals' general reward responsiveness to positive stimuli [25]. However, very little research has centered on how individuals suffering from depression are motivated to pursue reward and engage in video games. The capability of video games to enhance reward processing, motivation, and engagement could play a critical role in the development of app technology specifically grappling with anhedonia.

Mental health apps provide a potential way to reduce symptoms and increase adherence by dispensing psychoeducation and other therapeutic skills through an electronic, easily accessible format [26-34]. Some apps include reward-based features, in the form of money, games, or hearts (eg, SPARX-R) [35], while other apps do not mention any type of reward (eg, AI Tess) [36]. Yet, these apps appear effective overall; in an initial meta-analysis collecting data from nine randomized controlled trials (RCTs) with depression as a secondary concern (mean age 36.1 years; male, 34.8%), results indicated that mental health apps led to a large reduction in depressive symptoms [30]. A second meta-analysis (19 RCTs; mean age 30.7 years; female, 63.17%) examined the effect of smartphone mental health apps on a variety of disorders (eg, anxiety, substance use, and sleep problems) including depression, and results showed significant differences between groups in reducing depressive symptoms with a small to moderate effect size [37]. Another meta-analysis, which predominately recruited adults over the age of 16 years (93.3%), examined 45 RCTs with various technological interventions for depressive patients, like symptom tracking, online diaries, and email and phone reminders. Depressive symptoms showed significant reductions in comparison to either wait-list or treatment-as-usual controls [38]. Taken together, there is evidence that mental health apps are effective, but the variability in effect size between meta-analyses suggest there could be another mechanism. Given the UGM model, gamification is a logical next step to replicate prior meta-analyses and add further context. Mental health apps offer a novel, easily accessible way to combine therapy techniques and motivational reward elements, like a video game, yet it is unclear whether reward or gamification features uniquely offer additional advantages in reducing depressive symptoms. Pairing this novel app approach with traditional techniques that are effective for depression may be able to mitigate anhedonia symptoms.

The purpose of this systematic review and meta-analysis is to (1) provide a comprehensive and updated meta-analytical evaluation of the effectiveness of mental health apps in reducing depressive symptoms and (2) to assess whether mental health apps with gamification elements are more effective than those without. Prior research on video games, like apps, indicates high levels of reward motivation and pleasure [25,39]. Additionally, previous systematic reviews and meta-analyses have shown significant reductions in symptoms for depression and other mental disorders while using mental health apps, some of which included cognitive behavioral therapy (CBT) or gamification elements [26,36,37,40,41]. However, no study to date has explored the effectiveness of mental health apps with gamification components in mitigating depressive symptoms.

Previous research shows that gamification can increase motivation to engage with mental health apps [42-44], improve mood [45], and activate the ventral striatum, which can enhance individuals' general reward responsiveness to positive stimuli [21]. Building on this research, we propose that gamification may enhance the efficacy of therapeutic-based apps (eg, CBT) and reduce depressive symptoms through the following mechanism: (1) it might increase engagement with and adherence to mental health apps and (2) it may activate

reward-mediated neural pathways, eliciting positive feelings, which might counteract some negative feelings from depression. We hypothesize that mental health apps that include gamification elements will be more effective in reducing depressive symptoms and increasing adherence than those without such elements.

Methods

Overview

The systematic review and meta-analysis were conducted following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [46]. A protocol was designed and registered through the Open Science Framework after data collection and before data extraction and analysis began [47]. The quality of the studies included was assessed through the Cochrane Collaboration's risk of bias assessment tool [48].

Eligibility and Inclusion Criteria

Studies were included if they met the following criteria: (1) studies involving human participants; (2) mental health apps targeting depression as a primary, secondary, or tertiary outcome; (3) RCTs or experimental or quasi-experimental designs with an active, wait-list, or treatment-as-usual control group; (4) published between January 1, 2005, and December 31, 2020; (5) mental health app intervention groups contain elements of CBT, acceptance and commitment therapy (ACT), behavioral activation (BA), or mindfulness; and (6) provide a measure of depressive symptoms pre- and posttreatment. Studies were excluded if they (1) were books chapters, meta-analyses, reviews, case studies, or opinion pieces; (2) were not written in English; (3) included participants younger than 18 years of age; (4) included participants with a terminal or life-threatening illness (eg, patients with cancer who had depression) to avoid potential confounds of disentangling which condition (ie, life-threatening illness or depression) influenced outcomes; and (5) included a therapist or other mental health specialist's guidance for the mental health apps, as this could create a confound. No specific measure of depression was required, in order to allow the focus to remain on the mental health apps.

Literature Search Strategy

Studies were identified through a comprehensive literature search in PubMed, PsycInfo, Cochrane Clinical Trials Registry, Web of Science, and PsyArXiv (for publication bias) with no publication date restriction. The search was conducted in February 2021. Additionally, the authors conducted a manual search to locate studies that were not identified through databases. Search terms used three different concepts critical to the extant literature: app-based, mental health, and reward or gamification. Within each concept (eg, app-based), we identified multiple tags that reflected this concept (eg, "mental health app" and "MHapp"). The specific combination of operators can be found in Multimedia Appendix 1. This resulted in 129 unique combinations of search terms (eg, "MHapp-Money" and "Depression-Points").

Study Selection Procedure

During the identification phase, articles were identified and collected based on the search term combinations from the five databases. After duplicate removal, two researchers (SGS and KAB) independently conducted initial screening for eligible articles by assessing titles and abstracts for inclusion or exclusion criteria (Multimedia Appendix 2). After the initial screening, both researchers independently assessed the remaining full-text articles against the inclusion and exclusion criteria. Disagreements were resolved through re-examination of the articles in question and discussion among the screeners.

Data Extraction

Two independent reviewers independently coded the studies in a Microsoft Excel spreadsheet. The following data were extracted from each article: first author, year of publication, participants' characteristics (ie, gender and age), population, and study length (Multimedia Appendix 3), as well as app name, app classification (ie, mobile or internet), presence of gamification elements, type of gamification element (eg, digital rewards, challenge or game, and competition or challenges), app adherence, the instrument used to measure depression, and type of therapy (eg, CBT and ACT) offered (Table 1) [31,49-85].

Table 1. Study app classification, therapy, and gamification information.

First author, publication year, and app	App classification	Depression measure	Therapy intervention	Game elements, n	Adherence rate, %
Bakker, 2018 [54]					
MoodMission	Mobile	PHQ-9[a]	CBT[b]	2	69.6
MoodPrism	Mobile	PHQ-9	CBT	2	46.4
MoodKit	Mobile	PHQ-9	CBT	0	46.0
Berger, 2011 [85]					
Deprexis	Mobile	BDI-II[c]	BA[d], PST[e], and mindfulness	1	N/A[f]
Birney, 2016 [80]					
Moodhacker	Internet	PHQ-9	Mindfulness	0	N/A
Bosso, 2020 [67]					
Headspace	Mobile	DASS-21[g]	Mindfulness	3	58.0
Bostock, 2019 [68]					
Headspace	Mobile	HADS[h]	Mindfulness	3	2.0
Botella, 2016 [57]					
Smiling is Fun	Internet	BDI-II	CBT	1	86.4
Choi, 2012 [81]					
Brighten Your Mood	Internet	CBDI[i]	CBT	0	68.0
Collins, 2018 [51]					
MindWise	Internet	PHQ-9	CBT	0	41.7
Dahne, 2019 [66]					
Aptivate	Mobile	BDI-II	CBT	2	36.4
iCouch CBT	Mobile	BDI-II	CBT	0	N/A
Dahne, 2019 [77]					
Moodivate	Mobile	BDI-II	CBT	2	42.9
Moodkit	Mobile	BDI-II	BA	0	N/A
Deady, 2020 [78]					
HeadGear	Internet	PHQ-9	BA and mindfulness	2	10.1
de Graaf, 2009 [62]					
Colour Your Life	Internet	BDI-II	CBT	0	36.0
Fish, 2019 [53]					
Headspace	Mobile	PHQ-9	Mindfulness	3	N/A
Flett, 2018 [69]					
Headspace	Mobile	CES-D[j]	Mindfulness	3	16.4
Smiling Mind	Mobile	CES-D	Mindfulness	1	15.4
Evernote	Mobile	CES-D	N/A	0	N/A
Fuller-Tyszkiewicz, 2020 [73]					
StressLess	Mobile	DASS-21	Mindfulness-based CBT	1	19.0
Stress Monitor	Mobile	DASS-21	N/A	1	N/A
Gilbody, 2015 [63]					
Beating the Blues	Internet	PHQ-9	CBT	0	79.0

First author, publication year, and app	App classification	Depression measure	Therapy intervention	Game elements, n	Adherence rate, %
MoodGYM	Internet	PHQ-9	CBT	2	75.0
Ha, 2020 [49]					
Spring	Mobile	BDI-II	CBT	1	N/A
Howells, 2016 [56]					
Headspace	Mobile	CES-D	Mindfulness	3	29.8
Catch Notes	Mobile	CES-D	N/A	0	N/A
Hur, 2018 [72]					
Todac Todac	Mobile	BDI-II	CBT	3	N/A
Kladnitski, 2020 [71]					
iCBT[k] program	Mobile	PHQ-9	CBT and mindfulness	2	69.4
MEiCBT[l] program	Mobile	PHQ-9	CBT and mindfulness	2	69.7
iMT[m] program	Mobile	PHQ-9	CBT and mindfulness	2	67.6
Krafft, 2019 [58]					
Simple Matrix	Internet	DASS-21	ACT[n]	1	42.9
Complex Matrix	Internet	DASS-21	ACT	2	40.0
Levin, 2020 [75]					
Stop, Breathe, & Think	Internet	CCAPS-34[o]	Mindfulness	0	63.0
Lintvedt, 2013 [79]					
MoodGYM	Mobile	CES-D	CBT	2	N/A
Blue Pages	Mobile	CES-D	N/A	0	N/A
Löbner, 2018 [64]					
MoodGYM	Internet	PHQ-9	CBT	2	13.0
Lokman, 2017 [60]					
CDMIs[p]: Sleep Better, Worry Less, and Stress Less	Mobile	IDS-SR[q]	CBT	1	N/A
Lüdtke, 2018 [59]					
Be Good to Yourself	Internet	PHQ-9	CBT	3	79.6
Mantani, 2017 [52]					
Kokoro	Mobile	BDI-II	CBT	6	40.7
McCloud, 2020 [70]					
Feel Stress Free	Internet	HADS	CBT and mindfulness	2	7.0
Moberg, 2019 [61]					
Pacifica	Internet	DASS-21	CBT and mindfulness	3	N/A
Montero-Marín, 2016 [82]					
Smiling is Fun	Internet	BDI-II	CBT	1	84.3
Richards, 2020 [55]					
Space from Depression	Internet	PHQ-9	CBT	2	N/A
Richards, 2015 [83]					
Space from Depression	Internet	BDI-II	CBT	2	36.0
Roepke, 2015 [31]					
SuperBetter	Internet	CES-D	CBT	7	45.6
Rollman, 2018 [84]					

First author, publication year, and app	App classification	Depression measure	Therapy intervention	Game elements, n	Adherence rate, %
Beating the Blues	Internet	PROMIS[r]	CBT	0	85.8
Schure, 2019 [76]					
Thrive	Mobile	PHQ-9	CBT	3	58.6
Sethi, 2013 [65]					
MoodGYM	Mobile	DASS-21	CBT	2	N/A
Tighe, 2017 [74]					
ibobbly	Mobile	PHQ-9	Mindfulness and ACT	0	85.0
Twomey, 2014 [50]					
MoodGYM	Internet	DASS-21	CBT	2	27.3

[a]PHQ-9: 9-item Patient Health Questionnaire.
[b]CBT: cognitive behavioral therapy.
[c]BDI-II: Beck Depression Inventory-II.
[d]BA: behavioral activation.
[e]PST: problem-solving therapy.
[f]N/A: not applicable; values were not reported.
[g]DASS-21: 21-item Depression, Anxiety, and Stress Scale.
[h]HADS: Hospital Anxiety and Depression Scale.
[i]CBDI: Chinese version of the Beck Depression Inventory.
[j]CES-D: Center for Epidemiological Studies Depression Scale.
[k]iCBT: internet-delivered cognitive behavioral therapy.
[l]MEiCBT: mindfulness-enhanced internet-delivered cognitive behavioral therapy.
[m]iMT: internet-delivered mindfulness training.
[n]ACT: acceptance and commitment therapy.
[o]CCAPS-34: Counseling Center Assessment of Psychological Symptoms-34.
[p]CDMI: complaint-directed mini-intervention.
[q]IDS-SR: Inventory of Depressive Symptomatology Self-Report.
[r]PROMIS: Patient-Reported Outcomes Measurement Information System.

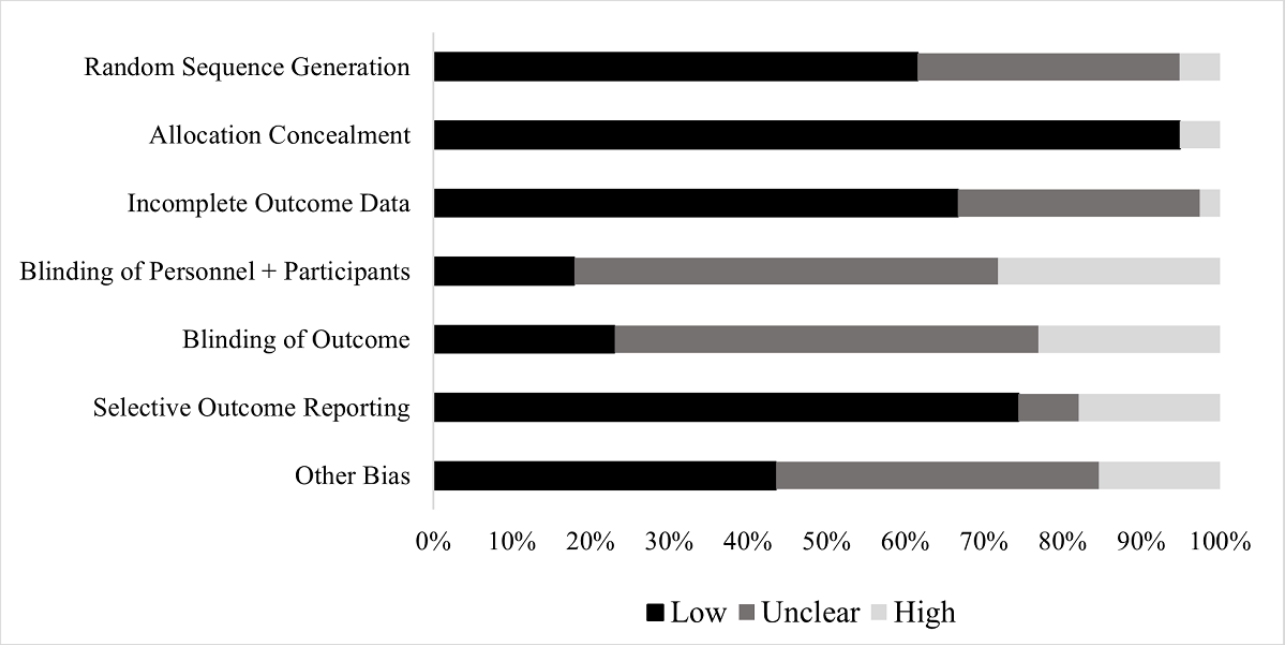
The conceptualization of gamification was modeled after previous research in which gamification was defined as having three components: a design feature that uses motivational affordances to influence psychological and behavioral outcomes [23]. Another literature review that focused exclusively on the health and well-being app domain has described very similar conceptualizations of gamification elements [86]. Modeled after these gamification literature reviews, gamification for this meta-analysis was defined using the following nine motivational affordance categories: points, achievements or badges, levels, narrative stories or themes, clear goals, performance-based feedback, rewards, progress metrics (eg, progress bars), and challenges [23,86]. While gamification can also include leaderboards [19], this element was excluded from the meta-analysis, as leaderboards in the context of mental health may promote social comparison, which can be counterproductive [87,88]. First authors of the studies were contacted via email to confirm conceptualization of the number of gamification elements. The number of gamification elements included in each intervention was included as a moderator in analyses.

Raw depression scores (mean and SD) at posttreatment for each study were extracted. If a study compared more than one mental health app intervention to a control group, or if more than one independent sample was examined in an article, both were included as separate comparisons. Studies used different, but convergently valid, measures of depression. If the articles met any of the exclusion criteria, specifically missing data, they were excluded from data analysis (n=12).

Quality Assessment

The quality of each study was assessed using the Cochrane Collaboration's risk of bias assessment tool, which provided seven basic criteria: random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting, and other biases [48]. Studies were scored on a scale ranging from 0 to 2, where 0 indicates "low or no bias," 1 indicates that the "level of bias is unclear," and 2 indicates "high bias" (Multimedia Appendix 4). In line with previous research, if a study did not address one of the categories, it was given a 1 for the lack of explanation [89]. Total scores for individual and all studies are presented in Figure 1.

Figure 1. Risk of bias across all studies included in the meta-analysis based on the Cochrane Collaboration's risk of bias assessment tool.



Data Analysis

Overview

The resulting depression questionnaires included in the analyzed studies were as follows: the 9-item Patient Health Questionnaire (PHQ-9) [90]; the Beck Depression Inventory-II (BDI-II) [91]; the Center for Epidemiological Studies Depression Scale (CES-D) [92]; the depression subscale of the 21-item Depression, Anxiety, and Stress Scale (DASS-21) [93]; the Hospital Anxiety and Depression Scale (HADS) [94]; the Inventory of Depressive Symptomatology Self-Report (IDS-SR) [95]; the depression scale from the Counseling Center Assessment of Psychological Symptoms-34 (CCAPS-34) [96]; and the depression scale from the Patient-Reported Outcomes Measurement Information System (PROMIS). All questionnaires involved a 4-point scale, ranging from 0 (very low levels of depressive symptoms) to 3 (very high levels of depressive symptoms). Some studies reported sum scores, while others reported average scores. To ensure that all questionnaire data were comparable along the same scale, average depression scores were computed for analysis.

The meta-analytic data were analyzed in RStudio, primarily using the meta (version 4.18) and metafor (version 3.0) packages in R (version 3.6.3; The R Foundation) to determine effect size and between-group differences. Means and variances were aggregated for studies that included two primary measures of depression to compute a single comparison [97]. This led to a total of 50 comparisons in the meta-analysis. From this data, the pooled SD, *t* test value, P value, degrees of freedom, SE, and Hedges g were calculated. The Hedges g effect size provides an index of the magnitude of the difference between two means and corrects for potential biases in small samples [97,98]. The bias correction was performed using the following formula:  where $n = n_1 + n_2$ [99]. An effect size of 0.2 represents a small effect size, 0.5 reflects a moderate effect size, and ≥0.80 represents a large effect size [98,100].

Following previous meta-analyses comparing intervention effects on depressive symptoms [26,37,41], the effectiveness of mental health apps was assessed using one outcome: difference in depressive symptoms between intervention and control groups at posttreatment. To test the hypothesis that mental health apps would be effective in reducing depressive symptoms, a random effects model was used to examine differences in the magnitude of depressive symptomology between those mental health app interventions compared to control conditions. The continuous variable of number of gamification elements was included in the random effects model as a moderator to test whether gamification elements influenced the effectiveness of mental health apps. The duration of the intervention, in months, was also included as a moderator variable. The I^2 statistic was computed to determine heterogeneity across studies: an I^2 value of ≤25% suggests low heterogeneity, ~50% suggests moderate heterogeneity, and ≥75% suggests high heterogeneity across studies [101]. While some articles provided follow-up time point data, only data from the postintervention period or data that were specified as the primary endpoint were analyzed in the primary meta-analysis.

Subgroup and Sensitivity Analyses

A sensitivity analysis using a random effects model with the gamification moderator was conducted to examine the effectiveness of mental health apps on depressive symptoms among the CBT-based apps that excluded ACT and mindfulness-based interventions. A secondary meta-regression analysis with the gamification moderator was performed for studies that included a measure of adherence rates (ie, percentage completion of all intervention modules or requirements) for the intervention condition. The adherence analysis included 28 studies with 37 comparisons.

Assessment of Publication Bias

A funnel plot was created to provide a visual of potential bias. The vertical line indicates the estimated effect of all studies. Pseudo-CIs were generated around this line in homogenous data sets to indicate 95% CI boundaries. Asymmetrical funnel plots suggest that the effects of an intervention in studies with small sample sizes are different—typically more impactful—than in studies with larger sample sizes and may indicate publication bias [102]. However, if model estimates suggest heterogeneity, a transformation manipulates these pseudo-CIs to take the heterogeneity into account: $\pm 1.96 \times \sqrt{(SE^2 + \tau^2)}$ where the τ^2 variable indicates the degree of heterogeneity. Its inclusion in the pseudo-CI calculation results in two curved lines asymptotic to the original estimated effect, a broader and wider funnel more inclusive of variance.

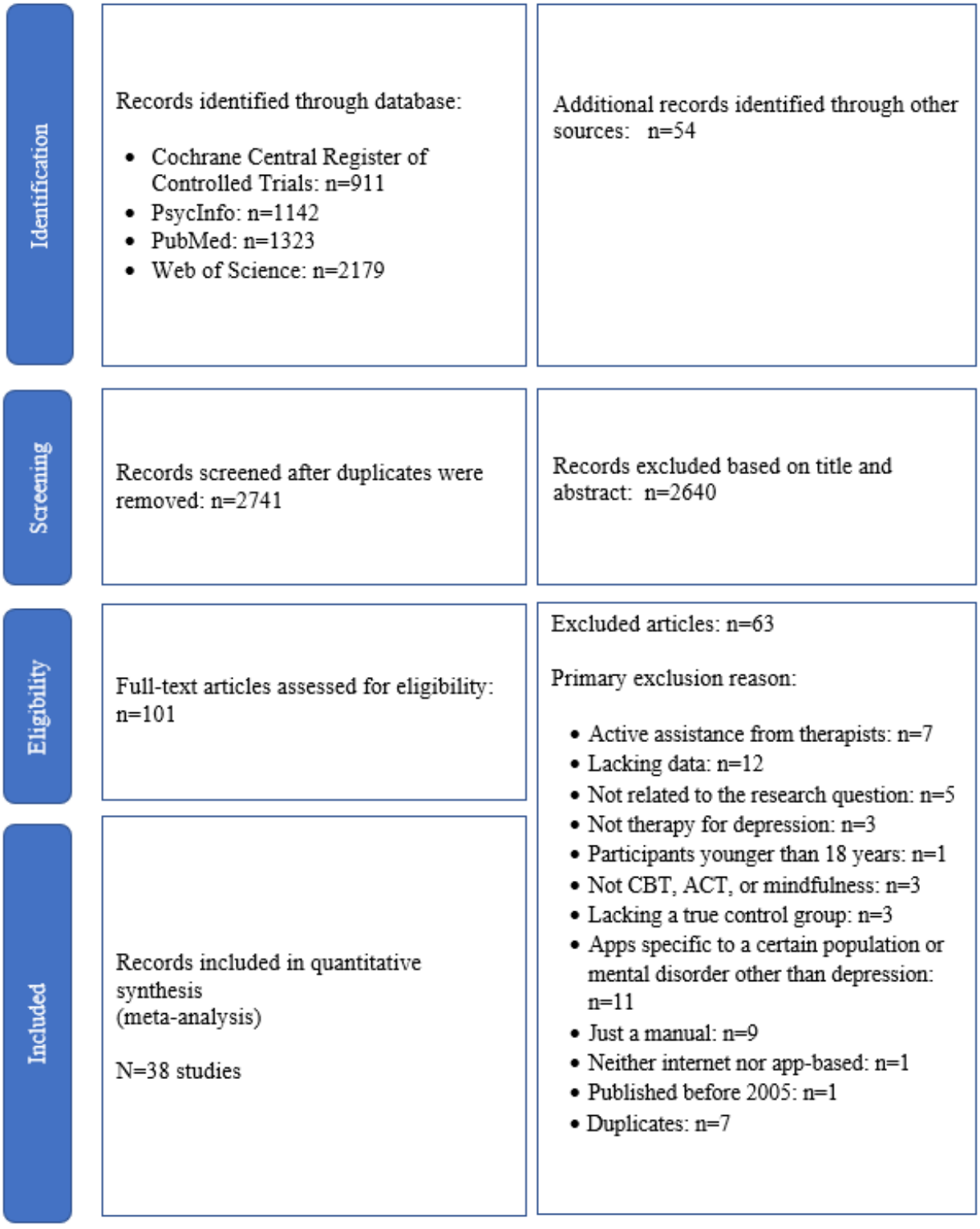
The presence of publication bias was also measured with the Egger test and the trim-and-fill approach by Duval and Tweedie. The Egger test [103] was performed to quantify whether there was significant small-study publication bias in the included studies. The trim-and-fill analysis by Duval and Tweedie was conducted to establish an unbiased estimate of the pooled effect size by correcting for funnel plot asymmetry due to publication bias [104]. Significant findings indicate whether the study sample is asymmetrical or "missing" publications that would positively or negatively bias the estimate.

Results

Results of the Review

Data collection commenced in February 2021 and ended in May 2021. As of June 2021, a total of 5597 articles were identified through the literature search. After duplicate removal, 2741 eligible articles remained for title and abstract screening. Two researchers (SGS and KAB) independently identified 101 articles as potentially eligible. Screeners had an agreement rate of 98.94% (Cohen *k*=0.85). Full-text screening of these 101 articles was conducted. Of the reviewed articles, 38 studies with 39 different mental health apps met the inclusion criteria and were therefore included. Figure 2 presents the study selection identification, screening, and eligibility process.

Figure 2. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram of the studies included in the systematic review and meta-analysis. ACT: acceptance and commitment therapy; CBT: cognitive behavioral therapy.



Participant and Study Characteristics

A total of 8110 participants provided analyzable data (4362, 53.8%, in the interventions and 3748, 46.2%, in the control conditions) for this meta-analysis, with the majority of participants being female (n=4728, 58.3%; mean age 35.6, SD 7.9, years). Multimedia Appendix 3 shows descriptive information for each study. Treatment duration ranged from 10 days to 4 months. Most of the studies (26/38, 68%) used mental health apps with CBT [31,45-48,50,51,53,55-62,66-68,72-75,77-79], 32% (12/38) used mindfulness [49,52,57,63-67,69-71,74,76,82], 8% (3/38) used ACT [62,73,82], 8% (3/38) did not have a therapy associated with the app, and only 5% (2/38) used BA [54,70]. In terms of the depressive symptom outcome measures, 32% (12/38) of the studies used the PHQ-9, 26% (10/38) used the

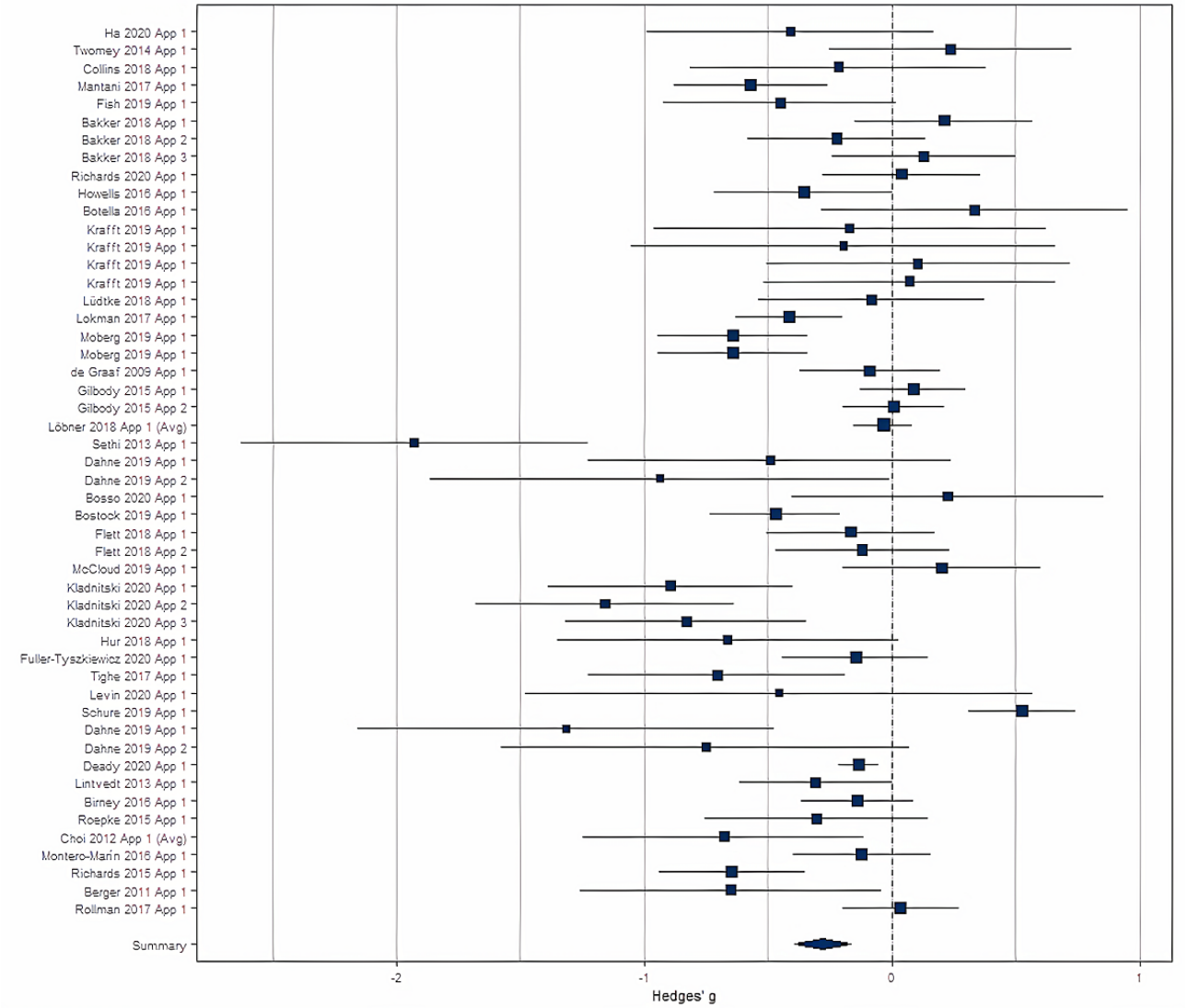
BDI-II, 16% (6/38) used the DASS-21, 11% (4/38) used the CES-D, 3% (1/38) used the HADS, 3% (1/38) used the IDS-SR, 3% (1/38) used the CCAPS-34, and 3% (1/38) used the PROMIS.

Of the 50 different comparisons used in the 38 different studies, 71% (27/38) contained gamification elements and 29% (11/38) did not. The number of gamification elements observed in each study are shown in Multimedia Appendix 5. Table 1 shows the type of intervention, app, and number of gamification elements for each article. Multimedia Appendix 6 reports supplemental meta-analytic results for long-term follow-up time points and control variables.

Primary Analysis

A forest plot for the postintervention differences between the mental health app intervention group and the control group is shown in Figure 3. The random effects model for all eligible studies (n=50 comparisons) revealed a small to medium effect of mental health apps in reducing depressive symptoms compared to controls (g=–0.27, 95% CI –0.36 to –0.17; P<.001). However, significant heterogeneity in the results were observed (I^2=0.76, τ^2=0.076; P<.001). The gamification moderator was not a significant predictor of depressive symptoms (β=–0.03, SE=0.04; P=.38); the intervention duration moderator was also not a significant predictor (β=–0.02, SE=0.04; P=.67).

Figure 3. Forest plot for all studies (n=50 comparisons) showing the effect sizes for each. Author and year for each study are listed on the y-axis.



Sensitivity Analysis: CBT-Only Studies

A sensitivity analysis was performed for app comparisons that involved CBT-based therapy (37/50, 74%). The overall effect size was similar to the overall analysis (g=–0.30, 95% CI –0.42 to –0.17; P<.001), and there was no significant effect of gamification elements as a moderator (β=–0.04, SE=0.04; P=.31) or of study duration as a moderator (β=–0.02, SE=0.05; P=.69). Multimedia Appendix 7 show the funnel plot for this analysis.

Secondary Analysis: Adherence

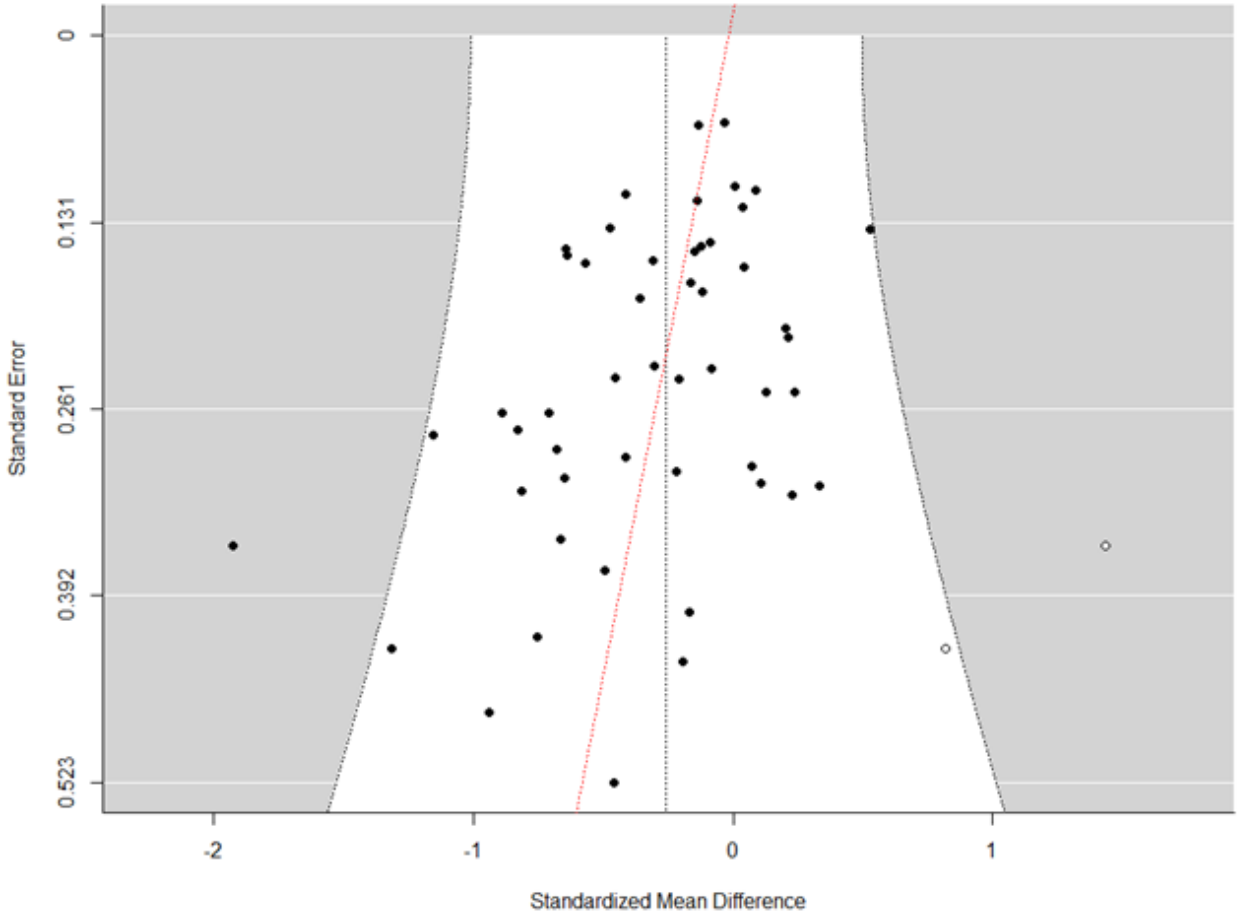
The secondary analysis examining comparisons that included a measure of adherence (36/50, 72%) failed to show a significant effect of gamification elements (β=–1.93, SE=2.28; P=.40) on adherence rates. However, intervention duration was a significant predictor such that longer interventions were associated with higher adherence rates (β=11.33, SE=3.57; P=.003). Similarly, when examining only the CBT-based mental health apps (25/50, 50%), there was no effect of gamification

elements on adherence (β=0.17, SE=2.51; P=.95), but intervention duration positively predicted adherence (β=12.23, SE=4.21; P=.008).

Funnel Plot Results

Examination of the funnel plot of the posttreatment effect SEs indicated heteroscedasticity. The Egger test of asymmetry was significant (Q_{50}=214.46, P<.001). A trim-and-fill analysis suggested a trending finding of three missing publications to the right, though this did not reach the level of statistical significance (β=2.45, P=.13). To be conservative, the pseudo-CIs in Figure 4 were adjusted per τ^2, with trim-and-fill studies as white dots. Controlling for the violations of these assumptions, the overall random effects model was significant (β=–0.26, SE=0.06; P<.001), indicating a significant improvement overall.

Figure 4. Funnel plot for all studies (n=50 comparisons) showing the heterogeneity for each.



Discussion

Principal Findings

This meta-analysis provides a comprehensive update about the effectiveness of mental health apps in reducing depressive symptoms, and it tests whether mental health apps with gamification elements are effective for reducing depressive symptoms. Results indicated that mental health apps are effective for reducing depressive symptoms, but gamification elements within mental health apps do not seem to reduce depressive symptoms or increase the adherence of using mental health apps.

Examination of mental health apps that exclusively employed computerized CBT (excluding mindfulness and ACT) similarly failed to show an effect of gamification on depressive symptom reduction. It should be noted that some of the included studies had a high risk of bias and, as such, the results should be interpreted with caution. Nevertheless, the evidence from this meta-analysis suggests that CBT effectively aids with the control, maintenance, and reduction of depressive symptoms, but gamification and reward elements embedded in mental health apps do not significantly alleviate such symptoms.

Consistent with prior systematic reviews and meta-analyses on digital interventions for depression, we found that mental health app interventions are moderately effective in reducing depressive symptoms compared to controls [26,37,38,41,89]. Effect sizes from these prior meta-analyses ranged from 0.33 to 0.58, which is slightly higher than the small to moderate effect size observed in this meta-analysis. While this research provides strong consensus of a general positive effect of mental health apps on alleviating depressive symptoms, the specific elements that contribute to this effect remain largely elusive. Previous systematic reviews have demonstrated that guidance and support from a professional can augment the effectiveness of mental health app interventions [37,105]. Other work suggests that incorporating reminders into digital mental health interventions can promote engagement and adherence with the intervention,

which may, in turn, enhance the intervention's therapeutic benefits [38]. However, the results of this meta-analysis did not show a significant relationship between gamification and adherence rates. Understanding how such features work together to mitigate depressive symptoms is critical for improving mental health app development.

One potential explanation of why gamification elements do not moderate depressive symptoms may be reward sensitivity. Depression generally has been associated with hyposensitivity to rewards, particularly among those with anhedonic symptoms [11-13,16,17]. This diminished reward-related neural activity may decrease motivation to obtain rewards or diminish the positive reward experience. While the UGM model proposes reward as a critical part of motivation to ensure engagement, it also suggests moderating effects, such as self-efficacy and locus of control. These variables have known negative relationships with depression [106,107] and may make gamification less effective. An alternative explanation for the results may be that gamification does not engender additive benefits. The included mental health apps all used strong, evidence-based therapeutic interventions, including CBT, ACT, and mindfulness. Based on the results of this meta-analysis, these interventions appear to be sufficient in mitigating depressive symptoms.

Overall, 30 studies in this meta-analysis showed a significant effect, while nine studies failed to produce significant reductions in depressive symptoms. In terms of gamification elements, four of the apps that were associated with nonsignificant findings included at least two elements of gamification: Kokoro, Headspace, MoodGYM, and Be Good to Yourself [50,52,59,67]. In contrast, two of these apps—Headspace and MoodGYM—were used in other studies, where they produced significant positive changes in depressive symptoms [31,53,64,65,68,69]. These mixed findings support the notion that while mental health apps may have a future in telemedicine and psychological settings [89,108], more research is needed to understand which mental health app features are integral to improving mental health symptoms. From there, these facets can be properly implemented as a methodologically reliable therapeutic technique.

The push for the creation of efficient and scientifically supported mental health apps grows each year as technology becomes more ubiquitous [109]. These apps could aid therapists who are unable to accept any new clients and people who may not have access to psychological centers or counselors due to financial strains, location, or disabilities. This meta-analysis adds to the current literature by suggesting against overreliance on reward and gamification elements as major reducers of depressive symptoms. These elements may be beneficial in mental health apps, but no more so than other evidence-based therapeutic features. Many of the mental health apps currently available to the general public lack valid testing of their efficacy; thus, there is a strong need for rigorous evaluation of such apps on psychological outcomes. Indeed, in this meta-analysis, nearly half of the included studies incorporated gamification elements, yet the results suggest that such elements exert minimal therapeutic benefits. Significantly more research is needed to identify which specific mental health app features maximize therapeutic effects.

Limitations

This systematic review and meta-analysis had some limitations, which were largely related to inadequacies in the studies available for analyses. Results were calculated based on aggregated samples, which may have caused a certain level of ecological bias. Due to the disparate sample of apps, heterogeneity was detected in all primary analyses. However, heterogeneity was nearly identical to past research on this topic, and we took statistical efforts to minimize this impact [38]. In addition, the meta-analysis was not exclusive to individuals with clinical levels of depression, but all experiencing depressive symptoms. No data sets were consistently available across all studies to assess clinical depression diagnoses or psychiatric comorbidities. Consequently, the findings may not generalize to individuals with severe depression or individuals with other mental health conditions.

Future Research Using Gamification Elements

The main findings of this study were that mental health apps are effective for reducing depressive symptoms, but gamification elements within these mental health apps do not seem to affect depressive symptoms. It is possible that mental health apps with gamification elements may influence patients managing anxiety, stress, or other conditions where anhedonia is not present. Thus, gamification may not be a promising app feature for depressive symptoms but may hold promise for other mental health conditions. Future research should consider examining the effectiveness of mental health apps with gamification on other mental health conditions. While previous research has investigated the effectiveness of mental health apps on anxiety and life satisfaction [50,56], the effect of gamification elements on these psychological factors remains underexplored. Moreover, there is a need for additional research to better characterize the usability benefits and user preferences of mental health apps. If gamification within mental health apps is not effective for individuals experiencing depressive symptoms, then it is important to identify other potential features. Designing specific features that may motivate users with depression toward continued mental health app adherence could lead to beneficial outcomes.

Conclusions

Mental health apps have proven to be a useful tool in reducing depressive symptoms with or without the inclusion of gamification elements. These results demonstrate that although there is a significant improvement in using mental health apps overall, there is no evidence to suggest that gamification makes outcomes significantly better or worse. Additional elements, such as personalization, motivational reminders, social support, and usability, need to be investigated. Mental health apps may provide a readily available option for global psychological care; however, supplementary research is needed on their effectiveness before reliable implementation into the health care system can occur.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms for the meta-analysis.
[DOCX File , 15 KB - mental_v8i11e32199_app1.docx]

Multimedia Appendix 2
Database screening tool.
[DOCX File , 15 KB - mental_v8i11e32199_app2.docx]

Multimedia Appendix 3
Study demographics extracted from articles: sample, length, population, age, and gender.
[DOCX File , 24 KB - mental_v8i11e32199_app3.docx]

Multimedia Appendix 4
Quality assessment screening.
[DOCX File , 21 KB - mental_v8i11e32199_app4.docx]

Multimedia Appendix 5
Gamification screening for publications.
[DOCX File , 25 KB - mental_v8i11e32199_app5.docx]

Multimedia Appendix 6
Sensitivity and supplemental analyses.
[DOCX File , 14 KB - mental_v8i11e32199_app6.docx]

Multimedia Appendix 7
Sensitivity analysis for cognitive behavioral therapy (CBT)–only studies.
[DOCX File , 29 KB - mental_v8i11e32199_app7.docx]

References

1. Major depression. National Institute of Mental Health. 2019. URL: https://www.nimh.nih.gov/health/statistics/major-depression.html [accessed 2021-07-12]
2. National Collaborating Centre for Mental Health (UK). Depression: The Treatment and Management of Depression in Adults (Updated Edition). Leicester, UK: British Psychological Society; 2010.
3. Weir K. APA offers new guidance for treating depression. Monitor on Psychology, American Psychological Association. 2019 Sep 01. URL: https://www.apa.org/monitor/2019/09/ce-corner-depression [accessed 2021-07-15]
4. Cartwright C, Gibson K, Read J, Cowan O, Dehar T. Long-term antidepressant use: Patient perspectives of benefits and adverse effects. Patient Prefer Adherence 2016;10:1401-1407 [FREE Full text] [doi: 10.2147/PPA.S110632] [Medline: 27528803]
5. Stein-Shvachman I, Karpas DS, Werner P. Depression treatment non-adherence and its psychosocial predictors: Differences between young and older adults? Aging Dis 2013 Sep 20;4(6):329-336 [FREE Full text] [doi: 10.14336/AD.2013.0400329] [Medline: 24307966]
6. Wood P, Burwell J, Rawlett K. New study reveals lack of access as root cause for mental health crisis in America. National Council for Mental Wellbeing. 2018 Oct 10. URL: https://www.thenationalcouncil.org/press-releases/new-study-reveals-lack-of-access-as- [accessed 2021-07-15]
7. Van Hees S, Siewe Fodjo JN, Wijtvliet V, Van den Bergh R, Faria de Moura Villela E, da Silva CF, et al. Access to healthcare and prevalence of anxiety and depression in persons with epilepsy during the COVID-19 pandemic: A multicountry online survey. Epilepsy Behav 2020 Nov;112:107350 [FREE Full text] [doi: 10.1016/j.yebeh.2020.107350] [Medline: 32920373]
8. Winerman L. By the numbers: The cost of treatment. Monitor on Psychology, American Psychological Association. 2017 Mar. URL: https://www.apa.org/monitor/2017/03/numbers [accessed 2021-07-15]
9. Neary M, Schueller SM. State of the field of mental health apps. Cogn Behav Pract 2018 Nov;25(4):531-537 [FREE Full text] [doi: 10.1016/j.cbpra.2018.01.002] [Medline: 33100810]
10. Mobile fact sheet. Pew Research Center. Washington, DC: Pew Research Center; 2021 Apr 07. URL: https://www.pewresearch.org/internet/factsheet/mobile/ [accessed 2021-07-15]

11. Admon R, Pizzagalli DA. Dysfunctional reward processing in depression. Curr Opin Psychol 2015 Aug 01;4:114-118 [FREE Full text] [doi: 10.1016/j.copsyc.2014.12.011] [Medline: 26258159]
12. Chen C, Takahashi T, Nakagawa S, Inoue T, Kusumi I. Reinforcement learning in depression: A review of computational research. Neurosci Biobehav Rev 2015 Aug;55:247-267. [doi: 10.1016/j.neubiorev.2015.05.005] [Medline: 25979140]
13. Eshel N, Roiser JP. Reward and punishment processing in depression. Biol Psychiatry 2010 Jul 15;68(2):118-124. [doi: 10.1016/j.biopsych.2010.01.027] [Medline: 20303067]
14. Safra L, Chevallier C, Palminteri S. Depressive symptoms are associated with blunted reward learning in social contexts. PLoS Comput Biol 2019 Jul;15(7):e1007224 [FREE Full text] [doi: 10.1371/journal.pcbi.1007224] [Medline: 31356594]
15. Vrijen C, Ackermans MA, Bosma A, Kretschmer T. Reward responsiveness, optimism, and social and mental functioning in children aged 6-7: Protocol of a cross-sectional pilot study. JMIR Res Protoc 2020 Sep 24;9(9):e18902 [FREE Full text] [doi: 10.2196/18902] [Medline: 32969825]
16. Foti D, Hajcak G. Depression and reduced sensitivity to non-rewards versus rewards: Evidence from event-related potentials. Biol Psychol 2009 Apr;81(1):1-8. [doi: 10.1016/j.biopsycho.2008.12.004] [Medline: 19162124]
17. Yang X, Huang J, Zhu C, Wang Y, Cheung EFC, Chan RCK, et al. Motivational deficits in effort-based decision making in individuals with subsyndromal depression, first-episode and remitted depression patients. Psychiatry Res 2014 Dec 30;220(3):874-882. [doi: 10.1016/j.psychres.2014.08.056] [Medline: 25262638]
18. Liu W, Chan RCK, Wang L, Huang J, Cheung EFC, Gong Q, et al. Deficits in sustaining reward responses in subsyndromal and syndromal major depression. Prog Neuropsychopharmacol Biol Psychiatry 2011 Jun 01;35(4):1045-1052. [doi: 10.1016/j.pnpbp.2011.02.018] [Medline: 21371518]
19. Blohm I, Leimeister JM. Gamification. Bus Inf Syst Eng 2013 Jun 14;5(4):275-278. [doi: 10.1007/s12599-013-0273-5]
20. King D, Greaves F, Exeter C, Darzi A. 'Gamification': Influencing health behaviours with games. J R Soc Med 2013 Mar;106(3):76-78 [FREE Full text] [doi: 10.1177/0141076813480996] [Medline: 23481424]
21. Lister C, West JH, Cannon B, Sax T, Brodegard D. Just a fad? Gamification in health and fitness apps. JMIR Serious Games 2014 Aug 04;2(2):e9 [FREE Full text] [doi: 10.2196/games.3413] [Medline: 25654660]
22. Schmidt-Kraepelin M, Toussaint PA, Thiebes S, Hamari J, Sunyaev A. Archetypes of gamification: Analysis of mHealth apps. JMIR Mhealth Uhealth 2020 Oct 19;8(10):e19280 [FREE Full text] [doi: 10.2196/19280] [Medline: 33074155]
23. Hamari J, Koivisto J, Sarsa H. Does gamification work? -- A literature review of empirical studies on gamification. In: Proceedings of the Annual Hawaii International Conference on System Sciences. 2014 Presented at: Annual Hawaii International Conference on System Sciences; January 6-9, 2014; Waikoloa, HI p. 3025-3034. [doi: 10.1109/hicss.2014.377]
24. Dykens I, Wetzel A, Dorton S, Batchelor E. Towards a unified model of gamification and motivation. In: Proceedings of the International Conference on Human-Computer Interaction. Cham, Switzerland: Springer International Publishing; 2021 Presented at: International Conference on Human-Computer Interaction; July 24-29, 2021; Virtual p. 53-70. [doi: 10.1007/978-3-030-77857-6_4]
25. Cole SW, Yoo DJ, Knutson B. Interactivity and reward-related neural activation during a serious videogame. PLoS One 2012;7(3):e33909 [FREE Full text] [doi: 10.1371/journal.pone.0033909] [Medline: 22442733]
26. Li J, Theng Y, Foo S. Game-based digital interventions for depression therapy: A systematic review and meta-analysis. Cyberpsychol Behav Soc Netw 2014 Aug;17(8):519-527 [FREE Full text] [doi: 10.1089/cyber.2013.0481] [Medline: 24810933]
27. Brandt LR, Hidalgo L, Diez-Canseco F, Araya R, Mohr DC, Menezes PR, et al. Addressing depression comorbid with diabetes or hypertension in resource-poor settings: A qualitative study about user perception of a nurse-supported smartphone app in Peru. JMIR Ment Health 2019 Jun 18;6(6):e11701 [FREE Full text] [doi: 10.2196/11701] [Medline: 31215511]
28. Menezes P, Quayle J, Garcia Claro H, da Silva S, Brandt LR, Diez-Canseco F, et al. Use of a mobile phone app to treat depression comorbid with hypertension or diabetes: A pilot study in Brazil and Peru. JMIR Ment Health 2019 Apr 26;6(4):e11698 [FREE Full text] [doi: 10.2196/11698] [Medline: 31025949]
29. Avalos LA, Aghaee S, Kurtovich E, Quesenberry C, Nkemere L, McGinnis MK, et al. A mobile health mindfulness intervention for women with moderate to moderately severe postpartum depressive symptoms: Feasibility study. JMIR Ment Health 2020 Nov 12;7(11):e17405 [FREE Full text] [doi: 10.2196/17405] [Medline: 33180028]
30. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
31. Roepke AM, Jaffee SR, Riffle OM, McGonigal J, Broome R, Maxwell B. Randomized controlled trial of SuperBetter, a smartphone-based/internet-based self-help tool to reduce depressive symptoms. Games Health J 2015 Jun;4(3):235-246. [doi: 10.1089/g4h.2014.0046] [Medline: 26182069]
32. Griffiths MD, Kuss DJ, Ortiz de Gortari AB. Videogames as therapy: An updated selective review of the medical and psychological literature. In: Alternative Pain Management: Solutions for Avoiding Prescription Drug Overuse. Hershey, PA: IGI Global; 2020:324-363.
33. Lau HM, Smit JH, Fleming TM, Riper H. Serious games for mental health: Are they accessible, feasible, and effective? A systematic review and meta-analysis. Front Psychiatry 2016;7:209 [FREE Full text] [doi: 10.3389/fpsyt.2016.00209] [Medline: 28149281]

34. 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]
35. Fleming T, Dixon R, Frampton C, Merry S. A pragmatic randomized controlled trial of computerized CBT (SPARX) for symptoms of depression among adolescents excluded from mainstream education. Behav Cogn Psychother 2012 Oct;40(5):529-541. [doi: 10.1017/S1352465811000695] [Medline: 22137185]
36. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: Randomized controlled trial. JMIR Ment Health 2018 Dec 13;5(4):e64 [FREE Full text] [doi: 10.2196/mental.9782] [Medline: 30545815]
37. Weisel KK, Fuhrmann LM, Berking M, Baumeister H, Cuijpers P, Ebert DD. Standalone smartphone apps for mental health-a systematic review and meta-analysis. NPJ Digit Med 2019;2:118 [FREE Full text] [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
38. Wahle F, Bollhalder L, Kowatsch T, Fleisch E. Toward the design of evidence-based mental health information systems for people with depression: A systematic literature review and meta-analysis. J Med Internet Res 2017 May 31;19(5):e191 [FREE Full text] [doi: 10.2196/jmir.7381] [Medline: 28566267]
39. Lorenz RC, Gleich T, Gallinat J, Kühn S. Video game training and the reward system. Front Hum Neurosci 2015;9:40 [FREE Full text] [doi: 10.3389/fnhum.2015.00040] [Medline: 25698962]
40. Vallury KD, Jones M, Oosterbroek C. Computerized cognitive behavior therapy for anxiety and depression in rural areas: A systematic review. J Med Internet Res 2015 Jun 05;17(6):e139 [FREE Full text] [doi: 10.2196/jmir.4145] [Medline: 26048193]
41. Wright JH, Owen JJ, Richards D, Eells TD, Richardson T, Brown GK, et al. Computer-assisted cognitive-behavior therapy for depression: A systematic review and meta-analysis. J Clin Psychiatry 2019 Mar 19;80(2):e1-e14 [FREE Full text] [doi: 10.4088/JCP.18r12188] [Medline: 30900849]
42. Lukas CA, Eskofier B, Berking M. A gamified smartphone-based intervention for depression: Randomized controlled pilot trial. JMIR Ment Health 2021 Jul 20;8(7):e16643 [FREE Full text] [doi: 10.2196/16643] [Medline: 34283037]
43. Dias LPS, Barbosa JLV, Vianna HD. Gamification and serious games in depression care: A systematic mapping study. Telematics Inform 2018 Apr;35(1):213-224. [doi: 10.1016/j.tele.2017.11.002]
44. Christie GI, Shepherd M, Merry SN, Hopkins S, Knightly S, Stasiak K. Gamifying CBT to deliver emotional health treatment to young people on smartphones. Internet Interv 2019 Dec;18:100286 [FREE Full text] [doi: 10.1016/j.invent.2019.100286] [Medline: 31890633]
45. Russoniello C, O'Brien K, Parks J. The effectiveness of casual video games in improving mood and decreasing stress. J CyberTherapy Rehabil 2009;2(1):53-66 [FREE Full text]
46. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. J Clin Epidemiol 2009 Oct;62(10):e1-e34 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.06.006] [Medline: 19631507]
47. Six S, Byrne KA. Rewards in mental health applications for aiding depression: A meta-analysis. Open Science Framework. Charlottesville, VA: Center for Open Science; 2021 May 27. URL: https://osf.io/ypm4x/ [accessed 2021-11-22]
48. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
49. Ha SW, Kim J. Designing a scalable, accessible, and effective mobile app based solution for common mental health problems. Int J Hum Comput Interact 2020 Apr 26;36(14):1354-1367. [doi: 10.1080/10447318.2020.1750792]
50. Twomey C, O'Reilly G, Byrne M, Bury M, White A, Kissane S, et al. A randomized controlled trial of the computerized CBT programme, MoodGYM, for public mental health service users waiting for interventions. Br J Clin Psychol 2014 Nov;53(4):433-450. [doi: 10.1111/bjc.12055] [Medline: 24831119]
51. Collins S, Byrne M, Hawe J, O'Reilly G. Evaluation of a computerized cognitive behavioural therapy programme, MindWise (2.0), for adults with mild-to-moderate depression and anxiety. Br J Clin Psychol 2018 Jun;57(2):255-269. [doi: 10.1111/bjc.12165] [Medline: 29197102]
52. Mantani A, Kato T, Furukawa TA, Horikoshi M, Imai H, Hiroe T, et al. Smartphone cognitive behavioral therapy as an adjunct to pharmacotherapy for refractory depression: Randomized controlled trial. J Med Internet Res 2017 Nov 03;19(11):e373 [FREE Full text] [doi: 10.2196/jmir.8602] [Medline: 29101095]
53. Fish MT, Saul AD. The gamification of meditation: A randomized-controlled study of a prescribed mobile mindfulness meditation application in reducing college students' depression. Simul Gaming 2019 Jun 04;50(4):419-435. [doi: 10.1177/1046878119851821]
54. Bakker D, Kazantzis N, Rickwood D, Rickard N. A randomized controlled trial of three smartphone apps for enhancing public mental health. Behav Res Ther 2018 Oct;109:75-83. [doi: 10.1016/j.brat.2018.08.003] [Medline: 30125790]
55. Richards D, Enrique A, Eilert N, Franklin M, Palacios J, Duffy D, et al. A pragmatic randomized waitlist-controlled effectiveness and cost-effectiveness trial of digital interventions for depression and anxiety. NPJ Digit Med 2020;3:85 [FREE Full text] [doi: 10.1038/s41746-020-0293-8] [Medline: 32566763]

56. Howells A, Ivtzan I, Eiroa-Orosa FJ. Putting the 'app' in happiness: A randomised controlled trial of a smartphone-based mindfulness intervention to enhance wellbeing. J Happiness Stud 2014 Oct 29;17(1):163-185. [doi: 10.1007/s10902-014-9589-1]
57. Botella C, Mira A, Moragrega I, García-Palacios A, Bretón-López J, Castilla D, et al. An internet-based program for depression using activity and physiological sensors: Efficacy, expectations, satisfaction, and ease of use. Neuropsychiatr Dis Treat 2016;12:393-406 [FREE Full text] [doi: 10.2147/NDT.S93315] [Medline: 27042067]
58. Krafft J, Potts S, Schoendorff B, Levin ME. Arandomized controlled trial of multiple versions of an acceptance and commitment therapy matrix app for well-being. Behav Modif 2019 Mar;43(2):246-272. [doi: 10.1177/0145445517748561] [Medline: 29262693]
59. Lüdtke T, Pult LK, Schröder J, Moritz S, Bücker L. A randomized controlled trial on a smartphone self-help application (Be Good to Yourself) to reduce depressive symptoms. Psychiatry Res 2018 Nov;269:753-762. [doi: 10.1016/j.psychres.2018.08.113] [Medline: 30273901]
60. Lokman S, Leone SS, Sommers-Spijkerman M, van der Poel A, Smit F, Boon B. Complaint-directed mini-interventions for depressive complaints: A randomized controlled trial of unguided web-based self-help interventions. J Med Internet Res 2017 Jan 04;19(1):e4 [FREE Full text] [doi: 10.2196/jmir.6581] [Medline: 28052840]
61. Moberg C, Niles A, Beermann D. Guided self-help works: Randomized waitlist controlled trial of Pacifica, a mobile app integrating cognitive behavioral therapy and mindfulness for stress, anxiety, and depression. J Med Internet Res 2019 Jun 08;21(6):e12556 [FREE Full text] [doi: 10.2196/12556] [Medline: 31199319]
62. de Graaf LE, Gerhards SAH, Arntz A, Riper H, Metsemakers JFM, Evers SMAA, et al. Clinical effectiveness of online computerised cognitive-behavioural therapy without support for depression in primary care: Randomised trial. Br J Psychiatry 2009 Jul;195(1):73-80. [doi: 10.1192/bjp.bp.108.054429] [Medline: 19567900]
63. Gilbody S, Littlewood E, Hewitt C, Brierley G, Tharmanathan P, Araya R, REEACT Team. Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): Large scale pragmatic randomised controlled trial. BMJ 2015 Nov 11;351:h5627 [FREE Full text] [doi: 10.1136/bmj.h5627] [Medline: 26559241]
64. Löbner M, Pabst A, Stein J, Dorow M, Matschinger H, Luppa M, et al. Computerized cognitive behavior therapy for patients with mild to moderately severe depression in primary care: A pragmatic cluster randomized controlled trial (@ktiv). J Affect Disord 2018 Oct 01;238:317-326. [doi: 10.1016/j.jad.2018.06.008] [Medline: 29902736]
65. Sethi S. Treating youth depression and anxiety: A randomised controlled trial examining the efficacy of computerised versus face‐to‐face cognitive behaviour therapy. Aust Psychol 2020 Nov 12;48(4):249-257. [doi: 10.1111/ap.12006]
66. Dahne J, Collado A, Lejuez CW, Risco CM, Diaz VA, Coles L, et al. Pilot randomized controlled trial of a Spanish-language behavioral activation mobile app (¡Aptívate!) for the treatment of depressive symptoms among United States Latinx adults with limited English proficiency. J Affect Disord 2019 May 01;250:210-217 [FREE Full text] [doi: 10.1016/j.jad.2019.03.009] [Medline: 30870770]
67. Bosso KB. The Effects of Mindfulness Training on BDNF Levels, Depression, Anxiety, and Stress Levels of College Students [undergraduate thesis]. Boca Raton, FL: Florida Atlantic University; 2020 May. URL: http://fau.digital.flvc.org/islandora/object/fau%3A42592 [accessed 2021-11-22]
68. Bostock S, Crosswell AD, Prather AA, Steptoe A. Mindfulness on-the-go: Effects of a mindfulness meditation app on work stress and well-being. J Occup Health Psychol 2019 Feb;24(1):127-138 [FREE Full text] [doi: 10.1037/ocp0000118] [Medline: 29723001]
69. Flett JAM, Hayne H, Riordan BC, Thompson LM, Conner TS. Mobile mindfulness meditation: A randomised controlled trial of the effect of two popular apps on mental health. Mindfulness 2018 Oct 31;10(5):863-876. [doi: 10.1007/s12671-018-1050-9]
70. McCloud T, Jones R, Lewis G, Bell V, Tsakanikos E. Effectiveness of a mobile app intervention for anxiety and depression symptoms in university students: Randomized controlled trial. JMIR Mhealth Uhealth 2020 Jul 31;8(7):e15418 [FREE Full text] [doi: 10.2196/15418] [Medline: 32735221]
71. Kladnitski N, Smith J, Uppal S, James MA, Allen AR, Andrews G, et al. Transdiagnostic internet-delivered CBT and mindfulness-based treatment for depression and anxiety: A randomised controlled trial. Internet Interv 2020 Apr;20:100310 [FREE Full text] [doi: 10.1016/j.invent.2020.100310] [Medline: 32140427]
72. Hur J, Kim B, Park D, Choi S. A scenario-based cognitive behavioral therapy mobile app to reduce dysfunctional beliefs in individuals with depression: A randomized controlled trial. Telemed J E Health 2018 Sep;24(9):710-716. [doi: 10.1089/tmj.2017.0214] [Medline: 29323626]
73. Fuller-Tyszkiewicz M, Richardson B, Little K, Teague S, Hartley-Clark L, Capic T, et al. Efficacy of a smartphone app intervention for reducing caregiver stress: Randomized controlled trial. JMIR Ment Health 2020 Jul 24;7(7):e17541 [FREE Full text] [doi: 10.2196/17541] [Medline: 32706716]
74. Tighe J, Shand F, Ridani R, Mackinnon A, De La Mata N, Christensen H. Ibobbly mobile health intervention for suicide prevention in Australian Indigenous youth: A pilot randomised controlled trial. BMJ Open 2017 Jan 27;7(1):e013518 [FREE Full text] [doi: 10.1136/bmjopen-2016-013518] [Medline: 28132007]
75. Levin ME, Hicks ET, Krafft J. J Am Coll Health 2020 Mar 09:1-9. [doi: 10.1080/07448481.2020.1728281] [Medline: 32150519]

76. Schure MB, Lindow JC, Greist JH, Nakonezny PA, Bailey SJ, Bryan WL, et al. Use of a fully automated internet-based cognitive behavior therapy intervention in a community population of adults with depression symptoms: Randomized controlled trial. J Med Internet Res 2019 Nov 18;21(11):e14754 [FREE Full text] [doi: 10.2196/14754] [Medline: 31738173]
77. Dahne J, Lejuez CW, Diaz VA, Player MS, Kustanowitz J, Felton JW, et al. Pilot randomized trial of a self-help behavioral activation mobile app for utilization in primary care. Behav Ther 2019 Jul;50(4):817-827 [FREE Full text] [doi: 10.1016/j.beth.2018.12.003] [Medline: 31208690]
78. Deady M, Glozier N, Calvo R, Johnston D, Mackinnon A, Milne D, et al. Preventing depression using a smartphone app: A randomized controlled trial. Psychol Med 2020 Jul 06:1-10. [doi: 10.1017/S0033291720002081] [Medline: 32624013]
79. Lintvedt OK, Griffiths KM, Sørensen K, Østvik AR, Wang CEA, Eisemann M, et al. Evaluating the effectiveness and efficacy of unguided internet-based self-help intervention for the prevention of depression: A randomized controlled trial. Clin Psychol Psychother 2013;20(1):10-27. [doi: 10.1002/cpp.770] [Medline: 21887811]
80. Birney AJ, Gunn R, Russell JK, Ary DV. MoodHacker mobile web app with email for adults to self-manage mild-to-moderate depression: Randomized controlled trial. JMIR Mhealth Uhealth 2016 Jan 26;4(1):e8 [FREE Full text] [doi: 10.2196/mhealth.4231] [Medline: 26813737]
81. Choi I, Zou J, Titov N, Dear BF, Li S, Johnston L, et al. Culturally attuned internet treatment for depression amongst Chinese Australians: A randomised controlled trial. J Affect Disord 2012 Feb;136(3):459-468. [doi: 10.1016/j.jad.2011.11.003] [Medline: 22177742]
82. Montero-Marín J, Araya R, Pérez-Yus MC, Mayoral F, Gili M, Botella C, et al. An internet-based intervention for depression in primary care in Spain: A randomized controlled trial. J Med Internet Res 2016 Aug 26;18(8):e231 [FREE Full text] [doi: 10.2196/jmir.5695] [Medline: 27565118]
83. Richards D, Timulak L, O'Brien E, Hayes C, Vigano N, Sharry J, et al. A randomized controlled trial of an internet-delivered treatment: Its potential as a low-intensity community intervention for adults with symptoms of depression. Behav Res Ther 2015 Dec;75:20-31. [doi: 10.1016/j.brat.2015.10.005] [Medline: 26523885]
84. Rollman BL, Herbeck Belnap B, Abebe KZ, Spring MB, Rotondi AJ, Rothenberger SD, et al. Effectiveness of online collaborative care for treating mood and anxiety disorders in primary care: A randomized clinical trial. JAMA Psychiatry 2018 Jan 01;75(1):56-64 [FREE Full text] [doi: 10.1001/jamapsychiatry.2017.3379] [Medline: 29117275]
85. Berger T, Hämmerli K, Gubser N, Andersson G, Caspar F. Internet-based treatment of depression: A randomized controlled trial comparing guided with unguided self-help. Cogn Behav Ther 2011;40(4):251-266. [doi: 10.1080/16506073.2011.616531] [Medline: 22060248]
86. Johnson D, Deterding S, Kuhn K, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: A systematic review of the literature. Internet Interv 2016 Nov;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]
87. Esteves J, Valogianni K, Greenhill A. Online social games: The effect of social comparison elements on continuance behaviour. Inf Manag 2021 Jun;58(4):103452. [doi: 10.1016/j.im.2021.103452]
88. Lian S, Sun X, Niu G, Zhou Z. Upward social comparison on SNS and depression: A moderated mediation model and gender difference. Acta Psychol Sinica 2017;49(7):941-952. [doi: 10.3724/sp.j.1041.2017.00941]
89. 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]
90. 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]
91. Beck AT, Brown GK, Steer RA. Beck Depression Inventory-II (BDI-II). Pearson. 1996. URL: https://www.pearsonclinical.ca/store/caassessments/en/Store/Professional-Assessments/Personality-%26-Biopsychosocial/Brief/Beck-Depression-Inventory-II/p/P100008037.html [accessed 2021-11-22]
92. Eaton W, Smith C, Ybarra M, Muntaner C, Tien A. Center for Epidemiologic Studies Depression Scale: Review and revision (CESD and CESD-R). In: Maruish ME, editor. The Use of Psychological Testing for Treatment Planning and Outcomes Assessment. 3rd edition. Volume 3: Instruments for Adults. Mahwah, NJ: Lawrence Erlbaum Associates; 2004:363-377.
93. Lovibond PF, Lovibond SH. The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995 Mar;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-u] [Medline: 7726811]
94. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
95. 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]
96. Locke BD, McAleavey AA, Zhao Y, Lei P, Hayes JA, Castonguay LG, et al. Development and initial validation of the Counseling Center Assessment of Psychological Symptoms–34. Meas Eval Couns Dev 2017 Mar 10;45(3):151-169. [doi: 10.1177/0748175611432642]
97. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to Meta-Analysis. Chichester, UK: John Wiley & Sons, Ltd; 2009.

98. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. Front Psychol 2013 Nov 26;4:863 [FREE Full text] [doi: 10.3389/fpsyg.2013.00863] [Medline: 24324449]
99. Hedges LV, Olkin I. Statistical Methods for Meta-Analysis. Orlando, FL: Academic Press, Inc; 1985.
100. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155] [Medline: 19565683]
101. Higgins JPT, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002 Jun 15;21(11):1539-1558. [doi: 10.1002/sim.1186] [Medline: 12111919]
102. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions. 2nd edition. Hoboken, NJ: John Wiley & Sons, Ltd; 2019.
103. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. BMJ 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
104. Duval S, Tweedie R. Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 2000 Jun;56(2):455-463. [doi: 10.1111/j.0006-341x.2000.00455.x] [Medline: 10877304]
105. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on internet-based mental health interventions — A systematic review. Internet Interv 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
106. Benassi VA, Sweeney PD, Dufour CL. Is there a relation between locus of control orientation and depression? J Abnorm Psychol 1988 Aug;97(3):357-367. [doi: 10.1037//0021-843x.97.3.357] [Medline: 3057032]
107. Bandura A. Self-efficacy mechanism in human agency. Am Psychologist 1982;37(2):122-147. [doi: 10.1037/0003-066x.37.2.122]
108. Hategan A, Giroux C, Bourgeois JA. Digital technology adoption in psychiatric care: An overview of the contemporary shift from technology to opportunity. J Technol Behav Sci 2019 Mar 7;4(3):171-177. [doi: 10.1007/s41347-019-00086-x]
109. Czeisler M, Lane RI, Petrosky E, Wiley JF, Christensen A, Njai R, et al. Mental health, substance use, and suicidal ideation during the COVID-19 pandemic - United States, June 24-30, 2020. MMWR Morb Mortal Wkly Rep 2020 Aug 14;69(32):1049-1057 [FREE Full text] [doi: 10.15585/mmwr.mm6932a1] [Medline: 32790653]

Abbreviations

ACT: acceptance and commitment therapy
BA: behavioral activation
BDI-II: Beck Depression Inventory-II
CBT: cognitive behavioral therapy
CCAPS-34: Counseling Center Assessment of Psychological Symptoms-34
CES-D: Center for Epidemiological Studies Depression Scale
DASS-21: 21-item Depression, Anxiety, and Stress Scale
HADS: Hospital Anxiety and Depression Scale
IDS-SR: Inventory of Depressive Symptomatology Self-Report
PHQ-9: 9-item Patient Health Questionnaire
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROMIS: Patient-Reported Outcomes Measurement Information System
RCT: randomized controlled trial
UGM: Unified Gamification and Motivation

Edited by J Torous, G Eysenbach; submitted 24.07.21; peer-reviewed by T Baranowski, R Nogueira-Arjona, M Schure; comments to author 15.08.21; revised version received 10.10.21; accepted 11.10.21; published 29.11.21.

Please cite as:
Six SG, Byrne KA, Tibbett TP, Pericot-Valverde I
Examining the Effectiveness of Gamification in Mental Health Apps for Depression: Systematic Review and Meta-analysis
JMIR Ment Health 2021;8(11):e32199
URL: https://mental.jmir.org/2021/11/e32199
doi:10.2196/32199
PMID:34847058

Original Paper

Effectiveness of Self-Guided Virtual Reality–Based Cognitive Behavioral Therapy for Panic Disorder: Randomized Controlled Trial

Bokyoung Shin[1], MA; Jooyoung Oh[2], MD, PhD; Byung-Hoon Kim[2], MD, PhD; Hesun Erin Kim[1], PhD; Hyunji Kim[1], MA; Suji Kim[1], MA; Jae-Jin Kim[2], MD, PhD

[1]Institute of Behavioral Sciences in Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea

[2]Department of Psychiatry, Yonsei University College of Medicine, Gangnam Severance Hospital, Yonsei University Health System, Seoul, Republic of Korea

Corresponding Author:
Jooyoung Oh, MD, PhD
Department of Psychiatry, Yonsei University College of Medicine
Gangnam Severance Hospital, Yonsei University Health System
211 Eonju-ro, Gangnam-gu
Seoul, 06273
Republic of Korea
Phone: 82 2 2019 3342
Fax: 82 2 2019 4926
Email: ojuojuoju@yuhs.ac

Abstract

Background: Virtual reality (VR) is as effective a technique as traditional cognitive behavioral therapy (CBT) and a promising tool for treating panic disorder symptoms because VR exposure can be safer and has better acceptability than in vivo exposure and is more immersive than exposure through imagination. CBT techniques can be delivered more effectively using VR as well. So far, VR has required high-quality devices, but the development of mobile VR technology has improved user availability. At the same time, a well-structured form of VR can be reproduced and used anywhere. This means that VR can be used to provide a self-guided form of treatment and address the high treatment costs of evidence-based therapy and the lack of professional therapists. This study aimed to investigate the potential of self-guided VR as an alternative to high-cost treatment.

Objective: The main goal of this study was to offer data about the efficacy of a mobile app-based self-led VR CBT in the treatment of panic disorder.

Methods: A total of 54 subjects with panic disorder were enrolled in this study and randomly assigned to either the VR treatment group or waitlist group. The VR treatment was designed to be total 12 sessions for 4 weeks. The VR treatment consists of 4 steps in which patients are gradually exposed to phobic stimuli while learning to cope with panic symptoms in each stage. The effectiveness of treatment was assessed through the Panic Disorder Severity Scale, Hamilton Rating Scale for Depression, Body Sensations Questionnaire, Albany Panic and Phobia Questionnaire, Anxiety Sensitivity Index, State-Trait Anxiety Inventory, Hospital Anxiety and Depression Scale, Korean Inventory of Social Avoidance and Distress Scale, Korean Inventory of Depressive Symptomatology, and Perceived Stress Scale. In addition, physiological changes using heart rate variability were evaluated.

Results: In within-group analyses, the VR treatment group exhibited improvements in panic disorder symptoms, anxiety, and depression after 4 weeks, while the waitlist group did not show any significant improvement. Compared to the waitlist group, the VR treatment group showed significantly greater improvements in the Panic Disorder Severity Scale in both completer analysis and intention-to-treat analysis. Heart rate variability in the VR treatment group showed improvement in normalized high frequency from baseline to postassessment with no significant differences in any outcome measure between groups.

Conclusions: The self-guided, mobile app-based VR intervention was effective in the treatment of panic symptoms and restoring the autonomic nervous system demonstrating the validity of the use of VR for self-guided treatment. VR treatment can be a cost-effective therapeutic approach.

Trial Registration: ClinicalTrials.gov NCT04985019; https://clinicaltrials.gov/ct2/show/NCT04985019

(JMIR Ment Health 2021;8(11):e30590) doi:10.2196/30590

KEYWORDS

virtual reality; panic disorder; cognitive behavioral therapy; exposure therapy; intervention

Introduction

Panic disorder [1], with or without agoraphobia, is one of the most common mental disorders in the general population [2]. Panic disorder is characterized by a sudden anxiety with physiological symptoms including palpitations, sweating, and choking sensations accompanied by cognitive symptoms such as catastrophizing and fear of dying, which can lead to avoidance of particular places or situations [3]. Avoiding public places to reduce fear or panic prompt negative consequences that decrease the quality of life of patients [4-7].

As demonstrated in several studies, the most empirically supported psychosocial treatment for panic disorder is cognitive behavioral therapy (CBT) [8-10]. The classic form of CBT for panic disorder, proposed by Clark and Wells [11], includes education on dysfunctional thoughts, exposure to the feared situation, interoceptive exposure, cognitive restructuring, breathing retraining, and applied relaxation. These cognitive and behavioral techniques are designed to help patients realize and correct dysfunctional thoughts and behavior related to their anxiety [12].

The core contents of CBT for panic disorder are repeated exposure to feared situations and sensations, supported by coping skill trainings including breathing training and progressive muscle relaxation [9,13]. Previously, exposure therapy has been performed by imagination or in vivo exposure [14]. However, imagination by itself may not make patients fully immersed [15]. In vivo exposure has the disadvantages of temporal/spatial constraints and high costs. Patients with severe symptoms also tend to misrecognize and avoid in vivo exposure as aversive [16-18]. As an alternative, virtual reality (VR) exposure can be safer with better acceptability than in vivo exposure and is more immersive than exposure through imagination [19]. VR not only makes it easy to reach diverse places, such as an airplane, but it also has the advantage of the participant being able to easily escape from difficult situations during treatment by taking off the head-mounted display. Delivering other CBT techniques like abdominal breathing and progressive muscle relaxation with VR can increase effectiveness. Since VR can be used repeatedly, psychoeducation about the panic disorder such as prevalence, prognosis, possible effect, and side effect of each type of medication is possible without any additional manpower by using a virtual therapist [20].

As already noted, breathing training and progressive muscle relaxation are important parts of CBT for panic disorder as coping skill training with the purpose of modifying pathological breathing, training abdominal breathing [21,22], reducing general tension, and decreasing the risk of panic [23-25]. VR has advantages not only in exposure but also in coping skill training. For instance, in VR, if a virtual therapist demonstrates and teaches breathing and progressive muscle relaxation training, patients will be able to understand it more easily. There is evidence that the additive effects of breathing training plus exposure yielded better outcomes than exposure without breathing training [25]. In addition, progressive muscle relaxation in particular reduces panic symptoms when used in combination with interceptive exposure [26]. Accordingly, this simultaneous use of coping skill training and exposure technique can be thought of as an effective way to manage the panic symptoms. Since VR seems to lend an advantage to both coping skill training and exposure treatment, it will be of great help in simultaneous use of them as well.

Increasing numbers of studies have found VR is as effective a technique as traditional CBT and a promising tool for treating panic disorder symptoms [18,20,27-29]. Many studies have reported that VR and in vivo exposure show the same therapeutic effects, both being significantly better than the control group [18,20,27-32]. VR-based CBT could significantly reduce the number of panic attacks, level of depression, state and trait anxiety, and agoraphobia symptoms [28]. Even though these studies were conducted on a relatively small sample size, the consistent results support the use of VR as an effective tool in the treatment of panic disorder [33].

Recently, the development of mobile VR technology with low-cost devices have improved user availability [34]. So far, VR has usually been implemented as a therapist-led intervention, which made the VR costly. Contrary to this, self-guided VR has potential as an alternative to high-cost treatment [35,36]. Previously, some studies have explored the feasibility of a self-help VR app [20,37-39]. Lindner et al [20] demonstrated that the smartphone version of self-led VR exposure was as effective for treating public speaking anxiety within a therapist-led treatment format. Although effectiveness of self-led VR exposure was shown, VR was only a tool for exposure, indicating that it was partially self-led therapy [20]. Donker et al [37] implemented a fully self-guided VR-based CBT app for acrophobia, and it could significantly reduce acrophobia symptoms compared with the control group. However, this study relied on only self-reported measurement, so the treatment effect was not sufficiently validated by objective evaluation. More importantly, there was still no study that verified the effectiveness of self-guided VR for the treatment of panic disorder.

In this study, we developed a mobile app-based self-led VR CBT that included comprehensive components of CBT and various VR exposure contents. Patients learned the CBT techniques from a virtual therapist and were repeatedly exposed to realistic virtual environments. We studied whether this VR app for a self-help treatment was efficacious for panic disorder. It was hypothesized that individuals using a VR app would show greater changes in posttreatment measurement scores than the waitlist group. The effectiveness was assessed by the trained psychologists and through self-reported measurements. In addition, we measured physiological changes using heart rate variability before and after treatments, which was not investigated in previous studies.

Methods

Participants

A total of 61 participants who were diagnosed with panic disorder by a psychiatrist according to diagnostic criteria in the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*, at a psychiatric outpatient clinic (Yonsei University Gangnam Severance Hospital, Seoul, South Korea) were recruited for this study. To be included, participants met the diagnostic criteria for panic disorder (with or without agoraphobia) based on the Mini-International Neuropsychiatric Interview, had no change in drug dosage during the study period, and were aged 19 to 60 years. Patients with a history of major neurological or significant medical illness or who met the diagnostic criteria of current substance misuse were excluded. Written informed consent was acquired from all participants at the first visit. The study design and protocol were approved by the institutional review board of Yonsei University Gangnam Severance hospital (3-2018-0292). The trial was registered at ClinicalTrials.gov [NCT04985019].

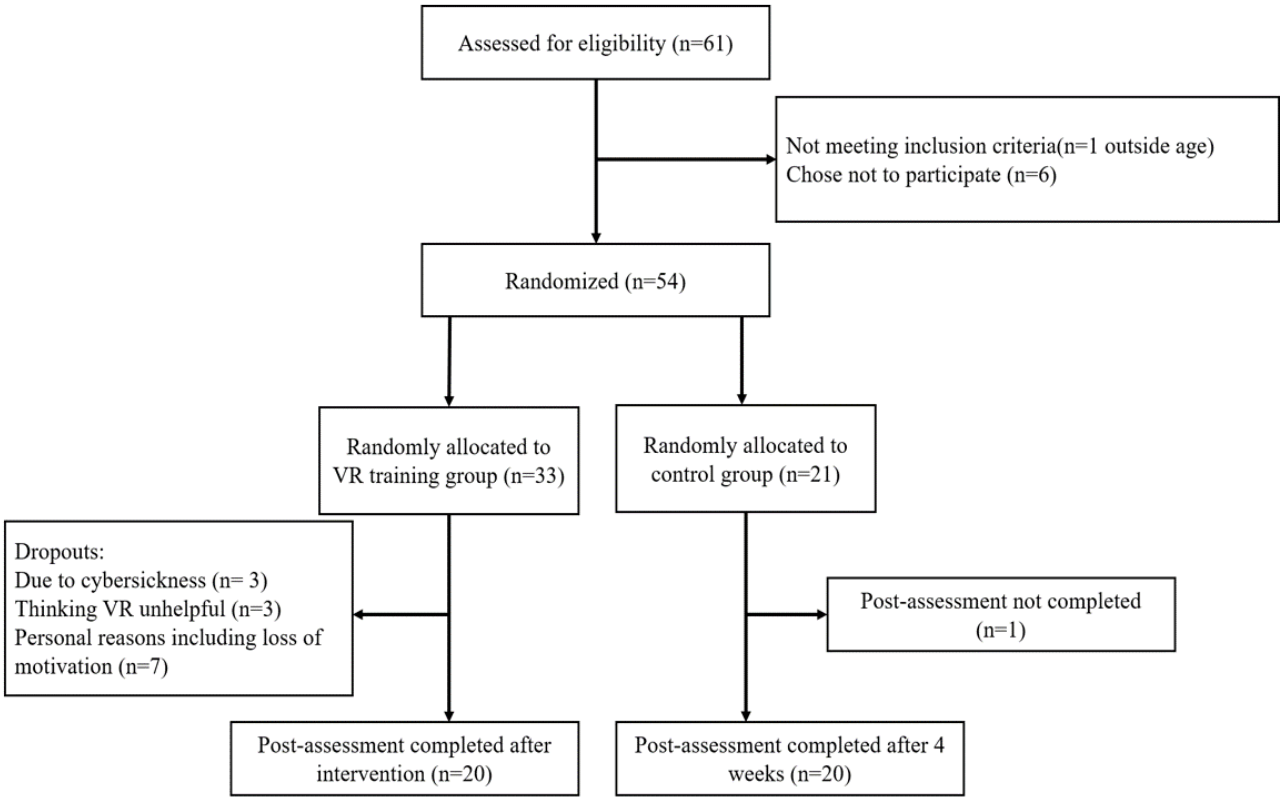
Procedure

Study Design

After participants were informed of the purpose of the study and consented to participate, for baseline assessment, a psychiatric interview by a psychiatrist or clinical psychologist was conducted and a self-rating questionnaire completed. A total of 54 participants were randomly assigned to either the VR treatment group or waitlist group after completion of the baseline assessment. Considering the dropout rates in previous VR studies were relatively high, we assigned patients to VR groups and waitlist groups in a ratio of 3:2 [40] in order to minimize the imbalance of group size. Randomization was carried out using R (version 4.0.2, R Foundation for Statistical Computing) with block sizes of 5. The allocation sequence was stratified for gender. Patients were naturally aware of the allocation. But the trained psychologist assessing clinician-administered scales was not able to know the patient's group information from the beginning to the end of the study.

Participants assigned to the VR group were asked to complete the 4-week VR-based CBT for panic disorder following the guideline. Before the start of treatment, they received the VR devices with a use description. After the treatment, participants were asked to complete questionnaires on VR presence and stimulator sickness. After 4 weeks, participants were given the same measures as pretreatment except the diagnostic interview. Participants in the waitlist group went through the 2 assessment sessions within 4 weeks without any other treatment except the medication between assessments. Figure 1 presents a flowchart of this study.

Figure 1. Flowchart of study process.



VR Treatment

The VR treatment aimed at learning how to cope with panic symptoms and exposed participants to feared situations in a gradual and planned manner. The mobile app was designed to be used without a supporting therapist, although it can be used alongside a therapist or in a clinical setting if needed. The treatment plan for 4 weeks was designed to be 3 times per week, for a total of 12 sessions. Each session was 15 to 30 minutes depending on the module. All participants completed at least

12 sessions, and we allowed them to do more sessions or repeat modules if they wanted within the 4 weeks.

The VR system was developed on a mobile-based platform. The content was designed on Unity 2018.3.11f1 software (Unity Technologies). The avatars and structures comprising the virtual environment were built using a 3Ds Max 2014 (Autodesk). The program was installed on a Galaxy 8+ (Samsung Electronics) smartphone for use with a Gear VR (Samsung Electronics). All in-app data were collected on the smartphone with basic demographic information (eg, birth date and name). Video recordings of actors were taken with an Insta360 Pro camera (Insta360) to create a video-based virtual environment. Because most VR contents were based on video recordings, interactions with environments were limited in VR. However, users chose difficulty levels and training sessions they wanted among various contents and exposure stimuli in VR, indicating that the VR was a partially user-driven program.

The VR consisted of 4 steps in which patients were gradually exposed to phobic stimuli while learning to cope with panic symptoms in each stage (Figure 2). The first was the psychoeducation step in which the patients learn how to breathe, relax, and be exposed to interoceptive stimuli and the reasons why these exercises reduce panic symptoms; they practiced what they learned in the practice step (Figure 3). Next were the two types of exposure therapy using VR in which participants experienced an immersive virtual environment comprising digital human avatars or video recordings of real people. In the exposure with guidance step, patients were exposed to various situations (eg, driving a car, taking an elevator) while using the skills they learned in previous steps under the virtual therapist's guidance (the actor's motion and the therapist's voice; Figure 4). For example, while a participant was exposed to the driving a car scenario, a virtual therapist encouraged the patient to use abdominal breathing with a model's demonstration and direct explanation. Each location had 2 or 3 levels of exposure manipulated by crowded density (eg, taking an elevator) or the time required to complete (eg, driving a car, getting on a plane or subway). After that, without a guide, for more than 5 minutes a patient was exposed to the situations that patients with agoraphobia are afraid of (eg, getting on a plane, driving on bridge; Figure 5). In the last step, 4 scenarios were divided into 2 levels, beginner and advanced, according to the difficulty such as length of time. After each VR session, participants completed questionnaires assessing their feeling of presence and their cybersickness symptoms.

Figure 2. The content design of a virtual reality treatment application with 4 cognitive behavioral therapy components: psychoeducation, practice, virtual reality exposure with guidance, and virtual reality exposure without guidance.

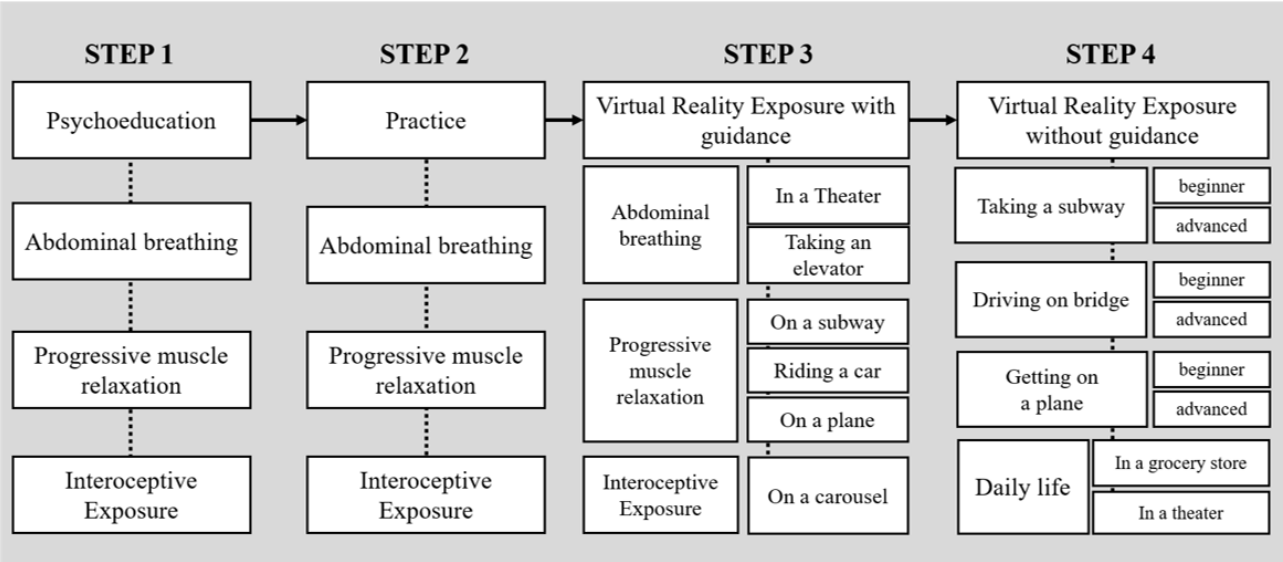


Figure 3. Virtual reality in the psychoeducation step: introduction to progressive muscle relaxation (left) and demonstration of interoceptive exposure training (right).

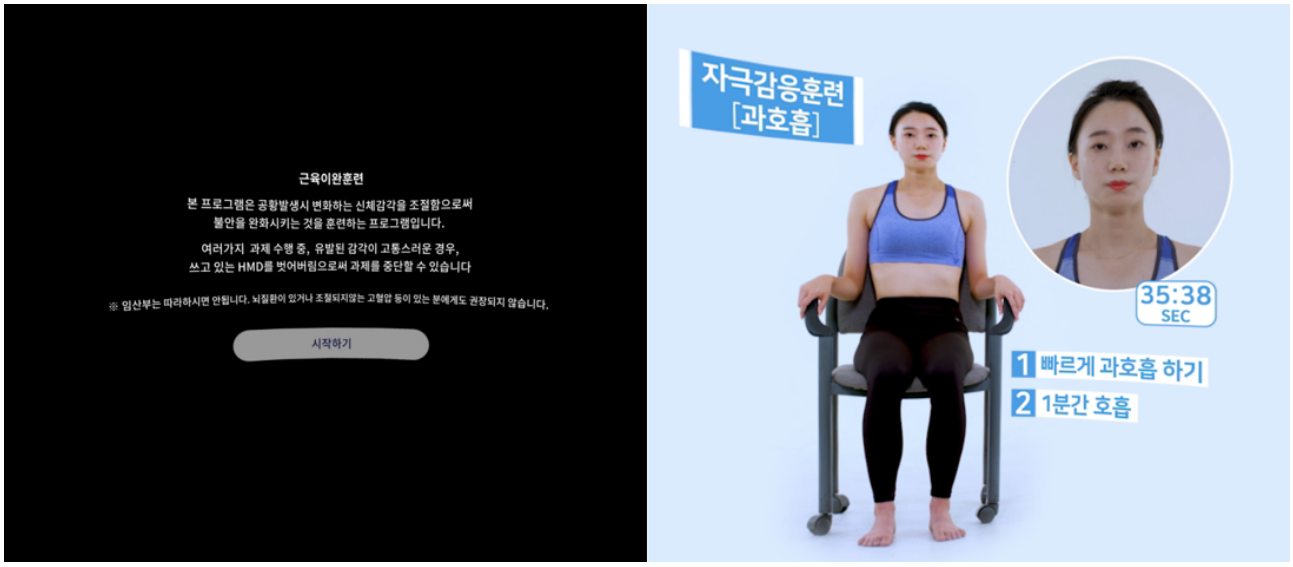


Figure 4. Virtual reality in the exposure with guidance step: driving in a tunnel (left) and taking a seat on a plane with the guidance of muscle relaxation (right).

Figure 5. Virtual reality in the virtual reality exposure without guidance step: taking a subway with animated avatars (left) and driving on a bridge without guidance (right).

Clinical Assessment

Clinical measures administered by trained psychologists at baseline and 4 weeks included the Panic Disorder Severity Scale (PDSS) [41], which consists of 7 items, each rated on a 5-point scale, assessing panic frequency, distress during panic, panic-focused anticipatory anxiety, phobic avoidance of situations, phobic avoidance of physical sensations, impairment in work functioning, and impairment in social functioning. The Hamilton Rating Scale for Depression (HRSD) [42], a depression assessment scale containing 17 items pertaining to symptoms of depression experienced over the past week, was also administered. Individual IQ scores were estimated using the short form of the Wechsler Adult Intelligence Scale–Revised [43].

Several measurements were conducted to assess the psychological state of participants. The Body Sensations Questionnaire (BSQ) [40] is a 17-item 5-point Likert scale that evaluates how afraid and anxious a patient is when they feel the physical sensation of each item in an unstable state. The Albany Panic and Phobia Questionnaire (APPQ) [44] is a 27-item 9-point Likert scale that was designed to measure the distinct dimension of fear in various situations with 3 subscales: social phobia, agoraphobia, and interoceptive fear. The Anxiety Sensitivity Index (ASI) [40] is a 16-item measure tapping the fear of anxiety sensations, known to be a risk factor for the development of panic. The State-Trait Anxiety Inventory (STAI) [45] has 20 items for assessing trait anxiety and 20 for state anxiety. The Hospital Anxiety and Depression Scale (HADS) [46] has 14 items with anxiety and depression subscales. The Korean Inventory of Social Avoidance and Distress Scale (K-SADS) [47] is a 28-item self-rated instrument used to measure various aspects of social anxiety including distress, discomfort, fear, anxiety, and avoidance of social situations. The Korean Inventory for Depressive Symptomatology (KIDS-SR) [48] is a self-report questionnaire that comprises symptoms of depression including melancholic, atypical, and anxious symptoms. The Perceived Stress Scale (PSS) [49] is a measure of the degree to which situations in one's life are appraised as stressful.

After exposure to VR treatment, the 16-item Simulator Sickness Questionnaire (SSQ) [50] was used to assess participants' subjective discomfort (disorientation, oculomotor symptoms, and nausea) to measure simulator sickness due to discrepancies between vision and motion after VR use.

Physiological Recording

Heart rate variability measurements of both groups were taken at baseline and after 4 weeks. Heart rate variability was assessed for 5 minutes in the frequency domain, recommended when examining autonomic nervous system activities [51], by using a SA-3000P arterial testing device (Medicore Co Ltd). During the measurement of heart rate variability, the participant stayed in the seated position without any movement. Electrodes on both wrists and the left ankle were used in this measurement procedure. The frequency domain measures were calculated as absolute and normalized powers of the power spectrum density in the high frequency (HF: 0.15 to 0.40 Hz) and low frequency (LF: 0.04 to 0.15 Hz) bands and LF/HF ratio. The HF bands are a marker of the parasympathetic tone, and LF bands correlate to sympathetic tone or to autonomic balance. LF/HF ratio is an index of the interaction between sympathetic and vagal activity [51].

Statistical Analysis

To compare the difference between groups existing at baseline in terms of demographic and clinical variables, independent *t* tests for continuous variables and chi-square analyses for categorical variables were conducted. To assess the effects of the intervention on the clinical scales, we used analysis of covariance (ANCOVA) with randomization group as the independent variable and postscores for each variable as the dependent variables, controlling for baseline values of each outcome. Analyses on the mean differences between the two groups and confidence intervals were conducted to compare the two groups for each outcome. Effect sizes (Cohen *d*) divided by baseline pooled standard deviation were calculated for

within- and between-group changes, with 0.2, 0.5, and 0.8 corresponding to small, medium, and large effect sizes [51]. We conducted completer analysis and intention-to-treat (ITT) analysis using multiple imputation with multivariate imputation by chained equations [52]. To examine the association between the use of VR and changes in the PDSS score, we conducted correlation analysis using the Pearson method. All statistical analyses were completed with R studio (R version 4.0.2, R Foundation for Statistical Computing). In all cases, a 2-tailed $P<.05$ was considered statistically significant. In addition, linear regression analysis was conducted to observe changes in SSQ scores over time, and Pearson correlation analysis was conducted to see the relationship between changes in SSQ and changes in clinical symptoms, including anxiety.

Results

Demographical and Clinical Characteristics

A total of 61 patients were screened for eligibility. Of these patients, 7 were excluded (see Figure 1) and 14 patients dropped out during the study (13 patients in the VR group and 1 patient in the waitlist group) because of various reasons; most of them did not complete the treatment according to the guideline. Three participants thought VR would not be helpful at the beginning of use, another 3 felt motion sickness while using VR, and the remaining 7 reported discontinuing due to personal reasons including loss of motivation. One participant in the waitlist group was absent at the follow-up assessment.

At baseline, the 2 groups did not differ significantly regarding demographic characteristics or initial clinical assessment scores (Table 1). Demographic information of completers was described in Table 2, and there were no significant differences except for the PDSS scores. The average PDSS score in 13 dropout patients in the VR group was 12.46; the average in patients who did not drop out was 14.8 in the VR group and 11.95 in the waitlist group. The average of use time acquired from in-app data was 245 minutes (SD 61.44, median 250, range 105-457). Patients completed the SSQ every session, and the average SSQ from all patients and sessions was 12.26 (SD 13.26, median 6, range 0-48). There was no relationship between the SSQ scores and any clinical scales at both baseline and final assessment. Figure 6 displays change in SSQ over time. The number of sessions performed was significantly associated with lower SSQ scores (β=–.17, t_{19}=–2.24, P=.03); however, there was no correlation between changes in SSQ and changes in any clinical scale we measured.

Table 1. Demographic and clinical characteristics at baseline by group.

Characteristics	VR[a] training group (n=33)	Waitlist group (n=21)	t/χ^2	*P* value
Gender, n (%)				
Male	13 (39)	7 (33)	0.02	.87
Female	21 (60)	14 (67)	0.05	.82
Education, n (%)				
<High	10 (30)	12 (57)	2.84	.09
>College	23 (70)	9 (43)	—[b]	—
Occupation, n (%)				
Employed	18 (54)	13 (60)	0.23	.87
Unemployed	15 (45)	8 (40)	—	—
Age (years), mean (SD)	35.84 (10.37)	37.14 (13.54)	0.30	.77
Duration of illness (months), mean (SD)	67.74 (65.70)	56.94 (66.34)	–0.25	.80
Intelligence score, mean (SD)	95.92 (17.40)	92.44 (11.01)	0.77	.44
Psychotropic medication use				
Antidepressants, n (%)	30 (91)	19 (91)	0	>.99
Anxiolytics, n (%)	23 (70)	16 (76)	0.04	.83
HRSD[c], mean (SD)	15.06 (8.85)	14.29 (5.87)	0.39	.70
PDSS[d], mean (SD)	13.88 (4.29)	12.29 (3.35)	1.52	.13
STAI[e]_TOTAL, mean (SD)	55.36 (12.54)	53.43 (10.29)	0.01	.98
STAI_S[f]	53.58 (13.03)	55.43 (8.24)	0.62	.54
STAI_T[g]	108.94 (24.75)	108.86 (17.32)	–0.64	.52
KIDS-SR[h], mean (SD)	15.55 (8.8)	17.43 (7.86)	–0.82	.42
PSS[i], mean (SD)	19.97 (4.51)	19.05 (4.01)	0.78	.43
K-SADS[j], mean (SD)	82.36 (21.24)	91.38 (19.23)	–1.61	.11
ASI[k], mean (SD)	69.91 (32.88)	72.1 (28.55)	–0.26	.80
HADS[l], mean (SD)	21.24 (9.12)	21.24 (7.08)	—	—
ANX[m]	11.27 (4.62)	11.38 (3.71)	–0.09	.92
DEP[n]	9.97 (4.83)	9.86 (3.64)	0.10	.92
APPQ[o], mean (SD)	85.15 (43.86)	86.86 (48.82)	–0.13	.90
AGORA[p]	31.42 (14.78)	29.29 (20.25)	0.41	.67
SOCIAL[q]	27.52 (19.14)	36.19 (19.21)	–1.62	.11
INTERO[r]	26.21 (14.71)	21.38 (15.37)	1.14	.26
BSQ[s], mean (SD)	58.03 (13.46)	59.71 (13.24)	–0.45	.65

[a]VR: virtual reality.

[b]Not applicable.

[c]HRSD: Hamilton Rating Scale for Depression.

[d]PDSS: Panic Disorder Severity Scale.

[e]STAI: State and Trait Anxiety questionnaire.

[f]STAI_S: state anxiety.

[g]STAI_T: trait anxiety.

[h]KIDS-SR: Korean Inventory of Depressive Symptomatology.

[i]PSS: Perceived Stress Scale.
[j]K-SADS: Korean Inventory of Social Avoidance and Distress Scale.
[k]ASI: Anxiety Sensitivity Index.
[l]HADS: Hospital Anxiety and Depression Scale.
[m]ANX: Anxiety subscale of HADS.
[n]DEP: depression subscale of HADS.
[o]APPQ: Albany Panic and Phobia Questionnaire.
[p]AGORA: agoraphobia subscale of APPQ.
[q]SOCIAL: social anxiety subscale of APPQ.
[r]NTERO: interoceptive fear subscale of APPQ.
[s]BSQ: Body Sensations Questionnaire.

Table 2. Demographic and clinical characteristics of completers by group.

Characteristics	VR[a] training group (n=20)	Waitlist group (n=20)	t/χ^2	*P* value
Gender, n (%)	—	—	0.06	.80
Male	8 (40)	7 (35)	—	—
Female	12 (60)	13 (65)	—	—
Education, n (%)	—	—	1.63	.20
<High	6 (30)	11 (55)	—	—
>College	14 (70)	9 (45)	—	—
Occupation, n (%)	—	—	0	>.99
Employed	12 (60)	12 (60)	—	—
Unemployed	8 (40)	8 (40)	—	—
Age (years), mean (SD)	35.84 (10.37)	37.14 (13.54)	0.15	.87
Duration of illness (months), mean (SD)	64.25 (54.60)	44.10 (58.16)	1.12	.26
Intelligence score, mean (SD)	95.90 (12.47)	92.44 (11.01)	0.92	.35
Psychotropic medication use				
Antidepressants, n (%)	19 (95)	18 (90)	0	>.99
Anxiolytics, n (%)	14 (70)	15 (75)	0	>.99
HRSD[b], mean (SD)	14.15 (8.72)	14.20 (6.01)	–0.02	.98
PDSS[c], mean (SD)	14.80 (4.18)	11.95 (3.05)	2.46	.01
STAI[d]_TOTAL, mean (SD)	108.30 (22.68)	109.35 (17.62)	0.20	.84
STAI_S[e]	54.35 (11.31)	53.65 (10.51)	–0.53	.59
STAI_T[f]	53.95 (12.02)	55.70 (8.35)	–0.16	.87
KIDS-SR[g], mean (SD)	15.60 (8.47)	17.60 (8.02)	–7.66	.44
PSS[h], mean (SD)	19.85 (4.52)	19.05 (4.11)	0.58	.56
K-SADS[i], mean (SD)	84.45 (21.00)	92.80 (18.57)	–1.33	.19
ASI[j], mean (SD)	73.85 (34.60)	74.65 (26.71)	–0.08	.93
HADS[k], mean (SD)				
ANX[l]	12.15 (4.25)	12.15 (4.25)	0.51	.60
DEP[m]	11.05 (4.21)	11.05 (4.21)	1.03	.30
APPQ[n], mean (SD)	87.05 (45.15)	90.75 (46.63)	–0.25	.80
AGORA[o]	33.10 (14.27)	30.45 (20.04)	0.48	.63
SOCIAL[p]	27.65 (20.85)	37.85 (18.09)	–1.65	.10
INTERO[q]	26.30 (15.21)	22.45 (14.95)	0.80	.42
BSQ[r], mean (SD)	60.55 (12.22)	61.40 (11.03)	–0.23	.81

[a]VR: virtual reality.

[b]HRSD: Hamilton Rating Scale for Depression.

[c]PDSS: Panic Disorder Severity Scale.

[d]STAI: State and Trait Anxiety questionnaire.

[e]STAI_S: state anxiety.

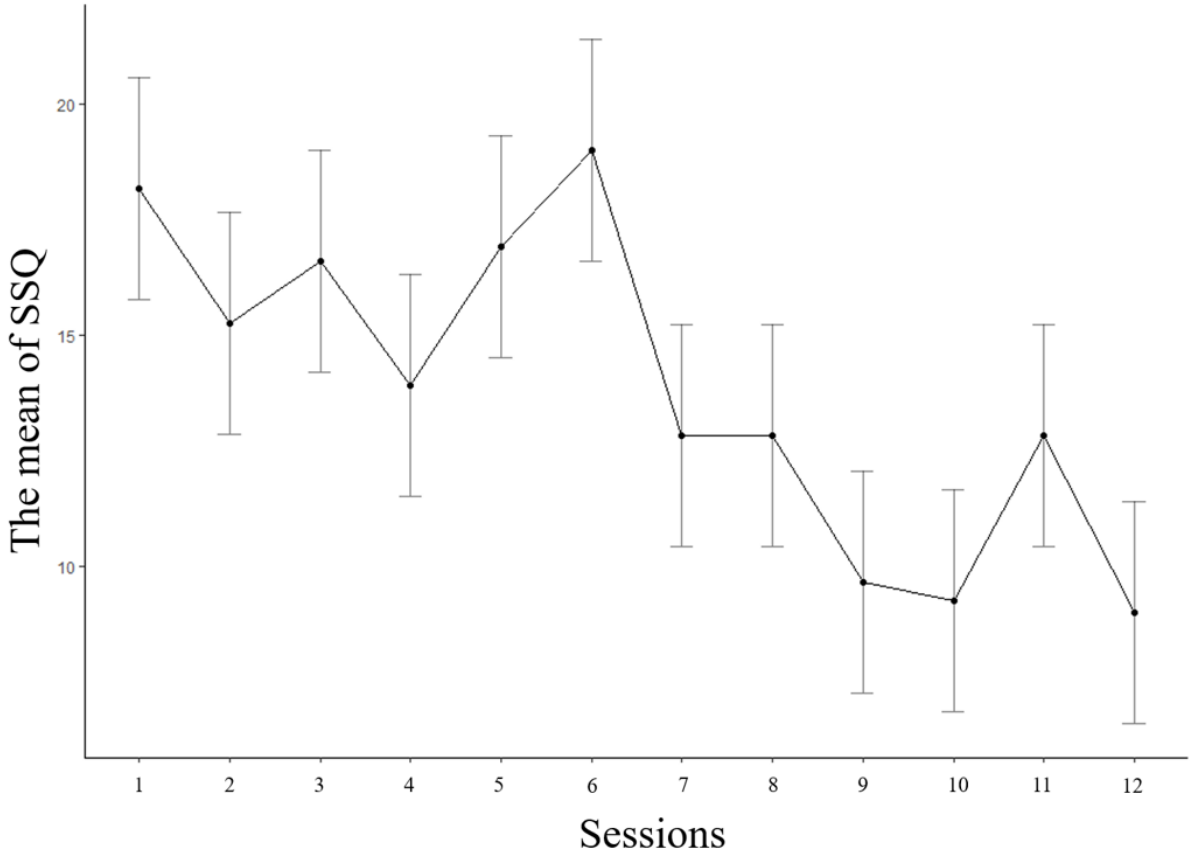
[f]STAI_T: trait anxiety.

[g]KIDS-SR: Korean Inventory of Depressive Symptomatology.

[h]PSS: Perceived Stress Scale.

[i]K-SADS: Korean Inventory of Social Avoidance and Distress Scale.
[j]ASI: Anxiety Sensitivity Index.
[k]HADS: Hospital Anxiety and Depression Scale.
[l]ANX: Anxiety subscale of HADS.
[m]DEP: depression subscale of HADS.
[n]APPQ: Albany Panic and Phobia Questionnaire.
[o]AGORA: agoraphobia subscale of APPQ.
[p]SOCIAL: social anxiety subscale of APPQ.
[q]NTERO: interoceptive fear subscale of APPQ.
[r]BSQ: Body Sensations Questionnaire.

Figure 6. Mean Simulator Sickness Questionnaire scores for each session over time (bars shown: standard error of the mean, n=20).



ITT Analysis

Clinical Assessments

We repeated all tests using multiple imputation. Univariate ANCOVAs on the HRSD, PDSS, state anxiety (STAI_S), and PSS posttreatment scores, controlling for pretreatment scores (Multimedia Appendix 1), showed that the VR treatment group had significantly lower posttreatment scores than the control group for the HRSD ($F_{1,49}$=5.96, *P*=.02,), PDSS ($F_{1,50}$=9.20, *P*=.003,), STAI_S ($F_{1,50}$=4.45 *P*=.04,), and PSS ($F_{1,49}$=4.56, *P*=.03,). Within-group effect sizes for the outcome measures are indicated in Multimedia Appendix 1. Large within-group effect sizes were found in the VR treatment group for the PDSS (*d*=1.05) and STAI_S (*d*=0.91). Moderate within-group effect sizes were found for the HRSD (*d*=0.68), STAI (*d*=0.65), anxiety subscale of HADS (*d*=0.59), and BSQ (*d*=0.57). Between-group differences of change were found in the PDSS (*P*<.01), STAI_S (*P*=.04), and PSS (*P*=.04).

Heart Rate Variability Frequency Domain Analysis

No significant difference was found at baseline in the mean values of total absolute power, normalized power, or LF/HF ratio between the two groups. As shown in Multimedia Appendix 2, only for the VR group, the mean normalized HF (nHF) power and LF/HF ratio were significantly increased after the VR treatment with moderate within-group effect sizes (LF/HF, *d*=0.53; nHF, *d*=0.52). But in the waitlist group, significant change and between-group difference of change were not found.

Completer Analysis

Clinical Assessments

Univariate ANCOVAs on the PDSS and HADS scores, controlling for pretreatment scores (Table 2), showed that the VR treatment group had significantly lower posttreatment scores than the control group for PDSS ($F_{1,36}$=6.16, *P*=.02,) and HADS ($F_{1,36}$=4.66, *P*=.04,). Within-group effect sizes for the outcome measures are shown in Multimedia Appendix 3. Large within-group effect sizes were found in the VR treatment group for the PDSS (*d*=1.05) and STAI (*d*=0.98), and moderate within-group effect sizes were found for the HADS (*d*=0.69) and depression subscale of HADS (*d*=0.68). Between-group differences of change were found in the PDSS (*P*=.02) and HADS (*P*=.04).

Heart Rate Variability Frequency Domain Analysis

No significant difference was found at baseline in the mean values of total absolute power, normalized power, or the LF/HF ratio between the two groups. As shown in Multimedia Appendix 4, only for the VR group, the mean nHF power level was significantly increased after the treatment with moderate within-group effect sizes (nHF, *d*=0.55). In the waitlist group, significant change was not found. Between-group difference of change was found in the absolute HF power (*P*=.04).

Discussion

Principal Findings

This study aimed to investigate the effectiveness of mobile app-based VR for panic disorder in a randomized controlled design. To our knowledge, this is the first study to demonstrate the effectiveness of self-help VR to treat patients with panic disorder, while including not only the exposure technique but most other components of CBT such as psychoeducation, abdominal breathing, and progressive muscle relaxation. Our study demonstrated that self-help VR was effective for reducing panic disorder symptoms as assessed by the PDSS, compared with the waitlist group. These findings are inconsistent with previous meta-analyses of therapist-led VR for anxiety [53,54].

Although a group difference in clinical symptoms was not found at baseline, the comparison between completers showed a difference in PDSS score. The difference may be attributed to the dropout patients. This suggests that only those with severe symptoms remained to the end, and as the result, the average initial PDSS score in the VR group was increased. The severity of panic disorder symptoms might have been a motivation for participants to continue and to complete intervention, while the dropout patients with relatively mild symptoms discontinued use. In other words, our system might have been particularly useful for patients with severe symptoms. Considering the most common reason for dropout in VR exposure treatment is the failure to immerse in the VR environment [55], more severe panic disorder symptoms were probably able to elicit more anxiety in the VR environment, which might have increased the effectiveness of the treatment. In addition, given that fear of exposure is another common reason patients withdraw from exposure treatment [55], gradual and systematic exposure with coping techniques in our VR-based CBT program may have encouraged patients with more severe symptoms to continue the treatment. Our dropout pattern related to the severity and motivation can cause self-selection bias when analyzing only the completers, so we also examined ITT results.

We also observed a reduction in the STAI, suggesting that the VR group had a decrease in overall anxiety after 4 weeks of the VR program. Considering the significant difference between the two groups on HADS total score and adjusted mean changes of the VR group, VR treatment seemed to improve psychological distress. The full-scale HADS score has been used as a global measure of distress in several studies [56,57]. On the other hand, there was no significant change in other anxiety and depression scales including the HRSD and KIDS-SR among completers. Given that the program was designed to treat panic and agoraphobia symptoms, the effect may be limited to only such symptoms.

Contrary to our expectation, VR did not exhibit any significant influence on the APPQ and BSQ, which are thought to be highly related to panic disorder symptoms. The PDSS scale is a global assessment including panic attack frequency, distress during panic attacks, and work and social impairment, while the APPQ evaluates the fear of interoceptive sensation and agoraphobia. The 4-week study period might have been insufficient to reduce the specific fear itself. Otherwise, considering ITT analysis, it can be due to the reduced statistical power [58]. The ITT analysis using multiple imputation generally further confirmed that VR treatment groups experienced significant improvements in panic symptoms, anxiety, and depression including significant changes in the HRSD and BSQ.

Our positive results on psychological assessments including decreases in PDSS score were in line with our findings on heart rate variability data. The within-group results showed an increase in HF power and a decrease in LF/HF in only the VR group. Associations between panic disorder and low HF power, high LF power, and elevated LF/HF have been observed in numerous studies [59-63]. In particular, HF power is considered to be related to the parasympathetic activity [64]. Our results suggest there may be an improvement of the balance in the autonomic nervous system of the VR group participants; in other words, overactivation of sympathetic nervous system seems to be normalized after the treatment.

Previous studies showed that physiological symptoms of anxiety and cybersickness can overlap [65,66]. However, in the correlation analysis between the SSQ and clinical scales, we could not find a significant relationship although we observed a decrease in SSQ scores over time. There was a limitation that we did not measure clinical symptoms as frequently as the SSQ. However, at the very least this study revealed that global improvements in clinical symptoms were not due to patients becoming familiar with the VR environments.

Until recently, despite advantages of VR, most therapists have shown little interest in applying VR in their clinical practice [67]. One concern when applying VR to therapy is technical difficulties [68]. In this study, many participants completed the whole treatment process although they had never used VR

devices, indicating the decent usability of VR. If patients are familiar with smartphones, they can easily overcome some inconvenience in mobile-based VR, and no additional training is required. In addition, this study demonstrated that mobile-based VR can be used by patients alone and exhibit positive results. It seems obvious that minimizing therapist effort is related to the cost effectiveness of the treatment [69].

In the context of minimizing therapist intervention, we used a virtual therapist who delivered psychoeducational contents and provided guided exposure to encourage the use of CBT techniques. Our result showed the possibility that patients eligible for VR could benefit from the immersive learning experience as well as exposure to feared stimuli under the guidance of a virtual therapist. Although its efficacy compared to the VR CBT without the presence of a virtual therapist remains to be investigated, a virtual therapist seems to be an essential component for fully self-guided VR treatment.

Additionally, it would be meaningful to try VR protocols including relaxation and breathing techniques and examine the therapeutic effects in future study. Based on recent studies describing the mechanisms of exposure therapy as inhibitory learning, the exposure to feared stimuli without coping skills may maximize the mismatch between expectations and experience and may enhance the learning effect rather than the exposure with coping skills [70]. Although traditional CBT protocol including coping skills have been used in this study, our VR protocol can be flexibly modified, so the recent protocol without coping skills can be applied as well. We could then conclude whether it shows better therapeutic effects.

Limitations

There are several limitations of this study. First, our dropout rate was higher than a previous VR study [55] and similar to self-help treatment approaches [71-73]. As we noted, during the study periods, VR was fully executed by patients, so there was no other way to increase the treatment adherence of patients except for patient's motivation. This lack of encouragement could be a limitation of a fully self-help VR approach, and in this study, it caused the initial difference in the clinical variable with high dropout rate. However, we tried to overcome this limitation by using ANCOVA and by also presenting the ITT analysis results. Second, due to the relatively small sample size, it is difficult to generalize the results even though we tried to provide the more objective physiological evidence as well [74]. It will be necessary to replicate our findings in larger samples. Third, given that we did not perform a follow-up assessment, long-term treatment effects remain unknown. Fourth, heart rate variability was measured only at rest. We would have been more informed if we had also measured heart rate variability in anxiety-provoking states. Finally, it is still not clear whether self-help VR is better than other treatment options since patients without any intervention were set as the waitlist group in this study.

Conclusions

In summary, our findings support the hypothesis that self-guided, mobile app-based VR can reduce panic symptoms and help restore the autonomic nervous system. This study also suggested the decent feasibility of self-help VR for panic disorder, demonstrating the validity of the use of this new technique in real-world treatment. Future studies with larger sample size, longer duration of follow-up, and comparison with other treatment options will be able to verify and expand our results.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021M3E5D9025019). This work was also supported by the Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety) (Project Number: KMDF_PR_20200901_0186, 9991006856).

Authors' Contributions

JO and JJ designed and conceptualized the study. BS and JO conducted the literature search. BS, JO, HK, and SK collected the data. BS, BK, and HEK analyzed the data. BS, JO, BK, and JJ interpreted the data. BS, JO, and JJ drafted and revised the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Changes in clinical variables at baseline and 4 weeks (intention-to-treat analysis).
[DOCX File , 20 KB - mental_v8i11e30590_app1.docx]

Multimedia Appendix 2
Estimated mean and standard error of heart rate variability items (intention-to-treat analysis).
[DOCX File , 16 KB - mental_v8i11e30590_app2.docx]

Multimedia Appendix 3
Changes of completers in clinical variables at baseline and 4 weeks (completer analysis).
[DOCX File , 20 KB - mental_v8i11e30590_app3.docx]

Multimedia Appendix 4
Estimated mean and standard error of heart rate variability items (completer analysis).
[DOCX File , 18 KB - mental_v8i11e30590_app4.docx]

Multimedia Appendix 5
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1246 KB - mental_v8i11e30590_app5.pdf]

References

1. Rothbaum BO, Hodges L, Watson BA, Kessler CD, Opdyke D. Virtual reality exposure therapy in the treatment of fear of flying: a case report. Behav Res Ther 1996;34(5-6):477-481. [doi: 10.1016/0005-7967(96)00007-1] [Medline: 8687369]
2. Barlow DH. Anxiety and Its Disorders: The Nature and Treatment of Anxiety and Panic, 2nd Edition. New York: Guilford Press; 2002.
3. Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Washington: American Psychiatric Association; 2013.
4. Alonso J, Angermeyer MC, Bernert S, Bruffaerts R, Brugha TS, Bryson H, et al. Prevalence of mental disorders in Europe: results from the European Study of the Epidemiology of Mental Disorders (ESEMeD) project. Acta Psychiatr Scand Suppl 2004(420):21-27. [doi: 10.1111/j.1600-0047.2004.00327.x] [Medline: 15128384]
5. Candilis PJ, McLean RY, Otto MW, Manfro GG, Worthington JJ, Penava SJ, et al. Quality of life in patients with panic disorder. J Nerv Ment Dis 1999 Jul;187(7):429-434. [doi: 10.1097/00005053-199907000-00006] [Medline: 10426463]
6. Goodwin RD, Faravelli C, Rosi S, Cosci F, Truglia E, Wittchen HU. The epidemiology of panic disorder and agoraphobia in Europe. Eur Neuropsychopharmacol 2005 Aug;15(4):435-443. [doi: 10.1016/j.euroneuro.2005.04.006] [Medline: 15925492]
7. Schmidt NB, Telch MJ. Nonpsychiatric medical comorbidity, health perceptions, and treatment outcome in patients with panic disorder. Health Psychol 1997;16(2):114-122. [doi: 10.1037/0278-6133.16.2.114]
8. Barlow DH, Gorman JM, Shear MK, Woods SW. Cognitive-behavioral therapy, imipramine, or their combination for panic disorder: a randomized controlled trial. JAMA 2000 May 17;283(19):2529-2536. [doi: 10.1001/jama.283.19.2529] [Medline: 10815116]
9. Craske M, Barlow D. Panic disorder and agoraphobia. In: Clinical Handbook of Psychological Disorders: A Step-By-Step Treatment Manual, 4th Edition. New York: Guilford Press; 2008:1-64.
10. Pompoli A, Furukawa TA, Imai H, Tajika A, Efthimiou O, Salanti G. Psychological therapies for panic disorder with or without agoraphobia in adults: a network meta-analysis. Cochrane Database Syst Rev 2016 Apr 13;4:CD011004 [FREE Full text] [doi: 10.1002/14651858.CD011004.pub2] [Medline: 27071857]
11. Clark D, Wells A. Cognitive therapy for anxiety disorder. In: Dickstein L, Riba MB, Oldham J, editors. Review of Psychiatry, 1st Edition. Washington: American Psychiatric Press; 1997.
12. Barlow DH. Anxiety and Its Disorders: The Nature and Treatment of Anxiety and Panic. New York: Guilford Press; 1988.
13. Meuret AE, Wolitzky-Taylor KB, Twohig MP, Craske MG. Coping skills and exposure therapy in panic disorder and agoraphobia: latest advances and future directions. Behav Ther 2012 Jun;43(2):271-284 [FREE Full text] [doi: 10.1016/j.beth.2011.08.002] [Medline: 22440065]
14. Kaczkurkin AN, Foa EB. Cognitive-behavioral therapy for anxiety disorders: an update on the empirical evidence. Dialogues Clin Neurosci 2015 Sep;17(3):337-346 [FREE Full text] [Medline: 26487814]
15. Wiederhold BK, Wiederhold MD. Lessons Learned From 600 Virtual Reality Sessions. CyberPsychol Behav 2000 Jun;3(3):393-400. [doi: 10.1089/10949310050078841]
16. Becker CB, Zayfert C, Anderson E. A survey of psychologists' attitudes towards and utilization of exposure therapy for PTSD. Behav Res Ther 2004 Mar;42(3):277-292. [doi: 10.1016/S0005-7967(03)00138-4] [Medline: 14975770]
17. Olatunji BO, Deacon BJ, Abramowitz JS. The cruelest cure? Ethical issues in the implementation of exposure-based treatments. Cognitive Behav Pract 2009 May;16(2):172-180. [doi: 10.1016/j.cbpra.2008.07.003]
18. Choi Y, Vincelli F, Riva G, Wiederhold BK, Lee J, Park K. Effects of group experiential cognitive therapy for the treatment of panic disorder with agoraphobia. Cyberpsychol Behav 2005 Aug;8(4):387-393. [doi: 10.1089/cpb.2005.8.387] [Medline: 16092896]
19. Botella C, Villa H, García-Palacios A, Baños RM, Perpiñá C, Alcañiz M. Clinically significant virtual environments for the treatment of panic disorder and agoraphobia. Cyberpsychol Behav 2004 Oct;7(5):527-535. [doi: 10.1089/cpb.2004.7.527] [Medline: 15667047]

20. Lindner P, Miloff A, Fagernäs S, Andersen J, Sigeman M, Andersson G, et al. Therapist-led and self-led one-session virtual reality exposure therapy for public speaking anxiety with consumer hardware and software: a randomized controlled trial. J Anxiety Disord 2019 Jan;61:45-54. [doi: 10.1016/j.janxdis.2018.07.003] [Medline: 30054173]
21. Barlow DH, Craske MG, Cerny JA, Klosko JS. Behavioral treatment of panic disorder. Behav Ther 1989 Mar;20(2):261-282. [doi: 10.1016/s0005-7894(89)80073-5]
22. Telch MJ, Lucas JA, Schmidt NB, Hanna HH, Jaimez T, Lucas RA. Group cognitive-behavioral treatment of panic disorder. Behav Res Ther 1993 Mar;31(3):279-287. [doi: 10.1016/0005-7967(93)90026-q]
23. Jacobson E. Progressive Relaxation, 2nd Edition. Chicago: University of Chicago Press; 1938.
24. Wolpe J. Psychotherapy by Reciprocal Inhibition. Palo Alto: Stanford University Press; 1958.
25. Bernstein D, Borkovec T. Progressive Relaxation Training: A Manual for the Helping Professions. Champaign: Research Press; 1973.
26. Ost LG. Applied relaxation vs progressive relaxation in the treatment of panic disorder. Behav Res Ther 1988;26(1):13-22. [doi: 10.1016/0005-7967(88)90029-0] [Medline: 3277613]
27. Botella C, García-Palacios A, Villa H, Baños RM, Quero S, Alcañiz M, et al. Virtual reality exposure in the treatment of panic disorder and agoraphobia: a controlled study. Clin Psychol Psychother 2007 May;14(3):164-175. [doi: 10.1002/cpp.524]
28. Vincelli F, Anolli L, Bouchard S, Wiederhold BK, Zurloni V, Riva G. Experiential cognitive therapy in the treatment of panic disorders with agoraphobia: a controlled study. Cyberpsychol Behav 2003 Jun;6(3):321-328. [doi: 10.1089/109493103322011632] [Medline: 12855090]
29. Pelissolo A, Zaoui M, Aguayo G, Yao S, Roche S, Ecochard R. Virtual reality exposure therapy versus cognitive behavior therapy for panic disorder with agoraphobia: a randomized comparison study. J CyberTher Rehab 2012:35-43 [FREE Full text] [doi: 10.3389/conf.neuro.14.2009.06.082]
30. Peñate W, Pitti C, Bethencourt J, de la Fuente J, Gracia R. The effects of a treatment based on the use of virtual reality exposure and cognitive-behavioral therapy applied to patients with agoraphobia. Int J Clin Health Psychol 2008:5-22 [FREE Full text] [doi: 10.1016/j.clinph.2008.04.222]
31. Pitti C, Peñate W, de la Fuente J, Bethencourt J, Acosta L, Villaverde M, et al. [Agoraphobia: combined treatment and virtual reality. Preliminary results]. Actas Esp Psiquiatr 2008;36(2):94-101 [FREE Full text] [Medline: 18365789]
32. Meyerbroeker K, Morina N, Kerkhof GA, Emmelkamp PMG. Virtual reality exposure therapy does not provide any additional value in agoraphobic patients: a randomized controlled trial. Psychother Psychosom 2013;82(3):170-176. [doi: 10.1159/000342715] [Medline: 23548832]
33. Peflate W, Pitti C, Bethencourt J, de la Fuente J, Gracia R. The effects of a treatment based on the use of virtual reality exposure and cognitive-behavioral therapy applied to patients with agoraphobia. Int J Clin Health Psychol 2008;8(1):5-22 [FREE Full text]
34. Bun P, Gorski F, Grajewski D, Wichniarek R, Zawadzki P. Low-cost devices used in virtual reality exposure therapy. Procedia Comput Sci 2017;104:445-451. [doi: 10.1016/j.procs.2017.01.158]
35. Cuijpers P, Schuurmans J. Self-help interventions for anxiety disorders: an overview. Curr Psychiatry Rep 2007 Aug;9(4):284-290. [doi: 10.1007/s11920-007-0034-6] [Medline: 17880859]
36. den Boer PCAM, Wiersma D, Van den Bosch RJ. Why is self-help neglected in the treatment of emotional disorders? A meta-analysis. Psychol Med 2004 Aug;34(6):959-971. [doi: 10.1017/s003329170300179x] [Medline: 15554567]
37. Donker T, Cornelisz I, van Klaveren C, van Straten A, Carlbring P, Cuijpers P, et al. Effectiveness of self-guided app-based virtual reality cognitive behavior therapy for acrophobia: a randomized clinical trial. JAMA Psychiatry 2019 Mar 20;76(7):682-690. [doi: 10.1001/jamapsychiatry.2019.0219] [Medline: 30892564]
38. Kim HE, Hong Y, Kim M, Jung YH, Kyeong S, Kim J. Effectiveness of self-training using the mobile-based virtual reality program in patients with social anxiety disorder. Comput Human Behav 2017 Aug;73:614-619. [doi: 10.1016/j.chb.2017.04.017]
39. Hong Y, Kim HE, Jung YH, Kyeong S, Kim J. Usefulness of the mobile virtual reality self-training for overcoming a fear of heights. Cyberpsychol Behav Soc Netw 2017 Dec;20(12):753-761. [doi: 10.1089/cyber.2017.0085] [Medline: 29211504]
40. Chambless DL, Caputo GC, Bright P, Gallagher R. Assessment of fear of fear in agoraphobics: the body sensations questionnaire and the agoraphobic cognitions questionnaire. J Consult Clin Psychol 1984 Dec;52(6):1090-1097. [doi: 10.1037//0022-006x.52.6.1090] [Medline: 6520279]
41. Shear MK, Brown TA, Barlow DH, Money R, Sholomskas DE, Woods SW, et al. Multicenter collaborative panic disorder severity scale. Am J Psychiatry 1997 Nov;154(11):1571-1575. [doi: 10.1176/ajp.154.11.1571] [Medline: 9356566]
42. Hamilton M. The Hamilton Rating Scale for Depression. In: Sartorius N, Ban TA, editors. Assessment of Depression. Berlin: Springer; 1986:143-152.
43. Kaufman A, Lichtenberger E, editors. Assessing Adolescent and Adult Intelligence. Boston: Allyn & Bacon; 1990.
44. Brown TA, White KS, Barlow DH. A psychometric reanalysis of the Albany Panic and Phobia Questionnaire. Behav Res Ther 2005 Mar;43(3):337-355. [doi: 10.1016/j.brat.2004.03.004] [Medline: 15680930]
45. Spielberger C. State-Trait Anxiety Inventory. In: Weiner IB, Craighead WE, editors. Corsini Encyclopedia of Psychology, Volume 4, 4th Edition. Hoboken: John Wiley & Sons; 2010:1698.

46. Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
47. Lee J, Choi C. Study of the reliability and the validity of the Korean versions of social phobia scales (K-SAD and K-BFNE). Korean J Clin Psychol 1997;16:64.
48. Yoon J, Jon D, Hong H, Hong N, Seok J. Reliability and validity of the Korean Version of Inventory for Depressive Symptomatology. Mood Emot 2012:131-151 [FREE Full text]
49. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [doi: 10.2307/2136404]
50. Kennedy RS, Lane NE, Berbaum KS, Lilienthal MG. Simulator Sickness Questionnaire: an enhanced method for quantifying simulator sickness. Int J Aviation Psychol 1993 Jul;3(3):203-220. [doi: 10.1207/s15327108ijap0303_3]
51. Cohen J. A power primer. Psychol Bull 1992 Jul;112(1):155-159. [doi: 10.1037//0033-2909.112.1.155] [Medline: 19565683]
52. Enders CK. Multiple imputation as a flexible tool for missing data handling in clinical research. Behav Res Ther 2017 Nov;98:4-18. [doi: 10.1016/j.brat.2016.11.008] [Medline: 27890222]
53. Parsons TD, Rizzo AA. Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: a meta-analysis. J Behav Ther Exp Psychiatry 2008 Sep;39(3):250-261. [doi: 10.1016/j.jbtep.2007.07.007] [Medline: 17720136]
54. Powers MB, Emmelkamp PMG. Virtual reality exposure therapy for anxiety disorders: a meta-analysis. J Anxiety Disord 2008;22(3):561-569. [doi: 10.1016/j.janxdis.2007.04.006] [Medline: 17544252]
55. Benbow AA, Anderson PL. A meta-analytic examination of attrition in virtual reality exposure therapy for anxiety disorders. J Anxiety Disord 2019 Jan;61:18-26. [doi: 10.1016/j.janxdis.2018.06.006] [Medline: 30646997]
56. Lewis G, Wessely S. Comparison of the General Health Questionnaire and the Hospital Anxiety and Depression Scale. Br J Psychiatry 1990 Dec 02;157(6):860-864. [doi: 10.1192/bjp.157.6.860] [Medline: 2289095]
57. Razavi D, Delvaux N, Farvacques C, Robaye E. Screening for adjustment disorders and major depressive disorders in cancer in-patients. Br J Psychiatry 1990 Jan 02;156(1):79-83. [doi: 10.1192/bjp.156.1.79] [Medline: 2297623]
58. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: Intention-to-treat versus per-protocol analysis. Perspect Clin Res 2016;7(3):144-146 [FREE Full text] [doi: 10.4103/2229-3485.184823] [Medline: 27453832]
59. Yeragani VK, Pohl R, Berger R, Balon R, Ramesh C, Glitz D, et al. Decreased heart rate variability in panic disorder patients: a study of power-spectral analysis of heart rate. Psychiatry Res 1993 Jan;46(1):89-103. [doi: 10.1016/0165-1781(93)90011-5] [Medline: 8464959]
60. Garakani A, Martinez JM, Aaronson CJ, Voustianiouk A, Kaufmann H, Gorman JM. Effect of medication and psychotherapy on heart rate variability in panic disorder. Depress Anxiety 2009;26(3):251-258. [doi: 10.1002/da.20533] [Medline: 18839407]
61. Kang E, Lee I, Park J, Kim K, Yu B. Platelet serotonin transporter function and heart rate variability in patients with panic disorder. J Korean Med Sci 2010 Apr;25(4):613-618 [FREE Full text] [doi: 10.3346/jkms.2010.25.4.613] [Medline: 20358007]
62. Martinez J, Garakani A, Kaufmann H, Aaronson C, Gorman J. Heart rate and blood pressure changes during autonomic nervous system challenge in panic disorder patients. Psychosomatic Med 2010;72(5):442-449. [doi: 10.1097/psy.0b013e3181d972c2]
63. Chang H, Chang C, Tzeng N, Kuo TBJ, Lu R, Huang S. Decreased cardiac vagal control in drug-naive patients with panic disorder: a case-control study in Taiwan. Asia Pac Psychiatry 2013 Jun 27;5(2):80-89. [doi: 10.1111/appy.12032] [Medline: 23857808]
64. Pagani M, Lombardi F, Guzzetti S, Rimoldi O, Furlan R, Pizzinelli P, et al. Power spectral analysis of heart rate and arterial pressure variabilities as a marker of sympatho-vagal interaction in man and conscious dog. Circ Res 1986 Aug;59(2):178-193. [doi: 10.1161/01.res.59.2.178] [Medline: 2874900]
65. Pot-Kolder R, Veling W, Counotte J, van der Gaag M. Anxiety partially mediates cybersickness symptoms in immersive virtual reality environments. Cyberpsychol Behav Soc Netw 2018 Mar;21(3):187-193. [doi: 10.1089/cyber.2017.0082] [Medline: 29356575]
66. Ling Y, Brinkman W, Nefs HT, Qu C, Heynderickx I. Cybersickness and anxiety in virtual environments. VTT Symposium 2011;1(1):80-82 [FREE Full text] [doi: 10.1162/pres_a_00111]
67. Schwartzman D, Segal R, Drapeau M. Perceptions of virtual reality among therapists who do not apply this technology in clinical practice. Psychol Serv 2012 Aug;9(3):310-315. [doi: 10.1037/a0026801] [Medline: 22867123]
68. Segal R, Bhatia M, Drapeau M. Therapists' perception of benefits and costs of using virtual reality treatments. Cyberpsychol Behav Soc Netw 2011;14(1-2):29-34. [doi: 10.1089/cyber.2009.0398] [Medline: 21329440]
69. Bower P, Gilbody S. Stepped care in psychological therapies: access, effectiveness and efficiency. Narrative literature review. Br J Psychiatry 2005 Jan;186:11-17 [FREE Full text] [doi: 10.1192/bjp.186.1.11] [Medline: 15630118]
70. Craske MG, Treanor M, Conway CC, Zbozinek T, Vervliet B. Maximizing exposure therapy: an inhibitory learning approach. Behav Res Ther 2014 Jul;58:10-23 [FREE Full text] [doi: 10.1016/j.brat.2014.04.006] [Medline: 24864005]
71. 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]

72. Haug T, Nordgreen T, Öst L, Havik OE. Self-help treatment of anxiety disorders: a meta-analysis and meta-regression of effects and potential moderators. Clin Psychol Rev 2012 Jul;32(5):425-445. [doi: 10.1016/j.cpr.2012.04.002] [Medline: 22681915]
73. Oromendia P, Orrego J, Bonillo A, Molinuevo B. Internet-based self-help treatment for panic disorder: a randomized controlled trial comparing mandatory versus optional complementary psychological support. Cogn Behav Ther 2016 Jun;45(4):270-286. [doi: 10.1080/16506073.2016.1163615] [Medline: 27007256]
74. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Circulation 1996 Mar 01;93(5):1043-1065. [Medline: 8598068]

Abbreviations

ANCOVA: analysis of covariance
APPQ: Albany Panic and Phobia Questionnaire
ASI: Anxiety Sensitivity Index
BSQ: Body Sensations Questionnaire
CBT: cognitive behavioral therapy
HADS: Hospital Anxiety and Depression Scale
HF: high frequency
HRSD: Hamilton Rating Scale for Depression
ITT: intention-to-treat
KIDS-SR: Korean Inventory of Depressive Symptomatology
K-SADS: Korean Inventory of Social Avoidance and Distress Scale
LF: low frequency
nHF: normalized high frequency
PDSS: Panic Disorder Severity Scale
PSS: Perceived Stress Scale
SSQ: Simulator Sickness Questionnaire
STAI: State-Trait Anxiety Inventory
STAI_S: state anxiety
VR: virtual reality

Edited by J Torous; submitted 21.05.21; peer-reviewed by R Ciorap, R Pot-Kolder, A Tajika; comments to author 10.06.21; revised version received 02.08.21; accepted 25.08.21; published 22.11.21.

Please cite as:
Shin B, Oh J, Kim BH, Kim HE, Kim H, Kim S, Kim JJ
Effectiveness of Self-Guided Virtual Reality–Based Cognitive Behavioral Therapy for Panic Disorder: Randomized Controlled Trial
JMIR Ment Health 2021;8(11):e30590
URL: https://mental.jmir.org/2021/11/e30590
doi:10.2196/30590
PMID:34813486

Original Paper

Prevalence and Temporal Trends Analysis of Screening and Diagnostic Instruments in Posttraumatic Stress Disorder: Text Mining Study

Hui Zong[1], PhD; Binyang Hu[2], MSc; Yang Han[2], PhD; Zuofeng Li[2], PhD; Xiaoyan Zhang[1], PhD

[1]Research Center for Translational Medicine, Shanghai East Hospital, School of Life Sciences and Technology, Tongji University, Shanghai, China
[2]Philips Research China, Shanghai, China

Corresponding Author:
Xiaoyan Zhang, PhD
Research Center for Translational Medicine, Shanghai East Hospital
School of Life Sciences and Technology
Tongji University
1239 Siping Rd
Shanghai, 200092
China
Phone: 86 02165980233
Email: xyzhang@tongji.edu.cn

Abstract

Background: Various instruments for patient screening and diagnosis have been developed for and applied in posttraumatic stress disorder (PTSD).

Objective: This study comprehensively investigates the prevalence and temporal trends of the most widely used instruments in PTSD-related studies.

Methods: A total of 1345 files of registered clinical trials from ClinicalTrials.gov and 9422 abstracts from the PubMed database from 2005 to 2020 were downloaded for this study. The instruments applied in clinical trials were manually annotated, and instruments in abstracts were recognized using exact string matching. The prevalence score of an instrument in a certain period was calculated as the number of studies divided by the number of instances of the instrument. By calculating the yearly prevalence index of each instrument, we conducted a trends analysis and compared the trends in index change between instruments.

Results: A total of 4178 instrument synonyms were annotated, which were mapped to 1423 unique instruments. In the 16 years from 2005 to 2020, only 10 instruments were used more than once per year; the 4 most used instruments were the PTSD Checklist, the Clinician-Administered PTSD Disorder Scale, the Patient Health Questionnaire, and the Beck Depression Inventory. There were 18 instruments whose yearly prevalence index score exceeded 0.1 at least once during the 16 years. The changes in trends and time points of partial instruments in clinical trials and PubMed abstracts were highly consistent. The average time duration of a PTSD-related trial was 1495.5 days or approximately 4 years from submission to ClinicalTrial.gov to publication in a journal.

Conclusions: The application of widely accepted and appropriate instruments can help improve the reliability of research results in PTSD-related clinical studies. With extensive text data obtained from real clinical trials and published articles, we investigated and compared the usage of instruments in the PTSD research community.

(JMIR Ment Health 2021;8(11):e33599) doi:10.2196/33599

KEYWORDS

posttraumatic stress disorder; instruments; prevalence; clinical trials; text mining

Introduction

Posttraumatic stress disorder (PTSD) is a mental health condition triggered by experiencing or witnessing a traumatic event [1,2]. Over 70% of adults worldwide have experienced a traumatic event at least once in their lifetime, with 30.5% have experiencing 4 or more events [3]. The most commonly reported traumatic events for individuals are the unexpected death of a loved one, witnessing death or serious injury, being robbed, and life-threatening automobile accidents [4]. Rapid and accurate assessment facilitates timely diagnosis and early intervention

in PTSD. Assessment tools comprise screening and diagnostic instruments, which vary in their format (self-reporting or structured interviews) depending on the population, target symptoms, or actions for which they are designed. With the advancement of modern medicine, many instruments have been developed and applied in scientific research and clinical trials. However, choosing the appropriate instrument for a PTSD study can be challenging without comprehensive comparison or evaluation.

Several studies have investigated and compared commonly used instruments in PTSD. In a previous study [5], researchers conducted a web-based survey on 277 traumatic stress professionals to assess traumatic event exposure and posttraumatic effects and revealed 7 commonly used instruments, including the Post-traumatic Stress Diagnostic Scale, the Trauma Symptom Inventory, the Life Events Checklist, the Clinician-Administered PTSD Scale (CAPS), the PTSD Checklist (PCL), the Impact of Event Scale-Revised, and the Trauma Symptom Checklist. In another study, researchers described the reliability and validity of common self-report instruments and structured clinical interviews used to assess depression [6] and PTSD after sepsis [7]. Some researchers also compared different versions of the PCL spanning the transition between the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) and DSM-V [8]. The above studies demonstrated the importance of determining the most widely used instruments. However, the number of participants, the institutions they belong to, the number of instruments included, and subjective or memory factors may have introduced bias in the results.

With the exponential growth of biomedical literature, text mining becomes increasingly promising for biomedical research, especially in the fields of public health and biomedical informatics. Extracting potentially useful information using keyword matching or advanced methods and investigating the prevalence trends of specific topics can help to gain better insight into a particular field and discover inconspicuous changes. Analysis of prevalence trends is a widespread practice of collecting information and attempting to spot trends in the information, such as cultural trends [9], cognitive distortion prevalence [10], research topic trends [11], and top popular questionnaires [12]. These studies have shed light on large-scale text data analyses for examining prevalence and trends.

There are growing numbers of published articles and ongoing registered clinical trials involving PTSD, in which several instruments have been applied to assess the symptoms, emotions, feelings, and actions of the participants [6]. Investigating and comparing the prevalence and temporal trends of these instruments can help determine the conventional assessment tools used in this field. In addition, this valuable knowledge can be provided to researchers when developing assessment criteria in clinical studies, which can particularly benefit clinicians and researchers who are new to the field.

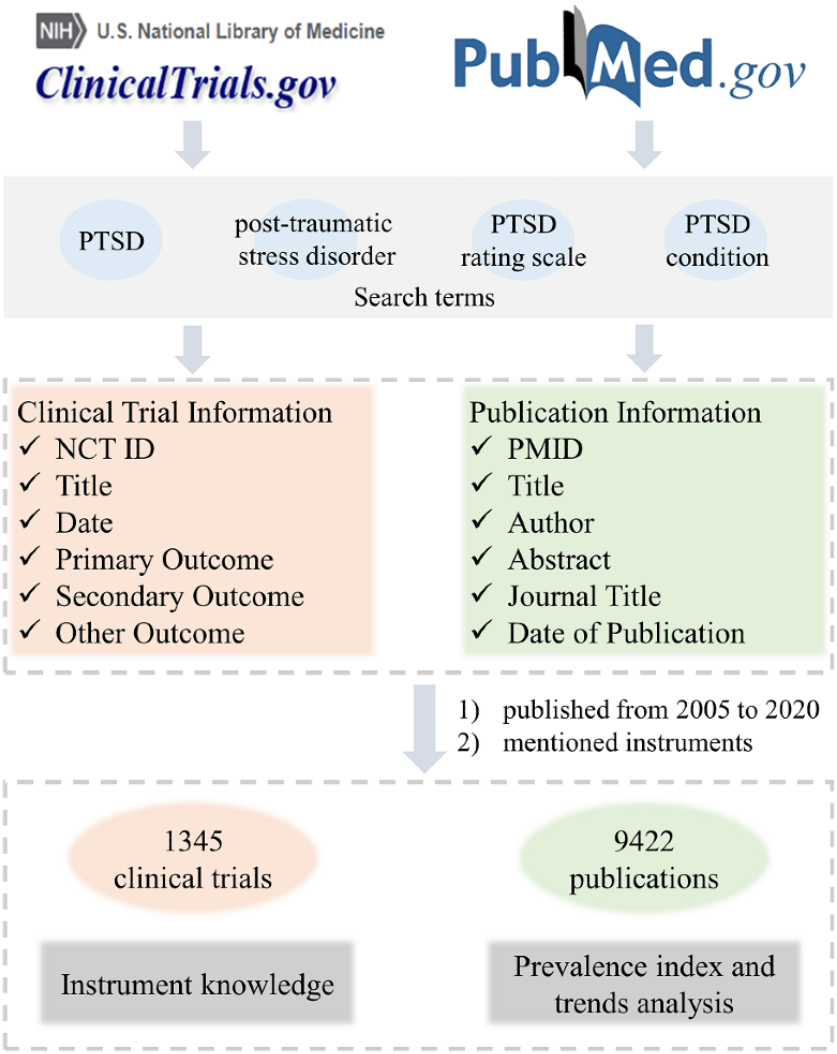
In this study, we conducted a comprehensive investigation on the prevalence trends of the most widely employed instruments using extensive text data from real clinical trials and published articles related to PTSD. It may help reveal the conventional assessment tools used for evaluating PTSD and provide valuable knowledge that might not be otherwise apparent. We believe the prevalence of instruments is an important index in measurement selection, and knowledge in this regard can serve as a reference for designing studies and trials.

Methods

Data and Annotation

A total of 2502 registered clinical trials were accessed and downloaded from ClinicalTrials.gov by searching with keywords “PTSD,” “PTSD rating scale,” “PTSD condition,” and “post-traumatic stress disorder,” as shown in Figure 1. Important details, such as the clinical trial identifier, title, brief summary, date, study description, and outcome measures, were extracted and manually reviewed to determine whether the clinical trials were related to PTSD. A custom Python script (Python Software Foundation) was used to retrieve clinical trials, download registration files, and extract previous information. We established 2 exclusion criteria during this step. First, trials that mentioned PTSD but focused on other diseases were eliminated; for example, a trial on psychogenic nonepileptic seizures that employed the Davidson Trauma Scale (DTS) and mentioned PTSD when introducing the DTS was excluded. Second, trials that were related to PTSD but did not include any instruments in their outcome measures were eliminated; for instance, a trial that focused on the treatment of PTSD with guanfacine was excluded since specific assessment instruments were not mentioned.

Figure 1. Flow diagram showing collection process of clinical trials and PubMed abstracts included in this study. NCT ID: National Clinical Trials identifier; PMID: PubMed identifier; PTSD: posttraumatic stress disorder.



Following the confirmation of relevance, the instruments in each study were annotated from the designed outcome measures. Instruments were restricted to rating scales, self-report inventories, and structured interviews, such as the CAPS [13] and the PCL [14]. Laboratory tests (such as heart rate variability), mentions of symptoms without a clear statement on the instrument applied (eg, weekly number of nightmares and depression symptoms), and other measures were ignored during the annotation. The annotations were performed using brat [15], a widely used web-based tool for text annotation in text mining.

Based on our annotations and the 2 exclusion criteria, 1120 clinical trials were excluded from our data set. As there were only 37 trials found from 1999 to 2004 and the small number of trials made it difficult to calculate the prevalence index, we excluded trials prior to 2005 in this study. In total, 1345 trials were included in our final clinical trial data set for prevalence analysis, and their first submission dates on ClinicalTrials.gov ranged from March 2005 to December 2020.

Instrument Knowledge Construction

A unified name mapping system was built to map the different original instrument names in the text to their corresponding normalized full names to distinguish different instruments and their abbreviations to indicate different versions of each instrument. For example, the PTSD Checklist–Civilian Version and the PTSD Checklist–Military Version were mapped to the normalized full name PTSD Checklist, whereas the abbreviations were assigned as PCL-C and PCL-M, respectively.

Prevalence Index

The prevalence of an instrument is indicated by its usage frequency, which is calculated as the total number of studies in a given period divided by the number of times an instrument is used in that period.

The strength of this index is its capacity to reduce the noise caused by different numbers of studies during different time periods. It quantifies the prevalence and enables comparison of the prevalence across different time periods. In this study, the prevalence index for each instrument was calculated for each year. For example, there were 34 clinical studies in 2006 and

the CAPS was used 20 times during that year. Therefore, the prevalence index of the CAPS in 2006 was 0.5882 (20/34). In 2008, the CAPS was employed in 34 of 57 studies and the prevalence index was 0.6071 (34/56). Although the total number of studies and the number of times each instrument was applied varied, there was no significant change in the prevalence of the CAPS.

To conduct trends analysis on instruments, we calculated and visualized the yearly prevalence index for each instrument and then compared the index change trends between instruments. After trends analysis, we were able to determine whether an instrument was still popular or its usage was decreasing, which instruments were more commonly used, and those that will be widely used in future studies.

Validation With PubMed

Considering the bias and inadequacy introduced by only using registered clinical trials in ClinicalTrials.gov, we retrieved and downloaded PTSD-related abstracts from PubMed using the previously mentioned keywords. A custom Python script, using the Entrez application program interface, helped us retrieve the PubMed IDs and extract publication information. A similar trends analysis was conducted on instruments mentioned in those publications to validate the analysis results from clinical trials. The instruments mentioned in the abstracts were automatically recognized by exact string matching using the various instrument names obtained during manual annotation. Abstracts that did not mention any instrument were excluded. In total, 9422 abstracts were included for prevalence and temporal trends analysis, as shown in Figure 1.

To evaluate the risk of a clinical study's measures being obsolete, the time durations between the submission and publication dates of a clinical trial were also compared. For a published study, the time duration was defined as the interval between the first posted date in ClinicalTrials.gov and the publication date in PubMed. If there were multiple papers published based on one clinical trial, the earliest publication date was used. This comparison was performed to infer the potential influence of knowledge updates on clinical studies.

Results

Overall Trends

A total of 4178 instrument synonyms were annotated, which were mapped to 1423 unique instruments. The number of trials, the number of applied instruments, and the number of applied unique instruments in each year, as well as the average number of instruments applied in one trial, are provided in Table 1. It should be noted that the number of scales applied each year shows an increasing trend over time.

From 2005 to 2020, only 10 instruments were used more than once per year (Table 2), 17 instruments were used ≥50 times, and 1255 instruments were employed less than 10 times in total. The most commonly used instruments were the PCL, the CAPS, the Patient Health Questionnaire (PHQ), and the Beck Depression Inventory (BDI). The list of instruments and their prevalence trends are available online [16].

Table 1. Statistical data for each year.

Year	Trials, n	Applied instruments, n	Unique instruments applied, n	Instruments used in one trial, mean
2005	44	170	63	3.86
2006	34	186	97	5.47
2007	34	112	57	3.29
2008	56	207	92	3.70
2009	58	228	107	3.93
2010	60	217	115	3.62
2011	68	272	133	4.00
2012	104	422	187	4.06
2013	87	408	209	4.69
2014	87	389	190	4.47
2015	89	424	223	4.76
2016	99	398	201	4.02
2017	106	538	277	5.08
2018	112	586	276	5.23
2019	136	725	318	5.33
2020	171	842	326	4.92

Table 2. Instruments applied more than once per year and their associated total prevalence index (2005-2020).

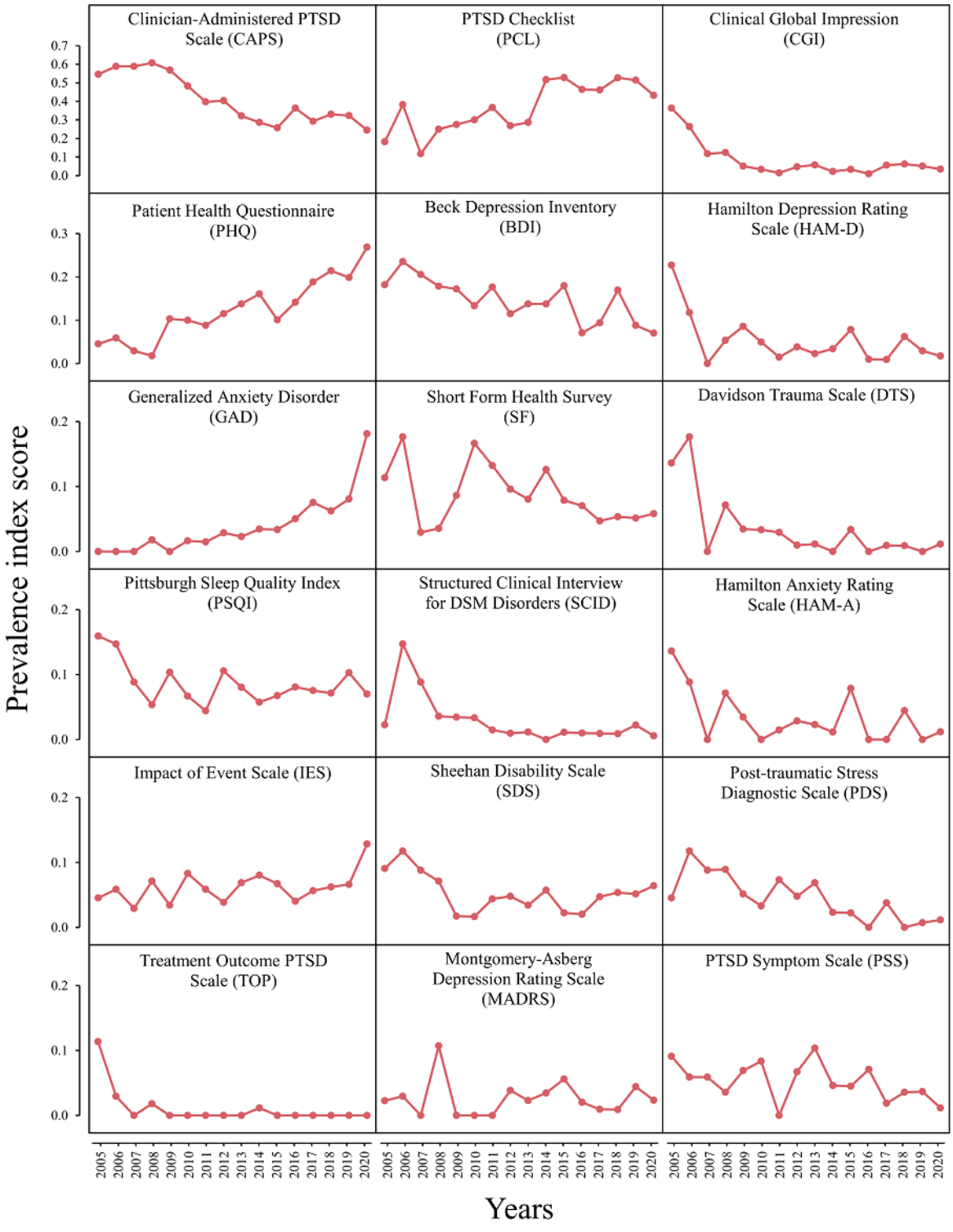
Instrument	Times applied, n	Total prevalence index[a]
PTSD[b] Checklist	541	0.4022
Clinician-Administered PTSD Scale	495	0.3680
Patient Health Questionnaire	202	0.1502
Beck Depression Inventory	175	0.1301
Pittsburgh Sleep Quality Index	110	0.0818
Short Form Health Survey	108	0.0803
Impact of Event Scale	91	0.0677
Clinical Global Impression	84	0.0625
Hospital Anxiety and Depression Scale	71	0.0528
Sheehan Disability Scale	66	0.0491

[a]The total prevalence index was calculated by dividing the number of times an instrument was used from 2005 to 2020 by the total number of clinical trials in those 16 years.

[b]PTSD: posttraumatic stress disorder.

To analyze the temporal trends of the instruments, the prevalence index for each instrument was calculated for each year. There were 18 instruments whose prevalence index exceeded 0.1 at least once in the 16 years, as observed in Figure 2. Among the 18 instruments with distinct variation trends, the prevalence index of the CAPS took 16 years (2005 to 2020) to decrease from 0.5455 to 0.2456, which is a relatively long period. In the 11 years from 2005 to 2015, the usage rate of the PCL increased from 0.1819 to 0.5281. The prevalence index of the PHQ increased 4 times after 2008, and the use of the Clinical Global Impression (CGI) decreased by approximately 68% (from 0.3636 to 0.1176) in 2 years beginning in 2005. The prevalence indexes of the Structured Clinical Interview for DSM Disorders (SCID), DTS, and PHQ also decreased by more than 50% over 2 or 3 years beginning in 2005 or 2006. The results showed that even the usage rates of the top instruments in this field were not very high. Scales such as the CAPS, the CGI, the BDI, and the Hamilton Depression Rating Scale (HAM-D) were popular in the early years, but their usage rate decreased over time.

Figure 2. Prevalence indexes for the top 18 instruments whose indexes exceeded 0.1 at least once between 2005 and 2020. DSM: Diagnostic and Statistical Manual of Mental Disorders; PTSD: posttraumatic stress disorder.

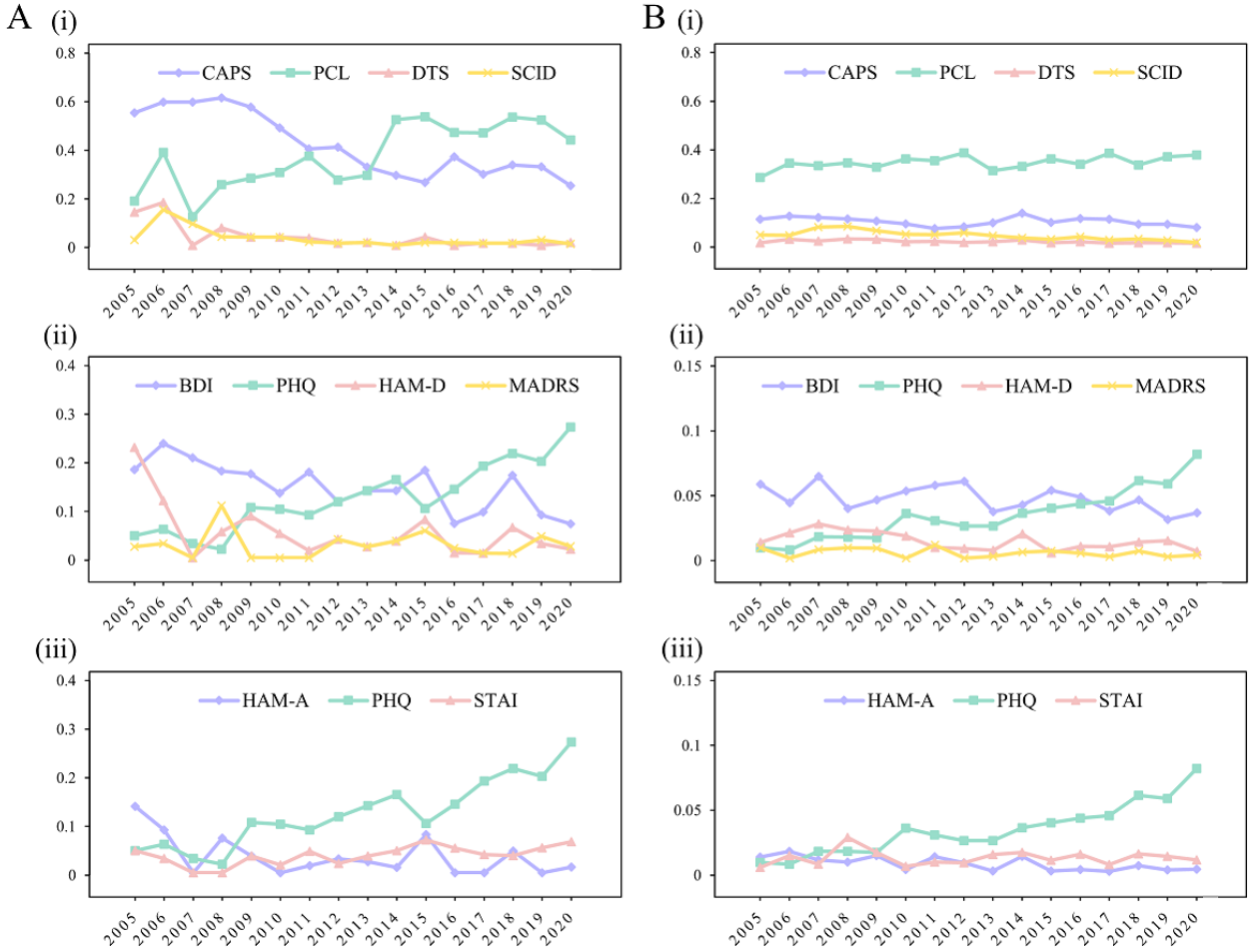


Trend Analysis Based on the Assessment Targets

It is difficult to determine the latent information behind temporal trends when all available instruments are considered. Therefore, we compared and analyzed the changing trends among the top instruments according to their assessment targets. By comparing the usage trends of rating scales with similar functions, changes in the conventional usage of assessment tools can be revealed.

The CAPS, PCL, DTS, and SCID were categorized as "comprehensive rating scales" that assess the symptoms of PTSD based on the DSM. The BDI, PHQ, HAM-D, and Montgomery-Asberg Depression Rating Scale were classified into the "depression" group. The Hamilton Anxiety Rating Scale (HAM-A), PHQ, and State-Trait Anxiety Inventory (STAI) were classified under the "anxiety" scale group. The usage rates of each group are shown in Figure 3A.

Figure 3. Temporal trends comparison of the instruments based on the same assessment target: (A) based on the clinical trials data and (B) based on the published abstracts data set. (i) Indicates the comprehensive scales; (ii) indicates the depression symptoms scales; (iii) indicate the anxiety symptoms scales. In both data sets, the PTSD Checklist (PCL) and the Patient Health Questionnaire (PHQ) show steady upward trends, whereas the use of the Clinician-Administered PTSD Disorder Scale (CAPS) and Beck Depression Inventory (BDI) is decreasing. DTS: Davidson Trauma Scale; HAM-A: Hamilton Anxiety Rating Scale; HAM-D: Hamilton Depression Rating Scale; MADRS: Montgomery-Asberg Depression Rating Scale; SCID: Structured Clinical Interview for Diagnostic and Statistical Manual Disorders; STAI: State-Trait Anxiety Inventory.

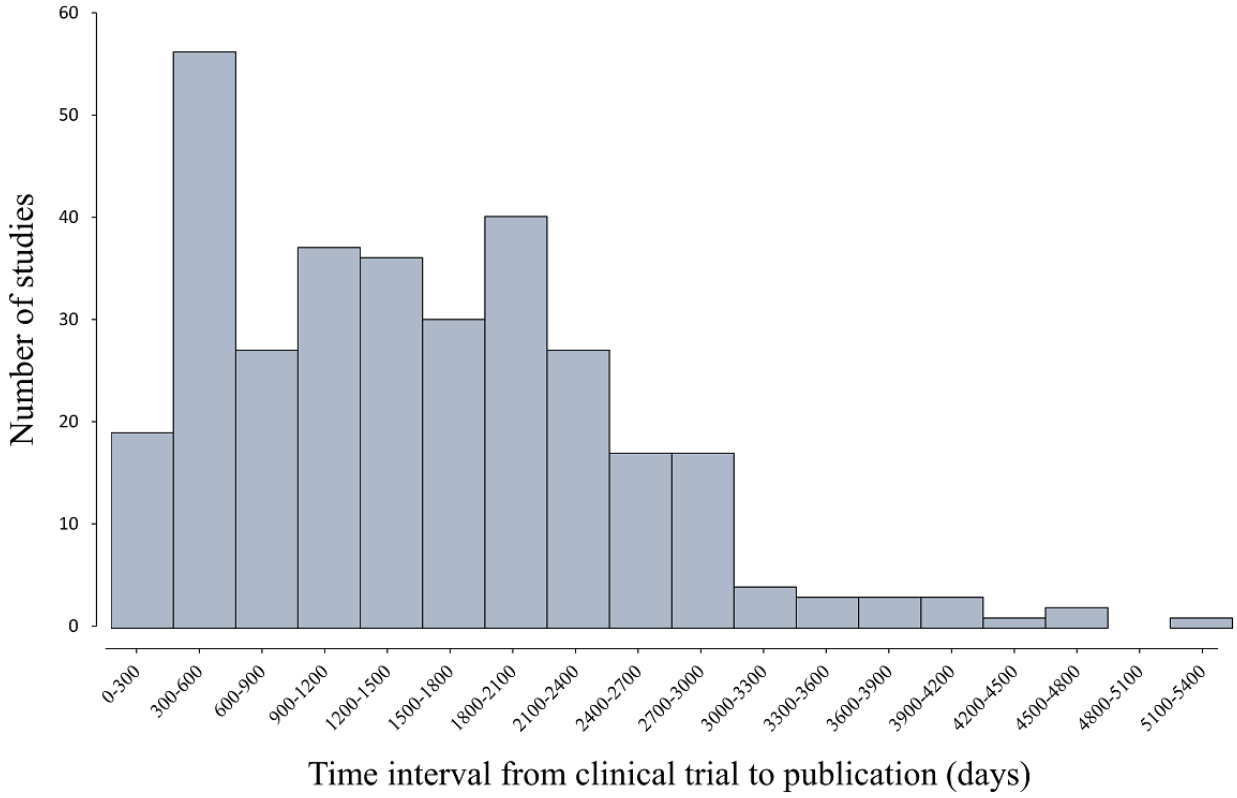


There were 9422 PTSD-related papers published from 2005 to 2020 with instruments mentioned in their abstracts. The top instruments were similar to those identified in the clinical trial data set, with some changes in their rank. In the PubMed data set, the PCL was the most used instrument (appearing 3244 times) and the Impact of Event Scale was the second most common instrument (appearing 1057 times). The trends in the prevalence indexes of the 3 groups are presented in Figure 3B, demonstrating that the changes in trends and time points are highly consistent with the results of the clinical trial data set.

Trial Time Duration and Changes in the Prevalence Rate

The average time duration of the clinical trials was compared with the time required for the prevalence of an instrument to significantly increase or decrease. We retrieved 487 papers from the PubMed database, which were derived from 323 clinical trials. The interval between the first submission date in ClinicalTrials.gov and the publication date of the earliest paper was regarded as the time duration. The results in Figure 4 show that the average time duration is 1495.5 days or approximately 4 years. Most of the trials (n=253) required 300 to 2400 days to publish a paper after they started, and the average time duration of these 253 trials was 1275 days (median 1281 days). In the 323 trials, the CAPS and the PCL were the most popular instruments, and both were employed 116 times.

Figure 4. Distribution of the publication time durations of the 323 clinical studies ranging from 78 to 5339 days.



Discussion

Principal Findings

Our results demonstrate that there are many available instruments with generally low usage rates, indicating that there is no conventional assessment tool for PTSD. It is meaningful to reveal the trends and knowledge based on which researchers select assessment tools to use in clinical trials.

As we observed the decreased prevalence of some top instruments, two reasons seemed appropriate to explain this phenomenon. One reason is the reduction in the number of instruments in each clinical trial. For example, 1 clinical trial may have used approximately 10 instruments early in the study, 7 of which were the top instruments; however, currently, the same clinical trial only uses 5 instruments, 3 of which are top instruments. This reduction can lead to decreased prevalence owing to the method used to calculate the prevalence index. The other probable reason is that a certain number of new instruments were introduced into the field each year, diluting the prevalence of the top instruments. To verify these possibilities, the number of instruments applied in each year and the average number of instruments applied in 1 trial were calculated. The latter reason was supported by the results presented in Table 1. It seems that the number of new, but not widely used, scales introduced into the field each year diluted the prevalence of all the instruments and reduced the usage of a particular conventional instrument.

Analysis of the temporal trends showed that the CAPS and the PCL were the most widely applied, comprehensive instruments, and their high usage rate demonstrated that most of the clinical trials required comprehensive diagnostic instruments. The popularity of the CAPS decreased (from 0.5455 in 2005 to 0.2456 in 2020), whereas that of the PCL increased (from 0.1818 in 2005 to 0.4327 in 2020) over time, which was also validated using the PubMed data. To determine the reason for this phenomenon, we retrieved and reviewed studies related to the CAPS and the PCL. As Fonkoue et al [17] point out, the CAPS and the PCL have excellent psychometric properties with high interrater reliability, test-retest reliability, and internal consistency [18]. The scores of the CAPS and the PCL were also highly correlated and showed high diagnostic agreement [19,20]. The most likely cause of this phenomenon could be the different formats of these instruments, rather than their quality. The CAPS employs a structured interview that can only be administered by trained and experienced clinicians, whereas the PCL is a more convenient self-report assessment instrument that requires less time and resources.

As for the depression assessments, the results from both data sets indicate that although the BDI is still the most widely used assessment instrument (increasing from approximately 0.07 to 0.24 over the 16-year period), usage of the PHQ is increasing (from 0.05 to 0.27) and it may become the most prevalent instrument to assess depression symptoms. Compared with the STAI and the HAM-A, the PHQ was also the most commonly used instrument for anxiety symptoms. These results show that since 2009, the PHQ has become a widely accepted instrument to assess depression and anxiety. One possible reason for these findings may be that the PHQ-9 is a relatively new, simple, and freely available instrument [21].

There were also rating scales that had no alternatives among the most applied instruments, such as the Pittsburgh Sleep

Quality Index to assess sleep, the Short Form Heath Survey to assess an individual's health status, and the Impact of Event Scale to assess the influence of traumatic events. These instruments are regarded as the accepted scales for their respective assessment targets.

Compared with the time duration required for clinical research to be published (4 years on average), variations in the prevalence of instruments occur rapidly. The popularity of instruments such as the PHQ, CGI, SCID, DTS, and HAM-A changed by more than 50% in less than 3 years. In contrast, trends in the prevalence of popular instruments such as the CAPS and the PCL indicate that they are more stable. Considering the reliability of research outcomes as well as the stability and high acceptance of study measures, employing the scales that are more prevalent or trending upward is recommended when there are several alternatives that can meet research interests equally. This can also help to build a consensus regarding the assessment tools used in the field of PTSD.

Although we have tried to include as many PTSD-related clinical trials as possible in this study, there are some limitations to using ClinicalTrials.gov as a data resource: (1) this data source may not be able to represent all researchers studying PTSD; and (2) the studies included in ClinicalTrials.gov change over time (such as the inclusion of more small-scale trials), which may also result in temporal changes.

To overcome these limitations and validate the results obtained from ClinicalTrials.gov, we used the large-scale PubMed data set to conduct similar trend analyses. We found that the results obtained using the PubMed data set were highly consistent with those obtained from the ClinicalTrials.gov data set. In this manner, the results and conclusions of our approach can be reciprocally verified by the clinical trials and their corresponding publications. Although we aimed to conduct a trends analysis on all presently used instruments, this task was difficult to complete. Most of the instruments were applied for less than 6 years with a low usage rate, which resulted in many zero values and fluctuating trends in the prevalence indexes. To achieve a reliable conclusion, the temporal trends analysis only focused on the most popular instruments. Therefore, some instruments known to be relevant to PTSD are not mentioned in our results and discussions, such as the Composite International Diagnostic Interview (CIDI), a diagnostic measure for PTSD that was only used 5 times from 2005 to 2020. There are 12 years during which all the clinical trials in that year did not employ the CIDI in their studies (ie, prevalence index=0). The prevalence indexes for the CIDI in 2007, 2011, 2014, and 2015 were 0.0589, 0.0147, 0.0115, and 0.0112, respectively. Other relevant instruments such as the Posttraumatic Stress Syndrome Inventory and Short Inventory of Problems are not mentioned in this paper owing to the abovementioned reason.

In the future, more accurate natural language processing methods should be developed to recognize and extract instruments automatically from more databases. More integrated and comprehensive knowledge should also be collected. Based on this data and knowledge, comprehensive studies on instruments focusing on one condition can be conducted. The study objective, target population, and functions of the instruments should also be considered when recommending an instrument. A system that can automatically recommend suitable instruments for a certain study using statistical methods and indexes can be developed.

Conclusions

Using widely accepted and applied instruments can help improve the reliability of research results in PTSD-related clinical studies. Considering the long duration of each study as well as the large variety of study instruments, it is challenging for researchers or clinicians to select the most appropriate instrument according to updated knowledge or trends. In this study, we investigated the prevalence indexes of various PTSD-related instruments and conducted a temporal trend analysis for PTSD-related studies using data from clinical trials and PubMed. Our work aimed to determine the most prevalently used instruments in PTSD-related clinical studies while considering the assessment target and to reveal knowledge and trends. Furthermore, we discuss the reasons for these trends to provide updated information and supportive knowledge to researchers to help reach a consensus regarding the use of assessment tools. Our results also demonstrate the feasibility of conducting a temporal trends analysis on clinical studies and its potential to support research design and implementation.

Acknowledgments

HZ acknowledges Jiaxue Cha, PhD candidate, of Tongji University in China, for discussing and providing helpful suggestions during this study. This work was supported by the National Natural Science Foundation of China (grants 81972914 and 81573023), the Fundamental Research Funds for the Central Universities (grant 22120200014), the Shanghai "Rising Stars of Medical Talent" Youth Development Program (grant 2019-72), and the National Key R&D Program of China (grant 2016YFC1303200).

Authors' Contributions

HZ and BH were involved in methodology development, experimental design, annotation, data analysis, and manuscript preparation. YH provided constructive suggestions and revised the manuscript. ZL and XZ supervised the research. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Pai A, Suris A, North C. Posttraumatic Stress Disorder in the DSM-5: Controversy, Change, and Conceptual Considerations. Behav Sci (Basel) 2017 Feb 13;7(1):7 [FREE Full text] [doi: 10.3390/bs7010007] [Medline: 28208816]
2. Vahia V. Diagnostic and statistical manual of mental disorders 5: A quick glance. Indian J Psychiatry 2013 Jul;55(3):220-223 [FREE Full text] [doi: 10.4103/0019-5545.117131] [Medline: 24082241]
3. Shalev A, Liberzon I, Marmar C. Post-Traumatic Stress Disorder. N Engl J Med 2017 Jun 22;376(25):2459-2469. [doi: 10.1056/nejmra1612499]
4. Benjet C, Bromet E, Karam EG, Kessler RC, McLaughlin KA, Ruscio AM, et al. The epidemiology of traumatic event exposure worldwide: results from the World Mental Health Survey Consortium. Psychol Med 2015 Oct 29;46(2):327-343. [doi: 10.1017/s0033291715001981]
5. Elhai JD, Gray MJ, Kashdan TB, Franklin CL. Which instruments are most commonly used to assess traumatic event exposure and posttraumatic effects?: A survey of traumatic stress professionals. J Trauma Stress 2005 Oct;18(5):541-545. [doi: 10.1002/jts.20062] [Medline: 16281252]
6. Steel JL, Dunlavy AC, Stillman J, Pape HC. Measuring depression and PTSD after trauma: common scales and checklists. Injury 2011 Mar;42(3):288-300 [FREE Full text] [doi: 10.1016/j.injury.2010.11.045] [Medline: 21216400]
7. Rosendahl J, Kisyova H, Gawlytta R, Scherag A. Comparative validation of three screening instruments for posttraumatic stress disorder after intensive care. J Crit Care 2019 Oct;53:149-154. [doi: 10.1016/j.jcrc.2019.06.016] [Medline: 31247513]
8. LeardMann CA, McMaster HS, Warner S, Esquivel AP, Porter B, Powell TM, Millennium Cohort Study Team. Comparison of Posttraumatic Stress Disorder Checklist Instruments From Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition vs Fifth Edition in a Large Cohort of US Military Service Members and Veterans. JAMA Netw Open 2021 Apr 01;4(4):e218072 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.8072] [Medline: 33904913]
9. Michel J, Shen YK, Aiden AP, Veres A, Gray MK, Google Books Team, et al. Quantitative analysis of culture using millions of digitized books. Science 2011 Jan 14;331(6014):176-182 [FREE Full text] [doi: 10.1126/science.1199644] [Medline: 21163965]
10. Bollen J, Ten Thij M, Breithaupt F, Barron ATJ, Rutter LA, Lorenzo-Luaces L, et al. Historical language records reveal a surge of cognitive distortions in recent decades. Proc Natl Acad Sci U S A 2021 Jul 27;118(30):e2102061118 [FREE Full text] [doi: 10.1073/pnas.2102061118] [Medline: 34301899]
11. Yu Y, Li Y, Zhang Z, Gu Z, Zhong H, Zha Q, et al. A bibliometric analysis using VOSviewer of publications on COVID-19. Ann Transl Med 2020 Jul;8(13):816-816 [FREE Full text] [doi: 10.21037/atm-20-4235] [Medline: 32793661]
12. Chou L, Chang K, Puspitasari I. Drug Abuse Research Trend Investigation with Text Mining. Comput Math Methods Med 2020 Feb 01;2020:1030815-1030818 [FREE Full text] [doi: 10.1155/2020/1030815] [Medline: 32076454]
13. Blake DD, Weathers FW, Nagy LM, Kaloupek DG, Gusman FD, Charney DS, et al. The development of a clinician-administered PTSD scale. J Traum Stress 1995 Jan;8(1):75-90. [doi: 10.1002/jts.2490080106]
14. 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 25;28(6):489-498. [doi: 10.1002/jts.22059] [Medline: 26606250]
15. Stenetorp P. brat: a Web-based Tool for NLP-Assisted Text Annotation. In: Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics. Avignon, France: Association for Computational Linguistics; 2012 Presented at: 13th Conference of the European Chapter of the Association for Computational Linguistics; April 2012; Avignon, France p. 102-107.
16. PTSD Instruments Prevalance and Trends. Tongji University. URL: http://bmtongji.cn:1236/scale/index [accessed 2021-11-08]
17. Fonkoue IT, Marvar PJ, Norrholm S, Li Y, Kankam ML, Jones TN, et al. Symptom severity impacts sympathetic dysregulation and inflammation in post-traumatic stress disorder (PTSD). Brain Behav Immun 2020 Jan;83:260-269 [FREE Full text] [doi: 10.1016/j.bbi.2019.10.021] [Medline: 31682970]
18. Zaninotto AL, Vicentini JE, Fregni F, Rodrigues PA, Botelho C, de Lucia MCS, et al. Updates and Current Perspectives of Psychiatric Assessments after Traumatic Brain Injury: A Systematic Review. Front Psychiatry 2016 Jul 14;7:95 [FREE Full text] [doi: 10.3389/fpsyt.2016.00095] [Medline: 27378949]
19. Blanchard EB, Jones-Alexander J, Buckley TC, Forneris CA. Psychometric properties of the PTSD checklist (PCL). Behaviour Research and Therapy 1996 Aug;34(8):669-673. [doi: 10.1016/0005-7967(96)00033-2]
20. Rubin DC, Deffler SA, Ogle CM, Dowell NM, Graesser AC, Beckham JC. Participant, rater, and computer measures of coherence in posttraumatic stress disorder. J Abnorm Psychol 2016 Jan;125(1):11-25 [FREE Full text] [doi: 10.1037/abn0000126] [Medline: 26523945]
21. Kroenke K, Spitzer RL, Williams JBW. 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]

Abbreviations

BDI: Beck Depression Inventory
CAPS: Clinician-Administered PTSD Scale

CGI: Clinical Global Impression
CIDI: Composite International Diagnostic Interview
DSM: Diagnostic and Statistical Manual of Mental Disorders
DTS: Davidson Trauma Scale
HAM-A: Hamilton Anxiety Rating Scale
HAM-D: Hamilton Depression Rating Scale
PCL: PTSD Checklist
PHQ: Patient Health Questionnaire
PTSD: posttraumatic stress disorder
SCID: Structured Clinical Interview for DSM Disorders
STAI: State-Trait Anxiety Inventory

Edited by G Eysenbach; submitted 15.09.21; peer-reviewed by KM Chang; comments to author 07.10.21; revised version received 12.10.21; accepted 17.10.21; published 17.11.21.

Please cite as:
Zong H, Hu B, Han Y, Li Z, Zhang X
Prevalence and Temporal Trends Analysis of Screening and Diagnostic Instruments in Posttraumatic Stress Disorder: Text Mining Study
JMIR Ment Health 2021;8(11):e33599
URL: https://mental.jmir.org/2021/11/e33599
doi:10.2196/33599
PMID:34666307

Original Paper

Burden and Help-Seeking Behaviors Linked to Problem Gambling and Gaming: Observational Quantitative and Qualitative Analysis

Amandine Luquiens[1,2], MD, PhD; Cora von Hammerstein[3,4], PhD; Amine Benyamina[3], MD, PhD, Prof Dr; Pascal Perney[1,2], MD, PhD, Prof Dr

[1]Universitary Hospital of Nîmes, University of Montpellier, Nîmes, France

[2]UVSQ, CESP, INSERM, Paris-Saclay University, University of Paris-Sud, Villejuif, France

[3]Addictions Department, Paul Brousse Hospital, APHP, Paris-Saclay University, Villejuif, France

[4]University Research Team EA4360 APEMAC (Health Adjustment, Measurement and Assessment, Interdisciplinary Approaches), University of Lorraine, Metz, France

Corresponding Author:
Amandine Luquiens, MD, PhD
Universitary Hospital of Nîmes
University of Montpellier
Place Robert Debré
Nîmes, 30000
France
Phone: 33 466022569
Email: amandineluquiens@gmail.com

Abstract

Background: Models based on the uniqueness of addiction processes between behavioral addictions are highly contentious, and the inclusion of gaming disorder in the addiction nosography remains controversial. An exploratory approach could clarify a hypothesized common and subjectively identifiable process in addictive behaviors and the necessarily different expressions of the disorder due to behavior specificities, in particular the sociocultural characteristics and profiles of users.

Objective: The aim of this study was to describe the nature of contacts to a help service by exploring commonality and specificities of burden and help-seeking for problem gambling or gaming.

Methods: This was an observational quantitative-qualitative study. We included all contacts (ie, online questions and contacts by phone or chat when the helper completed a summary) to a helpline for gamers, gamblers, and relatives over a 7-year period. We constituted a text corpus with online questions and summaries of contacts by phone or chat. We collected basic sociodemographic data, including the device used to contact the service (phone or internet), contacting the service for oneself ("user") or being a relative of a user and type of relative, gambling (yes/no), gaming (yes/no), and age and sex of the gambler/gamer. We describe the corpus descriptively and report the computerized qualitative analysis of online questions, chat, and summary of phone calls. We performed a descendant hierarchical analysis on the data.

Results: A total of 14,564 contacts were made to the helpline, including 10,017 users and 4547 relatives. The corpus was composed of six classes: (1) gaming specificities, (2) shared psychological distress and negative emotions, (3) the procedure for being banned from gambling, (4) the provided help, (5) gambling specificities, and (6) financial problems.

Conclusions: Negative emotions and shared distress linked to gambling and gaming support current scientific consensus that these behaviors can produce psychological distress in se; however, meaningful differences were observed in core symptoms of addiction between gamers and gamblers, beyond specificities related to the behavior itself: loss of control was elicited in the class corresponding to gambling specificities and not by gamers and their relatives.

(JMIR Ment Health 2021;8(11):e26521) doi:10.2196/26521

KEYWORDS

gambling; gaming; helpline; burden, relatives; qualitative research

Introduction

The International Classification of Diseases (ICD)-11 recognizes two behavioral addictions [1]: gaming and gambling disorders. The clinical descriptions of these disorders comprise three main dimensions: impaired control, increasing priority given to gambling/gaming, and continuation despite the occurrence of negative consequence. This classification implies common processes between different types of addiction in the development and maintenance of addiction. This commonality is proposed by models of behavioral addictions, such as the Interaction of Person-Affect-Cognition-Execution model [2], which presents addictive behaviors as consequences of the interactions between individual vulnerability factors, affective and cognitive responses to specific behavior, and executive functions. This model includes specificities related to the behavior itself that render them more attractive to individuals with certain psychological profiles. However, it does not include the social dimension of behaviors accounting for a social group's greater propensity to use the same behavior as a commodity associated with the group's identity, in contrast with other sociological approaches [3]. Models based on the uniqueness of addiction processes between behavioral addictions are highly contentious, and the inclusion of gaming disorder in the addiction nosography remains controversial [4]. Diagnosis criteria are debated [5], and some authors still consider problem gaming as a symptom of another mental disorder [6].

Beyond these models based on the uniqueness of addictive disorders, comparisons between addictive behaviors in the same study are scarce. Furthermore, the literature lacks a clear picture of the differences between potential addictive behaviors, particularly in the representation of the behavior itself, the associated disorder, and the inner experience and perception of the disorder by users and their relatives. Few studies have made comparisons between gaming and gambling. A previous qualitative study found significant overlap in the experience of people living with a substance use problem and a behavioral addiction/problem [7]. However, gaming and gambling differ at various levels: at the clinical level, assessing the importance of depression as a risk factor in gaming [8]; at the neuropsychological level, in terms of delay discounting and decision making [9]; at the neural level; and in the pharmacodynamic response to certain treatments [10]. However, there are also specificities linked to different profiles of users between the behaviors as commodities. Gaming is more widespread in adolescents and young adults, and the age of onset is earlier than that in gambling, where the average age of onset is around 34 years and the practice is prohibited among minors in most countries [11]. For instance, in a study among 824 adolescents, a prevalence of 3% was obtained of at-risk gaming, with a mean age of 14.5 years, and the majority reported a symptom course greater than 12 months [12]. This younger age range leads to specific cognitions in problem gamers [12].

These differences in profiles support the biopsychosocial model in addictions [13], where addictive behavior as a commodity is consumed by different populations and related disorders have a different representation in the general population [14] with different degrees of stigma, leading to specific clinical pictures and specific complaints from the addict population and relatives. However, these elements are surprisingly absent from the classifications, as illustrated by the absence of age specification in the clinical description of the ICD-11 for gaming disorder [15]. It is therefore important to focus on the difficulties experienced and described by the users themselves and their relatives [16]. The importance of exploratory approaches and qualitative research has also been emphasized by several authors [17,18].

The treatment gap is considerable in behavioral addictions, with fewer than 12% of pathological gamblers seeking help from a health professional [19]. The epidemiological data on gaming disorder are weak, but the prevalence is estimated to range from 0.5% to 10%. The care network and pathways are still being structured. No quantified data exist on the treatment gap in gaming disorder, but it is assumed to be even wider [20]. Nonface-to-face, telephone, and online support devices could help reduce this treatment gap [21], and could further help to collect information to describe a population to which clinicians and researchers have no other access. The debate over the criteria for gaming disorder could be redressed by the description of gaming difficulties by self-diagnosed problem gamers seeking help. There are also minimal data on nonface-to-face help-seekers in gambling and particularly in gaming. The current knowledge on gambling shows the preference of young gamblers for online devices [22], along with the greater reluctance among women to request help and their more difficult access to treatment [22-24]. A study of 168 gamers voluntarily contacting a help service [25] demonstrated the association of mood symptoms and severity of gaming disorder. However, no study has yet explored gaming- and gambling-related burdens collected through the same approach. An exploratory approach could clarify a hypothesized common and subjectively identifiable process in addictive behaviors and the necessarily different expressions of the disorder due to behavior specificities, in particular the sociocultural characteristics and profile of users (eg, age, sex).

We performed a quantitative-qualitative analysis of all contacts to a nonface-to-face help service over 7 years. All contactors voluntarily called or wrote to the help service asking for help or counseling regarding gaming or gambling. No diagnostic or screening assessment was performed or required to access the help service. Our aim was to explore, without a priori assumptions, the burden of gaming and gambling cited by contacts to a help service by phone and internet. Our hypothesis was that we would find both commonalities and specificities in the burden linked to gaming and gambling problems from the perspectives of users and their relatives.

Methods

Population and Data Collected

We report the computerized qualitative analysis of all contacts to a public national help service in France for gamers, gamblers, and their relatives since its creation in 2010 until 2018. This governmental help service, "Joueurs Info Service," is part of a global plan for information and prevention on alcohol, substance, gaming, and gambling addictive behaviors. This is

a national remote help service for gambling, gaming, and other addictions, and is also in charge of listing, updating, and making available to the public the national directory of specialized addiction structures. "Joueurs Info Service" is based on the rules of anonymity, confidentiality, neutrality, and nonjudgment in its missions of information, advice, support, and guidance of the public. Helpers are trained psychologists specialized in addictive behaviors. They empathically and actively listen to contactors and can refer them to another health facility if appropriate following their own clinical judgment. Summaries are written for phone calls and chats with helpers. These summaries do not follow a particular plan, and the level of detail and content of the summary is left to the discretion of the helpers, although they are encouraged to document the contact as exhaustively as possible. A single number and website for "Jeu Info Service" (in French, "jeu" encompasses both gambling and gaming) allows gamers, gamblers, and their relatives to talk to professional helpers by phone, chat, or written question with delayed answers. The same helpers respond to gambling and gaming contacts. We included all contacts from users and their relatives to the help service by phone when the helper completed a summary and online (ie, chat when the helper completed a summary and questions) when the helper considered contacts to be in the scope of the help service (ie, a problem with gaming and gambling: n=17,440 of a total 97,350 contacts). Nonincluded contacts were categorized by helpers as not within the scope of the helpline and were most commonly prank calls and errors (gamblers looking for the technical hotline of gambling websites). Included contacts were then people reporting having a problem with gambling, gaming, or a relative's gaming or gambling behavior, and seeking help or counseling.

The merged corpus was composed of summaries of calls and chat discussions written by helpers, and the content of questions written directly by contactors. In the manuscript, the word "user" will designate both gamblers and gamers. Basic sociodemographic data were collected, including the device used to contact the service (phone or internet), contacting the service for oneself ("user") or being a relative of a user and the type of relative, gambling (yes/no), gaming (yes/no), age of the gambler/gamer, age category (<25, 25 to <65, and >65 years), and sex of the gambler/gamer.

Data Analysis

A content text analysis was performed on the corpus using the open-source automated analysis software Iramuteq (version 0.7 alpha 2) according to the Reinart method [26], which is a descending hierarchical classification method to obtain stable classes of words. Classes of words were the most significant themes of the corpus, after a first lemmatization step, when all words (ie, nouns, verbs, adjectives, adverbs) were reduced to their radicals into "forms." The Reinart method was performed as follows: (i) text segmentation into text segments of approximatively 3 lines; (ii) definition of a contingency table of forms (as defined above) and text segments; and (iii) descending hierarchical classification with a double classification of text segments grouping in an iterative partition that maximizes interclass inertia and minimizes intraclass inertia, so that classes were as homogeneous as possible within the class (ie, having text segments with the common pattern of forms) and as heterogeneous as possible between the classes (having less in common). This iterative partition stopped when the extracted interclass inertia was not improved by a new partition of the corpus [27]. The final number of classes was a priori undetermined. A dendrogram was generated.

Once these classes were identified, associations between classes and "passive" independent variables were tested. Passive independent variables were sociodemographic variables as described above, including gaming and gambling. The strength of association between the forms and the classes, and between the passive independent variables and the classes was determined by χ^2 tests. Significant forms and variables are presented (P<.01). Given the size of the corpus, we chose to present only forms and variables with χ^2>100 within a class. This automated text-mining approach generates similar themes to traditional qualitative content analysis, and is thus considered to be a reliable text analysis [28]. In this study, we purposely used a computerized automated analysis to limit any a priori assumptions on the burden linked to gambling/gaming and to ignore gambling and gaming categorizations so as to explore commonalities and specificities of these two fields: gaming and gambling variables were the only passive variables as explained above. Classes were then described according to context.

Ethics

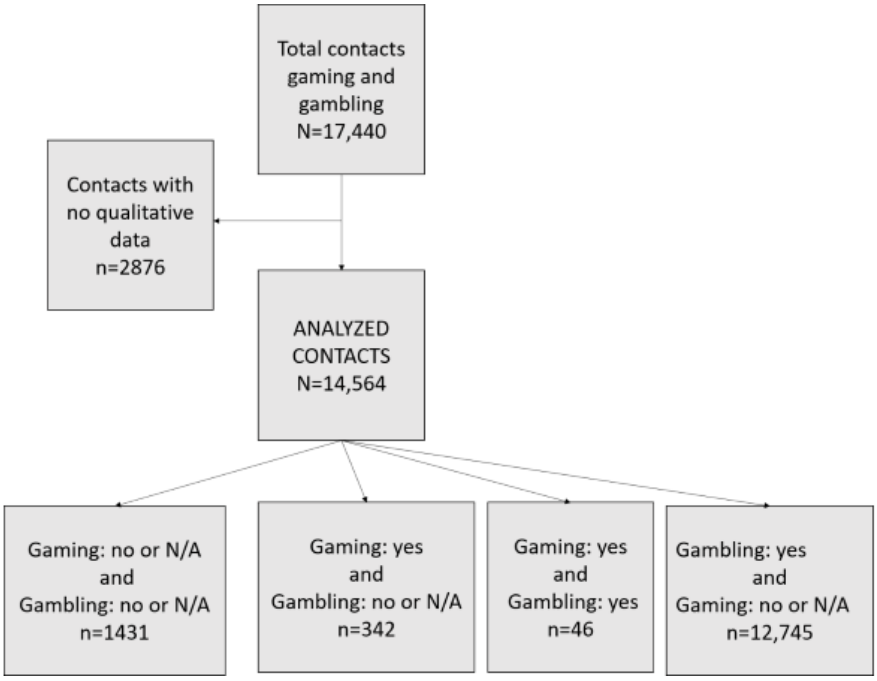
Data collection was anonymous and declared to the National Commission for Data Processing and Liberties (CNIL number 1433300). Contactors were informed of data collection and analysis in the Terms and Conditions section of the website.

Results

Sample

A total of 17,440 contacts were made during the study period, 14,564 (83.51%) of which had written summaries. Figure 1 presents the flow chart of inclusion of contacts in the study. Contactors were predominantly users (10,017/14,564, 68.78%), with 4547 relatives, including 1654 life partners, 1206 parents, 478 children, 398 siblings, 280 friends, 221 other members of the family (179 boy/girlfriends, 35 grandparents, 90 others); 6 were not completed. Users were 58.9% male (n=57 not completed). In total, 1144 were aged <25 years, 7363 were aged 25-64 years, and 703 were aged >65 years (n=5352 not completed). Only 330 contactors (2.27%) contacted the service via internet. Gaming was noted in 388 contacts and gambling in 12,791 contacts.

Figure 1. Flow chart of contacts. N/A: not available.



The mean age of gambling users was 40.9 (SD 17.1) years and that of gaming users was 22.8 (SD 9.1) years. Relatives made up 75.3% of contacts regarding gaming (n=292) and 29.60% regarding gambling (n=3786).

Statistics of the Corpus

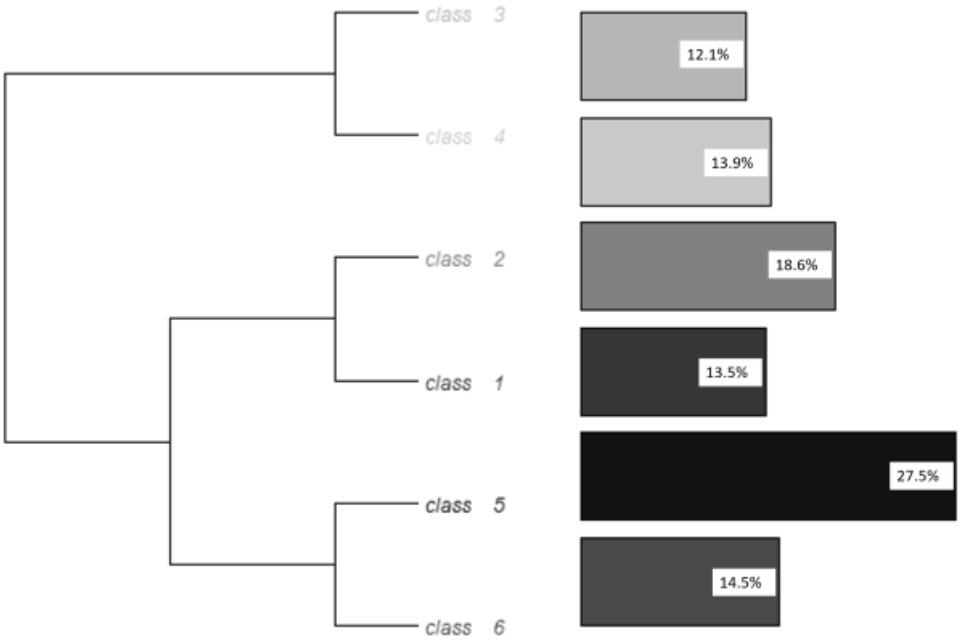
We counted 969,626 occurrences (ie, total number of words) in the corpus, and 19,119 forms and 18,428 analyzed forms. In total, 379 forms showed a frequency ≥100. The top 10 forms (play, game/gamble, to go to, to ask for, money, euro, call, casino, to speak, to wish) were the triggers for making the contact, and gambling was highly represented. The following forms were related to emotional distress: addiction, help-seeking and referral, family members and contexts, financial difficulties, work, and gambling/gaming practice in itself.

Reinart Descendant Hierarchical Analysis

Overview

The corpus was automatically divided into 68,658 text segments. The analysis retained six different classes; 72.7% of segments were classed into one of the six classes. Figure 2 represents the dendrogram of the Reinart descendant hierarchical analysis.

Figure 2. Dendrogram of the Reinart descendant hierarchical analysis.



Class 1: Gaming Specificities

This class of forms was mostly elicited by relatives of gamers (ie, χ^2 values for relatives, gaming, and no gambling were respectively 2629.1, 3869.0, and 1424.6; P<.001). The gaming specificities class is characterized by child adverse events such as the death (χ^2 for "death"=211.9, χ^2 for "dead"=176.6; P<.001) or suicide (χ^2=167.9, P<.001) of a loved one and the alcohol (χ^2=188.7, P<.001) dependence of a parent or divorce (χ^2=101.5, P<.001): "addicted to gaming for 8 years, which corresponds to his parents' divorce" (parent of a male adult gamer via phone contact). Concerns of relatives were driven by missing school (χ^2=184.9, P<.001) or a drop in grades (eg, "he's been missing school more and more, his grades are dropping"; parent of a male gamer via phone contact) and violence (χ^2 for violence=219.5, χ^2 for violent=479.3; P<.001). Violence was significantly associated to this class. Relatives of gamers reported verbal and physical assault on furniture by the gamer, and both hetero- and self-aggressivity:

> *He had a crisis and became violent toward her [...] he feels she's depriving him of his freedom. He threatened to hang himself or shoot himself in the head.* [Parent of male gamer via phone contact]
>
> *Daily violence toward her mother as soon as she tries to set limits.* [Grandparent of a male gamer via phone contact]

They also reported a global violent climate and violence from the parents of the gamer: "she stands between father and son fearing violence against each other" (parent of a male gamer via phone contact).

Although gambling was not associated to this class, some relatives of gamblers also reported violence: "her partner becomes violent and pressures her for money" (life partner of a male gambler via phone contact). Gamblers also reported suicidal thoughts but rather as the consequence of psychological exhaustion:

> *I swear if someone doesn't help me I'll hang myself.* [Male gambler via internet contact; question]
>
> *Do you really think that after 11 years of nightmares, I can stop playing? I can't take it anymore. We are really alone in this battle that we will never win. Obviously, I've wanted to kill myself several times.* [Male gambler via internet contact; question]

Class 2: Shared Psychological Distress and Negative Emotions

This class was not significantly associated to gambling or gaming but was associated with relatives, and especially life partners (χ^2=24.9 and 327.5, respectively; P<.001), who described both their own distress and the distress of the user (see Textbox 1 for representative quotes).

Textbox 1. Quotes related to psychological distress and negative emotions (class 2).

Distress of the user expressed by relatives/life partners:

"He feels [about his life partner who gambles] a gap, an incomprehension, a feeling of powerlessness" (life partner of a male gamer via phone contact).

"His parents can't help but support him [financially], as they think that he feels lonely" (sister of a gambler via phone contact).

Feelings of powerlessness of the user expressed by relatives/life partners:

"Her husband gambles. When she tries to talk about it, he gets into a terrible rage" (life partner of a male gambler via phone contact).

"She is very angry with her sister-in-law because she gave money to help her and she doesn't seem to care" (sibling of a female gambler via phone contact).

"The father wants to hit him and restrains himself [since he stole money for video games], the mother is angry and feels hatred toward him, they don't understand what's happening" (parent of a male gamer via phone contact).

Negative emotion and psychological distress expressed by users:

"He feels empty and feels like he's lost years just gaming and hasn't grown up" (male user, unspecified, via phone contact).

"This young woman is very concerned about normality and shows a strong feeling of guilt and shame" (female gambler via phone contact).

"She's aware that if she thinks she is [addicted to video games], there's suffering behind it. She can't figure out why she's feeling bad" (female gamer via phone contact).

"He feels difficulty in relating to others, inhibition, shame, and guilt about not being able to take part in a discussion. He played video games for 4 years, cutting himself off from others, now feels out of step, with concentration difficulties [...], depressed" (male gamer via phone contact).

"She feels lonely and isolated, the gambling, for the last year, fills her loneliness" (female gambler via phone contact).

Powerlessness is often linked with a lack of trust (χ^2=84.8, P<.001): "His life partner has no trust in him" (life partner of a male gambler via phone contact). Anger was also present in this class (χ^2=94.1, P<.001), manifested both by users and relatives, and characterizes communication and the relationship between users and their relatives as the outcome of lack of trust and powerlessness (see Textbox 1 for representative quotes).

However, a large number of text segments from this class was derived from contacts with gamblers and gamers themselves. The word "feeling" and several negative emotions such as shame, guilt, and loneliness were associated to this class. Textbox 1 includes quotes representing the panel of negative emotion and psychological distress felt by users.

Class 3: Procedure for Being Banned From Gambling

This class was elicited by gamblers (overall χ^2=475.0, to ban χ^2=4199.0, *P*<.001; users, gambling, and no gaming χ^2=325.7, 145.8, and 286.3, respectively; *P*<.001): "He has questions about gambling ban" (male gambler via phone contact).

Class 4: Provided Help

The class related to provided help was mostly elicited in contacts with users themselves (overall: χ^2=189.7, *P*<.001; referral, help, address χ^2=2991.4, 973.4, and 1600.3, respectively; *P*<.001): "I support and refer this user toward the clinical outpatient center […] because he says he bets more and more on horse races with less and less control" (male gambler via internet contact; question).

Class 5: Gambling Specificities

This class was mostly elicited by gamblers (ie, users and gambling: yes, χ^2=2669.8 and 338.4, respectively; *P*<.001). The words "adrenalin," "[not being able to] stop," "can't," "control," "desire to make money," and "chasing" were significantly associated to this class and reflect symptoms of gambling disorder, especially the core symptom of addiction that is loss of control, and gambling expectations focused on excitement and making money. Textbox 2 includes quotes that illustrate this association.

Textbox 2. Quotes related to gambling specificities expressed by users (class 5).

"Says she's addicted to gambling, can't avoid gambling when she walks past a tobacconist's shop" (female gambler via phone contact).

"He can't get over the money he's lost and keeps chasing losses" (male gambler via phone contact).

"He realizes that he is losing control. He is looking for the desire to make money but also the excitement of the risk of losing" (male gambler via phone contact).

"He explains very well what gambling provides: adrenalin, feeling of being surrounded" (male gambler via phone contact).

"Keeps chasing losses, and that's all he thinks about all the time. He's fed up and wants to stop" (male gambler via internet contact; chat).

"I'm exhausted, I've spent all my savings on gambling and I can't stop it's stronger than me... please help me..." (male gambler via internet contact; question).

The words "win" and "lose" were also significantly associated to this class, along with words related to the gambling practice in itself (eg, horse betting, sport betting, poker, scratch card, soccer, money, and to bet).

Class 6: Financial Problems

This class was more strongly elicited by relatives than by users themselves (χ^2=887.3, *P*<.001) and barely in gambling (χ^2=28.2, *P*<.001 for gambling and χ^2=13.0, *P*=.003 for no gaming), especially life partners (χ^2=820.0, *P*<.001): "The women is calling about her brother, who she just discovered has a casino addiction and a huge debt. The family is in shock" (sibling of a male gambler via phone contact). Critical financial situations and debts due to gambling were reported repeatedly along with suicidal thoughts: "says he has so many debts [due to gambling] that he feels there is no other solution than suicide" (male gambler via internet contact; chat).

Gamer spending remains anecdotal and repeatedly financed by theft, but does not fall into the category of financial damage: "this young man says that he is addicted to video games, he spends a lot of money and has stolen his grandfather's credit card" (male gamer via phone contact).

Multimedia Appendix 1 presents the χ^2 values (>100) between forms and the classes, and between the passive independent variables and the classes (all *P*<.001).

Discussion

Our quantitative-qualitative explorative study allowed for exploration without any a priori burdens linked to gaming and gambling in a large and exhaustive sample of contacts to a help service by phone or online. As hypothesized, some classes grouped both gaming and gambling, but others illustrated their specificities, both for the profiles of users and for core clinical symptoms.

Negative emotions described in class 2 regarded gaming and gambling for users and relatives. This class highlights that users or relatives of gamers and gamblers can experience psychological distress related to these behaviors. Current American Psychiatric Association and World Health Organization classifications mention as a prerequisite for gambling and gaming disorder that these behaviors lead "to clinically significant impairment or distress" [1,29], and align with expert consensus that these behaviors produce psychological distress in some people [30]. Our findings support this scientific consensus and the relevance of structuring care and facilitating access to care for these populations, while countering the assumptions of moral panic on gambling and gaming particularly. These negative emotions (especially guilt and loneliness) have been consistently described in addictions, particularly with regard to gambling disorder [31]. The feeling of incomprehension of the user's own behavior and mental state, and of the user's behavior regarding relatives had also been previously described in the context of gambling [31]. Our study illustrates that this feeling is shared in gamers and relatives, and is similar to the powerlessness to change or help. Interestingly, "loss of control," as an escalation of behavior and failure of inhibition, was present in another class specific to gamblers, but was not associated to gaming. Thus, the specificities of gambling go beyond the practical aspect of games as a consumer product, touching on the addictive symptomatology. Loss of control, the core symptom of addiction, was not found to be associated with gaming in this study. This finding questions the recent assertion that gaming disorder is part of the addictive disorder framework in the ICD-11. Although loss of control

had been previously described as part of the disease in gaming disorder in qualitative research with a very small sample of nine gamblers using the grounded theory approach [32], our computerized analysis on a very large sample allowed us to move away from a priori models of addiction for understanding gaming and gambling disorders.

The difference in the onset of suicidal thoughts also puts into question the commonality between gaming and gambling disorder. In gambling, suicidal thoughts seemed to occur concomitantly to psychological exhaustion and the feeling of being pushed to the wall due to financial damage. These feelings have been consistently described in the literature [23,33]. In gaming, the suicide theme appeared indirectly during conflicts with parents. This finding supports previous studies suggesting predominant roles in gaming disorder development and maintenance of parental rejection, poor attachment modalities, and resulting poor self-esteem [34]. Moreover, childhood adverse events and a difficult family climate were broadly reported in class 1, which reflects gaming specificities, namely in younger users. Emotional trauma was recently described as an indirect path to gaming disorder through depressive symptoms [35]. The hypothesis of high involvement in gaming as an effort to cope has also been previously debated [36]. No causal inference can be drawn from this cross-sectional study; however, relatives repeatedly described a chronological link between an adverse event and the onset of gaming disorder. This could also reflect a reporting bias in that relatives, and particularly parents who comprised the majority of the gaming specificities class, are more prone to share family history, or that the helpline operatives asked more often about the familial context when the user was a gamer. Communication difficulties with parents were also widely reported. Parents' anxiety and own depression, previously reported to be highly correlated to gaming disorder in children [37], could contribute to these particularities and especially to communication difficulties.

Relatives of gambling users frequently complained about financial difficulties, grouped in class 6, that often become their shared burden. This result supports previous findings that financial difficulties and familial conflicts were key help-seeking precipitators among gamblers [38]. In contrast, specific help-seeking precipitators in gamers were externalized behavior problems and drop in grades, reported in class 1. Externalized behavior problems have been previously linked to cumulative childhood trauma [39]. A future exploration of this link in gaming disorder could reinforce therapeutic strategies and highlight prevention.

Our study has several limitations. The imbalance between gaming and gambling and the proportion of relatives in the respective subsamples could have influenced the hierarchical classification; however, both gaming and gambling samples were very large compared with those reported in classical qualitative studies. Some contactors could have contacted the help service several times, although the anonymity of the facility prevented documentation of repeat contacts. Moreover, no structured diagnostic or clinical assessments were collected during the calls, and we cannot assume that all contacts regarded people with a clinical addictive disorder. We can only assume that the contactors felt the need to seek help regarding gambling or gaming.

In conclusion, negative emotions and shared distress linked to gambling and gaming support their classification as mental disorders in se, but meaningful differences were observed in core symptoms of addiction between gamers and gamblers, beyond specificities related to the behavior itself. Our findings support differences in the development and maintenance of gaming and gambling disorder and should been further explored in a process-oriented approach.

Acknowledgments

We are grateful to Sarah Kabani for English editing, and to Addictions Drogues Alcool Info Service (ADALIS) for having shared the data. No financial support was received for this study.

Authors' Contributions

All authors were responsible for the study concept and design, analysis and interpretation of data, and statistical analysis. CH contributed to analysis and interpretation of data, and writing of the manuscript. AB and PP contributed to interpretation of data and review of the manuscript. All authors had full access to all data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Conflicts of Interest

None declared.

Multimedia Appendix 1

χ^2 values (>100, all <italic>P</italic><.001) between forms and the classes, and between passive independent variables and the classes.

[DOCX File , 34 KB - mental_v8i11e26521_app1.docx]

References

1. CD-11 for Mortality and Morbidity Statistics. Disorders due to addictive behaviours. World Health Organization. URL: https://icd.who.int/browse11/l-m/en#/http%3a%2f%2fid.who.int%2ficd%2fentity%2f499894965 [accessed 2018-01-01]
2. Brand M, Wegmann E, Stark R, Müller A, Wölfling K, Robbins TW, et al. The Interaction of Person-Affect-Cognition-Execution (I-PACE) model for addictive behaviors: update, generalization to addictive behaviors beyond internet-use disorders, and specification of the process character of addictive behaviors. Neurosci Biobehav Rev 2019 Sep;104:1-10 [FREE Full text] [doi: 10.1016/j.neubiorev.2019.06.032] [Medline: 31247240]
3. Bourdieu P. La Distinction: Critique sociale du jugement. France: Les Éditions de Minuit; Jan 01, 1983:103.
4. Stein DJ, Billieux J, Bowden-Jones H, Grant JE, Fineberg N, Higuchi S, et al. Balancing validity, utility and public health considerations in disorders due to addictive behaviours. World Psychiatry 2018 Oct 07;17(3):363-364. [doi: 10.1002/wps.20570] [Medline: 30192089]
5. Billieux J, Flayelle M, Rumpf H, Stein DJ. High involvement versus pathological involvement in video games: a crucial distinction for ensuring the validity and utility of gaming disorder. Curr Addict Rep 2019 May 23;6(3):323-330. [doi: 10.1007/s40429-019-00259-x]
6. Maroney N, Williams BJ, Thomas A, Skues J, Moulding R. A stress-coping model of problem online video game use. Int J Ment Health Addiction 2018 Apr 16;17(4):845-858. [doi: 10.1007/s11469-018-9887-7]
7. Kim HS, Hodgins DC, Kim B, Wild TC. Transdiagnostic or disorder specific? Indicators of substance and behavioral addictions nominated by people with lived experience. J Clin Med 2020 Jan 24;9(2):334 [FREE Full text] [doi: 10.3390/jcm9020334] [Medline: 31991652]
8. Richard J, Fletcher É, Boutin S, Derevensky J, Temcheff C. Conduct problems and depressive symptoms in association with problem gambling and gaming: A systematic review. J Behav Addict 2020 Oct 12;9(3):497-533. [doi: 10.1556/2006.2020.00045] [Medline: 32750033]
9. Wölfling K, Duven E, Wejbera M, Beutel M, Müller KW. Discounting delayed monetary rewards and decision making in behavioral addictions - A comparison between patients with gambling disorder and internet gaming disorder. Addict Behav 2020 Sep;108:106446. [doi: 10.1016/j.addbeh.2020.106446] [Medline: 32408114]
10. Bae S, Hong JS, Kim SM, Han DH. Bupropion shows different effects on brain functional connectivity in patients with internet-based gambling disorder and internet gaming disorder. Front Psychiatry 2018 Apr 10;9:130. [doi: 10.3389/fpsyt.2018.00130] [Medline: 29692743]
11. Black DW, Shaw M, Coryell W, Crowe R, McCormick B, Allen J. Age at onset of DSM-IV pathological gambling in a non-treatment sample: Early- versus later-onset. Compr Psychiatry 2015 Jul;60:40-46 [FREE Full text] [doi: 10.1016/j.comppsych.2015.04.007] [Medline: 25956751]
12. King DL, Delfabbro PH. The cognitive psychopathology of internet gaming disorder in adolescence. J Abnorm Child Psychol 2016 Nov 15;44(8):1635-1645. [doi: 10.1007/s10802-016-0135-y] [Medline: 26875565]
13. Griffiths M. A 'components' model of addiction within a biopsychosocial framework. J Subst Use 2009 Jul 12;10(4):191-197. [doi: 10.1080/14659890500114359]
14. Konkolÿ Thege B, Colman I, el-Guebaly N, Hodgins DC, Patten SB, Schopflocher D, et al. Social judgments of behavioral versus substance-related addictions: a population-based study. Addict Behav 2015 Mar;42:24-31. [doi: 10.1016/j.addbeh.2014.10.025] [Medline: 25462650]
15. Inclusion of "gaming disorder" in ICD-11. World Health Organization. 2018 Sep 14. URL: https://www.who.int/news/item/14-09-2018-inclusion-of-gaming-disorder-in-icd-11 [accessed 2021-11-14]
16. Colder Carras M, Carras M, Labrique AB. Stakeholders' consensus on strategies for self- and other-regulation of video game play: a mixed methods study. Int J Environ Res Public Health 2020 May 28;17(11):3846 [FREE Full text] [doi: 10.3390/ijerph17113846] [Medline: 32481721]
17. Colder Carras M, Porter AM, Van Rooij AJ, King D, Lange A, Carras M, et al. Gamers' insights into the phenomenology of normal gaming and game "addiction": A mixed methods study. Comput Human Behav 2018 Feb;79:238-246 [FREE Full text] [doi: 10.1016/j.chb.2017.10.029] [Medline: 29456287]
18. Billieux J, Schimmenti A, Khazaal Y, Maurage P, Heeren A. Are we overpathologizing everyday life? A tenable blueprint for behavioral addiction research. J Behav Addict 2015 Sep;4(3):119-123 [FREE Full text] [doi: 10.1556/2006.4.2015.009] [Medline: 26014667]
19. Costes J. Le point sur .. La demande d'aide et le recours aux soins des joueurs problematiques en France. Observatoires des jeux. 2019 Feb 01. URL: https://www.ofdt.fr/odj/Le_point_sur._Ndeg2.pdf [accessed 2021-11-14]
20. van den Brink W. ICD-11 gaming disorder: needed and just in time or dangerous and much too early? J Behav Addict 2017 Sep 01;6(3):290-292 [FREE Full text] [doi: 10.1556/2006.6.2017.040] [Medline: 28816496]
21. Bastiani L, Fea M, Potente R, Luppi C, Lucchini F, Molinaro S. National helpline for problem gambling: a profile of its users' characteristics. J Addict 2015;2015:659731. [doi: 10.1155/2015/659731] [Medline: 26064772]
22. Darbeda S, Aubin H, Lejoyeux M, Luquiens A. Characteristics of gamblers who use the French national problem gambling helpline and real-time chat facility: longitudinal observational study. JMIR Form Res 2020 May 20;4(5):e13388 [FREE Full text] [doi: 10.2196/13388] [Medline: 32432554]
23. Carr MM, Ellis JD, Ledgerwood DM. Suicidality among gambling helpline callers: a consideration of the role of financial stress and conflict. Am J Addict 2018 Sep 16;27(6):531-537. [doi: 10.1111/ajad.12787] [Medline: 30113105]

24. Sullivan SG, McCormick R, Sellman JD. Increased requests for help by problem gamblers: data from a gambling crisis telephone hotline. N Z Med J 1997 Oct 10;110(1053):380-383. [Medline: 9364187]
25. King DL, Adair C, Saunders JB, Delfabbro PH. Clinical predictors of gaming abstinence in help-seeking adult problematic gamers. Psychiatry Res 2018 Mar;261:581-588. [doi: 10.1016/j.psychres.2018.01.008] [Medline: 29407726]
26. Ratinaud P, Déjean S. IRaMuTeQ: Implémentation de la méthode ALCESTE d'analyse de texte dans un logiciel libre "Iramuteq". 2009. URL: http://repere.no-ip.org/repere/Members/pratinaud/mes-documents/articles-et-presentations/presentation_mashs2009.pdf.1 [accessed 2009-01-01]
27. Tichit A, Mathonnat C, Landivar D. Classifying non-banking monetary systems using web data. Haute Autorité de Santé - archives ouvertes. 2015. URL: https://halshs.archives-ouvertes.fr/halshs-01230219/document [accessed 2015-11-08]
28. Manchaiah V, Amlani AM, Bricker CM, Whitfield CT, Ratinaud P. Benefits and shortcomings of direct-to-consumer hearing devices: analysis of large secondary data generated from Amazon customer reviews. J Speech Lang Hear Res 2019 May 21;62(5):1506-1516. [doi: 10.1044/2018_jslhr-h-18-0370]
29. DSM-5 Fact Sheets: Substance related and addictive disorders. American Psychiatric Association. URL: https://www.psychiatry.org/psychiatrists/practice/dsm/educational-resources/dsm-5-fact-sheets [accessed 2021-11-14]
30. Snodgrass JG, Zhao W, Lacy MG, Zhang S, Tate R. The cross-cultural expression of internet gaming distress in North America, Europe, and China. Addict Behav Rep 2019 Jun;9:100146 [FREE Full text] [doi: 10.1016/j.abrep.2018.100146] [Medline: 31193753]
31. Bonfils NA, Grall-Bronnec M, Caillon J, Limosin F, Benyamina A, Aubin H, et al. Giving room to subjectivity in understanding and assessing problem gambling: A patient-centered approach focused on quality of life. J Behav Addict 2019 Mar 01;8(1):103-113 [FREE Full text] [doi: 10.1556/2006.7.2018.137] [Medline: 30663328]
32. Beranuy M, Carbonell X, Griffiths MD. A qualitative analysis of online gaming addicts in treatment. Int J Ment Health Addiction 2012 Oct 27;11(2):149-161. [doi: 10.1007/s11469-012-9405-2]
33. Guillou-Landreat M, Guilleux A, Sauvaget A, Brisson L, Leboucher J, Remaud M, et al. Factors associated with suicidal risk among a French cohort of problem gamblers seeking treatment. Psychiatry Res 2016 Jun 30;240:11-18 [FREE Full text] [doi: 10.1016/j.psychres.2016.04.008] [Medline: 27078754]
34. Throuvala MA, Janikian M, Griffiths MD, Rennoldson M, Kuss DJ. The role of family and personality traits in Internet gaming disorder: A mediation model combining cognitive and attachment perspectives. J Behav Addict 2019 Mar 01;8(1):48-62 [FREE Full text] [doi: 10.1556/2006.8.2019.05] [Medline: 30739463]
35. Kircaburun K, Griffiths MD, Billieux J. Psychosocial factors mediating the relationship between childhood emotional trauma and internet gaming disorder: a pilot study. Eur J Psychotraumatol 2019 Jan 14;10(1):1565031 [FREE Full text] [doi: 10.1080/20008198.2018.1565031] [Medline: 30693081]
36. Kardefelt-Winther D. Conceptualizing internet use disorders: addiction or coping process? Psychiatry Clin Neurosci 2017 Jul 02;71(7):459-466. [doi: 10.1111/pcn.12413] [Medline: 27278653]
37. Wartberg L, Kriston L, Kramer M, Schwedler A, Lincoln T, Kammerl R. Internet gaming disorder in early adolescence: Associations with parental and adolescent mental health. Eur Psychiatry 2017 Jun 23;43:14-18. [doi: 10.1016/j.eurpsy.2016.12.013] [Medline: 28365463]
38. Valdivia-Salas S, Blanchard KS, Lombas AS, Wulfert E. Treatment-seeking precipitators in problem gambling: analysis of data from a gambling helpline. Psychol Addict Behav 2014 Mar;28(1):300-306. [doi: 10.1037/a0035413] [Medline: 24731119]
39. Hébert M, Langevin R, Oussaïd E. Cumulative childhood trauma, emotion regulation, dissociation, and behavior problems in school-aged sexual abuse victims. J Affect Disord 2018 Jan 01;225:306-312 [FREE Full text] [doi: 10.1016/j.jad.2017.08.044] [Medline: 28843081]

Abbreviations

ICD: International Classification of Diseases

Edited by J Torous; submitted 15.12.20; peer-reviewed by D Hodgins; comments to author 29.04.21; revised version received 22.06.21; accepted 23.08.21; published 26.11.21.

Please cite as:
Luquiens A, von Hammerstein C, Benyamina A, Perney P
Burden and Help-Seeking Behaviors Linked to Problem Gambling and Gaming: Observational Quantitative and Qualitative Analysis
JMIR Ment Health 2021;8(11):e26521
URL: https://mental.jmir.org/2021/11/e26521
doi:10.2196/26521
PMID:34842562

Original Paper

A Text Messaging Intervention (StayWell at Home) to Counteract Depression and Anxiety During COVID-19 Social Distancing: Pre-Post Study

Adrian Aguilera[1,2], PhD; Rosa Hernandez-Ramos[1], BA; Alein Y Haro-Ramos[3], MPH; Claire Elizabeth Boone[3], MPH; Tiffany Christina Luo[1], MSW; Jing Xu[4,5], PhD; Bibhas Chakraborty[5,6,7], PhD; Chris Karr[8], BSc, MSc; Sabrina Darrow[2], PhD; Caroline Astrid Figueroa[1], MD, PhD

[1]School of Social Welfare, University of California, Berkeley, Berkeley, CA, United States

[2]Department of Psychiatry and Behavioral Sciences, University of California San Francisco, San Francisco, CA, United States

[3]School of Public Health, University of California, Berkeley, Berkeley, CA, United States

[4]Data Science Programme, Division of Science and Technology, Beijing Normal University-Hong Kong Baptist University United International College, Zhuhai, China

[5]Center for Quantitative Medicine, Duke National University of Singapore, Singapore, Singapore

[6]Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, United States

[7]Department of Statistics and Applied Probability, National University of Singapore, Singapore, Singapore

[8]Audacious Software, Chicago, IL, United States

Corresponding Author:
Adrian Aguilera, PhD
School of Social Welfare
University of California, Berkeley
120 Haviland Hall, MC7400
School of Social Welfare
Berkeley, CA, 94720
United States
Phone: 1 (510) 642 8564
Email: aguila@berkeley.edu

Abstract

Background: Social distancing and stay-at-home orders are critical interventions to slow down person-to-person transmission of COVID-19. While these societal changes help contain the pandemic, they also have unintended negative consequences, including anxiety and depression. We developed StayWell, a daily skills-based SMS text messaging program, to mitigate COVID-19–related depression and anxiety symptoms among people who speak English and Spanish in the United States.

Objective: This paper describes the changes in StayWell participants' anxiety and depression levels after 60 days of exposure to skills-based SMS text messages.

Methods: We used self-administered, empirically supported web-based questionnaires to assess the demographic and clinical characteristics of StayWell participants. Anxiety and depression were measured using the 2-item Generalized Anxiety Disorder (GAD-2) scale and the 8-item Patient Health Questionnaire-8 (PHQ-8) scale at baseline and 60-day timepoints. We used 2-tailed paired *t* tests to detect changes in PHQ-8 and GAD-2 scores from baseline to follow-up measured 60 days later.

Results: The analytic sample includes 193 participants who completed both the baseline and 60-day exit questionnaires. At the 60-day time point, there were significant reductions in both PHQ-8 and GAD-2 scores from baseline. We found an average reduction of –1.72 (95% CI –2.35 to –1.09) in PHQ-8 scores and –0.48 (95% CI –0.71 to –0.25) in GAD-2 scores. These improvements translated to an 18.5% and 17.2% reduction in mean PHQ-8 and GAD-2 scores, respectively.

Conclusions: StayWell is an accessible, low-intensity population-level mental health intervention. Participation in StayWell focused on COVID-19 mental health coping skills and was related to improved depression and anxiety symptoms. In addition to improvements in outcomes, we found high levels of engagement during the 60-day intervention period. Text messaging interventions could serve as an important public health tool for disseminating strategies to manage mental health.

Trial Registration: ClinicalTrials.gov NCT04473599; https://clinicaltrials.gov/ct2/show/NCT04473599

International Registered Report Identifier (IRRID): RR2-10.2196/23592

(JMIR Ment Health 2021;8(11):e25298) doi:10.2196/25298

KEYWORDS

mobile health; COVID-19; text messaging; cognitive behavioral therapy; anxiety; depression; microrandomized trials; mHealth; intervention; mental health; SMS

Introduction

The COVID-19 pandemic is a significant public health crisis that has caused devastating physical illness and concurrent mental health challenges [1]. Public health measures, including stay-at-home orders and the closure of nonessential businesses, have been necessary to reduce transmission but have also disrupted social life by limiting social activities and physical interactions with one's networks [2,3].

Societal changes to contain the COVID-19 pandemic have caused significant psychological distress worldwide in people of various backgrounds [1,4-6]. Studies show lowered psychological well-being and increased depressive and anxiety symptoms among the general public compared to prepandemic rates [1]. In the United States, the risk of depression among adults increased 3-fold during the COVID-19 pandemic compared to before the pandemic [7]. Stressors associated with social distancing and loss of usual routines, including infection fear, financial insecurity, frustration, and a sense of isolation, had negative psychological impacts, including increased depression and anxiety symptoms [4]. These stressors also increased insomnia [8], decreased physical activity [9], and increased alcohol and substance use [10] in diverse global samples.

While the prevalence of anxiety and depression has increased in the general population, certain groups are at a higher risk of mental health disorders. Individuals with a greater risk for depression during the pandemic include those from lower socioeconomic backgrounds with insufficient economic resources, inadequate social support, and greater exposure to social stressors, such as pandemic-related job loss [7]. Furthermore, the pandemic has disproportionately affected the health of already at-risk individuals such as those from low-income backgrounds, communities of color, and non–English-speaking groups [11].

Text messaging is a promising tool to deliver interventions that address the detrimental mental health effects of the COVID-19 pandemic [12-15], especially for underserved populations [16,17]. Texting, often viewed as the "workhorse" of digital and mobile health, is a widely used communication strategy that has been leveraged to deliver mental health interventions by relaying health information, skills-based messages, and self-monitoring messages to participants [16]. Text messaging interventions can help fill the gap between the need for and availability of mental health services, including behavioral health appointments, a gap that worsened during the pandemic [18]. Because 85% of all US adults and 76% of lower income adults own a smartphone, texting interventions have the capacity to reach a large and diverse group of people [19]. Expanding the reach of mental health programs is especially crucial since depression and anxiety symptoms have increased in the general population and intensified among those with existing mental health disorders and vulnerable communities.

Incorporating cognitive behavioral therapy (CBT) and skills-based text messaging interventions have proved feasible and acceptable among a diverse group of patients with affective disorders [20]. A significant body of research has established the effectiveness of CBT as an evidence-based, first-line treatment for mental health conditions such as depression and anxiety [21]. CBT has been implemented in diverse populations, including communities of color and individuals of lower socioeconomic status. Additionally, CBT is a focused, directive, and structured form of psychotherapy, making it well suited for delivery via digital platforms such as text messaging. Electronically delivered CBT is at least as effective as face-to-face CBT at reducing mental health symptoms [22].

In April 2020, we initiated the StayWell at Home intervention (ClinicalTrials.gov, NCT04473599), a 60-day skills-based daily text messaging program in accordance with principles of CBT [23]. This paper assesses the effects of StayWell on symptoms of depression and anxiety in a broad adult population living in the United States during the COVID-19 pandemic. Texts were based on two core components of CBT: behavioral activation (BA) and psychoeducation. BA aims to help people engage in enjoyable activities, reduce reliance on unhealthy coping mechanisms, and decrease avoidance of anxiety-provoking situations. By directing individuals to pleasurable and meaningful activities, BA can improve mood and decrease loneliness and isolation related to the pandemic. Providing psychoeducation around thoughts, feelings, and behaviors is also an important part of CBT. Messages focused on promoting adaptive cognitive approaches to pandemic-related stress and encouraged BA within the limits of social distancing. Maladaptive thoughts and behaviors can be identified and replaced to reduce the frequency and intensity of negative emotions. Further, information and reminders related to self-care, sleep, physical activity, and mindfulness may promote positive health behavior change and have beneficial effects on individuals' psychological well-being. We hypothesize that participants in the intervention will report fewer depression and anxiety symptoms at the end of the intervention.

Methods

StayWell Trial Design

StayWell is a fully remote trial and has various designs: (1) a pre-post comparison, in which we assessed depression and anxiety symptoms for all patients before and after the intervention and (2) a randomized controlled trial with two

groups: Uniform Random (UR) messaging and a Reinforcement Learning (RL) messaging. Owing to a coding error in the algorithm, all participants received messages randomly (UR condition). Therefore, we altered the study's main aims by focusing on the pre-post effects of participating in StayWell on depression and anxiety symptoms. The institutional review board of the University of California (UC) Berkeley reviewed and approved all study procedures.

Participants were enrolled in the StayWell trial using the HealthySMS platform—an automated text messaging platform developed by the authors [20,24]. HealthySMS has been successfully used with various low-income English- and Spanish-speaking populations to send automated text messages and manage participant responses using a secure, Health Insurance Portability and Accountability Act–compliant platform [20]. Participants received 2 messages daily for 60 days: 1 skills-based message and 1 message inquiring about their mood. The skills-based messages included tips on how to deal with worry and stress brought on by the COVID-19 pandemic. Half of these messages were based on BA strategies, and half were based on skill-based strategies. We developed the messages to highlight evidence-based practices used in depression and anxiety interventions that promote behavior change. Templates for these messages were developed from previous work conducted by authors SMD and AA [16]. The study team edited the original messages to fit the COVID-19 context and improve readability. The team also translated and culturally adapted messages in Spanish to expand the reach of the intervention. Messages were sent daily at a random time between 9 AM and 6 PM. Three hours following the delivery of the skills-based messages, participants were sent a message inquiring about their mood on a scale of 1-9, with 9 being the best mood.

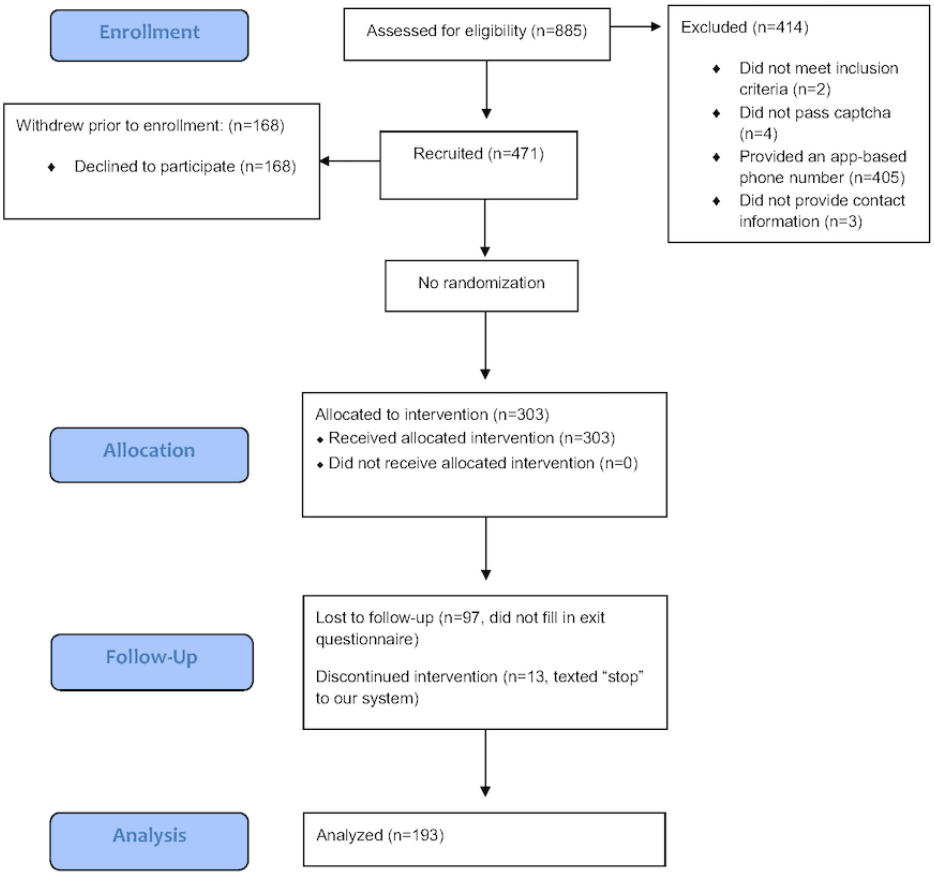
UR Message Arm

Participants received messages uniform randomly (ie, a microrandomized trial design [16]), where every day during the study treatment allocation was characterized by a full factorial design with two factors: skills-based messages (M) and the time frame (T) when the message was sent. M has 2 levels (BA and skill-based), and T has 3 levels (9 AM-12 PM, 12 PM-3 PM, and 3 PM-6 PM). Participants received 1 daily skills-based message and 1 daily mood message that did not vary.

Data Collection

Adult Spanish and English speakers aged 18 years and older who had a mobile phone were recruited via web-based media advertisements on Facebook, Craigslist, and university websites (UC Berkeley and UC San Francisco) to participate. Data were collected through a web-based Qualtrics survey and HealthySMS. Participants were excluded if they used a web-based text messaging app or were outside of the United States. Web-based text messaging apps are more prone to scams on the internet and facilitate the creation of multiple phone numbers for 1 participant. The study lasted from April to December 2020. The CONSORT (Consolidated Standards for Reporting Trials) flow diagram for data collection is shown in Figure 1.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram of the data collection process.



Recruitment

We designed web-based advertisements to target vulnerable populations, including low-income groups and people of color, who are disproportionately impacted by the COVID-19 pandemic in the United States. Using a user-centered design, we created 16 user personas. A persona is a fictional characterization of a user, which includes specific characteristics and demographics found in the target population [25]. The personas informed the title, picture, and reason for participating in the study used on each advertisement and the characteristics used for detailed targeted advertising, which is available on Facebook advertisements.

Enrollment

To prevent web-based scams and fraud, interested subjects were sent to a Qualtrics survey to assess eligibility criteria and human identity through a built-in *captcha*. Eligible participants were then sent a unique weblink to a baseline assessment. Using a different Qualtrics survey, participants consented and answered demographic questions and other measures of interest. Upon completion, participants were enrolled in an automated text messaging intervention for 60 days. On day 61, participants were sent an exit assessment where they were asked the measures of interests originally asked at baseline. Participants were paid US $20 at the end of the study for completing study questionnaires.

Outcome Measures

Our primary outcomes, including the 8-item Patient Health Questionnaire 8 [26] and the 2-item Generalized Anxiety Disorder 2 [27] (GAD-2) scale, were collected through a Qualtrics survey at pre- and postintervention.

Secondarily, we were interested in assessing engagement in the intervention by measuring response rates to mood rating messages and calculating how many participants stopped the text messaging.

Hypotheses

We will conduct a pre-post comparison among all participants. The depression score measured using the 8-item Patient Health Questionnaire (PHQ-8) and the anxiety score measured using the 2-item Generalized Anxiety Disorder (GAD-2) scale will be improved over the 60-day study.

Power Analysis

The sample size calculation was performed in a previous protocol paper [23], which includes 2 aims. This paper only considers the primary aim. At a medium standardized effect size (ie, Cohen d=0.5), a sample size of 64 participants is required to detect an improvement of either the depression or anxiety score from baseline to 60-day at 80% power and 5% level of significance. The sample size of this study was 193 participants, which is based on the secondary aim. The secondary outcome focuses on the proximal effect of daily improvement on the mood rating. However, the analysis will be presented in a separate manuscript.

Statistical Analysis

Main Analyses

To detect the change in depression (PHQ-8) and anxiety (GAD-2) scores from baseline to follow-up measured 60 days later, we used 2-tailed paired t tests. The normality assumptions for the change in each score are validated by their corresponding histogram plots, which are relatively symmetric. The goodness of fit to normal distribution was validated using the Anderson-Darling test, while skewness for normality was validated using the Shapiro-Wilk skewness test [28].

Exploratory Analyses

We used simple linear regression analysis to model the change in PHQ-8 and GAD-2 scores as a function of participants' response rates (ie, the proportion of mood messages answered) to determine whether the improvements in depression and anxiety are predicted by engagement with the intervention. Furthermore, to determine whether any other covariates predict the effects of the StayWell intervention, we used multivariable linear regression analysis to model the change in PHQ-8 and GAD-2 scores (ie, scores at follow-up minus scores at baseline) as a function of demographic predictors, response rates, self-rated health, and the change in COVID-19 weekly rolling average case rates in each participants' county of residence. Demographic predictors include education (at least high school, some college, college, or graduate degree), age, gender (female, male, or other), language (English or Spanish), and employment (full-time, part-time, unemployed, or other). Self-rated health includes poor/fair, good, very good, and excellent. The weekly rolling average of daily new confirmed cases per 100,000 population is calculated for the day in which participants enroll and exit the program by averaging the values of that day, 3 days prior, and 3 days thereafter. We then used the change in COVID-19 case rates at 60-day follow-up from the baseline date for each participant.

Results

Results Overview

A total of 303 people entered the study and received text messages. Of these, 12 were recruited via ClinicalTrials.gov, 75 via Craigslist, 184 via Facebook, and 32 by texting the StayWell phone number. We show the distribution of baseline characteristics in Table 1. Most baseline respondents were female (76.0%) and spoke English (88.4%). While almost half of the respondents identified as White or Caucasian (47.9%), the sample was relatively diverse with 20.5% Latinx, 13.2% Asian or Pacific Islander, 11.5% multiethnic, and 6.6% Black or African American participants. Of the baseline participants, 193 also completed an exit questionnaire and were included in the main analysis.

Table 1. Baseline demographic and clinical characteristics (N=303).

Characteristics	Value
Age (years), mean (SD)	33.3 (11.0)
Females, n (%)	230 (76.0)
Language, n (%)	
English	268 (88.4)
Spanish	35 (11.3)
Employment, n (%)	
Full-time (greater than or equal to 35 hours/week)	137 (45.2)
Part-time (less than 35 hours)	61 (20.1)
Homemaker	28 (9.2)
Unemployed	50 (16.5)
Disabled/on disability	12 (4.0)
Retired	1 (0.3)
Other	14 (4.6)
Race/ethnicity, n (%)	
Asian or Pacific Islander	40 (13.2)
Black or African American	20 (6.6)
White or Caucasian	145 (47.9)
Latino(a) or Hispanic	62 (20.5)
Multiethnic	35 (11.5)
Unknown	1 (0.3)
Education, n (%)	
Between 6th and 8th grade	1 (0.3)
Some high school	10 (3.3)
High school graduate	29 (9.6)
Some college or technical school	94 (31.0)
College graduate	103 (34.0)
Graduate degree	66 (21.8)
Paying for basics (eg, food, housing, medical care, and heating) is, n (%)	
Very hard	119 (39.3)
Sometimes hard	137 (45.2)
Not hard at all	47 (15.5)
Self-reported health, n (%)	
Excellent	8 (2.6)
Very good	41 (13.5)
Good	89 (29.4)
Fair	103 (34.0)
Poor	61 (20.1)
Psychological outcomes, mean (SD)	
Depression (PHQ-8[a])	9.41 (5.79)
Anxiety (GAD-2[b])	2.71 (1.87)
Impact of COVID-19 (1=completely disagree, 5=completely agree), mean (SD)	

Characteristics	Value
I feel lonelier	3.55 (1.17)
I am running into financial issues	3.20 (1.36)
I feel more stressed	4.07 (0.97)
I feel more anxious	3.89 (1.09)

[a]PHQ-8: 8-item Patient Health Questionnaire.

[b]GAD-2: 2-item Generalized Anxiety Disorder.

Table 2 displays the raw scores and the distributions of change in depression (PHQ-8) and anxiety (GAD-2) scores at 60 days from baseline for respondents who completed the baseline and exit surveys, respectively. The data in Table 2 indicate that average PHQ-8 and GAD-2 scores decreased significantly from baseline to the end of the study, suggesting improvements in depression and anxiety symptoms. There was a reduction in the mean PHQ-8 and GAD-2 scores of 18.5% and 17.2%, respectively, at 60 days compared to the baseline scores. The normality assumption of each score change is valid.

To evaluate the generalizability of our data in terms of anxiety and depression prevalence and symptoms in our baseline samples, we compared the clinical parameters between participants who only responded to the baseline survey versus those who responded to both the baseline and 60-day surveys (Tables 3 and 4). Likely major depressive disorder and likely GAD were assessed using cutoff scores of ≥10 on the PHQ-8 and ≥3 on the GAD-2, respectively. There was no significant difference (for all, $P>.05$) in clinical parameters between people who only responded to the baseline survey and those who responded to the baseline and 60-day assessment. This suggests that the mental health burden was similar between our study sample and individuals who did not complete the 60-day survey.

Table 2. Changes in the 8-item Patient Health Questionnaire and 2-item Generalized Anxiety Disorder scale scores for individuals who completed both the baseline and 60-day assessment (n=193).

Measure	Scores				*P* value
	Baseline score, mean (SD)	60-day score, mean (SD)	Change from baseline, %	Mean difference (95% CI)	
8-item Patient Health Questionnaire	9.30 (5.70)	7.58 (5.27)	18.50	–1.72 (–2.35 to –1.09)	<.001
2-item Generalized Anxiety Disorder scale	2.80 (1.89)	2.32 (1.83)	17.20	–0.48 (–0.71 to –0.25)	<.001

Table 3. Comparison of the prevalence rates of the risk for generalized anxiety disorder and likely major depressive disorder between subscribers who only completed the baseline survey and those who completed both the baseline and 60-day surveys.

Condition	Prevalence at baseline, n/total responses (%)		Chi-square (*df*)	*P* value
	Subscribers who completed the baseline assessment but not the 60-day assessment (n=303)	Subscribers who completed both the baseline and 60-day assessments (n=193)		
Likely major depressive disorder (8-item Patient Health Questionnaire score ≥ 10)	137/303 (45.21)	89/193 (46.11)	0.039 (1)	.84
At risk for generalized anxiety disorder (2-item Generalized Anxiety Disorder scale score ≥ 3)	135/303 (44.55)	88/193 (45.60)	0.052 (1)	.82

Table 4. Comparison of the mean scores on the 2-item Generalized Anxiety Disorder scale and the 8-item Patient Health Questionnaire between participants who only completed the baseline survey and subscribers who completed both the baseline and 60-day surveys.

Scale	Score at baseline, mean (SD)		Independent samples *t* test (*df*)	*P* value
	Subscribers who completed the baseline assessment but not the 60-day assessment (n=303)	Subscribers who completed both the baseline and 60-day assessments (n=193)		
8-item Patient Health Questionnaire	9.41 (5.79)	9.30 (5.70)	0.208 (414)	0.84
2-item Generalized Anxiety Disorder scale	2.71 (1.87)	2.80 (1.89)	0.520 (406)	0.60

To determine whether the changes in PHQ-8 and GAD-2 scores remain constant after accounting for participants' engagement in the intervention, we used a simple linear regression model (Table 5) with participants' fraction of mood messages answered (ie, response rates) as the main predictor. The outcome was the change in GAD-2 and PHQ-8 scores at 60 days from the baseline; thus, a negative coefficient indicates a greater improvement (larger decrease in scores) in anxiety and depressive symptoms than the reference group. We found that even when accounting for engagement in StayWell, the average improvements in both PHQ-8 and GAD-2 scores remain significant. The average improvements in PHQ-8 and GAD-2 scores controlling for engagement are –2.7 points (*P*=.001) and –0.78 points (*P*=.01), respectively.

To assess the influence of other factors (COVID-19 infection rates, self-rated health, and other demographic variables) on GAD-2 and PHQ-8 score improvements, we conducted a post hoc exploratory analysis (Table 6). This analysis also adjusts for response rates. Eight individuals lacked a valid zip code and were excluded from the analysis. Compared to females, the change in PHQ-8 score at 60 days from baseline was 2.4 points larger (*P*=.01) among males, adjusting for all other covariates; this suggests that males experienced relatively lesser improvement in depression symptoms. Having very good self-rated health was associated with less improvement in anxiety symptoms (increase of 0.83 in the GAD-2 score, *P*=.04) at 60 days from baseline compared to those with poor health and adjusting for covariates.

As a sensitivity analysis, we conducted 2 separate 1-way analysis of variance to assess the differences in the average change in PHQ-8 scores between genders and explore the association between self-rated health and the average change in GAD-2 scores. Our results show that gender had a significant effect on the change in PHQ-8 scores at 60 days from baseline ($F_{2,190}$=4.106; *P*=.02). This suggests that there are true differences in the average improvement in PHQ-8 scores among male- and female-identifying participants. However, the mean change in GAD-2 scores did not differ between self-rated health categories ($F_{3,189}$=2.954; *P*=.37). Thus, we cannot conclude that there are differences in the average improvement in GAD-2 scores by self-rated health.

The results in Table 2 show that both the depression and anxiety scores decreased over the 60-day period. Table 5 shows that the changes in both scores remained significant, adjusting for the response rate, as determined using a simple linear regression model (PHQ-8, *P*=.01 and GAD-2, *P*=.50). We also found that participants' response rates are not significant predictors of improved PHQ-8 and GAD-2 scores (Table 5 and Table 6). In Table 6, we adjusted for the response rate and other demographic and clinical variables using a multivariable linear regression model, and we observed that the change in depression scores remained significant, but the change in anxiety score was not significant. However, we had previously modeled the change in PHQ-8 and GAD-2 scores by only adjusting for demographic and clinical variables, and the change in GAD-2 scores was also not significant (Multimedia Appendix 1). Therefore, the nonsignificant results for the change in GAD-2 scores are not necessarily attributed to the response rates.

Table 5. Simple linear regression model: responding predictor of the changes in the 8-item Patient Health Questionnaire and 2-item Generalized Anxiety Disorder scores at the 60-day exit from baseline for participants who completed both surveys.

Change in scores	Intercept, coefficient (95% CI)	Response rate, % (95% CI)	Observations, n	Adjusted R^2	SE (*df*)	*F* test (*df*)
8-item Patient Health Questionnaire	–2.7 (–4.3 to –1.1)[a]	0.01 (–0.01 to 0.04)[b]	193	0.003	4.42 (191)	1.66 (1,191)
2-item Generalized Anxiety Disorder Scale	–0.78 (–1.40 to –0.16)[c]	0.00 (0.00 to 0.01)[d]	193	0.0002	1.67 (191)	1.03 (1,191)

[a]*P*=.001.

[b]*P*=.20.

[c]*P*=.01.

[d]*P*=.30.

Table 6. Multivariable linear regression models: demographic, clinical, and engagement predictors of the changes in score in the 8-item Patient Health Questionnaire and 2-item Generalized Anxiety Disorder scale at the 60-day exit from baseline for participants who completed both surveys.

Characteristics	Change in the 8-item Patient Health Questionnaire score[a]		Change in the 2-item Generalized Anxiety Disorder scale score[b]	
	Coefficient (95% CI)	*P* value	Coefficient (95% CI)	*P* value
Intercept	–4.84 (–8.4 to –1.3)	.01	–0.46 (–1.8 to 0.87)	.50
Weekly COVID-19 case rates per 100,000 population	–0.01 (–0.04 to 0.02)	.60	0.00 (–0.01 to 0.01)	.80
Education				
At least high school	—[c]	—	—	—
Some college	0.68 (–1.60 to 2.90)	.60	–0.62 (–1.5 to 0.23)	.20
College	1.60 (–0.65 to 3.80)	.20	–0.63 (–1.5 to 0.22)	.15
Graduate school	1.20 (–1.3 to 3.70)	.30	–0.93 (–1.9 to 0.02)	.06
Self-rated health				
Poor/fair	—	—	—	—
Good	1.30 (–0.92 to 3.50)	.30	0.69 (–0.15 to 1.50)	.11
Very good	2.00 (–0.11 to 4.10)	.06	0.83 (0.04 to 1.60)	.04
Excellent	1.40 (–1.10 to 3.90)	.30	0.42 (–0.52 to 1.40)	.40
Age	–0.01 (–0.07 to 0.05)	.80	–0.02 (–0.04 to 0.01)	.20
Gender				
Female	—	—	—	—
Male	2.40 (0.52 to 4.20)	.01	0.45 (–0.24 to 1.10)	.20
Other	–2.80 (–8.00 to 2.40)	.30	0.08 (–1.90 to 2.00)	>.90
Employment				
Full-time	—	—	—	—
Part-time	0.44 (–1.40 to 2.30)	.60	0.45 (–0.25 to 1.20)	.20
Unemployed	–0.44 (–2.40 to 1.50)	.70	0.13 (–0.60 to 0.86)	.70
Other	0.00 (–1.90 to 1.90)	>.90	–0.18 (–0.89 to 0.54)	.60
Language				
English	—	—	—	—
Spanish	1.00 (–1.30 to 3.30)	.40	0.63 (–0.23 to 1.50)	.15

[a]Response rate=0.01% (185 observations); adjusted R^2=0.02; SE 4.41 (*df*=169); $F_{15,169}$=1.24.

[b]Response rate=0.00% (185 observations); adjusted R^2=0.03; SE 1.66 (*df*=169); $F_{15,169}$=1.36.

[c]—: not determined.

Engagement With the Text Messages

Participants answered the mood text messages on average 60.0% of the time (ranging from 0% to 100%). Furthermore, 21 people did not respond to any mood-related messages, and 13 participants opted out of text messaging by texting "STOP" to our system. The 303 baseline participants were in the study for a range of 2 to 72 days (mean 59 days). Seventy participants went beyond the 60-day time frame owing to a system glitch; these participants were in the study for an average number of 63 days.

Discussion

Principal Findings

Participants who received the StayWell text messaging program showed improved depression and anxiety symptoms at completion of the program (60 days) on average. These results are similar to those of previous studies utilizing text messaging as a public mental health intervention to counteract the deleterious emotional and mental health effects of the COVID-19 pandemic in Canada [12-15]. Additionally, engagement with our texting study (2 messages per day) was relatively high. Response rates averaged to 60% in the daily

mood check-in, and only 4% of participants opted out of the text messages during the study. This study supports the use of text messaging as a broad-based tool for improving mental health, especially in the context of a global pandemic when in-person behavioral health visits are inaccessible.

It was important to assess whether other factors (local COVID-19 infection rates, self-rated health, and other demographic variables) influenced the positive outcomes; however, we found that improvements in GAD-2 scores were not related to other measured variables. We found a greater improvement in PHQ-8 scores for female-identifying participants, but improvements held despite local infection rates or other demographic factors. These findings suggest that women may experience greater benefits in their depression symptoms from participating in the StayWell program. Nonetheless, it is particularly notable that weekly local infection rates were not related to changes in outcomes. For example, it is possible that decreases in symptoms could be influenced by reductions in local infection rates and accompanying lowered concerns of infection or reduction in policies such as stay-at-home orders. However, improvements in PHQ-8 scores persisted after accounting for changes in weekly infection rates and other covariates, suggesting that the intervention improved symptoms beyond the influence of any change in the severity of the COVID-19 pandemic in participants' area of residence.

Over a third (36.3%) of participants did not complete the final assessment, and we took steps to assess how noncompletion impacted our results. First, we assessed whether reductions in depression and anxiety symptoms remained after controlling for participants' response rates, and our findings suggest that the intervention may be beneficial overall even if participants do not respond. It is possible that reading messages can be beneficial even if participants do not respond to subsequent mood ratings. On average, participants who completed the study had a significantly higher response rate to mood rating messages (70%, 42/60 responses) than those who dropped out (35%, 21/60 responses) (P<.001). Despite no significant differences in baseline mood ratings, PHQ-8, or GAD-2 scores between completers and noncompleters, those who completed the final assessment had increased daily mood ratings (0.0037; P<.001), whereas those who did not complete the final assessment reported decreases in mood ratings (difference of 0.0041 from 0.0078 to 0.0037; P<.001). Therefore, it is possible that people with worsening symptoms respond less often or that less engagement is related with worse outcomes. It is possible that the study outcomes are biased on the basis of the mood rating post hoc analyses (attrition bias); however, this is in contrast with significant differences in outcomes remaining after controlling for the response rate. Ideally, we would be able to assess any differences in primary outcomes (PHQ-8 and GAD-2 scores), but we do not have those data for participants who dropped out of the study. This provides a mixed picture but suggests that it is more likely that participants experiencing worsening symptoms drop out of the study more often.

As stated in the trial protocol, an additional aim of this study was to test whether a reinforcement learning algorithm could improve personalization and outcomes beyond randomly selecting messages within different categories (behavioral activation and coping skills). Considering technical difficulties, we could not randomize participants into the distinct conditions, and all participants received the same intervention (random message condition). To prevent technical errors, future studies using RL should incorporate pilot work before commencing the trial to check for errors; however, we were unable to do so because we needed a quick roll-out during the pandemic. Further studies are needed to compare the effectiveness of personalizing messages using a reinforcement learning algorithm to that of a random message condition. Other studies in progress may be better suited to answer questions related to the impact of reinforcement learning models' utility for improving personalization [23].

Strengths

This study had a racially/ethnically diverse sample in the United States, which included 11.3% of Spanish speakers. On average, participants entered the study with mild/moderate symptoms of depression, which improved to mild symptoms at the end of the intervention period. While reductions in PHQ-8 and GAD-2 were not large, the low-intensity intervention approach is highly scalable and can accommodate a large number of people. This study also shows that participants are open to receiving 2 messages per day, including one mood response, while maintaining a high response rate (60%). Lastly, the longitudinal data collected provides opportunities for further inquiry into daily mood ratings and response rates.

Limitations

The main limitation of this study is the lack of a control group with no or an inactive intervention. Given the impacts of the COVID-19 pandemic on mental health, the study team felt that it was unethical to withhold mental health support at this time. In addition, we were interested in assessing whether the intervention might be improved by applying a reinforcement learning algorithm to personalize messages, but technical errors prevented us from examining this. Lastly, a significant portion of the sample did not complete the final assessment, which may result in attrition bias, and further exploration of the engagement-outcome relationship is merited.

Implications for Future Studies

Our study provides support for low-intensity text messaging interventions to improve mental health at the population level. Our data show the feasibility of sending 2 messages a day and asking for daily mood responses. Texting and other mobile interventions could serve as ways to identify individuals in need of more intensive intervention on the basis of reported daily mood ratings. Future studies should continue to assess the impacts of text messaging and related mobile health interventions for mental health and continue to assess methods including machine learning to improve personalization.

Conclusions

Participation in a CBT-based text messaging program focused on COVID-19 cognitive flexibility, and behavioral activation and acceptance were related to improved depression and anxiety symptoms. In addition to improvements in outcomes, this study reported high levels of engagement during a 60-day intervention that sent 2 messages per day. Text messaging interventions

could serve as an important public health tool to disseminate strategies for managing mental health.

Acknowledgments

This study was funded by the Health Resources & Services Administration (HRSA) Coronavirus Aid, Relief, and Economic Security (CARES) Grant No. T1NHP39159.

Authors' Contributions

AA provided the platform, developed the intervention content, design, and writing of the manuscript. RHR contributed to intervention development, content and implementation, writing, and editing of the manuscript. AYHR contributed to intervention conceptualization, data analysis, writing, and editing of the manuscript. CEB contributed to intervention conceptualization, data analysis, and editing of the manuscript. TL helped develop user-centered personas. JX advised on the experimental design and the statistical analysis plan and conducted analyses. BC helped conceptualize the trial design and advised on the analysis plan. SD contributed to the development of the intervention content. CF contributed to intervention development, design, writing, and analysis. CK contributed to software development and intervention deployment.

Conflicts of Interest

AA is the creator and owner of HealthySMS.

Multimedia Appendix 1
Multivariable linear regression modeling the change in PHQ-8 and GAD-2 scores at 60-day exit from baseline for participants who completed both surveys.
[DOCX File , 17 KB - mental_v8i11e25298_app1.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. Bavel JJV, Baicker K, Boggio PS, Capraro V, Cichocka A, Cikara M, et al. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav 2020 May;4(5):460-471. [doi: 10.1038/s41562-020-0884-z] [Medline: 32355299]
3. Pedrosa AL, Bitencourt L, Fróes ACF, Cazumbá MLB, Campos RGB, de Brito SBCS, et al. Emotional, Behavioral, and Psychological Impact of the COVID-19 Pandemic. Front Psychol 2020;11:566212 [FREE Full text] [doi: 10.3389/fpsyg.2020.566212] [Medline: 33117234]
4. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet 2020 Mar 14;395(10227):912-920 [FREE Full text] [doi: 10.1016/S0140-6736(20)30460-8] [Medline: 32112714]
5. Leigh-Hunt N, Bagguley D, Bash K, Turner V, Turnbull S, Valtorta N, et al. An overview of systematic reviews on the public health consequences of social isolation and loneliness. Public Health 2017 Nov;152:157-171. [doi: 10.1016/j.puhe.2017.07.035] [Medline: 28915435]
6. Dsouza DD, Quadros S, Hyderabadwala ZJ, Mamun MA. Aggregated COVID-19 suicide incidences in India: Fear of COVID-19 infection is the prominent causative factor. Psychiatry Res 2020 Aug;290:113145 [FREE Full text] [doi: 10.1016/j.psychres.2020.113145] [Medline: 32544650]
7. 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]
8. Cellini N, Canale N, Mioni G, Costa S. Changes in sleep pattern, sense of time and digital media use during COVID-19 lockdown in Italy. J Sleep Res 2020 Aug;29(4):e13074 [FREE Full text] [doi: 10.1111/jsr.13074] [Medline: 32410272]
9. Tison GH, Avram R, Kuhar P, Abreau S, Marcus GM, Pletcher MJ, et al. Worldwide Effect of COVID-19 on Physical Activity: A Descriptive Study. Ann Intern Med 2020 Nov 03;173(9):767-770 [FREE Full text] [doi: 10.7326/M20-2665] [Medline: 32598162]
10. Rajkumar RP. COVID-19 and mental health: A review of the existing literature. Asian J Psychiatr 2020 Aug;52:102066 [FREE Full text] [doi: 10.1016/j.ajp.2020.102066] [Medline: 32302935]
11. Williams DR, Cooper LA. COVID-19 and Health Equity-A New Kind of "Herd Immunity". JAMA 2020 Jun 23;323(24):2478-2480. [doi: 10.1001/jama.2020.8051] [Medline: 32391852]
12. Agyapong VIO. Coronavirus Disease 2019 Pandemic: Health System and Community Response to a Text Message (Text4Hope) Program Supporting Mental Health in Alberta. Disaster Med Public Health Prep 2020 Oct;14(5):e5-e6 [FREE Full text] [doi: 10.1017/dmp.2020.114] [Medline: 32317038]

13. Agyapong VIO, Hrabok M, Vuong W, Shalaby R, Noble JM, Gusnowski A, et al. Changes in Stress, Anxiety, and Depression Levels of Subscribers to a Daily Supportive Text Message Program (Text4Hope) During the COVID-19 Pandemic: Cross-Sectional Survey Study. JMIR Ment Health 2020 Dec 18;7(12):e22423 [FREE Full text] [doi: 10.2196/22423] [Medline: 33296330]
14. Agyapong VIO, Hrabok M, Shalaby R, Vuong W, Noble JM, Gusnowski A, et al. Text4Hope: Receiving Daily Supportive Text Messages for 3 Months During the COVID-19 Pandemic Reduces Stress, Anxiety, and Depression. Disaster Med Public Health Prep 2021 Feb 08:1-5 [FREE Full text] [doi: 10.1017/dmp.2021.27] [Medline: 33551009]
15. Agyapong VIO, Shalaby R, Hrabok M, Vuong W, Noble JM, Gusnowski A, et al. Mental Health Outreach via Supportive Text Messages during the COVID-19 Pandemic: Improved Mental Health and Reduced Suicidal Ideation after Six Weeks in Subscribers of Text4Hope Compared to a Control Population. Int J Environ Res Public Health 2021 Feb 23;18(4):2157 [FREE Full text] [doi: 10.3390/ijerph18042157] [Medline: 33672120]
16. Berrouiguet S, Baca-García E, Brandt S, Walter M, Courtet P. Fundamentals for Future Mobile-Health (mHealth): A Systematic Review of Mobile Phone and Web-Based Text Messaging in Mental Health. J Med Internet Res 2016 Jun 10;18(6):e135 [FREE Full text] [doi: 10.2196/jmir.5066] [Medline: 27287668]
17. Figueroa CA, Aguilera A. The Need for a Mental Health Technology Revolution in the COVID-19 Pandemic. Front Psychiatry 2020;11:523 [FREE Full text] [doi: 10.3389/fpsyt.2020.00523] [Medline: 32581891]
18. Pfefferbaum B, North CS. Mental Health and the Covid-19 Pandemic. N Engl J Med 2020 Aug 06;383(6):510-512. [doi: 10.1056/nejmp2008017]
19. Mobile Fact Sheet. Pew Research Center. 2021 Apr 07. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-04-08]
20. Aguilera A, Bruehlman-Senecal E, Demasi O, Avila P. Automated Text Messaging as an Adjunct to Cognitive Behavioral Therapy for Depression: A Clinical Trial. J Med Internet Res 2017 May 08;19(5):e148 [FREE Full text] [doi: 10.2196/jmir.6914] [Medline: 28483742]
21. Klasnja P, Hekler EB, Shiffman S, Boruvka A, Almirall D, Tewari A, et al. Microrandomized trials: An experimental design for developing just-in-time adaptive interventions. Health Psychol 2015 Dec;34S:1220-1228 [FREE Full text] [doi: 10.1037/hea0000305] [Medline: 26651463]
22. Chan A, Tetzlaff J, Altman D, Laupacis A, Gøtzsche PC, Krleža-Jerić K, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. Ann Intern Med 2013 Feb 05;158(3):200-207 [FREE Full text] [doi: 10.7326/0003-4819-158-3-201302050-00583] [Medline: 23295957]
23. Figueroa CA, Hernandez-Ramos R, Boone CE, Gómez-Pathak L, Yip V, Luo T, et al. A Text Messaging Intervention for Coping With Social Distancing During COVID-19 (StayWell at Home): Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2021 Jan 14;10(1):e23592 [FREE Full text] [doi: 10.2196/23592] [Medline: 33370721]
24. HealthySMS: -- StayWell. URL: https://staywell.healthysms.org/accounts/login/?next=/dashboard/ [accessed 2021-10-13]
25. Norman D, Draper S. User Centered System Design: New Perspectives on Human-computer Interaction. Boca Raton, FL: CRC Press; 1986.
26. 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]
27. Hughes AJ, Dunn KM, Chaffee T, Bhattarai J, Beier M. Diagnostic and Clinical Utility of the GAD-2 for Screening Anxiety Symptoms in Individuals With Multiple Sclerosis. Arch Phys Med Rehabil 2018 Oct;99(10):2045-2049 [FREE Full text] [doi: 10.1016/j.apmr.2018.05.029] [Medline: 29964000]
28. Shapiro SS, Wilk MB, Chen HJ. A Comparative Study of Various Tests for Normality. J Am Stat Assoc 1968 Dec;63(324):1343-1372. [doi: 10.1080/01621459.1968.10480932]

Abbreviations

BA: behavioral activation
CBT: cognitive behavioral therapy
CONSORT: Consolidated Standards of Reporting Trials
GAD-2: 2-item Generalized Anxiety Disorder scale
PHQ-8: 8-item Patient Health Questionnaire
RL: Reinforcement Learning
UC: University of California
UR: Uniform Random

Edited by J Torous; submitted 07.05.21; peer-reviewed by J Zulueta, V Agyapong; comments to author 26.05.21; revised version received 07.07.21; accepted 13.09.21; published 01.11.21.

Please cite as:
Aguilera A, Hernandez-Ramos R, Haro-Ramos AY, Boone CE, Luo TC, Xu J, Chakraborty B, Karr C, Darrow S, Figueroa CA
A Text Messaging Intervention (StayWell at Home) to Counteract Depression and Anxiety During COVID-19 Social Distancing: Pre-Post Study
JMIR Ment Health 2021;8(11):e25298
URL: https://mental.jmir.org/2021/11/e25298
doi:10.2196/25298
PMID:34543230

Viewpoint

Opening the Black Box of Daily Life in Nonsuicidal Self-injury Research: With Great Opportunity Comes Great Responsibility

Glenn Kiekens[1,2], PhD; Kealagh Robinson[3], MSc; Ruth Tatnell[4], PhD; Olivia J Kirtley[2], PhD

[1]Faculty of Psychology and Educational Sciences, Clinical Psychology, KU Leuven, Leuven, Belgium

[2]Department of Neurosciences, Center for Contextual Psychiatry, KU Leuven, Leuven, Belgium

[3]School of Psychology, Te Herenga Waka-Victoria University of Wellington, Wellington, New Zealand

[4]Faculty of Health, School of Psychology, Deakin University, Melbourne, Australia

Corresponding Author:
Glenn Kiekens, PhD
Faculty of Psychology and Educational Sciences, Clinical Psychology
KU Leuven
Tiensestraat 102
Leuven, 3720
Belgium
Phone: 32 16372852
Email: glenn.kiekens@kuleuven.be

Abstract

Although nonsuicidal self-injury (NSSI)—deliberate damaging of body tissue without suicidal intent—is a behavior that occurs in interaction with real-world contexts, studying NSSI in the natural environment has historically been impossible. Recent advances in real-time monitoring technologies have revolutionized our ability to do exactly that, providing myriad research and clinical practice opportunities. In this viewpoint paper, we review new research pathways to improve our ability to understand, predict, and prevent NSSI, and provide critical perspectives on the responsibilities inherent to conducting real-time monitoring studies on NSSI. Real-time monitoring brings unique opportunities to advance scientific understanding about (1) the dynamic course of NSSI, (2) the real-time predictors thereof and ability to detect acute risk, (3) the ecological validity of theoretical models, (4) the functional mechanisms and outcomes of NSSI, and (5) the promotion of person-centered care and novel technology-based interventions. By considering the opportunities of real-time monitoring research in the context of the accompanying responsibilities (eg, inclusive recruitment, sound and transparent research practices, participant safety and engagement, measurement reactivity, researcher well-being and training), we provide novel insights and resources to open the black box of daily life in the next decade(s) of NSSI research.

(JMIR Ment Health 2021;8(11):e30915) doi:10.2196/30915

KEYWORDS

real-time monitoring; nonsuicidal self-injury; NSSI; experience sampling; ecological momentary assessment; digital psychiatry

Introduction

Nonsuicidal self-injury (NSSI), defined as the direct and deliberate damage of one's body tissue without suicidal intent (eg, cutting and hitting oneself) [1], is a behavior seemingly at odds with the principles of minimizing pain and maximizing pleasure, which guide most human behaviors. One in 5 people engage in NSSI at least once before the age of 25 years [2,3], and doing so increases their risk for future suicidal thoughts and behaviors and mental health conditions [4,5] and other adverse developmental outcomes [6-8]. Unfortunately, few individuals access support for their NSSI [9], with many young people who self-injure not finding their way to treatment [10]. Together, these findings highlight NSSI as behavior that warrants greater awareness and a better understanding—a viewpoint that the American Psychiatric Association formally emphasized by including NSSI as a "condition requiring further study" in the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders [11].

Taking stock of the research published in the past decade reveals substantial advances in our understanding of the epidemiology, phenomenology, and developmental course of NSSI [2,3,10,12]. Longitudinal cohort studies have substantially advanced knowledge regarding intrapersonal and interpersonal risk and protective factors that clarify who is at the highest risk for developing [13,14] and continuing NSSI behavior during

adolescence and emerging adulthood [10,15,16]. Unfortunately, our understanding of when young people are at risk of NSSI in everyday life has not progressed similarly. We see 3 main reasons hindering this knowledge progression. First, it is an ecological fallacy to believe that a nomothetic approach that provides between-group knowledge about who is relatively at high risk throughout adolescence and emerging adulthood can be translated to the here and now at the individual level [17,18]. Indeed, knowing that someone is developmentally at risk to engage in NSSI (eg, due to a history of victimization) [19] tells us little about *when* that person is most likely to self-injure in everyday life. Second, nearly all longitudinal research studies used observation windows from months to years to clarify developmental risk [20]. However, retrospectively aggregating data over months to years (eg, *Have you self-injured since last year?*) lacks the temporal precision to detect individual risk within minutes-to-hours. Third, and perhaps most importantly, researchers have historically been constrained by practical restrictions that rendered frequent assessments of NSSI in individuals' daily life virtually impossible. Nevertheless, in a new era of precision medicine, if we are to enable individualized intervention *when* and *where* it is most needed, then research needs to take an idiographic approach in which risk stratification repeatedly occurs in the natural environment with individuals serving as their own control [18].

Out of the Laboratory and Into Everyday Life

Recent advances in digital technology now make it possible for researchers to take such an idiographic approach, shifting research from the laboratory into the everyday environment where NSSI thoughts, urges, and behaviors occur. Real-time monitoring (also called experience sampling or ecological momentary assessment) is a structured self-report diary technique in which individuals provide information on their situational context, feelings, thoughts, and behavioral patterns in the flow of daily life [21,22]. Self-report questionnaires are completed *multiple times* throughout the day for several days or weeks. Participants are prompted to fill in questionnaires either during predetermined intervals of time (eg, every 2 hours, interval-contingent sampling) at random unpredictable moments (ie, signal-contingent sampling) or following an event of interest (ie, event-contingent sampling) [23]. Daily diaries are a particular case of interval-contingent sampling in which assessments occur only once and typically at the end of each day. Real-time monitoring methods are not a new methodology [24], with roots in ecological psychology, which argues that behavior can only be understood when investigated in the context in which it occurs [25]. Although initial real-time monitoring studies of NSSI relied upon pagers and personal digital assistants [26,27], the ubiquity of mobile smartphones in today's society [28] has made it practically feasible for the broader research community to study NSSI and its contextual determinants in everyday life. The increased practicality of real-time monitoring methods offers a promising avenue to answer critical questions and engage researchers and clinicians in collaborative discussions. However, real-time monitoring methods, which focus on NSSI, also present significant ethical and practical challenges.

Given that real-time monitoring research is burgeoning [29-31], it is timely to consider the valuable new directions in which the field could be heading when studying NSSI outside the laboratory, in everyday life. In 2019, the International Society for the Study of Self-Injury established a *Consortium for Research on Self-Injury in Everyday Life* to help build expertise and capacity in a rapidly growing field [32]. In this Viewpoint paper, arising from the work of the consortium, we (1) review new research pathways that use real-time monitoring methods to improve our ability to understand, predict, and prevent NSSI, and (2) provide critical perspectives on the responsibilities inherent to conducting real-time monitoring studies on NSSI. In doing so, we identified crucial open questions that require further investigation and offer guidance and concrete recommendations for future studies.

Opening the Black Box of Daily Life Brings Exciting New Opportunities for Science and Practice

In this section, we outline 5 key opportunities that real-time monitoring provides for advancing our ability to understand, predict, and prevent NSSI thoughts, urges, and behaviors in the lives of those at risk (Textbox 1).

Textbox 1. Five key opportunities of real-time monitoring.

New opportunities created by opening the black box of daily life in nonsuicidal self-injury research:

- Better understanding of the short-term course of nonsuicidal self-injury thoughts, urges, and behavior through direct observation and precise measurement.
- Advance knowledge of individual-level predictors of nonsuicidal self-injury thoughts, urges, and behavior and the ability to accurately detect idiographic risk.
- Test existing theories and develop new models that bridge the idiographic and nomothetic divide and explain who is at risk and when.
- Generate insights into the functional mechanisms and relationship of dynamic patterns with day-to-day and meaningful longer term developmental changes and outcomes.
- Promote person-centered care and deployment of personalized prevention and novel digital interventions.

Opportunity 1: A Better Understanding of the Short-term Course of NSSI Thoughts, Urges, and Behavior Through Direct Observation and Precise Measurement

Real-time monitoring enables rigorous descriptive research about the course of NSSI thoughts, urges, and behaviors. Initial work has demonstrated that NSSI thoughts frequently occur among individuals who self-injure but are usually short-lived and are of moderate intensity [27,33,34]. The propensity to experience intense and persistent NSSI thoughts has been found to increase throughout the day [34], with thought intensity fluctuating considerably from hour to hour for some individuals [35]. However, future work is needed to replicate these findings and many questions remain, including dynamics over even shorter timescales (ie, within seconds/minutes), whether different thought profiles can be identified in terms of intensity, duration, controllability, and persistence, and the degree to which changes in dynamic thought patterns relate to urges and behaviors.

Different qualitative aspects of NSSI thoughts and urges may combine to increase risk, such that the likelihood of NSSI behavior may increase in situations characterized by more intense persistent thoughts [33]. In this respect, real-time monitoring offers the opportunity to capture a fast-moving thought-to-action process through precise measurement in real time. Research suggests that it typically takes people 1-30 minutes to transition from NSSI thoughts to behavior [27,33], meaning that in most instances, there is a brief window of opportunity to intervene and interrupt the transition from thoughts to behavioral action. Better characterization of the thought profiles and behavioral patterns of NSSI as well as the extent to which these can change both within and across individuals are an essential first step in identifying individual-level predictors for risk screening and preventive intervention.

Opportunity 2: Advancing Knowledge of Individual-Level Predictors of NSSI Thoughts, Urges, and Behavior, and the Ability to Accurately Detect Idiographic Risk

Daily life research provides a contextualized understanding of the momentary factors that explain variability in the short-term course of NSSI. Using real-time monitoring, researchers can study theoretically relevant situational, emotional, and cognitive factors to advance knowledge of individual-level predictors for developing NSSI thoughts and urges, and for making the transition to behavior. Initial findings suggest that the likelihood of these outcomes may increase when people are alone [27], after negative social appraisals and perceived conflict [36,37], or following increased negative and decreased positive affect [35,38]. Studies have also observed an increased risk of NSSI thoughts, urges, and behavior in the presence of high self-criticism and negative repetitive thinking [39,40], or low momentary self-efficacy to resist NSSI [35]. Despite this knowledge, future research is needed to clarify the relative importance of these situational, emotional, and cognitive factors at each stage of the NSSI process and their specificity in predicting NSSI compared to co-occurring behaviors (eg, eating disorder behaviors, suicidal thoughts) [27,41]. Worth mentioning in this context is that real-time monitoring also provides a unique opportunity to clarify the relationship with these comorbid behaviors in daily life [42], thereby offering meaningful information to further diagnostic understanding of NSSI. Finally, the timescale in which factors exert an effect and how their interplay can be understood mathematically (ie, linear or nonlinear effects) warrants further clarification.

Building upon empirically derived answers to these critical questions, the next fundamental step is developing risk prediction models that can accurately detect *when* someone is at imminent risk for engaging in NSSI. Using each individual's longitudinal data, researchers can select and combine risk and protective factors to create risk stratification indices of NSSI thoughts, urges, and behavior in the natural environment for a particular person (eg, in the case of smoking behavior) [43]. Statistical classification approaches (also known as machine learning) and validation techniques can be employed to identify the most suitable person-specific combination of risk factors [44,45]. However, 2 caveats should be acknowledged for future research in this area. First, although idiographic risk prediction models will scale up the ability to identify individuals at acute risk for NSSI thoughts, urges, and behavior in daily life, making better use of mobile technologies' growing capacities will be pivotal to ensure that individuals identified as at-risk are not left without the necessary support (*see opportunity 5*). Second, because real-time monitoring for prolonged periods becomes burdensome, it will be crucial to capitalize on the ever growing technological capacities and explore the utility and integration of passively collected information in these models [46]. For example, smartphones continuously track a wealth of "in the moment" information (eg, call or SMS logs, location recording), and initial investigation supports the feasibility of using wearables to measure psychophysiology among high-risk adolescents [47]. Importantly, these 2 caveats illustrate that real-time monitoring of NSSI thoughts and behaviors also brings considerable ethical, legal, and practical challenges regarding inclusivity, informed consent, and participant safety and burden (*see responsibilities section for a discussion of these challenges*).

Opportunity 3: Test Existing Theories and Develop New Models That Bridge the Idiographic and Nomothetic Divide and Explain Who Is at Risk and When

Real-time monitoring provides researchers with the opportunity to put existing theories to the test in daily life. Contemporary theories of NSSI posit that the joint influence of social, affective, and cognitive vulnerabilities cause risk for NSSI via idiographic microprocesses that play out in the realm of ordinary life. However, these psychological processes are typically evaluated using cross-sectional and traditional longitudinal surveys—designs that do not have the necessary temporal granularity or ecological validity to reliably assess these theories' dynamic real-life components. Real-time monitoring overcomes this limitation. For instance, Hughes and colleagues [40] observed that momentary negative affect and repetitive negative thinking synergistically predict NSSI in daily life, thereby providing evidence for the Emotional Cascade Model

[48]. A limitation of this model is that it does not address *why* someone chooses to engage in NSSI instead of other dysregulated behaviors. In this respect, the Benefits and Barriers Model argues for the unique role of self-criticism in developing NSSI [49], whereas the Cognitive-Emotional Model argues for an expanded role of NSSI-specific cognitions [50]. Consistent with the Cognitive-Emotional Model, cross-sectional evidence suggests that behavior-specific beliefs (eg, self-efficacy to resist NSSI) explain why individuals use NSSI instead of risky alcohol use or disordered eating when distressed [51]. Investigations of these models in daily life are currently ongoing [35,39,52]. Emerging evidence, for instance, suggests that momentary belief in one's ability to resist NSSI is a robust short-term predictor of NSSI behavior among young adults in daily life [35]. However, more work is required to replicate and extend initial findings, including whether behavior-specific beliefs can explain engagement across different behaviors—NSSI and non-NSSI—for everyone.

Notably, existing theories of NSSI do not explicitly differentiate nomothetic and idiographic risk processes, thereby implicitly assuming that what causes risk is the same across individuals. Nevertheless, we can expect that variation in risk processes will be the rule rather than the exception [17,53]. As in most psychology areas [54], existing models are verbal theories, which formulate a narrative of how NSSI behavior manifests rather than translating the theory's tenets and assumptions into a formal model using mathematical notations. Emerging work underscores the need for novel models in psychology to make formal predictions [54-56], which would allow researchers to precisely estimate what a theory predicts at different measurement levels in computational models and compare this with real-world data. Real-time monitoring can facilitate the generation of theoretical models that conceptualize NSSI as a complex system of contextualized dynamic processes. When formalized, dynamic and contextually informed theories could predict concretely and precisely *when* NSSI thoughts, urges, and behaviors are likely to occur and for *whom*. Such theory construction would progress understanding of factors that increase/decrease the risk for everyone, a subgroup of individuals, or a specific individual [57], and help overcome the research practice gap by allowing practitioners to consider what causes risk for an individual while still enabling the scalability and generalizability of these predictions to be evaluated [58].

Opportunity 4: Insight Into the Functional Mechanisms, Day-to-day Outcomes, and Relationship of Dynamic Patterns With Meaningful Long-term Developmental Changes and Outcomes

By providing the opportunity to track the dynamic processes in the moments that lead up to and follow self-injurious behavior, real-time monitoring allows investigation of the functional mechanisms that maintain NSSI in daily life. According to the Four-Function Model [59], NSSI may be used to mitigate negative or unwanted thoughts and feelings (ie, intrapersonal negative reinforcement), to generate emotion as a form of stimulation (ie, intrapersonal positive reinforcement), to escape from uncomfortable social situations (ie, interpersonal negative reinforcement), or to seek support from others (ie, interpersonal positive reinforcement). Although the Four-Function Model has received considerable empirical support in cross-sectional studies [12,60], longitudinal measurement in real-life is needed to model the temporal contingencies of interpersonal and intrapersonal processes. A recent review of daily life studies of NSSI revealed the most evidence for intrapersonal negative reinforcement but also found substantial inconsistencies [30]. For example, although some studies observed an increase in negative affect before and a decrease following NSSI behavior [61], others failed to replicate this pattern [62], and some even found increased, rather than decreased, negative affect following NSSI behavior [63]. Although investigations of the other reinforcement processes are scarce, findings were also mixed [30]. An important recommendation for future work is to consider timeframes more carefully. Real-time monitoring studies that add brief follow-up surveys to their protocol when people report momentary NSSI thoughts and urges provide a unique opportunity to unravel contingencies that unfold across shorter (ie, seconds, minutes) and longer (eg, hours, days) time intervals [64].

Apart from providing insight into the functional mechanisms, such studies would also clarify the psychosocial outcomes of NSSI in daily life. For example, engagement in NSSI may lead to interpersonal conflict as well as increased social support [65-67], feelings of shame [68], and experiencing stigma (especially when scars are visible) [69,70], which may, in turn, increase social withdrawal and the likelihood of future NSSI. Hence, much could be learned from future investigations that adopt a transactional framework in which NSSI outcomes and psychosocial experiences might influence each other reciprocally in daily life. Incorporating real-time monitoring within prospective cohort studies (ie, measurement burst designs) [71] can uniquely inform how short-term patterns relate to long-term developmental change and outcomes. Although already employed in depression and substance use research [72,73], these measurement burst designs are currently an untapped resource for NSSI research. For instance, the degree to which NSSI thoughts are self-sustaining in daily life (ie, auto-correlation) could signal a more challenging recovery process [74] or help explain why some individuals (eg, those with depression) are at risk of a more chronic NSSI course [75]. Given the relationship between NSSI and suicidal thoughts and behaviors throughout development [4,76], a critical question is clarifying whether a dynamic blueprint of NSSI can help gauge the future risk of suicidal forms of self-injury. Providing greater clarity regarding potential day-to-day and long-term developmental outcomes would aid scientific understanding and provide valuable information for prevention efforts and clinical risk assessment.

Opportunity 5: Promotion of Person-Centered Care, Personalized Prevention, and Novel Technology-Based Interventions

Over the last decades, mental health care has gradually shifted from hospital-based to community-based care and changed focus from symptom reduction to patient-defined recovery [21]. Through repeated observation of emotions, thoughts, symptoms,

NSSI outcomes, and contextual determinants thereof in patients' lives, real-time monitoring can help respond to the call for more person-centered care in the treatment of NSSI [74]. For instance, through easy to interpret visualizations of real-time monitoring data, information on individual functioning and patient-defined outcomes can be fed back into the therapy room. This way, real-time monitoring could facilitate psychoeducation about relevant processes—that patients may otherwise be unaware of—and give clinicians and patients a valuable tool to monitor and tailor treatment according to patients' dynamic therapy needs. For an example of such a real-time monitoring tool, see the KU Leuven m-Path app and platform [77]. However, to enable the use of real-time monitoring as a therapeutic tool in the treatment of NSSI, pilot studies are required to address the barriers (eg, burden and fear of reactivity) [78] and requirements for successful implementation (eg, availability of an accessible and reliable platform) [79] of real-time monitoring as a blended care tool. Building upon this, randomized controlled trials are needed to determine how, when (eg, unguided in the moment or guided during a clinical session), and which type of feedback (eg, overall functioning, activities, social interactions, or NSSI-specific triggers and risk processes) should be offered. Codeveloping answers to these open questions with all stakeholders involved (ie, people with lived experience, researchers, clinicians, software developers) represents a critical step to harness the potential of real-time monitoring for NSSI treatment.

Finally, real-time monitoring provides scientist-practitioners the opportunity not only to observe but also to deliver support in people's everyday lives, taking mental health care beyond the clinical setting and into daily life. Ecological momentary interventions (EMIs) are delivered in real time through a smartphone app or a wearable (eg, smartwatch) and can be offered as a self-help mobile health intervention or to augment and extend the reach of existing treatments [21,80]. Initial findings indicate the acceptability and potential of EMIs and mobile apps that target NSSI [81-83], but this remains a largely underexplored area of research. A sophisticated EMI that currently shows promising results in mental health research is just-in-time adaptive interventions (JITAIs) [84], which helps people resist the urge to self-injure when needed most in daily life. JITAIs tailor interventions to the risk status (eg, low, medium, high) and the receptivity of the people within the environmental context, thereby enabling timely and contextually informed interventions for behaviors that are highly dynamic [85]. Give these possibilities, the use of JITAIs is already emerging in suicide research [86], with similar efforts needed to develop, evaluate, and integrate these new treatment methods into a stepped care model for NSSI.

Summary of Opportunities

Opening the black box of daily life in NSSI research has considerable potential to advance scientific understanding about (1) the short-term course of NSSI thoughts, urges, and behavior; (2) the individual-level predictors thereof and ability to accurately detect imminent risk; (3) the ecological validity of theoretical models and the possibility to explain *when* NSSI thoughts and behaviors are most likely to occur and *for whom;* (4) the functional mechanisms of NSSI and relationship of dynamic patterns with day-to-day and meaningful long-term change and outcomes; and (5) the implementation of real-time monitoring to prevent key NSSI outcomes and support individuals in distress when they need it the most. However, studying NSSI "in the wild" outside a controlled laboratory environment also presents unique challenges for which there are no established gold standard solutions.

With Great Opportunity Comes Great Responsibility

In the following section, we outline vital responsibilities when planning and carrying out real-time monitoring research of NSSI thoughts and behaviors (Textbox 2). Although some considerations are universally applicable to real-time monitoring research, here we focus on seven issues that have particular relevance in the context of NSSI and point to open questions in these domains for future research.

Textbox 2. Seven issues that have particular relevance in the context of nonsuicidal self-injury (NSSI).

Ethical and practical considerations when opening the black box of daily life in NSSI research:

- Recruitment should be inclusive from study inception to completion and actively include more vulnerable individuals, with representatives from any vulnerable group at every stage.
- The informed consent process should be fully transparent regarding the study demands, the safety protocol, whether data will be passively collected, reimbursement, researchers' responsibility to respond to risk, and potential implications of this responsibility.
- A proper safety protocol should be developed with all stakeholders that matches participants' needs (especially in the event of suicide risk), but that does not inadvertently defeat the study's observational purpose.
- Although there is no reason to expect that repeated questioning in everyday life will lead to measurement reactivity in nonsuicidal self-injury outcomes, researchers are responsible for evaluating whether this holds for all participants in their study.
- Study designs must be carefully balanced to appropriately answer the research question(s) while not unnecessarily burdening participants. Sufficient resources should be allocated to pilot all aspects of the protocol. Researchers are encouraged to preregister their protocol and be aware of the relevant privacy laws in their home country before commencing data collection.
- Participants should be recognized as valued contributors to the research and receive financial incentives and information about the overall findings. Where feasible, participants should receive feedback on their own data.
- Research staff should receive good quality training in responding to risk and continued supervision and mentoring. A lone researcher should never be the only person responsible for participants' safety.

Responsibility 1: Recruitment and Inclusivity

When the goal is to understand the dynamic course of NSSI, sample diversity—without becoming tokenistic—should be prioritized to safeguard against falsely generalizing from one individual's (or a subgroup's) experience to the entire population. Therefore, we recommend actively engaging with members of more vulnerable groups, where the risk of NSSI and suicide may be higher than that in the general population (eg, LGBTQIA+ [lesbian, gay, bisexual, transgender, queer, intersex, asexual], Black, Indigenous, and other people of color, people facing homelessness) [87,88], and utilizing their input on how best to approach the research. Inclusive research is always essential [89], but especially when dealing with sensitive topics such as NSSI. For instance, working with people from different cultural and linguistic backgrounds requires flexibility in the way themes such as NSSI, suicide, and death are considered and discussed owing to differences in cultural norms and language use [90,91]. Some people may also not have access to a smartphone with a 4G connection or might share 1 smartphone in a household or family, conferring additional privacy concerns [92,93]. Therefore, researchers might aim for a budget that allows devices or data bundles to be provided to participants who need them, rather than excluding them. Importantly, however, if the real-time monitoring protocol involves deploying EMIs, researchers should be aware that participants may have come to rely upon the device and the EMI during the study period and that withdrawing these at the end of the study may leave participants without crucial support. Flexible compensation schedules, in which participants can choose to keep the smartphone as compensation for their participation, may be one solution. If practically unfeasible—either because of logistical constraints on the researchers or because the EMI requires a mobile data plan that participants cannot access—participants should not be left without support and could be offered alternative interventions (not requiring mobile data access) following the completion of the study.

When planning to recruit school-aged individuals to real-time monitoring studies, extra consideration should be given to data collection within the school context. For instance, schools may prohibit access to devices during the school day. van Roekel and colleagues [94] provide useful recommendations for working in school contexts, such as ensuring that there is a strong alliance with schools, teachers, and parents by using participation cards so that students are allowed to use their phones when prompted, and making sure that schools also benefit from the research. Having a specific person who is the "face" of the study within the school can also be useful. For a detailed discussion of the challenges of conducting NSSI research generally within schools, see [95].

Responsibility 2: Informed Consent and Participant Briefing

Given that real-time monitoring research takes place in daily life without the researcher being present, additional consideration of the informed consent process and participant briefing is needed [96]. Information regarding the study's often intensive nature, such as the study's time course, the number of surveys per day, and the periods during which participants can expect prompts, should be made clear to potential participants before study enrollment. Given that participant compliance rates in real-time monitoring research can vary [97], participant briefing should cover whether financial compensation or other benefits of research participation are compliance-dependent and, if so, how many reminders will be sent. When studying NSSI, in particular, it is paramount that participants are informed about the safety procedures (especially when this involves human-led intervention contingent upon a survey response) and the potential consequences of these safety procedures (eg, when will the duty of care override the confidentiality principle and who will then be informed). The informed consent process should also clarify whether additional data will be passively collected (eg, location coordinates, incoming and outgoing SMS messages and calls, app usage statistics, accelerometer data) and make participants aware of the detailed level of data that can be collected *without* their active engagement. Poor digital literacy may threaten adequate informed consent [98], especially for passively collected data. Jacobson and colleagues [96] provide several valuable suggestions to ensure that participants have a complete and detailed understanding of the study, such as highlighting essential information, using comprehension quizzes, and preventing participants from scrolling through the informed consent without reading it (when provided online). Considering the extensive amount of information participants receive, it could be worthwhile to request consent for each part of the study separately (eg, data collection schedule, intervention component, safety protocols/plan). Providing real-world examples utilizing interactive videos or apps that can read information aloud could also be used to facilitate comprehension and mitigate the risk of poor digital literacy.

When working with minors, both the young person's *assent* and informed *consent* from their parent or caregiver will typically be required. However, as a highly stigmatized behavior, NSSI is often hidden from others [99]. Although disclosure to parents and caregivers can facilitate help-seeking and improve coping, it can also negatively impact the parent-child relationship and the wider family system [66,100] and lead the young person to worry about the involvement of parents or caregivers [101]. Although parents will often be informed when recruiting young people within a clinical setting, we recommend explaining the study's purpose in general (eg, to study interactions, emotions, thoughts, and behavior in daily life) instead of using NSSI-specific terms to avoid forced disclosures. This framing method also means that individuals may avoid reflecting upon their participation through a disease perspective or NSSI-labeled identity [102,103]

Responsibility 3: Participant Safety and Risk Monitoring

In real-time monitoring studies of suicidal thoughts and behaviors, ethical considerations regarding participant safety are, justifiably, a recurring concern [96,104]. In contrast, very few NSSI real-time monitoring studies report procedures for safeguarding and supporting participants during the study [101]. This may be because high suicide risk is sometimes an exclusion criterion for participation [33,105,106], and participant safety

procedures are generally to safeguard participants at high or imminent risk of making a suicide attempt [27,38]. Real-time monitoring studies tread a fine line between research and intervention and there must be a "goodness of fit" between a study's objectives and the design of the safety procedures [101]. For example, in a study of NSSI behavior, contacting participants every time they report engaging in NSSI would defeat the study's purpose and may even discourage participants from reporting NSSI during the study period [101]. If a participant scores highly on a momentary measure of suicidal intent, contacting the participant may be appropriate and would not compromise the study's goal of assessing NSSI. A critical ethical issue underlying participant safety procedures is that even though an increase in suicidal intent is unlikely to be caused by study participation [107,108], the individual's status as a participant in a real-time monitoring study creates an opportunity for intervention that would otherwise not exist. Therefore, we recommend assessing suicidal intent in real-time monitoring research on NSSI thoughts and behaviors and advise against the exclusion of people at risk of suicide.

The first consensus statement on ethical and safety procedures for real-time monitoring studies with individuals at risk of suicide has emerged recently [109]. The recommendations include collecting contact information for participants and a close contact, completing a safety plan at study enrollment, monitoring responses at least once per day, and in the event of a participant being at imminent risk of suicide, for a researcher to contact them. Interestingly, our experiences have been somewhat different, with clinicians expressing concern that a researcher may be the "first responder" to a suicidal crisis. In this regard, contacting the participant's clinician may be better. However, no consensus was reached regarding whether the researcher should contact a participant's clinician in the event of high or imminent suicide risk [109]. The logistical challenges of actively monitoring participants' responses and potentially intervening should not be underestimated, especially for large studies where multiple participants may require intervention simultaneously. Ensuring that adequate staffing and resources are available to carry out the study's safety procedure is essential. In the interests of transparency and to evaluate safety procedures in real-time monitoring studies of NSSI, we recommend reporting details regarding participant safety protocols as standard. Moreover, qualitative research should substantively investigate participants' and clinicians' preferences for safety procedures concerning NSSI outcomes.

Responsibility 4: Measurement Reactivity

A particular concern in all NSSI and suicide research is that asking individuals questions about their self-harm–related thoughts and behaviors may inadvertently increase the likelihood of the individual thinking about or engaging in self-harm. However, evidence suggests that this is not the case [110,111], prospectively across young adult [112], adult, and adolescent samples [113,114], and when using various NSSI-related stimuli (eg, images, words) [113]. In contrast, findings indicate that participants find their participation in research on NSSI and suicidal behavior to be beneficial [110,112,114]. Although it appears that asking people about NSSI and suicide at a single time point has no impact on self-harm–related thoughts and behaviors, real-time monitoring requires repeated questioning on these topics. To date, no research has tested the potential iatrogenic effects of real-time monitoring research specifically for NSSI, but evidence from the suicide literature is promising. Early work by Husky and colleagues [107] used real-time monitoring to assess depression, mood, and thoughts of suicide and self-harm 5 times a day for 1 week in 4 samples: people with a recent suicide attempt, people with a past suicide attempt, people with mood disorders but no suicidal behavior, and healthy controls. In this study, there was no reactivity to the repeated questioning about self-harming thoughts across any of the 4 groups. Law and colleagues [105] demonstrated similar outcomes with a longitudinal design assessing 248 adults (30% of whom reported a borderline personality disorder diagnosis, which confers additional suicide risk). In this study, the authors found no increase in suicidal thoughts and behaviors during the initial 2-week data collection phase nor at the 6-month follow-up, including for people with a borderline personality disorder diagnosis. Recently, Coppersmith and colleagues [108] found no association between the frequency of asking about suicidal ideation and intent in a real-time monitoring study and the severity of suicidal thoughts over time. They also observed no change in survey responses when ideation was severe, where a decrease might be expected if participants were reactive to questioning. Currently, the evidence suggests no iatrogenic effects of repeated questioning about suicide and suicide-related behaviors. Further research is required to confirm whether this pattern also holds true for NSSI and is robust across different subgroups.

Responsibility 5: Balancing Scientific Accuracy Against Participant Burden and Ensuring the Research Is Feasible, Transparent, and Safe

An important responsibility when designing a real-time monitoring study is the selection of an appropriate sampling design (ie, fixed, interval, [semi-] random, event-based, mixed), sampling density (ie, number of assessments per day), sampling duration (ie, number of days/weeks), and sample size [45,94]. It is crucial that the selected sample shows sufficient variability in the outcomes of interest to allow investigation of the research question in daily life [45]. For example, when the aim is to clarify the transition from NSSI thoughts to behavior, base rates of thoughts and behavior must be high enough during the real-time monitoring period. To ensure this is the case, researchers need to consider the inclusion criteria carefully (eg, by including individuals with more than 5 acts of NSSI behavior in the 2 weeks before study onboarding). The selected protocol should allow answering the prespecified research question(s) without unnecessarily burdening the participants. Balancing these needs may mean that protocols are not interchangeable across studies—there is no "one-size-fits-all" protocol. Researchers should be cautious about, for example, adopting sampling densities used in previous studies, as these may not be suitable for addressing different research questions. In some cases, it will be perfectly justifiable—even essential—to ask participants to complete a more intense or extended real-time monitoring protocol. However, to ensure divergence between protocols across studies is not arbitrary, researchers must justify

their protocol [115] and communicate expectations to potential participants during study onboarding.

Compared with other research methodologies, real-time monitoring studies involve a higher workload for and burden on participants. Therefore, we recommend piloting the feasibility and acceptability of real-time monitoring protocols by using a quality improvement procedure, where protocols are first tested extensively by members of the research team and then iteratively with a selected group of participants. In our experience, this approach requires more time and planning but safeguards participants' (and researchers') investment by allowing the protocol to be modified and optimized, if needed, in response to qualitative and quantitative feedback (eg, the time it takes to complete the questionnaire). Emerging evidence suggests that the questionnaire length, rather than sampling frequency, is associated with increased participant burden and reduced data quality [116]. Although these findings are hopeful for researchers wanting to use more intense protocols, they also underscore the necessity for careful item conceptualization and selection. To avoid a methodological "Wild West," we advise researchers to make their items publicly available (see the Experience Sampling Method Item Repository) [117] and report (where possible) the multilevel reliability and validity of operationalized NSSI outcomes. Against the backdrop of the replication crisis, preregistration and data sharing (open data) are also increasingly used to increase reproducibility and avoid wasting public resources [118]. These considerations are especially relevant considering the high demands of a real-time monitoring study on participants and researchers. For a tutorial and template for the preregistration of real-time monitoring studies, see Kirtley and colleagues [119]. Finally, there is consensus among researchers and clinicians that real-time monitoring platforms should be secure and compliant with the relevant privacy laws [109]. In some cases, data collected with real-time monitoring apps may be stored in another country and thus subject to different privacy legislation from the researcher's home country. Researchers should thus be well-informed before commencing a real-time monitoring study and when unsure, should contact their institution's research governance or data management department for clarification.

Responsibility 6: Compensation and Recognition of Participant Engagement

Consideration should be given to distributive justice principles, such that participants benefit from their engagement in the study. Prior work found that the majority of young people who take part in traditional survey research investigating sensitive topics, such as NSSI, report that their participation allowed them the opportunity to develop greater self-awareness and had altruistic value [114]. However, real-time monitoring studies tend to require more time and effort from participants. To give some initial insight into the experience and expectations of participants who self-injure, we present additional results from a recent real-time monitoring study in which emerging adults were prompted every 90 minutes during waking hours for 12 days [35].

Following the monitoring period (n=29), approximately 4 of 5 participants indicated increased self-awareness of feelings and thoughts (Table 1). About half considered the protocol demands to be tiring, but most participants also described their participation as positive. Notably, all participants reported being interested in receiving information on the key findings of the study when it concluded. Although not a primary reason for participation for most participants, they also expected feedback on their own data (21/29, 72%) and a monetary incentive (23/29, 79%). Participants considered €60 (€1=US $1.15) to be a fair compensation for the study protocol's demands (8 beeps/day for 12 days, 96 assessments in total, median compliance 79.2%, IQR 70.3%-91.7%), when they also received feedback about the overall findings and their own data. In the absence of feedback on their data, participants expected higher financial compensation for their investment. In this study, participants were reimbursed according to a structured financial scheme to encourage participation. Rather than paying a fixed amount per completed survey, a structured incentive scheme has the advantage that it allows participants to miss some surveys without direct financial consequences. For some groups (eg, adolescents, people with low socioeconomic status), avoiding direct financial consequences of missed surveys may be particularly important to prevent some participants from changing their daily routines to respond to each survey.

These findings are consistent with those of the traditional survey research [114] and provide the first indication that real-time monitoring offers self-awareness opportunities for participants. Further, it highlights to researchers that they should actively recognize participants with lived experience of NSSI as valued partners in the research by also giving informational support. Although the majority experienced participation as positive, future work should clarify why this may not be the case for everyone so that resources can be provided to those for whom participation may have increased burden and discomfort. In this respect, it is advisable to organize a debriefing session that allows participants to share their experiences. Moreover, future work could explore the perceived meaningfulness of several strategies, such as the tailoring of sampling schedules and including a temporary "suspend" button to offer participants flexibility and greater control over notifications [115], the randomization of items to reduce response fatigue (called planned missing data designs) [120,121], and allowing for catch-up days so that participants can reach the desired level of compliance [94].

Table 1. Experience and expectation of individuals with lived experience of nonsuicidal self-injury (n=29).[a]

Experiences and expectations	Disagree to disagree completely	Agree to agree completely
Subjective experience, n (%)		
By completing the questions in everyday life, I became more aware of how I felt	4 (14)	23 (79)
By completing the questions in everyday life, I became more aware of my thoughts	3 (10)	21 (72)
Participating in the smartphone study was tiring	7 (24)	14 (48)
I would describe my participation as a positive experience	3 (10)	19 (66)
The overall importance of receiving study findings, personal feedback on own data, and financial compensation, n (%)		
I am interested in the overall results of the study	0 (0)	29 (100)
Receiving personal feedback is an important reason for me to participate	8 (28)	13 (45)
Receiving financial compensation is an important reason for me to participate	4 (14)	12 (41)
The relative importance of general and personal feedback and financial compensation, n (%)		
If I receive feedback on the overall results, financial compensation is not necessary	23 (79)	1 (3)
If I receive feedback on the overall results, personal feedback is not necessary	21 (72)	5 (17)
If I receive financial compensation, personal feedback is not necessary	23 (79)	2 (7)
If I receive personal feedback, financial compensation is not necessary	18 (62)	5 (17)
Expected financial compensation (Euros), €1=US $1.15, median (IQR)		
What amount do you consider fair as compensation when you also receive general and personal feedback?	N/A[b]	60 (45.0-75.5)
What amount do you consider fair as compensation when you receive general but not personal feedback?	N/A	70 (55.0-81.5)

[a]Unpublished data of 29 young adults with lived experience following participation in a 12-day real-time monitoring protocol with 96 semirandom longitudinal assessments (8/day, median compliance 79.2%; IQR 70.3%-91.7%; 29/30, 97% retention [35]). The response category "neutral" is not shown in the table.

[b]N/A: not applicable.

Responsibility 7: Researcher Well-being and Training

Finally, we want to draw explicit attention to researcher well-being and training considerations. Real-time monitoring data are quite literally an individual's real-life experiences, occurring in real-time, which may lead researchers to feel especially close to participants and their experiences. Although studies have explored participants' experiences of taking part in research on self-harm and suicide more broadly [122], as well as in real-time monitoring studies [123,124], to our knowledge, no studies have investigated researchers' experiences of conducting real-time monitoring studies of NSSI or suicidal behaviors. However, the stakes are high when the research focus is on self-injurious behavior; monitoring participants' responses for signs of imminent suicidal crisis lays great responsibility on researchers' shoulders. Where safety protocols involve routine telephone check-ins, a missed check-in may cause the researcher to fear for the participant's safety even though there may be an innocuous explanation (eg, the participant was driving or in the shower). Researchers may also begin to feel a high level of responsibility for participants not engaging in NSSI or attempting suicide (ie, feeling they are keeping the participant alive or safe) [125,126]. Studies that employ real-time alerts when participants indicate escalating suicidal intent may lead researchers to become hypervigilant and worry that intervention may be required at any moment. Such warnings may come outside of working hours, thereby increasing work-life balance challenges. For the researcher, all of these situations underscore a lack of controllability, which may prove highly stressful, especially when experienced over a sustained period.

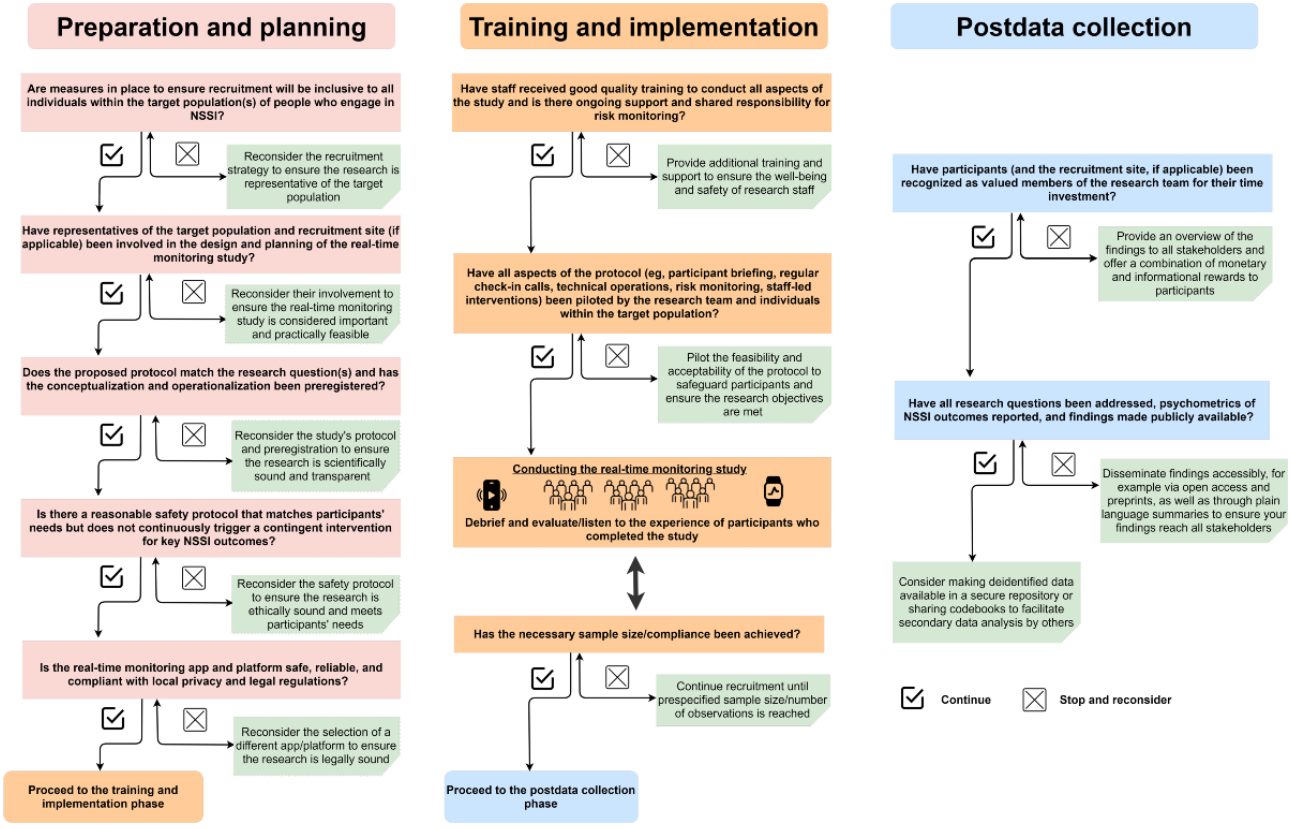
Good quality training in working with individuals who engage in NSSI and suicidal behaviors is essential when conducting real-time monitoring research. Much research is carried out by doctoral students, trainees, and research assistants, who may be less experienced in collecting sensitive data and managing the accompanying emotional labor [126]. Researchers often work alone and lack a good support network [127]. Therefore, training should cover supporting a participant during an acute suicidal crisis as well as less overtly "intense" situations (eg, where researchers are coding open-text responses about participants' reasons for engaging in NSSI). A lone researcher should never carry out safety protocols involving real-time risk-monitoring and intervention and responsibility should be shared among a

team, whereby multiple researchers are "on-call" for a specified period. Although planning is essential, some researcher well-being challenges may only become known once the study is underway [126]. Continued supervision and mentoring of researchers (of all career stages) are crucial to ensuring researcher well-being. This may involve regular debriefings with supervisors or colleagues or via external, independent counseling support. Qualitative studies with researchers who work on sensitive topics also highlight the importance of self-care and actively engaging in positive, non–work-related activities as valuable buffers against emotional distress [126,127].

Summary of Responsibilities

Real-time monitoring technologies give researchers the practical tools to investigate NSSI thoughts, urges, and behavior as they occur, without the need to be physically present. Although this provides immense opportunities for NSSI research, real-time monitoring research is not without challenges regarding recruitment, study enrolment and planning, and participant engagement. Focusing on NSSI also creates great responsibility for privacy and data security, participant safety and risk-monitoring, and researcher well-being and training. To guide researchers who want to study the experiences of individuals who engage in NSSI, we summarize the responsibilities and ways of overcoming challenges in a functional flowchart (Figure 1).

Figure 1. Flowchart of the critical considerations when opening the black box of daily life in nonsuicidal self-injury research. NSSI: nonsuicidal self-injury.



Where Do We Go From Here?

Since Nock et al's seminal study in 2009 [27], which demonstrated the feasibility of studying NSSI in adolescents' everyday life, researchers now have a toolbox full of smartphone apps and wearable technology that can readily capture real-time experiences of people who engage in NSSI in the real-world context. These advances produce a rapidly growing literature that could positively shape the field's future trajectory by facilitating a radical shift of focus from the group to the individual, from the research lab and clinic to the everyday life environment, and from traditional generalized treatment to person-centered prevention and intervention. This paper sets an ambitious agenda for new research pathways to realize such a shift as we move into the next decade of NSSI research. We also offered critical perspectives on the inevitable ethical and practical challenges that come with these research pathways. In this respect, opening the black box of daily life in NSSI research is truly a double-edged sword that requires responsibility and leads to new questions. Few studies to date have specifically considered ethical issues within real-time monitoring studies of NSSI. However, these questions regarding ethical practices must be addressed with substantive empirical research rather than being based upon the precedent of "if it ain't broke, don't fix it." In particular, research co-design and qualitative studies to capture rich information on participants' experiences of taking part in real-time monitoring research on NSSI offer promise. Real-time monitoring research on NSSI will advance better and more rapidly when all stakeholders' interests (ie, individuals with lived experience, their families,

researchers, and clinicians) are considered. Only by considering both the opportunities and the challenges will we be able to use real-time monitoring techniques to their full potential.

Acknowledgments

The authors want to thank Dr Laurence Claes for providing comments on the earlier version of this manuscript. This research was supported by postdoctoral fellowships from the Research Foundation Flanders awarded to the first and last author (12ZZM21N [GK], 1257821N [OJK]) and by the Royal Society of New Zealand Marsden Fund (18-VUW-162) awarded to the second author (KR).

Authors' Contributions

All authors contributed to the conception and design of the manuscript. GK drafted the opportunity section and is responsible for the results presented in this manuscript. GK, KR, RT, and OJK drafted the responsibility section and provided critical revisions to the manuscript. All authors approved the final version of the manuscript for submission.

Conflicts of Interest

OJK reports grants from the Union Chimique Belge Community Health Fund (managed and disbursed by the King Baudouin Foundation, Belgium) outside the submitted work. The other authors have no competing interests to declare.

References

1. What is nonsuicidal self-injury? International Society for the Study of Self-Injury. URL: https://itriples.org/about-self-injury/what-is-self-injury [accessed 2021-05-31]
2. Gandhi A, Luyckx K, Baetens I, Kiekens G, Sleuwaegen E, Berens A, et al. Age of onset of non-suicidal self-injury in Dutch-speaking adolescents and emerging adults: An event history analysis of pooled data. Compr Psychiatry 2018 Jan;80:170-178. [doi: 10.1016/j.comppsych.2017.10.007] [Medline: 29121554]
3. Gillies D, Christou MA, Dixon AC, Featherston OJ, Rapti I, Garcia-Anguita A, et al. Prevalence and Characteristics of Self-Harm in Adolescents: Meta-Analyses of Community-Based Studies 1990-2015. J Am Acad Child Adolesc Psychiatry 2018 Oct;57(10):733-741. [doi: 10.1016/j.jaac.2018.06.018] [Medline: 30274648]
4. Kiekens G, Hasking P, Boyes M, Claes L, Mortier P, Auerbach R, et al. The associations between non-suicidal self-injury and first onset suicidal thoughts and behaviors. J Affect Disord 2018 Oct 15;239:171-179. [doi: 10.1016/j.jad.2018.06.033] [Medline: 30014957]
5. Wilkinson PO, Qiu T, Neufeld S, Jones PB, Goodyer IM. Sporadic and recurrent non-suicidal self-injury before age 14 and incident onset of psychiatric disorders by 17 years: prospective cohort study. Br J Psychiatry 2018 Apr;212(4):222-226 [FREE Full text] [doi: 10.1192/bjp.2017.45] [Medline: 29514726]
6. Buelens T, Luyckx K, Gandhi A, Kiekens G, Claes L. Non-Suicidal Self-Injury in Adolescence: Longitudinal Associations with Psychological Distress and Rumination. J Abnorm Child Psychol 2019 Sep;47(9):1569-1581. [doi: 10.1007/s10802-019-00531-8] [Medline: 30900112]
7. Daukantaitė D, Lundh L, Wångby-Lundh M, Claréus B, Bjärehed J, Zhou Y, et al. What happens to young adults who have engaged in self-injurious behavior as adolescents? A 10-year follow-up. Eur Child Adolesc Psychiatry 2021 Mar;30(3):475-492 [FREE Full text] [doi: 10.1007/s00787-020-01533-4] [Medline: 32318877]
8. Robinson K, Garisch JA, Kingi T, Brocklesby M, O'Connell A, Langlands RL, et al. Reciprocal Risk: the Longitudinal Relationship between Emotion Regulation and Non-suicidal Self-Injury in Adolescents. J Abnorm Child Psychol 2019 Feb;47(2):325-332. [doi: 10.1007/s10802-018-0450-6] [Medline: 29923162]
9. Whitlock J, Muehlenkamp J, Purington A, Eckenrode J, Barreira P, Baral Abrams G, et al. Nonsuicidal self-injury in a college population: general trends and sex differences. J Am Coll Health 2011;59(8):691-698. [doi: 10.1080/07448481.2010.529626] [Medline: 21950249]
10. Steinhoff A, Ribeaud D, Kupferschmid S, Raible-Destan N, Quednow BB, Hepp U, et al. Self-injury from early adolescence to early adulthood: age-related course, recurrence, and services use in males and females from the community. Eur Child Adolesc Psychiatry 2021 Jun;30(6):937-951 [FREE Full text] [doi: 10.1007/s00787-020-01573-w] [Medline: 32572615]
11. The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Arlington, VA: American Psychiatric Association; 2013.
12. Taylor PJ, Jomar K, Dhingra K, Forrester R, Shahmalak U, Dickson JM. A meta-analysis of the prevalence of different functions of non-suicidal self-injury. J Affect Disord 2018 Feb;227:759-769 [FREE Full text] [doi: 10.1016/j.jad.2017.11.073] [Medline: 29689691]
13. Kiekens G, Hasking P, Claes L, Boyes M, Mortier P, Auerbach RP, et al. Predicting the incidence of non-suicidal self-injury in college students. Eur Psychiatry 2019 Jun;59:44-51. [doi: 10.1016/j.eurpsy.2019.04.002] [Medline: 31035219]

14. Victor SE, Hipwell AE, Stepp SD, Scott LN. Parent and peer relationships as longitudinal predictors of adolescent non-suicidal self-injury onset. Child Adolesc Psychiatry Ment Health 2019;13:1 [FREE Full text] [doi: 10.1186/s13034-018-0261-0] [Medline: 30622642]
15. Kiekens G, Hasking P, Bruffaerts R, Claes L, Baetens I, Boyes M, et al. What Predicts Ongoing Nonsuicidal Self-Injury?: A Comparison Between Persistent and Ceased Self-Injury in Emerging Adults. J Nerv Ment Dis 2017 Oct;205(10):762-770. [doi: 10.1097/NMD.0000000000000726] [Medline: 28817427]
16. Fox KR, Huang X, Linthicum KP, Wang SB, Franklin JC, Ribeiro JD. Model complexity improves the prediction of nonsuicidal self-injury. J Consult Clin Psychol 2019 Aug;87(8):684-692. [doi: 10.1037/ccp0000421] [Medline: 31219275]
17. Fisher AJ, Medaglia JD, Jeronimus BF. Lack of group-to-individual generalizability is a threat to human subjects research. Proc Natl Acad Sci U S A 2018 Jul 03;115(27):E6106-E6115 [FREE Full text] [doi: 10.1073/pnas.1711978115] [Medline: 29915059]
18. Piccirillo ML, Rodebaugh TL. Foundations of idiographic methods in psychology and applications for psychotherapy. Clin Psychol Rev 2019 Jul;71:90-100. [doi: 10.1016/j.cpr.2019.01.002] [Medline: 30665765]
19. Liu RT, Scopelliti KM, Pittman SK, Zamora AS. Childhood maltreatment and non-suicidal self-injury: a systematic review and meta-analysis. Lancet Psychiatry 2018 Jan;5(1):51-64 [FREE Full text] [doi: 10.1016/S2215-0366(17)30469-8] [Medline: 29196062]
20. Fox KR, Franklin JC, Ribeiro JD, Kleiman EM, Bentley KH, Nock MK. Meta-analysis of risk factors for nonsuicidal self-injury. Clin Psychol Rev 2015 Dec;42:156-167 [FREE Full text] [doi: 10.1016/j.cpr.2015.09.002] [Medline: 26416295]
21. Myin-Germeys I, Kasanova Z, Vaessen T, Vachon H, Kirtley O, Viechtbauer W, et al. Experience sampling methodology in mental health research: new insights and technical developments. World Psychiatry 2018 Jun;17(2):123-132 [FREE Full text] [doi: 10.1002/wps.20513] [Medline: 29856567]
22. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
23. Bolger N, Laurenceau J. Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research. New York: Guilford Press; 2013.
24. Larson R, Csikszentmihalyi M. The experience sampling method. In: Flow and the Foundations of Positive Psychology. Dordrecht: Springer; 2014:21-34.
25. Barker RG. Ecological psychology: Concepts and Methods for Studying the Environment of Human Behavior. Stanford: Stanford University Press; 1968.
26. Muehlenkamp JJ, Engel SG, Wadeson A, Crosby RD, Wonderlich SA, Simonich H, et al. Emotional states preceding and following acts of non-suicidal self-injury in bulimia nervosa patients. Behav Res Ther 2009 Jan;47(1):83-87 [FREE Full text] [doi: 10.1016/j.brat.2008.10.011] [Medline: 19027892]
27. Nock MK, Prinstein MJ, Sterba SK. Revealing the form and function of self-injurious thoughts and behaviors: A real-time ecological assessment study among adolescents and young adults. J Abnorm Psychol 2009 Nov;118(4):816-827 [FREE Full text] [doi: 10.1037/a0016948] [Medline: 19899851]
28. Smartphones-statistics and facts. Statista. URL: https://www.statista.com/topics/840/smartphones/ [accessed 2021-05-31]
29. Rodríguez-Blanco L, Carballo JJ, Baca-García E. Use of Ecological Momentary Assessment (EMA) in Non-Suicidal Self-Injury (NSSI): A systematic review. Psychiatry Res 2018 May;263:212-219. [doi: 10.1016/j.psychres.2018.02.051] [Medline: 29574356]
30. Hepp J, Carpenter RW, Störkel LM, Schmitz SE, Schmahl C, Niedtfeld I. A systematic review of daily life studies on non-suicidal self-injury based on the four-function model. Clin Psychol Rev 2020 Dec;82:101888. [doi: 10.1016/j.cpr.2020.101888] [Medline: 32949907]
31. Gee BL, Han J, Benassi H, Batterham PJ. Suicidal thoughts, suicidal behaviours and self-harm in daily life: A systematic review of ecological momentary assessment studies. Digit Health 2020;6:2055207620963958 [FREE Full text] [doi: 10.1177/2055207620963958] [Medline: 33224516]
32. Mission statement. International Society for the Study of Self-Injury. URL: https://itriples.org/ [accessed 2021-05-31]
33. Fitzpatrick S, Kranzler A, Fehling K, Lindqvist J, Selby EA. Investigating the role of the intensity and duration of self-injury thoughts in self-injury with ecological momentary assessment. Psychiatry Res 2020 Feb;284:112761. [doi: 10.1016/j.psychres.2020.112761] [Medline: 31951869]
34. Turner BJ, Baglole JS, Chapman AL, Gratz KL. Experiencing and Resisting Nonsuicidal Self-injury Thoughts and Urges in Everyday Life. Suicide Life Threat Behav 2019 Oct;49(5):1332-1346. [doi: 10.1111/sltb.12510] [Medline: 30152181]
35. Kiekens G, Hasking P, Nock MK, Boyes M, Kirtley O, Bruffaerts R, et al. Fluctuations in Affective States and Self-Efficacy to Resist Non-Suicidal Self-Injury as Real-Time Predictors of Non-Suicidal Self-Injurious Thoughts and Behaviors. Front Psychiatry 2020;11:214 [FREE Full text] [doi: 10.3389/fpsyt.2020.00214] [Medline: 32265760]
36. Turner BJ, Yiu A, Claes L, Muehlenkamp JJ, Chapman AL. Occurrence and co-occurrence of nonsuicidal self-injury and disordered eating in a daily diary study: Which behavior, when? Psychiatry Res 2016 Dec 30;246:39-47. [doi: 10.1016/j.psychres.2016.09.012] [Medline: 27657802]

37. Victor SE, Scott LN, Stepp SD, Goldstein TR. I Want You to Want Me: Interpersonal Stress and Affective Experiences as Within-Person Predictors of Nonsuicidal Self-Injury and Suicide Urges in Daily Life. Suicide Life Threat Behav 2019 Aug;49(4):1157-1177 [FREE Full text] [doi: 10.1111/sltb.12513] [Medline: 30159910]
38. Andrewes HE, Hulbert C, Cotton SM, Betts J, Chanen AM. Ecological momentary assessment of nonsuicidal self-injury in youth with borderline personality disorder. Personal Disord 2017 Oct;8(4):357-365. [doi: 10.1037/per0000205] [Medline: 27505188]
39. Burke TA, Fox K, Kautz M, Siegel DM, Kleiman E, Alloy LB. Real-time monitoring of the associations between self-critical and self-punishment cognitions and nonsuicidal self-injury. Behav Res Ther 2021 Feb;137:103775. [doi: 10.1016/j.brat.2020.103775] [Medline: 33421892]
40. Hughes CD, King AM, Kranzler A, Fehling K, Miller A, Lindqvist J, et al. Anxious and Overwhelming Affects and Repetitive Negative Thinking as Ecological Predictors of Self-Injurious Thoughts and Behaviors. Cogn Ther Res 2019 Jan 9;43(1):88-101. [doi: 10.1007/s10608-019-09996-9]
41. Kiekens G, Claes L. Non-Suicidal Self-Injury and Eating Disordered Behaviors: An Update on What We Do and Do Not Know. Curr Psychiatry Rep 2020 Oct 10;22(12):68 [FREE Full text] [doi: 10.1007/s11920-020-01191-y] [Medline: 33037934]
42. Czyz EK, Glenn CR, Arango A, Koo HJ, King CA. Short-term associations between nonsuicidal and suicidal thoughts and behaviors: A daily diary study with high-risk adolescents. J Affect Disord 2021 Sep 01;292:337-344. [doi: 10.1016/j.jad.2021.05.104] [Medline: 34139406]
43. Fisher A, Soyster P. Generating accurate personalized predictions of future behavior: A smoking exemplar. PsyArXiv Preprints 2021 [FREE Full text] [doi: 10.31234/osf.io/e24v6]
44. Burke TA, Ammerman BA, Jacobucci R. The use of machine learning in the study of suicidal and non-suicidal self-injurious thoughts and behaviors: A systematic review. J Affect Disord 2019 Feb 15;245:869-884. [doi: 10.1016/j.jad.2018.11.073] [Medline: 30699872]
45. Stange JP, Kleiman EM, Mermelstein RJ, Trull TJ. Using ambulatory assessment to measure dynamic risk processes in affective disorders. J Affect Disord 2019 Dec 01;259:325-336 [FREE Full text] [doi: 10.1016/j.jad.2019.08.060] [Medline: 31610996]
46. Allen NB, Nelson BW, Brent D, Auerbach RP. Short-term prediction of suicidal thoughts and behaviors in adolescents: Can recent developments in technology and computational science provide a breakthrough? J Affect Disord 2019 May 01;250:163-169 [FREE Full text] [doi: 10.1016/j.jad.2019.03.044] [Medline: 30856493]
47. Kleiman E, Millner AJ, Joyce VW, Nash CC, Buonopane RJ, Nock MK. Using Wearable Physiological Monitors With Suicidal Adolescent Inpatients: Feasibility and Acceptability Study. JMIR Mhealth Uhealth 2019 Sep 24:e0 [FREE Full text] [doi: 10.2196/13725] [Medline: 31586364]
48. Selby EA, Anestis MD, Joiner TE. Understanding the relationship between emotional and behavioral dysregulation: emotional cascades. Behav Res Ther 2008 May;46(5):593-611. [doi: 10.1016/j.brat.2008.02.002] [Medline: 18353278]
49. Hooley JM, Franklin JC. Why Do People Hurt Themselves? A New Conceptual Model of Nonsuicidal Self-Injury. Clinical Psychological Science 2017 Dec 28;6(3):428-451. [doi: 10.1177/2167702617745641]
50. Hasking P, Whitlock J, Voon D, Rose A. A cognitive-emotional model of NSSI: using emotion regulation and cognitive processes to explain why people self-injure. Cogn Emot 2017 Dec;31(8):1543-1556. [doi: 10.1080/02699931.2016.1241219] [Medline: 27702245]
51. Hasking P, Boyes M, Greves S. Self-efficacy and emotionally dysregulated behaviour: An exploratory test of the role of emotion regulatory and behaviour-specific beliefs. Psychiatry Res 2018 Dec;270:335-340. [doi: 10.1016/j.psychres.2018.09.045] [Medline: 30292086]
52. Burke TA, Fox K, Kautz MM, Rodriguez-Seijas C, Bettis AH, Alloy LB. Self-Critical and Self-Punishment Cognitions Differentiate Those With and Without a History of Nonsuicidal Self-Injury: An Ecological Momentary Assessment Study. Behav Ther 2021 May;52(3):686-697. [doi: 10.1016/j.beth.2020.08.006] [Medline: 33990242]
53. Wright AGC, Woods WC. Personalized Models of Psychopathology. Annu Rev Clin Psychol 2020 May 07;16:49-74. [doi: 10.1146/annurev-clinpsy-102419-125032] [Medline: 32070120]
54. Borsboom D, van der Maas HLJ, Dalege J, Kievit RA, Haig BD. Theory Construction Methodology: A Practical Framework for Building Theories in Psychology. Perspect Psychol Sci 2021 Jul;16(4):756-766. [doi: 10.1177/1745691620969647] [Medline: 33593167]
55. Millner AJ, Robinaugh DJ, Nock MK. Advancing the Understanding of Suicide: The Need for Formal Theory and Rigorous Descriptive Research. Trends Cogn Sci 2020 Sep;24(9):704-716 [FREE Full text] [doi: 10.1016/j.tics.2020.06.007] [Medline: 32680678]
56. Guest O, Martin AE. How Computational Modeling Can Force Theory Building in Psychological Science. Perspect Psychol Sci 2021 Jul;16(4):789-802. [doi: 10.1177/1745691620970585] [Medline: 33482070]
57. Lane ST, Gates KM, Pike HK, Beltz AM, Wright AGC. Uncovering general, shared, and unique temporal patterns in ambulatory assessment data. Psychol Methods 2019 Mar;24(1):54-69 [FREE Full text] [doi: 10.1037/met0000192] [Medline: 30124300]

58. Wright AGC, Zimmermann J. Applied ambulatory assessment: Integrating idiographic and nomothetic principles of measurement. Psychol Assess 2019 Dec;31(12):1467-1480 [FREE Full text] [doi: 10.1037/pas0000685] [Medline: 30896209]
59. Nock MK, Prinstein MJ. A functional approach to the assessment of self-mutilative behavior. J Consult Clin Psychol 2004 Oct;72(5):885-890. [doi: 10.1037/0022-006X.72.5.885] [Medline: 15482046]
60. Bentley KH, Nock MK, Barlow DH. The Four-Function Model of Nonsuicidal Self-Injury. Clinical Psychological Science 2014 Jan 03;2(5):638-656. [doi: 10.1177/2167702613514563]
61. Kranzler A, Fehling KB, Lindqvist J, Brillante J, Yuan F, Gao X, et al. An Ecological Investigation of the Emotional Context Surrounding Nonsuicidal Self-Injurious Thoughts and Behaviors in Adolescents and Young Adults. Suicide Life Threat Behav 2018 Apr;48(2):149-159. [doi: 10.1111/sltb.12373] [Medline: 28758704]
62. Snir A, Rafaeli E, Gadassi R, Berenson K, Downey G. Explicit and inferred motives for nonsuicidal self-injurious acts and urges in borderline and avoidant personality disorders. Personal Disord 2015 Jul;6(3):267-277 [FREE Full text] [doi: 10.1037/per0000104] [Medline: 25867834]
63. Koenig J, Klier J, Parzer P, Santangelo P, Resch F, Ebner-Priemer U, et al. High-frequency ecological momentary assessment of emotional and interpersonal states preceding and following self-injury in female adolescents. Eur Child Adolesc Psychiatry 2021 Aug;30(8):1299-1308 [FREE Full text] [doi: 10.1007/s00787-020-01626-0] [Medline: 32862250]
64. Voelkle MC, Gische C, Driver CC, Lindenberger U. The Role of Time in the Quest for Understanding Psychological Mechanisms. Multivariate Behav Res 2018;53(6):782-805. [doi: 10.1080/00273171.2018.1496813] [Medline: 30668172]
65. Burke TA, Hamilton JL, Abramson LY, Alloy LB. Non-suicidal self-injury prospectively predicts interpersonal stressful life events and depressive symptoms among adolescent girls. Psychiatry Res 2015 Aug 30;228(3):416-424 [FREE Full text] [doi: 10.1016/j.psychres.2015.06.021] [Medline: 26165966]
66. Waals L, Baetens I, Rober P, Lewis S, Van Parys H, Goethals ER, et al. The NSSI Family Distress Cascade Theory. Child Adolesc Psychiatry Ment Health 2018;12:52 [FREE Full text] [doi: 10.1186/s13034-018-0259-7] [Medline: 30568727]
67. Turner BJ, Cobb RJ, Gratz KL, Chapman AL. The role of interpersonal conflict and perceived social support in nonsuicidal self-injury in daily life. J Abnorm Psychol 2016 May;125(4):588-598 [FREE Full text] [doi: 10.1037/abn0000141] [Medline: 26845256]
68. Taylor PJ, McDonald J, Smith M, Nicholson H, Forrester R. Distinguishing people with current, past, and no history of non-suicidal self-injury: Shame, social comparison, and self-concept integration. J Affect Disord 2019 Mar 01;246:182-188. [doi: 10.1016/j.jad.2018.12.033] [Medline: 30583143]
69. Burke TA, Piccirillo ML, Moore-Berg SL, Alloy LB, Heimberg RG. The stigmatization of nonsuicidal self-injury. J Clin Psychol 2019 Mar;75(3):481-498 [FREE Full text] [doi: 10.1002/jclp.22713] [Medline: 30368804]
70. Staniland L, Hasking P, Boyes M, Lewis S. Stigma and nonsuicidal self-injury: Application of a conceptual framework. Stigma and Health 2021 Aug;6(3):312-323. [doi: 10.1037/sah0000257]
71. Nestler S. Modeling Intraindividual Variability in Growth with Measurement Burst Designs. Structural Equation Modeling: A Multidisciplinary Journal 2020 Jul 09;28(1):28-39. [doi: 10.1080/10705511.2020.1757455]
72. Buu A, Cai Z, Li R, Wong S, Lin H, Su W, et al. The association between short-term emotion dynamics and cigarette dependence: A comprehensive examination of dynamic measures. Drug Alcohol Depend 2021 Jan 01;218:108341. [doi: 10.1016/j.drugalcdep.2020.108341] [Medline: 33268228]
73. Bean CAL, Heggeness LF, Kalmbach DA, Ciesla JA. Ruminative Inertia and Its Association With Current Severity and Lifetime Course of Depression. Clinical Psychological Science 2020 Sep 15;8(6):1007-1016. [doi: 10.1177/2167702620949174]
74. Lewis SP, Hasking PA. Self-injury recovery: A person-centered framework. J Clin Psychol 2021 Apr;77(4):884-895. [doi: 10.1002/jclp.23094] [Medline: 33296508]
75. Giletta M, Prinstein MJ, Abela JRZ, Gibb BE, Barrocas AL, Hankin BL. Trajectories of suicide ideation and nonsuicidal self-injury among adolescents in mainland China: Peer predictors, joint development, and risk for suicide attempts. J Consult Clin Psychol 2015 Apr;83(2):265-279. [doi: 10.1037/a0038652] [Medline: 25688812]
76. Robinson K, Garisch J, Wilson M. Nonsuicidal self-injury thoughts and behavioural characteristics: Associations with suicidal thoughts and behaviours among community adolescents. J Affect Disord 2021 Mar 01;282:1247-1254. [doi: 10.1016/j.jad.2020.12.201] [Medline: 33601703]
77. m-Path – free tool for blended care by KU Leuven. 2021. URL: https://m-path.io/landing/ [accessed 2021-11-11]
78. Bos FM, Snippe E, Bruggeman R, Wichers M, van der Krieke L. Insights of Patients and Clinicians on the Promise of the Experience Sampling Method for Psychiatric Care. Psychiatr Serv 2019 Nov 01;70(11):983-991. [doi: 10.1176/appi.ps.201900050] [Medline: 31434558]
79. Dewa LH, Lavelle M, Pickles K, Kalorkoti C, Jaques J, Pappa S, et al. Young adults' perceptions of using wearables, social media and other technologies to detect worsening mental health: A qualitative study. PLoS One 2019;14(9):e0222655 [FREE Full text] [doi: 10.1371/journal.pone.0222655] [Medline: 31532786]
80. Armey MF. Ecological Momentary Assessment and Intervention in Nonsuicidal Self-Injury: A Novel Approach to Treatment. J Cogn Psychother 2012 Jan 01;26(4):299-317. [doi: 10.1891/0889-8391.26.4.299]

81. Arshad U, Farhat-Ul-Ain, Gauntlett J, Husain N, Chaudhry N, Taylor PJ. A Systematic Review of the Evidence Supporting Mobile- and Internet-Based Psychological Interventions For Self-Harm. Suicide Life Threat Behav 2020 Feb;50(1):151-179 [FREE Full text] [doi: 10.1111/sltb.12583] [Medline: 31448847]
82. Kruzan KP, Whitlock J, Bazarova NN. Examining the Relationship Between the Use of a Mobile Peer-Support App and Self-Injury Outcomes: Longitudinal Mixed Methods Study. JMIR Ment Health 2021 Jan 28;8(1):e21854 [FREE Full text] [doi: 10.2196/21854] [Medline: 33507154]
83. Rizvi SL, Hughes CD, Thomas MC. The DBT Coach mobile application as an adjunct to treatment for suicidal and self-injuring individuals with borderline personality disorder: A preliminary evaluation and challenges to client utilization. Psychol Serv 2016 Nov;13(4):380-388. [doi: 10.1037/ser0000100] [Medline: 27797571]
84. Wang L, Miller LC. Just-in-the-Moment Adaptive Interventions (JITAI): A Meta-Analytical Review. Health Commun 2020 Nov;35(12):1531-1544. [doi: 10.1080/10410236.2019.1652388] [Medline: 31488002]
85. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in Mobile Health: Key Components and Design Principles for Ongoing Health Behavior Support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
86. Coppersmith D, Dempsey W, Kleiman E, Bentley K, Murphy S, Nock M. Just-in-time adaptive interventions for suicide prevention: Promise, challenges, and future directions. PsyArXiv Preprints 2021 [FREE Full text] [doi: 10.31234/osf.io/eg9fx]
87. Liu RT, Sheehan AE, Walsh RFL, Sanzari CM, Cheek SM, Hernandez EM. Prevalence and correlates of non-suicidal self-injury among lesbian, gay, bisexual, and transgender individuals: A systematic review and meta-analysis. Clin Psychol Rev 2019 Dec;74:101783 [FREE Full text] [doi: 10.1016/j.cpr.2019.101783] [Medline: 31734440]
88. Kuentzel JG, Arble E, Boutros N, Chugani D, Barnett D. Nonsuicidal self-injury in an ethnically diverse college sample. Am J Orthopsychiatry 2012 Jul;82(3):291-297. [doi: 10.1111/j.1939-0025.2012.01167.x] [Medline: 22880967]
89. Rad MS, Martingano AJ, Ginges J. Toward a psychology of : Making psychological science more representative of the human population. Proc Natl Acad Sci U S A 2018 Nov 06;115(45):11401-11405 [FREE Full text] [doi: 10.1073/pnas.1721165115] [Medline: 30397114]
90. An S, Cruwys T, Lee H, Chang MX. Cultural Differences in Reactions to Suicidal Ideation: A Mixed Methods Comparison of Korea and Australia. Arch Suicide Res 2020;24(3):415-434. [doi: 10.1080/13811118.2019.1624668] [Medline: 31159683]
91. Gandhi A, Luyckx K, Adhikari A, Parmar D, Desousa A, Shah N, et al. Non-suicidal self-injury and its association with identity formation in India and Belgium: A cross-cultural case-control study. Transcult Psychiatry 2021 Feb;58(1):52-62. [doi: 10.1177/1363461520933759] [Medline: 32873190]
92. Hunter BH, Radoll PJ. Dynamics of Digital Diffusion and Disadoption. AJIS 2020 Aug 03;24:1-21. [doi: 10.3127/ajis.v24i0.1805]
93. Vangeepuram N, Mayer V, Fei K, Hanlen-Rosado E, Andrade C, Wright S, et al. Smartphone ownership and perspectives on health apps among a vulnerable population in East Harlem, New York. Mhealth 2018;4:31 [FREE Full text] [doi: 10.21037/mhealth.2018.07.02] [Medline: 30221166]
94. van Roekel E, Keijsers L, Chung JM. A Review of Current Ambulatory Assessment Studies in Adolescent Samples and Practical Recommendations. J Res Adolesc 2019 Sep;29(3):560-577. [doi: 10.1111/jora.12471] [Medline: 31573762]
95. Hasking PA, Lewis SP, Robinson K, Heath NL, Wilson MS. Conducting research on nonsuicidal self-injury in schools: Ethical considerations and recommendations. School Psychology International 2019 Feb 06;40(3):217-234. [doi: 10.1177/0143034319827056]
96. Jacobson NC, Bentley KH, Walton A, Wang SB, Fortgang RG, Millner AJ, et al. Ethical dilemmas posed by mobile health and machine learning in psychiatry research. Bull World Health Organ 2020 Apr 01;98(4):270-276 [FREE Full text] [doi: 10.2471/BLT.19.237107] [Medline: 32284651]
97. Rintala A, Wampers M, Myin-Germeys I, Viechtbauer W. Response compliance and predictors thereof in studies using the experience sampling method. Psychol Assess 2019 Mar;31(2):226-235. [doi: 10.1037/pas0000662] [Medline: 30394762]
98. Nebeker C, Torous J, Bartlett Ellis RJ. Building the case for actionable ethics in digital health research supported by artificial intelligence. BMC Med 2019 Jul 17;17(1):137 [FREE Full text] [doi: 10.1186/s12916-019-1377-7] [Medline: 31311535]
99. Hasking P, Rees CS, Martin G, Quigley J. What happens when you tell someone you self-injure? The effects of disclosing NSSI to adults and peers. BMC Public Health 2015 Oct 09;15:1039 [FREE Full text] [doi: 10.1186/s12889-015-2383-0] [Medline: 26453187]
100. Whitlock J, Lloyd-Richardson E, Fisseha F, Bates T. Parental Secondary Stress: The Often Hidden Consequences of Nonsuicidal Self-Injury in Youth. J Clin Psychol 2018 Jan;74(1):178-196. [doi: 10.1002/jclp.22488] [Medline: 28493555]
101. Bai S, Babeva KN, Kim MI, Asarnow JR. J Clin Child Adolesc Psychol 2021;50(1):141-153. [doi: 10.1080/15374416.2020.1815208] [Medline: 33047987]
102. Lewis SP. I cut therefore I am? Avoiding labels in the context of self-injury. Med Humanit 2017 Sep;43(3):204. [doi: 10.1136/medhum-2017-011221] [Medline: 28264901]
103. Hasking P, Boyes M. Cutting Words: A Commentary on Language and Stigma in the Context of Nonsuicidal Self-Injury. J Nerv Ment Dis 2018 Nov;206(11):829-833. [doi: 10.1097/NMD.0000000000000899] [Medline: 30371637]
104. Davidson CL, Anestis MD, Gutierrez PM. Ecological Momentary Assessment is a Neglected Methodology in Suicidology. Arch Suicide Res 2017 Jan 02;21(1):1-11. [doi: 10.1080/13811118.2015.1004482] [Medline: 26821811]

105. Law MK, Furr RM, Arnold EM, Mneimne M, Jaquett C, Fleeson W. Does assessing suicidality frequently and repeatedly cause harm? A randomized control study. Psychol Assess 2015 Dec;27(4):1171-1181 [FREE Full text] [doi: 10.1037/pas0000118] [Medline: 25894705]
106. Selby EA, Kranzler A, Lindqvist J, Fehling KB, Brillante J, Yuan F, et al. The Dynamics of Pain During Nonsuicidal Self-Injury. Clinical Psychological Science 2018 Oct 24;7(2):302-320. [doi: 10.1177/2167702618807147]
107. Husky M, Olié E, Guillaume S, Genty C, Swendsen J, Courtet P. Feasibility and validity of ecological momentary assessment in the investigation of suicide risk. Psychiatry Res 2014 Dec 15;220(1-2):564-570. [doi: 10.1016/j.psychres.2014.08.019] [Medline: 25155939]
108. Coppersmith DDL, Fortgang RG, Kleiman EM, Millner AJ, Yeager AL, Mair P, et al. Effect of frequent assessment of suicidal thinking on its incidence and severity: high-resolution real-time monitoring study. Br J Psychiatry 2021 Jul 09:1-3. [doi: 10.1192/bjp.2021.97]
109. Nock M, Kleiman E, Abraham M, Bentley K, Brent D, Buonopane R, et al. Consensus Statement on Ethical & Safety Practices for Conducting Digital Monitoring Studies with People at Risk of Suicide and Related Behaviors. Psychiatr Res Clin Pract 2021;3(2):57-66 [FREE Full text] [doi: 10.1176/appi.prcp.20200029] [Medline: 34414359]
110. Blades CA, Stritzke WGK, Page AC, Brown JD. The benefits and risks of asking research participants about suicide: A meta-analysis of the impact of exposure to suicide-related content. Clin Psychol Rev 2018 Aug;64:1-12. [doi: 10.1016/j.cpr.2018.07.001] [Medline: 30014862]
111. DeCou CR, Schumann ME. On the Iatrogenic Risk of Assessing Suicidality: A Meta-Analysis. Suicide Life Threat Behav 2018 Oct;48(5):531-543. [doi: 10.1111/sltb.12368] [Medline: 28678380]
112. Muehlenkamp JJ, Swenson LP, Batejan K, Jarvi SM. Emotional and Behavioral Effects of Participating in an Online Study of Nonsuicidal Self-Injury. Clinical Psychological Science 2014 May 14;3(1):26-37 [FREE Full text] [doi: 10.1177/2167702614531579]
113. Cha CB, Glenn JJ, Deming CA, D'Angelo EJ, Hooley JM, Teachman BA, et al. Examining potential iatrogenic effects of viewing suicide and self-injury stimuli. Psychol Assess 2016 Nov;28(11):1510-1515. [doi: 10.1037/pas0000280] [Medline: 26821197]
114. Hasking P, Tatnell RC, Martin G. Adolescents' reactions to participating in ethically sensitive research: a prospective self-report study. Child Adolesc Psychiatry Ment Health 2015;9:39 [FREE Full text] [doi: 10.1186/s13034-015-0074-3] [Medline: 26300966]
115. Trull TJ, Ebner-Priemer UW. Ambulatory assessment in psychopathology research: A review of recommended reporting guidelines and current practices. J Abnorm Psychol 2020 Jan;129(1):56-63. [doi: 10.1037/abn0000473] [Medline: 31868388]
116. Eisele G, Vachon H, Lafit G, Kuppens P, Houben M, Myin-Germeys I, et al. The Effects of Sampling Frequency and Questionnaire Length on Perceived Burden, Compliance, and Careless Responding in Experience Sampling Data in a Student Population. Assessment 2020 Sep 10:1073191120957102. [doi: 10.1177/1073191120957102] [Medline: 32909448]
117. Kirtley O, Hiekkaranta A, Kunkels Y, Eisele G, Verhoeven D, Van NM. The experience sampling method (ESM) item repository. OSF 2018 Dec 04:1. [doi: 10.17605/osf.io/kg376]
118. Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, du Sert NP, et al. A manifesto for reproducible science. Nat Hum Behav 2017 Jan 10;1:0021 [FREE Full text] [doi: 10.1038/s41562-016-0021] [Medline: 33954258]
119. Kirtley OJ, Lafit G, Achterhof R, Hiekkaranta AP, Myin-Germeys I. Making the Black Box Transparent: A Template and Tutorial for Registration of Studies Using Experience-Sampling Methods. Advances in Methods and Practices in Psychological Science 2021 Mar 02;4(1):251524592092468. [doi: 10.1177/2515245920924686]
120. Horstmann K. Experience sampling and daily diary studies: Basic concepts, designs, and challenges. In: The Handbook of Personality Dynamics and Processes. Cambridge, Massachusetts: Academic Press; 2021:791-814.
121. Silvia PJ, Kwapil TR, Walsh MA, Myin-Germeys I. Planned missing-data designs in experience-sampling research: Monte Carlo simulations of efficient designs for assessing within-person constructs. Behav Res Methods 2014 Mar;46(1):41-54 [FREE Full text] [doi: 10.3758/s13428-013-0353-y] [Medline: 23709167]
122. Biddle L, Cooper J, Owen-Smith A, Klineberg E, Bennewith O, Hawton K, et al. Qualitative interviewing with vulnerable populations: individuals' experiences of participating in suicide and self-harm based research. J Affect Disord 2013 Mar 05;145(3):356-362. [doi: 10.1016/j.jad.2012.08.024] [Medline: 23021191]
123. Littlewood DL, Harris K, Gooding P, Pratt D, Haddock G, Peters S. Using my Demons to Make Good: The Short- and Long-Term Impact of Participating in Suicide-Related Research. Arch Suicide Res 2021;25(2):315-339. [doi: 10.1080/13811118.2019.1663330] [Medline: 31544686]
124. Kleiman EM, Glenn CR, Liu RT. Real-Time Monitoring of Suicide Risk among Adolescents: Potential Barriers, Possible Solutions, and Future Directions. J Clin Child Adolesc Psychol 2019;48(6):934-946 [FREE Full text] [doi: 10.1080/15374416.2019.1666400] [Medline: 31560584]
125. Boden ZVR, Gibson S, Owen GJ, Benson O. Feelings and Intersubjectivity in Qualitative Suicide Research. Qual Health Res 2016 Jul;26(8):1078-1090. [doi: 10.1177/1049732315576709] [Medline: 25794522]
126. Kumar S, Cavallaro L. Researcher Self-Care in Emotionally Demanding Research: A Proposed Conceptual Framework. Qual Health Res 2018 Mar;28(4):648-658. [doi: 10.1177/1049732317746377] [Medline: 29224510]

127. Dickson-Swift V, James EL, Kippen S, Liamputtong P. Risk to researchers in qualitative research on sensitive topics: issues and strategies. Qual Health Res 2008 Jan;18(1):133-144. [doi: 10.1177/1049732307309007] [Medline: 18174541]

Abbreviations

EMI: ecological momentary intervention
JITAI: just-in-time adaptive intervention
NSSI: nonsuicidal self-injury

Edited by J Torous; submitted 03.06.21; peer-reviewed by C Hamza, K Kruzan; comments to author 02.07.21; revised version received 18.08.21; accepted 23.08.21; published 19.11.21.

Please cite as:
Kiekens G, Robinson K, Tatnell R, Kirtley OJ
Opening the Black Box of Daily Life in Nonsuicidal Self-injury Research: With Great Opportunity Comes Great Responsibility
JMIR Ment Health 2021;8(11):e30915
URL: https://mental.jmir.org/2021/11/e30915
doi:10.2196/30915
PMID:34807835

Original Paper

Machine Learning–Based Predictive Modeling of Anxiety and Depressive Symptoms During 8 Months of the COVID-19 Global Pandemic: Repeated Cross-sectional Survey Study

Katrina Hueniken[1,2], MPH; Nibene Habib Somé[2,3,4,5,6], PhD; Mohamed Abdelhack[1], PhD; Graham Taylor[7,8], PhD; Tara Elton Marshall[2,3,4,6,9,10], PhD; Christine M Wickens[2,3,4,11,12], PhD; Hayley A Hamilton[2,3,4], PhD; Samantha Wells[2,3,4,6,13,14], PhD; Daniel Felsky[1,13,15,16], PhD

[1]Krembil Centre for Neuroinformatics, Centre for Addiction and Mental Health, Toronto, ON, Canada
[2]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[3]Institute for Mental Health Policy Research, Centre for Addiction and Mental Health, Toronto, ON, Canada
[4]Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada
[5]Institute for Clinical Evaluative Sciences, Toronto, ON, Canada
[6]Department of Epidemiology and Biostatistics, Schulich School of Medicine and Dentistry, Western University, London, ON, Canada
[7]School of Engineering, University of Guelph, Guelph, ON, Canada
[8]Vector Institute for Artificial Intelligence, Toronto, ON, Canada
[9]School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa, Ottawa, ON, Canada
[10]Department of Health Sciences, Lakehead University, Thunder Bay, ON, Canada
[11]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada
[12]Department of Pharmacology and Toxicology, University of Toronto, Toronto, ON, Canada
[13]Department of Psychiatry, University of Toronto, Toronto, ON, Canada
[14]School of Psychology, Deakin University, Burwood, Australia
[15]Department of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
[16]Institute of Medical Sciences, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Daniel Felsky, PhD
Krembil Centre for Neuroinformatics
Centre for Addiction and Mental Health
250 College St, 12th Floor
Toronto, ON, M5T 1R8
Canada
Phone: 1 (416) 535 8501 ext 33587
Email: Daniel.Felsky@camh.ca

Abstract

Background: The COVID-19 global pandemic has increased the burden of mental illness on Canadian adults. However, the complex combination of demographic, economic, and lifestyle factors and perceived health risks contributing to patterns of anxiety and depression has not been explored.

Objective: The aim of this study is to harness flexible machine learning methods to identify constellations of factors related to symptoms of mental illness and to understand their changes over time during the COVID-19 pandemic.

Methods: Cross-sectional samples of Canadian adults (aged ≥18 years) completed web-based surveys in 6 waves from May to December 2020 (N=6021), and quota sampling strategies were used to match the English-speaking Canadian population in age, gender, and region. The surveys measured anxiety and depression symptoms, sociodemographic characteristics, substance use, and perceived COVID-19 risks and worries. First, principal component analysis was used to condense highly comorbid anxiety and depression symptoms into a single data-driven measure of emotional distress. Second, eXtreme Gradient Boosting (XGBoost), a machine learning algorithm that can model nonlinear and interactive relationships, was used to regress this measure on all included explanatory variables. Variable importance and effects across time were explored using SHapley Additive exPlanations (SHAP).

Results: Principal component analysis of responses to 9 anxiety and depression questions on an ordinal scale revealed a primary latent factor, termed "emotional distress," that explained 76% of the variation in all 9 measures. Our XGBoost model explained a substantial proportion of variance in emotional distress (r^2=0.39). The 3 most important items predicting elevated emotional distress were increased worries about finances (SHAP=0.17), worries about getting COVID-19 (SHAP=0.17), and younger age (SHAP=0.13). Hopefulness was associated with emotional distress and moderated the impacts of several other factors. Predicted emotional distress exhibited a nonlinear pattern over time, with the highest predicted symptoms in May and November and the lowest in June.

Conclusions: Our results highlight factors that may exacerbate emotional distress during the current pandemic and possible future pandemics, including a role of hopefulness in moderating distressing effects of other factors. The pandemic disproportionately affected emotional distress among younger adults and those economically impacted.

(JMIR Ment Health 2021;8(11):e32876) doi:10.2196/32876

KEYWORDS

mental health; machine learning; COVID-19; emotional distress; emotion; distress; prediction; model; anxiety; depression; symptom; cross-sectional; survey

Introduction

The emergence of the novel coronavirus SARS-CoV-2 in late 2019 and the resulting COVID-19 pandemic have caused social and economic upheaval worldwide. Public health measures to limit the spread of the virus have been linked to negative mental health outcomes, such as depression and anxiety [1-3]. Emotionally distressing symptoms are common, including nonspecific anxiety, fear of illness, loneliness, frustration, and boredom [4], and they are worsened by social isolation due to current lockdown policies [5]. Although these policies effectively limit the spread of infection, a deeper understanding of their effects on mental health is necessary to inform public health interventions.

The COVID-19 pandemic has had disproportionate impacts on some groups compared to others [6]. Millions of North Americans lost work as governments forced closure of businesses and imposed stay-at-home orders. Canadians in the lowest earnings quartile have been particularly affected, accounting for one-half of all job losses in early 2020 [7]. Job insecurity during the pandemic has been associated with symptoms of depression [8]. Furthermore, demographic factors such as female gender [9,10] and younger age [11] have been associated with higher rates of emotional distress during the COVID-19 pandemic. Identifying putative drivers of emotional distress during the COVID-19 pandemic can improve our understanding of population-wise patterns of mental health during a large-scale crisis and aid policy making to support those in need.

Previous literature examining predictive modeling of anxiety and depression has largely focused on classifying patients by anxiety or depression status. Studies predicting anxiety and depression diagnosis from clinical and demographic factors have achieved moderate to high predictive accuracy [12,13], although few studies have predicted symptoms of anxiety and depression at the population level.

The constellation of factors contributing to symptoms of depression and anxiety are appreciably complex. Therefore, meaningful conclusions on the importance of individual factors should be considered in the context of many available data types, as well as over time. We conducted an exploratory study to uncover relationships between predictive factors and self-reported depression and anxiety symptoms in Canadian adults during the COVID-19 pandemic. This study's first aim was to identify the most important factors predicting a composite score of depression and anxiety symptoms. The second aim was to characterize how associations between demographic and environmental factors and symptom scores changed over time. The third aim was to identify predictors that moderated or exacerbated the effects of others on depression and anxiety by examining two-way variable interactions in our model.

In this study, we applied a flexible decision tree–based machine learning method, eXtreme Gradient Boosting (XGBoost), to model a composite score of depression and anxiety using 50 explanatory factors related to sociodemographic characteristics, substance use, employment, and perceived COVID-19 risk. Data were collected from a cross-sectional survey administered to Canadians between May and December 2020. The XGBoost algorithm allowed inclusion of many input variables and simultaneous consideration of nonlinear and interactive effects between all inputs. We identified the most important predictive factors related to depression and anxiety, and we assessed changes in these effects over time.

Methods

Data Collection

Data were collected via repeated cross-sectional surveys between May 8 and December 1, 2020. A total of 6 waves of data were collected using a web-based panel administered by the research and data collection company Delvinia [14]. The sampling waves occurred from May 8 to 12 (n=1005), May 29 to June 1 (n=1002), June 19 to 23 (n=1005), July 10 to 14 (n=1003), September 18 to 22 (n=1003), and November 27 to December 1 (n=1003). Participants were sampled independently in each wave. The overall response rate was 16.1% (6021/38,987).

Quota sampling based on age, gender, and region was used to obtain a sample that is proportional to the English-speaking population of Canada. Canadians aged ≥18 years were eligible. Respondents provided written informed consent electronically prior to participation. Research ethics approval was obtained

from the Centre for Addiction and Mental Health research ethics board.

Survey questions included information on demographics, anxiety, depression, substance use, employment changes, perceived risks, and worries related to COVID-19. A full list of variables considered for analysis is included in Table S1 in Multimedia Appendix 1. Respondents' anxiety levels were captured using the Generalized Anxiety Disorder–7 questionnaire (GAD-7) [15], a validated inventory measuring the frequency of anxiety symptoms over the past 2 weeks. Raw item scores were used, ranging from 0 (not at all) to 3 (nearly every day). Depressive symptoms were measured using 3 modified questions from the Centre for Epidemiologic Studies Depression Scale (CES-D) [16]. Feelings of depression, loneliness, and hopefulness over the past week were reported on a Likert-style scale between 0 ("Rarely or none of the time [less than 1 day]") and 3 ("Most or all of the time [5-7 days]"). For further details on the included explanatory variables, see Text S1 in Multimedia Appendix 1.

Data Preparation and Quality Control

Given the strong comorbidity of population-level depression and anxiety symptoms [17], as well as their shared neurobiological underpinnings [18], we first examined pairwise correlations among all anxiety questionnaire items (GAD-7) and the 3 available mood/depression (CES-D) questionnaire items, and we applied principal component analysis (PCA). The questions from both scales were similar in scale. Correlations were high between the GAD-7 and CES-D items, allowing for the combination of items from both questionnaires into a single measure of emotional distress using symptoms of both anxiety and depression.

PCA with varimax rotation was used to reduce the number of mood and anxiety variables needed for modeling while retaining as much information as possible from all anxiety and depression variables. PCA is widely used to identify principal axes of variation in psychometric questionnaires [19,20]; in the absence of a validated method of combining items across the GAD-7 and CES-D, PCA was used to retain the largest amount of useful information in a single score incorporating both scales. To properly account for the ordinal nature of the GAD-7 and CES-D variables, polychoric correlations—measuring associations between ordinal variables assumed to be realizations of underlying latent Gaussian distributions [21]—were used. PCA was applied to our outcomes on all observations prior to train and test spitting.

For our data-inclusive approach, all survey questions were considered for inclusion as model predictors, excluding mood and anxiety variables used in our outcome measure. Questions not asked in all 6 survey waves were excluded. Categorical variables were one-hot encoded (1=yes, 0=no for category membership). "Prefer not to answer" responses were treated as missing (see the *Predictive Modeling* section below).

Statistical Analysis

Predictive Modeling

The XGBoost R package [22,23] was used to train and test gradient-boosted regularized tree-based models. The core XGBoost function predicts outcomes by fitting a series of decision trees, each building upon the information from all previous trees to improve predictive performance. XGBoost was chosen to model anxiety and depression symptoms, as its extremely flexible approach can enable modeling of linear, nonlinear, and interactive effects between all inputs simultaneously, allowing for more insight into complex interdependencies within inputs that may not be captured by simpler regression methods.

We withheld 20% of the observations from model training, randomly selected within each survey wave. Optimal model hyperparameters were selected using a random grid search and 10-fold cross-validation on the remaining 80% of observations.

As our latent outcome of interest was continuous, the root mean squared error was used as a loss function. Out-of-sample predictive performance was tested by computing Pearson correlations between predicted and observed distress values. Squared Pearson correlation coefficients (r^2) were calculated to describe the proportion of variance in the observed outcome captured by the predictive model.

XGBoost imputed missing variables by assigning a default direction to each decision node. In the sensitivity analysis, observations with missing values in any inputs were removed from the model training and validation data sets. Out-of-sample performance was compared between the main model (all observations) versus the sensitivity analysis model (only complete observations).

To assess the improvement in predictive performance of XGBoost over a less complex approach that does not account for interactions and nonlinear effects, least absolute shrinkage and selection operator (LASSO) regression was also tested. Regularized regression methods such as LASSO tend to exhibit improved predictive performance compared to unregularized, traditional regression methods via the introduction of a penalty parameter to control overfitting.

Our LASSO model included the same variables predicting distress and was trained on the same set of observations, excluding observations with missing data. The LASSO regularization penalty parameter was optimized via 10-fold cross-validation. Out-of-sample prediction was compared between the LASSO and XGBoost models trained on complete observations only, as well as on the full model using single imputation with predictive mean matching to impute data for LASSO.

Variable Importance and Interactions

To understand the relative contribution of each variable to model predictions, we computed importance of each variable using SHapley Additive exPlanations (SHAP) [24]. SHAP values measure the relative strength of each variable's marginal contribution to an individual's predicted outcome value,

conditioning on all other explanatory variables for that individual [25].

Overall variable importance was defined as the mean absolute value of all SHAP values for a given variable. Negative SHAP values indicate that predicted distress was reduced by that variable, while positive SHAP values indicate a positive influence on predicted distress. Relationships between predicted values and time were examined via partial dependence plots, adjusted for all other explanatory variables. SHAP values for each two-way interaction between variables were also computed and plotted [26].

In the absence of formal hypothesis tests for interaction SHAP values, and to provide a comparable regression-based framework for the interpretation of our XGBoost-identified interactions, we identified statistically significant two-way variable interactions by fitting separate linear regression models to each pair of input variables. Global *P* values for the overall significance of interaction terms were computed via likelihood-ratio tests. The model with both individual variables plus their interaction was compared to a model with the interaction term removed. Benjamini-Hochberg corrections were applied to all interaction global *P* values to constrain the false discovery rate (FDR) to 5%. Interactions were deemed statistically significant if their FDR-adjusted *P* values were <.05.

All analyses were conducted in R, version 3.6.3 (R Foundation for Statistical Computing; see Text S2 in Multimedia Appendix 1).

Results

Survey Respondents

A total of 6021 respondents provided complete surveys for analysis. The characteristics of the respondents are summarized in Table 1. Demographic distributions of age, sex, and region were representative of the English-speaking Canadian adult population [27,28].

Table 1. Baseline characteristics of the survey respondents (N=6021).

Characteristic	Responses by survey wave, n (%)						*P* value (Fisher test)
	1 (May 8-12, n=1005)	2 (May 29-June 1, n=1002)	3 (June 19-23, n=1005)	4 (July 10-14, n=1003)	5 (Sept 18-22, n=1003)	6 (Nov 27-Dec 1, n=1003)	
Region							>.99
Alberta	140 (13.9)	140 (14)	140 (13.9)	133 (13.3)	137 (13.7)	141 (14.1)	
British Columbia	152 (15.1)	146 (14.6)	150 (14.9)	151 (15.1)	148 (14.8)	152 (15.2)	
Ontario	418 (41.6)	418 (41.7)	415 (41.3)	421 (42)	419 (41.8)	419 (41.8)	
Quebec/Atlantic Canada	182 (18.1)	190 (19)	192 (19.1)	192 (19.1)	191 (19)	187 (18.6)	
Saskatchewan/Manitoba	111 (11)	108 (10.8)	104 (10.3)	105 (10.5)	106 (10.6)	102 (10.2)	
Yukon/Northwest Territories/Nunavut	2 (0.2)	0 (0)	4 (0.4)	1 (0.1)	2 (0.2)	2 (0.2)	
Age (years)							>.99
18-39	394 (39.2)	389 (38.8)	394 (39.2)	388 (38.7)	390 (38.9)	392 (39.1)	
40-59	306 (30.4)	312 (31.1)	307 (30.5)	309 (30.8)	305 (30.4)	305 (30.4)	
≥60	305 (30.3)	301 (30)	304 (30.2)	306 (30.5)	308 (30.7)	306 (30.5)	
Gender							.62
Female	498 (49.6)	497 (49.6)	499 (49.7)	492 (49.1)	498 (49.7)	503 (50.1)	
Male	504 (50.1)	492 (49.1)	501 (49.9)	501 (50)	497 (49.6)	492 (49.1)	
Other	3 (0.3)	13 (1.3)	5 (0.5)	10 (1)	8 (0.8)	8 (0.8)	
Has children							.80
No	776 (77.2)	766 (76.4)	768 (76.4)	761 (75.9)	769 (76.7)	787 (78.5)	
Yes	229 (22.8)	236 (23.6)	237 (23.6)	242 (24.1)	234 (23.3)	216 (21.5)	
Education							.55
High school or less	111 (11)	104 (10.4)	129 (12.8)	122 (12.2)	119 (11.9)	99 (9.9)	
Some post–high school education	159 (15.8)	165 (16.5)	148 (14.7)	162 (16.2)	147 (14.7)	150 (15)	
University or college	728 (72.4)	727 (72.6)	720 (71.6)	706 (70.4)	731 (72.9)	742 (74)	
Prefer not to answer	7 (0.7)	6 (0.6)	8 (0.8)	13 (1.3)	6 (0.6)	12 (1.2)	
Marital status							.84
Married/living with partner	613 (61)	605 (60.4)	622 (61.9)	634 (63.2)	638 (63.6)	653 (65.1)	
Never married	251 (25)	251 (25)	253 (25.2)	233 (23.2)	239 (23.8)	216 (21.5)	
Separated/divorced/widowed	128 (12.7)	132 (13.2)	119 (11.8)	122 (12.2)	113 (11.3)	118 (11.8)	
Prefer not to answer	13 (1.3)	14 (1.4)	11 (1.1)	14 (1.4)	13 (1.3)	16 (1.6)	
Race/ethnicity							.88
White (European, North American)	698 (69.5)	702 (70.1)	691 (68.8)	697 (69.5)	699 (69.7)	691 (68.9)	
Asian	200 (19.9)	175 (17.5)	201 (20)	188 (18.7)	190 (18.9)	202 (20.1)	
Black (African, Caribbean, North American)	16 (1.6)	18 (1.8)	19 (1.9)	24 (2.4)	23 (2.3)	13 (1.3)	
Other	71 (7.1)	78 (7.8)	68 (6.8)	66 (6.6)	60 (6)	70 (7)	
Not sure/prefer not to answer	20 (2)	29 (2.9)	26 (2.6)	28 (2.8)	31 (3.1)	27 (2.7)	
Household income (CAD $)[a]							.61
Less than 40,000	128 (12.7)	121 (12.1)	136 (13.5)	118 (11.8)	116 (11.6)	110 (11)	

Characteristic	Responses by survey wave, n (%)						*P* value (Fisher test)
	1 (May 8-12, n=1005)	2 (May 29-June 1, n=1002)	3 (June 19-23, n=1005)	4 (July 10-14, n=1003)	5 (Sept 18-22, n=1003)	6 (Nov 27-Dec 1, n=1003)	
40,000-79,000	268 (26.7)	236 (23.6)	238 (23.7)	235 (23.4)	247 (24.6)	236 (23.5)	
80,000-119,000	226 (22.5)	229 (22.9)	220 (21.9)	213 (21.2)	237 (23.6)	241 (24)	
120,000 or more	217 (21.6)	259 (25.8)	247 (24.6)	252 (25.1)	228 (22.7)	251 (25)	
Prefer not to answer	166 (16.5)	157 (15.7)	164 (16.3)	185 (18.4)	175 (17.4)	165 (16.5)	
Locality							.97
Rural area	158 (15.7)	164 (16.4)	151 (15)	171 (17)	164 (16.4)	164 (16.4)	
Suburban area	382 (38)	379 (37.8)	369 (36.7)	365 (36.4)	376 (37.5)	365 (36.4)	
Urban area	465 (46.3)	459 (45.8)	485 (48.3)	467 (46.6)	463 (46.2)	474 (47.3)	
Heavy alcohol use (past 7 days)[b]							.67
None or light alcohol use	765 (76.1)	753 (75.1)	736 (73.2)	726 (72.4)	744 (74.2)	743 (74.1)	
Heavy alcohol use	238 (23.7)	247 (24.7)	267 (26.6)	271 (27)	255 (25.4)	257 (25.6)	
Prefer not to answer	2 (0.2)	2 (0.2)	2 (0.2)	6 (0.6)	4 (0.4)	3 (0.3)	
Cannabis use							.18
No cannabis use	889 (88.5)	869 (86.7)	878 (87.4)	870 (86.7)	881 (87.8)	840 (83.7)	
Used cannabis	115 (11.4)	130 (13)	124 (12.3)	131 (13.1)	119 (11.9)	160 (16)	
Prefer not to answer	1 (0.1)	3 (0.3)	3 (0.3)	2 (0.2)	3 (0.3)	3 (0.3)	

[a]CAD $1=US $0.80.

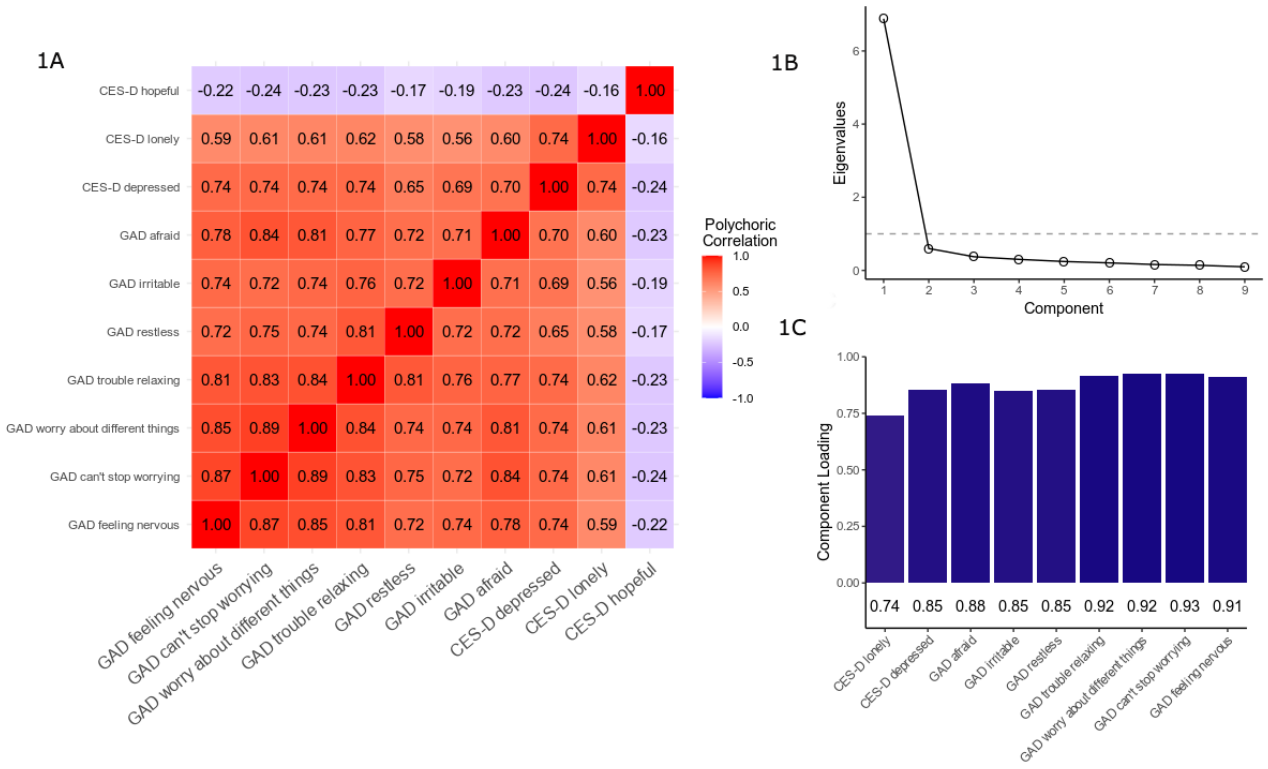
[b]Heavy alcohol use was defined as 5 or more standard drinks for men and 4 or more standard drinks for women in a given day.

Calculation of Latent Feature Representing Anxiety and Depression

The PCA was initially fit using 10 survey items: 7 from the GAD-7 (anxiety) and 3 from the CES-D (depressive symptoms). The polychoric correlations between the anxiety and depression variables are presented in Figure 1A. All items were moderately to strongly positively correlated (r=0.55-0.89), with the exception of hopefulness (CES-D item; correlation coefficients –0.18 to –0.27). All items except for hopefulness were loaded strongly onto principal component 1 (PC1) (loadings 0.74-0.93), with hopefulness loading weakly onto PC1 (loading –0.28) and strongly onto principal component 2 (PC2) (loading 0.96); this finding supported our choice to model a single latent outcome combining both scales. Given the weak negative correlation between hopefulness and the remaining mood and anxiety variables, hopefulness was dropped from the PCA and instead included as an explanatory variable in downstream modeling. Distributions of the GAD and CES-D scores and PC1 are included in Figure S1 (Multimedia Appendix 1).

The final PCA included 9 items. In the remainder of this analysis, we refer to PC1 as “emotional distress.” Figure 1B displays a scree plot for this 9-item PCA; the point of inflection occurs at the second PC, indicating that the PCs after the first do not add substantial additional information. The loadings on PC1, emotional distress, are presented in Figure 1C. The item loadings ranged from 0.74 to 0.93. Emotional distress explained 76% of the variance in all 9 variables.

Figure 1. Results of the principal component analysis. (A) Heat map of the polychoric correlations; (B) scree plot of the principal component analysis and loadings onto principal component 1; (C) Variable Loadings onto Principal Component 1. CES-D: Centre for Epidemiologic Studies Depression Scale; GAD: Generalized Anxiety Disorder–7 questionnaire.



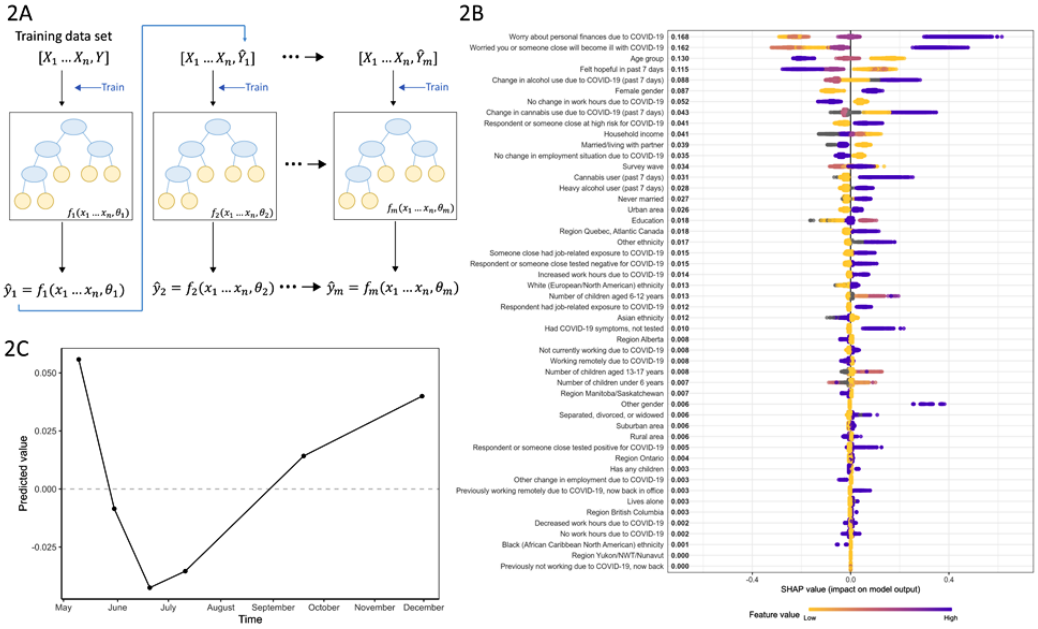
Fitting of the XGBoost Model and Comparison to LASSO

A total of 50 predictor variables were included in our primary XGBoost model. We included 4819 respondents in the training data set; out-of-sample prediction was tested on the remaining 1202 respondents. For the results of the hyperparameter selection, see Text S3 in Multimedia Appendix 1. The final model explained 38.7% of the variance in distress in the out-of-sample prediction (r^2=0.387). A scatterplot of the predicted and realized values of emotional distress is presented in Figure S2, Multimedia Appendix 1. In training, the r^2 value of the model was 0.394, which is close to the r^2 value for our holdout test set; this indicates that there was no substantial overfitting. For a LASSO model fit to the full test set, r^2 was 0.354 in the withheld validation set, which is slightly lower than the value for XGBoost in predictive accuracy.

A graphical depiction of model training using gradient-boosted trees [29] is presented in Figure 2A. A visual representation of the first 6 gradient-boosted trees in the fitted model is included in Figure S3 (Multimedia Appendix 1). A partial dependence plot overlaid with COVID-19 positivity rates [30] is presented in Figure S4 (Multimedia Appendix 1).

To determine if the model fit was sensitive to the XGBoost imputation algorithm for variables with missing values, XGBoost was refit with reoptimized hyperparameters and validated using only observations with complete data (n=3689). On the withheld data, this model explained 36.9% of the variance in emotional distress (r^2=0.369). Results of a LASSO model fit to the same set of complete observations indicated slightly lower performance compared to the XGBoost fit (LASSO r^2=0.346), indicating that distress was reasonably approximated by a linear fit, although at a loss of approximately 2% of the explained variance when nonlinear and interaction relationships were not considered.

Figure 2. (A) Gradient-boosted tree model training diagram; (B) variable importance plot ranked by mean absolute SHAP value; (C) partial dependence plot of predicted emotional distress values across survey waves. NWT: Northwest Territories; SHAP: SHapley Additive exPlanations.



Identification of Variables Most Strongly Associated with Emotional Distress

The SHAP values are presented in Figure 2B. To aid the direct interpretability of our SHAP value analysis, we modeled each questionnaire item as a function of our latent emotional distress outcome using linear regression. Each 1-unit change in our outcome corresponded to a difference in the question response value of between 0.68 and 0.84 (mean 0.78) across all 9 questions, meaning that a SHAP value of 0.61 (the largest value reported for individuals with high levels of reported financial worry) would translate to an average increase of 0.61*0.78=0.48 across the original question scales. Given that each question was measured as integers ranging from 1 to 4—representing an underlying quantitative scale mapping onto the number of recent days when symptoms were experienced—this maximum SHAP value would represent a predicted marginal change of 16% of the entire spectrum of symptom burden across all questions in the holdout test population: 0.48/(4-1)*100.

The variables with the greatest importance, in descending order, were worry about personal finances due to COVID-19 (mean absolute SHAP value=0.168), worry that oneself or loved ones will become ill with COVID-19 (0.162), age group (0.130), hopefulness (0.115), change in alcohol use due to the pandemic (0.088), and female gender (0.087).

For both financial and illness-related worries, low to moderate levels of worry were associated with decreased or average predicted distress, while severe worries increased predicted distress (0.41 for “very worried” about finances; 0.34 for “very worried” about illness). To a lesser extent, greater hopefulness was associated with decreased predicted distress. The relationship between age group and predicted emotional distress was somewhat linear: membership in the youngest age group (18 to 39 years) increased predicted distress, while membership in the age groups of 40 to 59 years and ≥60 years decreased predicted distress. Female gender increased predicted distress, and both increased and decreased alcohol intake compared to prepandemic intake had higher mean SHAP values compared to no change (mean SHAP values of 0.19 for increased use 0.12 for decreased use; –0.07 no change), indicating a nonlinear effect. Change in cannabis use due to the pandemic was directionally similar (mean SHAP values of 0.25 for increased use, 0.11 for decreased use, and –0.02 for no change), although its overall importance was lower.

Following the top 6 ranked variables, we found a heuristic “elbow point” separating the most important variables from those with lesser importance. For all remaining variables, the mean SHAP values were at or below 0.05. Of note, some important variables demonstrated low mean SHAP values, but their range of values was large. Notably, only a small number of respondents answered “other” (nonbinary) for gender identity (n=47); although these responses had a strong influence on their individual predicted values, mean SHAP values for the feature as a whole remained relatively low (mean absolute SHAP value 0.006, range –0.04 to 0.38).

To examine the effects of time, we explored predicted emotional distress values across survey waves. The survey wave variable ranked 14th out of 50 variables with respect to variable importance (mean absolute SHAP value 0.024), indicating that the passing of time was not as influential in our model as other time-independent variables. A partial dependence plot is shown in Figure 2C; when all other variables were held constant, predicted emotional distress was highest in wave 1 (May 8-12; adjusted mean 0.06). After wave 1, the values decreased (wave 2 adjusted mean –0.01). The predicted values decreased most by membership in wave 3 (mean –0.04, June 19-23) and wave 4 (mean –0.03, July 10-14), then increased again in waves 5 and 6 (mean 0.01 for September 18-22 and 0.04 for November 27-December 1).

Pairwise Interactions of Features in Predicting Emotional Distress

We next performed exploratory analysis of the importance of two-way variable interactions in our model. Mean SHAP values for two-way variable interactions from the top 15 most important variables are presented in Figure 3; all interactions are shown in Figure S5 (Multimedia Appendix 1). Overall, interactions between features were not substantially important in determining model predictions; the mean absolute SHAP values for the interaction terms ranged from 0 to 0.01, which represents a maximum of approximately 1/17 of the mean contribution of the most important variable (worry about finances). A total of 5 interactions had SHAP values above 0.008 (the visual elbow point); these top interactions included change in alcohol use × worry about finances (SHAP 0.010), worry about getting COVID-19 × worry about finances (SHAP 0.010), hopefulness × high risk for COVID-19 (0.009), hopefulness × female gender (0.009), and worry about getting COVID-19 × cannabis use (0.008).

Figure 4 displays relationships between pairs of variables in the top 5 most important interactions. Plotting both COVID-19–related worries (financial and illness-related) against predicted emotional distress, the distressing effects of illness-related worry were stronger at lower levels of financial worry, where individuals with severe financial worry had the greatest distress regardless of illness-related worries. Additionally, greater hopefulness mitigated differences in distress levels between those who were at high risk for COVID-19 (or had loved ones at high risk) versus those who were not, as well as differences in distress between female and nonfemale respondents. Finally, those who used cannabis had a steeper increase in distress as illness-related worries increased.

To validate our findings, unregularized regression analyses were performed to test the statistical significance of pairwise interaction models with realized distress values. Out of 1225 possible interactions between pairs of explanatory variables in bivariate linear regression on emotional distress, 58 interaction terms had significant associations after Benjamini-Hochberg correction. Of these 58 interactions, 10 involved hopefulness. Of all 5 interactions with SHAP variable importance above the elbow point of 0.008, the top 4 were also statistically significant in the regression analysis (see Table S2 in Multimedia Appendix 1).

Figure 3. SHAP variable importance of two-way variable interactions for the top 15 most important variables. SHAP: SHapley Additive exPlanations.

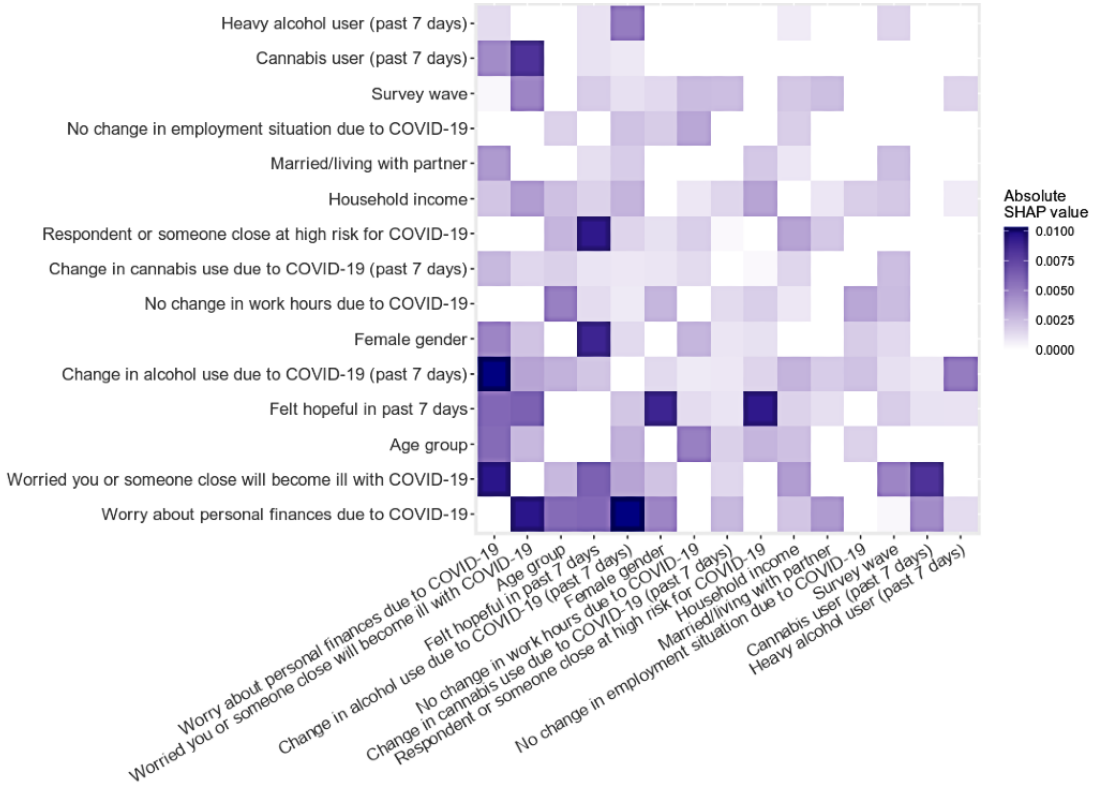
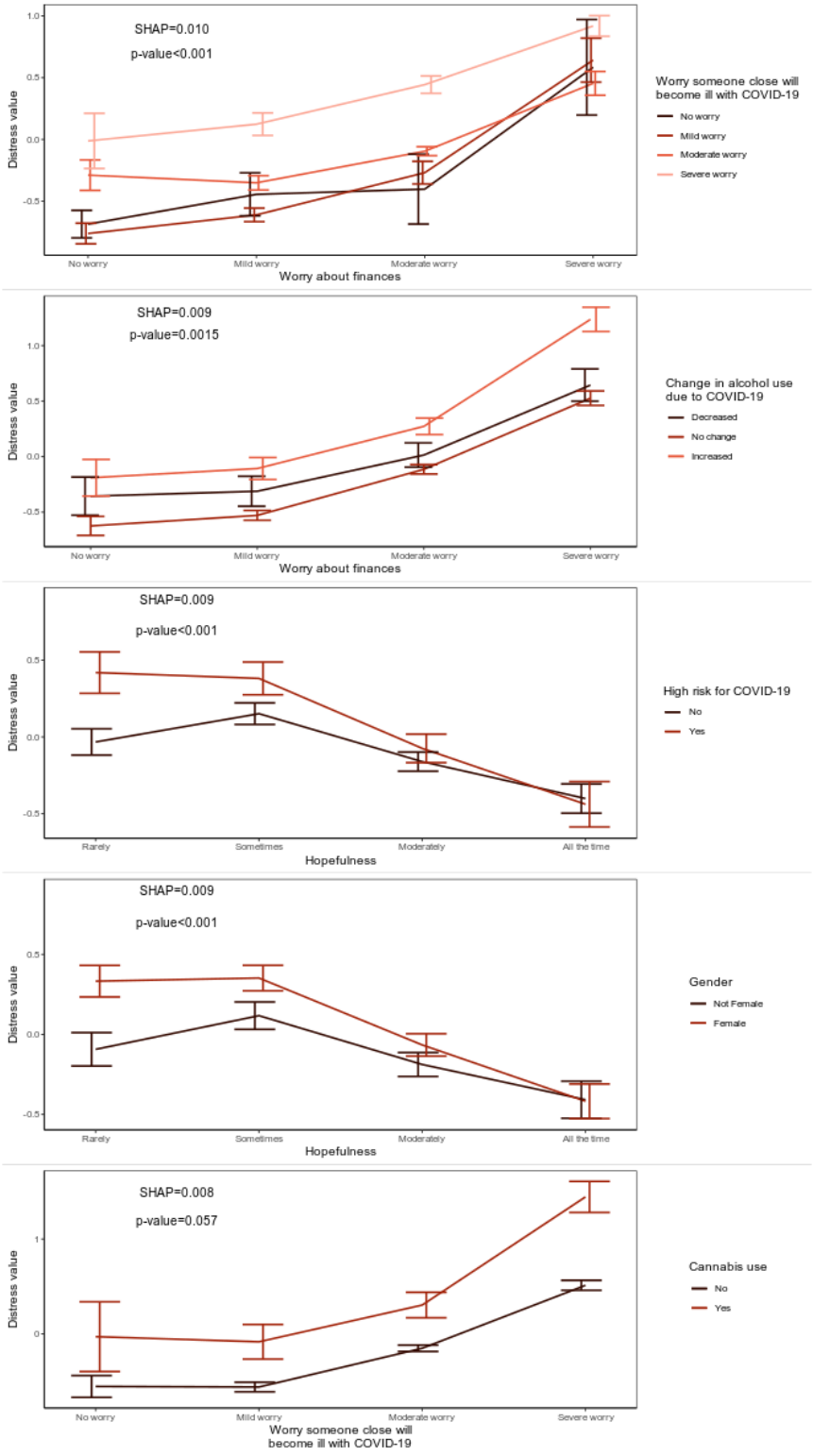


Figure 4. Relationships between pairs of variables in five most important variable interactions and predicted distress.



Discussion

Principal Findings

In this study, we used machine learning to examine factors associated with emotional distress during the COVID-19 pandemic, informed by self-reported levels of anxiety and depression symptoms, using data from a large national survey. We explored relationships between a wide range of sociodemographic characteristics, substance use patterns, and COVID-19–related perceived risks and worries with distress by examining nonlinear patterns in variable importance and by characterizing the importance of variable interactions.

Our findings that the top predictive factors for emotional distress included worries related to COVID-19 are consistent with a recent study of COVID-19–related anxieties [10], suggesting that the participants experienced substantial health and financial concerns. Female gender contributed substantially to increased

distress; this finding is consistent with evidence that even before the pandemic, both general anxiety [31] and depression [32] as well as COVID-19–specific anxieties [10] have been shown to be greater in women than in men. Our top predictive features also included change in alcohol use, which showed a nonlinear effect whereby any change (consuming either more or less alcohol compared to before the pandemic) was associated with an increase in emotional distress. This suggests that individual attempts to mitigate above-average levels of pandemic-related distress by drinking less have not been successful and that those who increased alcohol consumption due to the pandemic have experienced subsequent heightened distress.

Hopefulness was only weakly to moderately correlated with other anxiety and depression questionnaire items, and it was loaded onto its own principal component in PCA. These findings suggest that hopefulness may not measure the same latent construct captured by the remaining anxiety and depression survey items. This relative lack of cohesion between hopefulness and the other CES-D questions (feelings of loneliness and depression) in latent variable analysis is supported by previous work finding structural inconsistencies among CES-D items [33,34] Given the sudden and temporary nature of imposed pandemic restrictions, we hypothesized that the hopefulness responses represent a more trait-like positive affect rather than more situationally influenced responses to social isolation, such as lonely and depressed feelings. In predicting emotional distress, hopefulness was the fourth most important contributor, with higher hopefulness decreasing overall predicted distress. Hopefulness has been linked to lower emotional distress [35], and it may indicate greater resilience to adversities experienced during the COVID-19 pandemic.

One main aim was to assess changes in predicted emotional distress over time. The survey wave was not among the strongest contributors to emotional distress, with an importance approximately one seventh of that of the top variable. Despite this, variation in predicted values was present across time; the highest predicted emotional distress was observed during survey wave 1 in May 2020, amid nationwide lockdowns and just after the peak 7-day average case count [30] in Canada's "first wave" of COVID cases. Predicted values declined through waves 2 to 4, when case counts were decreasing and lockdown policies were relaxed. Waves 5 to 6 saw a second increase in predicted emotional distress levels through the fall and winter of 2020, as Canada began its second wave of cases. This pattern mirrored the trajectory of COVID-19 case counts nationwide [30]. Our findings are consistent with results from the Canadian Community Health Survey, which found that mental health worsened due to the pandemic, increasing from September to December in 2020 [36]. These results suggest that Canadians experienced a spike of emotional distress at the start of the pandemic, amid fearful public health messaging and great uncertainty. Following this, as case counts decreased and lockdown measures were lifted, we speculate that increased optimism or a reduction in the perceived threat of COVID-19 may have lowered collective distress. This summer period was followed by the fall and winter months, when seasonal changes and increasing case counts again led to increased distress.

Finally, we conducted an exploratory analysis to examine the importance of variable interactions in predicting emotional distress. Although the importance of these interactions was relatively low (1/17th that of the most important single variable), they played a role in determining the predicted values. In particular, the distressing effects of severe COVID-19–related worries were most pronounced when other worries were not present. Notably, high hopefulness also mitigated the effect of several other factors that increased predicted distress, including female gender and high risk for COVID-19.

This study has several strengths. First, it was conducted on a large national survey sample that was designed to be representative of the Canadian population in age, gender, and region. Second, few other studies have examined changes in mental health outcomes across time throughout the first 10 months of the COVID-19 pandemic. Third, we employed flexible, interpretable machine learning methods to detect nonlinear and interactive relationships of many predictors of anxiety and depression symptoms.

A limitation of this study was its cross-sectional design, which meant that we were not able to track changes in mental health outcomes within the same individuals longitudinally, nor were we able to determine the temporality or direction of associations. However, the repeated cross-sectional study design allowed us to track patterns in mental health outcomes over time. A second limitation was that the survey was not administered prior to the COVID-19 pandemic, so a direct comparison with prepandemic mental health was not possible. A third limitation was that the survey was administered on the web, meaning that Canadians who are not comfortable with technology may have been less likely to participate; furthermore, the complete response rate was low (16.1%), indicating potential selection bias. However, quota sampling techniques were used to represent the Canadian adult population as accurately as possible in the complete analysis data set, and this response rate is similar to that expected for population surveys of this length administered on the web and without financial incentive [37]. Finally, information on mental health histories or prior clinical diagnoses was not available. However, these findings provide insight into emotional distress of the Canadian population at large, independent of clinical diagnoses.

Conclusion

Demographic and COVID-19–related factors were associated with a substantial amount of the variation in emotional distress during the global COVID-19 pandemic. These associations were most strongly driven by COVID-19–related fears, namely worries about personal finance and worries about contracting the illness, and were partly mitigated by high levels of hopefulness.

Rates of negative mental health outcomes such as eating disorders [38], substance use [39], overdose [40], and suicide attempts [40] have risen over the course of the COVID-19 pandemic. Although public health policy has been and continues to be vital to curb the spread of SARS-CoV-2, policy makers should also prioritize the provision of population-level supports to address elevated depression and anxiety symptoms among certain groups. Although we cannot infer causation in our study

design, our results indicate that initiatives to mitigate financial worry, alleviate illness-related fear, and promote hopefulness may be effective against symptoms of anxiety and depression in the wake of this and potential future pandemics.

Data Availability

All study data have been made publicly available [41].

Acknowledgments

DF is supported by The Koerner Family Foundation New Scientist Award, The Krembil Foundation, The Centre for Addiction and Mental Health (CAMH) Discovery Fund, and the Canadian Institutes of Health Reseach (CIHR). Delvinia, a research technology firm, provided in-kind support for data collection.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary material.
[DOCX File , 935 KB - mental_v8i11e32876_app1.docx]

References

1. Shigemura J, Ursano RJ, Morganstein JC, Kurosawa M, Benedek DM. Public responses to the novel 2019 coronavirus (2019-nCoV) in Japan: Mental health consequences and target populations. Psychiatry Clin Neurosci 2020 Apr;74(4):281-282 [FREE Full text] [doi: 10.1111/pcn.12988] [Medline: 32034840]
2. Brooks SK, Webster RK, Smith LE, Woodland L, Wessely S, Greenberg N, et al. The psychological impact of quarantine and how to reduce it: rapid review of the evidence. Lancet 2020 Mar 14;395(10227):912-920 [FREE Full text] [doi: 10.1016/S0140-6736(20)30460-8] [Medline: 32112714]
3. Pfefferbaum B, North CS. Mental health and the Covid-19 pandemic. N Engl J Med 2020 Aug 06;383(6):510-512. [doi: 10.1056/NEJMp2008017] [Medline: 32283003]
4. Serafini G, Parmigiani B, Amerio A, Aguglia A, Sher L, Amore M. The psychological impact of COVID-19 on the mental health in the general population. QJM 2020 Jun 22;113(8):531-537 [FREE Full text] [doi: 10.1093/qjmed/hcaa201] [Medline: 32569360]
5. Khan S, Siddique R, Li H, Ali A, Shereen MA, Bashir N, et al. Impact of coronavirus outbreak on psychological health. J Glob Health 2020 Jun;10(1):010331 [FREE Full text] [doi: 10.7189/jogh.10.010331] [Medline: 32355556]
6. Peng M, Mo B, Liu Y, Xu M, Song X, Liu L, et al. Prevalence, risk factors and clinical correlates of depression in quarantined population during the COVID-19 outbreak. J Affect Disord 2020 Oct 01;275:119-124 [FREE Full text] [doi: 10.1016/j.jad.2020.06.035] [Medline: 32658813]
7. Lemieux T, Milligan K, Schirle T, Skuterud M. Initial impacts of the COVID-19 pandemic on the Canadian labour market. Can Public Policy 2020 Jul 01;46(S1):S55-S65. [doi: 10.3138/cpp.2020-049]
8. Wilson JM, Lee J, Fitzgerald HN, Oosterhoff B, Sevi B, Shook NJ. Job insecurity and financial concern during the COVID-19 pandemic are associated with worse mental health. J Occup Environ Med 2020 Sep;62(9):686-691. [doi: 10.1097/JOM.0000000000001962] [Medline: 32890205]
9. Kimhi S, Marciano H, Eshel Y, Adini B. Resilience and demographic characteristics predicting distress during the COVID-19 crisis. Soc Sci Med 2020 Nov;265:113389 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113389] [Medline: 33039732]
10. Maaravi Y, Heller B. Not all worries were created equal: the case of COVID-19 anxiety. Public Health 2020 Aug;185:243-245 [FREE Full text] [doi: 10.1016/j.puhe.2020.06.032] [Medline: 32688099]
11. Bruine de Bruin W. Age differences in COVID-19 risk perceptions and mental health: evidence from a national U.S. survey conducted in March 2020. J Gerontol B Psychol Sci Soc Sci 2021;76(2):E. [doi: 10.1093/geronb/gbaa074]
12. Muhammad A, Ashjan B, Ghufran M, Taghreed S, Nada A, Nada A, et al. Classification of anxiety disorders using machine learning methods: a literature review. Insights Biomed Res 2020 Nov 07;4(1). [doi: 10.36959/584/455]
13. Nemesure MD, Heinz MV, Huang R, Jacobson NC. Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence. Sci Rep 2021 Jan 21;11(1):1980 [FREE Full text] [doi: 10.1038/s41598-021-81368-4] [Medline: 33479383]
14. AskingCanadians. Delvinia. URL: https://www.delvinia.com/solutions/askingcanadians/ [accessed 2021-03-22]
15. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
16. Orme JG, Reis J, Herz EJ. Factorial and discriminant validity of the center for epidemiological studies depression (CES-D) scale. J Clin Psychol 1986 Jan;42(1):28-33. [doi: 10.1002/1097-4679(198601)42:1<28::aid-jclp2270420104>3.0.co;2-t]

17. Saha S, Lim CCW, Cannon DL, Burton L, Bremner M, Cosgrove P, et al. Co-morbidity between mood and anxiety disorders: a systematic review and meta-analysis. Depress Anxiety 2021 Mar 22;38(3):286-306 [FREE Full text] [doi: 10.1002/da.23113] [Medline: 33225514]
18. Goodwin GM. The overlap between anxiety, depression, and obsessive-compulsive disorder. Dialogues Clin Neurosci 2015 Oct;17(3):249-260. [doi: 10.31887/dcns.2015.17.3/ggoodwin]
19. Borsboom D. The attack of the psychometricians. Psychometrika 2006 Sep;71(3):425-440 [FREE Full text] [doi: 10.1007/s11336-006-1447-6] [Medline: 19946599]
20. Furr RM. Scale Construction and Psychometrics for Social and Personality Psychology. Thousand Oaks, CA: Sage Publications; 2011.
21. Ekström J. A generalized definition of the polychoric correlation coefficient. Department of Statistics, UCLA. 2011. URL: https://escholarship.org/uc/item/583610fv [accessed 2021-11-09]
22. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: KDD '16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Presented at: KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17, 2016; San Francisco, CA p. 785-794. [doi: 10.1145/2939672.2939785]
23. Chen T, Guestrin C, XGBoost contributors. XGBoost documentation. dmlc XGBoost. URL: https://xgboost.readthedocs.io/en/latest/index.html [accessed 2021-03-22]
24. Lundberg S, Lee SI. A unified approach to interpreting model predictions. ArXiv. Preprint posted online on May 22, 2017. [FREE Full text]
25. Molnar C. SHAP (SHapley Additive exPlanations). GitHub. URL: https://christophm.github.io/interpretable-ml-book/shap.html [accessed 2021-11-09]
26. Vaid A, Somani S, Russak AJ, De Freitas JK, Chaudhry FF, Paranjpe I, et al. Machine learning to predict mortality and critical events in COVID-19 positive New York City patients. medRxiv. Preprint posted online on April 29, 2020. [doi: 10.1101/2020.04.26.20073411]
27. Population estimates on July 1st, by age and sex (table 17-10-0005-01). Statistics Canada. URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710000501 [accessed 2021-07-24]
28. Population estimates, quarterly. Statistics Canada. URL: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1710000901 [accessed 2021-07-24]
29. Wang M, Huang D, Wang G, Li D. SS-XGBoost: a machine learning framework for predicting newmark sliding displacements of slopes. J Geotech Geoenviron Eng 2020 Sep;146(9):04020074. [doi: 10.1061/(asce)gt.1943-5606.0002297]
30. COVID-19 daily epidemiology update. Government of Canada. URL: https://health-infobase.canada.ca/covid-19/epidemiological-summary-covid-19-cases.html [accessed 2021-03-30]
31. McLean CP, Asnaani A, Litz BT, Hofmann SG. Gender differences in anxiety disorders: prevalence, course of illness, comorbidity and burden of illness. J Psychiatr Res 2011 Aug;45(8):1027-1035 [FREE Full text] [doi: 10.1016/j.jpsychires.2011.03.006] [Medline: 21439576]
32. Kuehner C. Gender differences in unipolar depression: an update of epidemiological findings and possible explanations. Acta Psychiatr Scand 2003 Sep;108(3):163-174. [doi: 10.1034/j.1600-0447.2003.00204.x] [Medline: 12890270]
33. Thorson JA, Powell FC. The CES-D: four or five factors? Bull Psychon Soc 2013 Nov 14;31(6):577-578. [doi: 10.3758/bf03337359]
34. Cole JC, Rabin AS, Smith TL, Kaufman AS. Development and validation of a Rasch-derived CES-D short form. Psychol Assess 2004 Dec;16(4):360-372. [doi: 10.1037/1040-3590.16.4.360] [Medline: 15584795]
35. Ong AD, Standiford T, Deshpande S. Hope and stress resilience. In: Gallagher MW, Lopez SJ, editors. The Oxford Handbook of Hope. Oxford, UK: Oxford Handbooks Online; 2018.
36. Canadians' health and COVID-19, by age and gender. Statistics Canada. URL: https://doi.org/10.25318/1310080601-eng [accessed 2021-03-30]
37. Guo Y, Kopec JA, Cibere J, Li LC, Goldsmith CH. Population survey features and response rates: a randomized experiment. Am J Public Health 2016 Aug;106(8):1422-1426. [doi: 10.2105/ajph.2016.303198]
38. Solmi F, Downs JL, Nicholls DE. COVID-19 and eating disorders in young people. Lancet Child Adolesc Health 2021 May;5(5):316-318. [doi: 10.1016/s2352-4642(21)00094-8]
39. Panchal N, Kamal R, Orgera K. The implications of COVID-19 for mental health and substance use. Kaiser Family Foundation. 2021 Feb 10. URL: https://www.kff.org/coronavirus-covid-19/issue-brief/the-implications-of-covid-19-for-mental-health-and-substance-use/ [accessed 2021-11-09]
40. Holland KM, Jones C, Vivolo-Kantor AM, Idaikkadar N, Zwald M, Hoots B, et al. Trends in US emergency department visits for mental health, overdose, and violence outcomes before and during the COVID-19 pandemic. JAMA Psychiatry 2021 Apr 01;78(4):372-379. [doi: 10.1001/jamapsychiatry.2020.4402] [Medline: 33533876]
41. Ongoing study to understand the mental health of Canadians during COVID-19. Delvinia. URL: https://www.delvinia.com/camh-coronavirus-mental-health/ [accessed 2021-03-22]

Abbreviations

CES-D: Centre for Epidemiologic Studies Depression Scale
FDR: false discovery rate
GAD-7: Generalized Anxiety Disorder–7 questionnaire
LASSO: least absolute shrinkage and selection operator
PCA: principal component analysis
PC1: principal component 1
PC2: principal component 2
SHAP: SHapley Additive exPlanations
XGBoost: eXtreme Gradient Boosting

Edited by J Torous; submitted 12.08.21; peer-reviewed by M Nemesure, S Fleming; comments to author 03.10.21; revised version received 19.10.21; accepted 20.10.21; published 17.11.21.

Please cite as:
Hueniken K, Somé NH, Abdelhack M, Taylor G, Elton Marshall T, Wickens CM, Hamilton HA, Wells S, Felsky D
Machine Learning–Based Predictive Modeling of Anxiety and Depressive Symptoms During 8 Months of the COVID-19 Global Pandemic: Repeated Cross-sectional Survey Study
JMIR Ment Health 2021;8(11):e32876
URL: https://mental.jmir.org/2021/11/e32876
doi:10.2196/32876
PMID:34705663

Original Paper

Technostress and Digital Competence Among Health Professionals in Swiss Psychiatric Hospitals: Cross-sectional Study

Christoph Golz[1], MSc; Karin Anne Peter[1], PhD; Thomas Jörg Müller[2,3], MD, PhD; Jochen Mutschler[2], MD, PhD; Sandra M G Zwakhalen[4], PhD; Sabine Hahn[1], PhD

[1]Department of Health Professions, Bern University of Applied Sciences, Bern, Switzerland

[2]Private Clinic Meiringen, Bern, Switzerland

[3]Translational Research Center, University Hospital of Psychiatry and Psychotherapy, University of Bern, Bern, Switzerland

[4]Department of Health Services Research, Care and Public Health Research Institute, Maastricht University, Maastricht, Netherlands

Corresponding Author:
Christoph Golz, MSc
Department of Health Professions
Bern University of Applied Sciences
Murtenstrasse 10
Bern, 3008
Switzerland
Phone: 41 0318484591
Email: christoph.golz@bfh.ch

Abstract

Background: Psychiatric hospitals are becoming increasingly digitized because of the disruptive rise in technical possibilities. This digitization leads to new tasks and demands for health professionals, which can have an impact on technostress. It is unclear whether digital competence reduces technostress and how technostress affects health professionals' mental and physical health.

Objective: This study aims to assess the association between digital competence and technostress, considering individual characteristics and the association between technostress and long-term consequences for health professionals.

Methods: Cross-sectional data from 3 Swiss psychiatric hospitals were analyzed using multiple linear regression. The dependent variables for the models were digital competence, technostress, and long-term consequences (intention to leave the organization or the profession, burnout symptoms, job satisfaction, general health status, quality of sleep, headaches, and work ability). One model was calculated for each long-term consequence. The mean scores for technostress and digital competence could range between 0 (*fully disagree*) and 4 (*fully agree*), where a high value for technostress indicated high technostress and a high value for digital competence indicated high digital competence.

Results: The sample comprised 493 health professionals in psychiatric hospitals. They rated their technostress as moderate (mean 1.30, SD 0.55) and their digital competence as high (mean 2.89, SD 0.73). Digital competence was found to be significantly associated with technostress (β=−.20; *P*<.001). Among the individual characteristics, age (β=.004; *P*=.03) and profession were significantly associated with both digital competence and technostress. Technostress is a relevant predictor of burnout symptoms (β=10.32; *P*<.001), job satisfaction (β=−6.08; *P*<.001), intention to leave the profession (β=4.53; *P*=.002), organization (β=7.68; *P*<.001), general health status (β=−4.47; *P*<.001), quality of sleep (β=−5.87; *P*<.001), headaches (β=6.58; *P*<.001), and work ability (β=−1.40; *P*<.001).

Conclusions: Physicians and nurses who have more interaction with digital technologies rate their technostress higher and their digital competence lower than those in other professions. Health professionals with low interaction with digital technologies appear to overestimate their digital competence. With increasing digitization in psychiatric hospitals, an increase in the relevance of this topic is expected. Educational organizations and psychiatric hospitals should proactively promote the digital competence of health professionals to manage expected disruptive changes.

(JMIR Ment Health 2021;8(11):e31408) doi:10.2196/31408

KEYWORDS

technostress; digital competence; psychiatry; health professionals; multiple regression

Introduction

Background

Psychiatric hospitals are increasingly becoming digitized because of the disruptive rise in technical possibilities [1,2] and legal requirements, such as the obligation to use nationally shared electronic health records [3]. Moreover, the COVID-19 pandemic has underlined the need for additional digital services such as telemedicine or remote monitoring in mental health to avoid social exclusion through lockdowns or because of living situations in remote regions [4,5]. Health professionals are thus increasingly confronted with digital technologies for clinical practice, interaction with patients, and administrative tasks.

Therefore, digitalization creates new tasks for health professionals and places demands on them that are not part of their education and training. These include, for example, the management of data privacy [1] or digital competences to enhance appropriate patient communication via internet [6]. In addition, new tasks make demands such as increasing time spent with documentation [7,8] or with low usability electronic health records [9] and technical support among colleagues [10], which were previously beyond the scope of work of health professionals.

The demands for digital competences and associated changes in the role of health professionals also require a change in the perception of and attitude toward digital resources in everyday work [11]. Consequently, this transformation may have a stress-inducing effect on health professionals, especially because psychiatric health professionals tend to be hesitant regarding new technologies because of the expected deleterious effects on the relationship between health professionals and patients [12,13]. For example, they may feel more disturbed by the digitization of their daily work than their colleagues in settings that are traditionally more digitized, such as acute care with intensive care units.

The phenomenon called technostress is “a reflection of one’s discomposure, fear, tenseness and anxiety when one is learning and using computer technology” [14]. The term was introduced in 1984 by Brod [15] as “a modern disease of adaptation caused by an inability to cope with the new computer technologies in a healthy manner” during the rapid emergence of technology in everyday life. Studies on technostress among health professionals are scarce [16,17]. A recent study revealed that psychiatric health professionals experience a moderate level of technostress [16].

Technostress is known to have an effect not only on the working life of professionals [10], such as reduced job satisfaction [18,19], but also on their private life, such as psychophysiological reactions such as headaches and fatigue [20,21] or burnout symptoms [22]. Exposure to stress-inducing technology can even result in reduced ability to work and an intention to leave the job, which could exacerbate the already-existing shortage of health professionals [23].

An important factor in technostress is expected to be an individual’s digital competence, as higher digital competence has been identified as having a mitigating association with technostress [10,24]. However, it was found that professionals with high digital competence tended to feel particularly stressed by the nonavailability or unreliability of the technologies used at work [24]. Research on digital competence among health professionals has quite a strong focus on the knowledge and skills of using digital technologies at work [25] or specific subgroups in nursing, such as nurse leaders [26,27]. The TIGER Nursing Informatics Competencies Model, for example, consists of 3 parts: basic computer competences (eg, using the computer and managing files), information literacy (eg, evaluating information and its sources critically), and information management (eg, using electronic health records) [25]. However, additional factors, such as attitude, motivation, and experience of using digital technologies, are also thought to be relevant in the context of digital competence. A recent review of research on health professionals’ digital competence summarized the key areas of this competence as “sufficient knowledge and skills [...], social and communication skills [...], motivation and willingness [...] and support for positive experiences in digitalization” [28]. Therefore, besides insufficient knowledge and skills for proper implementation and use of digital technologies, a lack of motivation and prejudice against digitalization are, for example, associated with reduced technology use. Moreover, health professionals must adapt their communication style, depending on whether they communicate face to face or via telemedicine [28]. Therefore, behavioral determinants are crucial for enhancing digital competence in addition to knowledge and skills [29].

Unfortunately, findings on digital competence and its association with technostress are not specific to health professionals in psychiatric hospitals. However, it is especially important for health professionals that information on their digital competence and technostress is needed, as they are considered to be reluctant adapters of digitization, despite increasing calls for adaptation to new tasks and requirements to keep up with their profession. These contradictions of reluctance and ongoing change need to be addressed at an early stage.

Objective

This paper, therefore, aims to answer the following research questions:

1. How do health professionals in psychiatric hospitals rate their digital competence?
2. How do health professionals in psychiatric hospitals rate their technostress?
3. What is the association between health professionals’ digital competence and their technostress, considering the individual characteristics of health professionals?
4. What is the association between technostress and long-term consequences for health professionals?

Methods

This cross-sectional study was conducted in 3 psychiatric hospitals in the German-speaking part of Switzerland as part of the Work-Related Stress Among Health Professionals in Switzerland (STRAIN) study [23]. This study is based on a cluster randomized controlled trial (Clinical Trials registration NCT03508596) consisting of 3 measurements (baseline, first, and second) and investigating work-related stress among health professionals in Switzerland.

Sample and Recruitment

The study sample of the STRAIN study included acute care and rehabilitation hospitals, psychiatric hospitals, nursing homes, and home care organizations. Detailed information on the STRAIN study sample has been published elsewhere [23]. For this study, a request to participate was sent to 12 psychiatric hospitals that had already participated in the STRAIN study. The internal coordinators of the psychiatric hospitals were contacted by email and asked whether their institution's health professionals might participate in this study, which would focus on technostress and digital competences. The project was then presented to decision makers at the psychiatric hospitals. Health professionals from the following work categories were included in this study: nursing staff, physicians, psychologists, medical therapeutic professionals, and social workers. Participants who labeled themselves as *researcher* or secretariat in the additional free text field were excluded. Overall, 1767 health professionals were eligible for participation in the study.

Data Collection

The study was conducted along with the second measurement of the STRAIN study between June and September 2020. The questionnaires for health professionals from the institutions that had agreed to participate were expanded to include topic-specific scales measuring technostress and digital competence.

The internal coordinator of the participating psychiatric hospitals disseminated the information for the participants and the survey to health professionals. Participation in the study was possible via paper or web-based questionnaires in German. For the paper questionnaires, a prestamped envelope was enclosed to return the questionnaire to the project team. For the web-based questionnaire, the link to the web-based survey using SurveyMonkey and UmfrageOnline was either sent individually by email or published on the organization's intranet by the coordinator. A reminder to complete the questionnaire was sent electronically or on paper 2 weeks afterward by the internal coordinator.

The Questionnaires

The 3 questionnaires used in this study comprised a technostress questionnaire [24], an in-house-developed digital competence questionnaire, and the STRAIN questionnaire [23]. The questionnaires were estimated to take 45 minutes overall to complete.

Technostress Questionnaire

For the measurement of technostress, the scale created by Gimpel et al [24] was used. The scale, which shows satisfactory reliability (Cronbach α=.91), is based on the technostress model of Ayyagari et al [30]—a model widely used in research on technostress. It consists of 12 items using a 5-point Likert scale, with the end points 0 (*fully disagree*) and 4 (*fully agree*). For interpretation of the data, the mean score was calculated (min=0; max=4), where a high score indicates high technostress. The questionnaire covers the following 12 items, which are derived from the theory's dimensions: uncertainty (ongoing changes lead to uncertainty and constant learning), insecurity (feeling threatened about losing one's job), unreliability (unreliability of technology used), overload (technology forces users to work faster and longer), invasion (employees can be reached anytime), complexity (users feel inadequate regarding their competences), performance control (feeling of being monitored and compared), ambiguity of the role (technical problems must be solved by oneself), interruptions (malfunctions and unstable systems), nonavailability (lack of technology that can reduce workload), no sense of achievement (feeling of lack of progress at work), and invasion of private life (feeling one's private life is affected).

Digital Competence Questionnaire

To measure digital competence among health professionals, no suitable and compact questionnaire was available that focused on the 5 key areas of digital competence (knowledge, skills, communication, experience, and attitude) for health professionals [28]. Moreover, to not lengthen the already-long questionnaire excessively, thereby negatively influencing the response rate, a short self-assessment scale measuring digital competence was needed. Therefore, for each of the 5 key areas, an item was developed in-house. The 5 items covered the following topics: knowledge (eg, one's own knowledge of digital technologies at work), skills (confidence in using digital technologies at work), communication (eg, confidence in communication using digital technologies at work), motivation (eg, motivation to use digital technologies in everyday work), and attitude (eg, attitude toward potential improvements through digital technologies at work). Items were scored on a 5-point Likert scale ranging from 0 (*fully disagree*) and 4 (*fully agree*). For interpretation, the mean score was calculated (min=0; max=4), with a high score again indicating high digital competence.

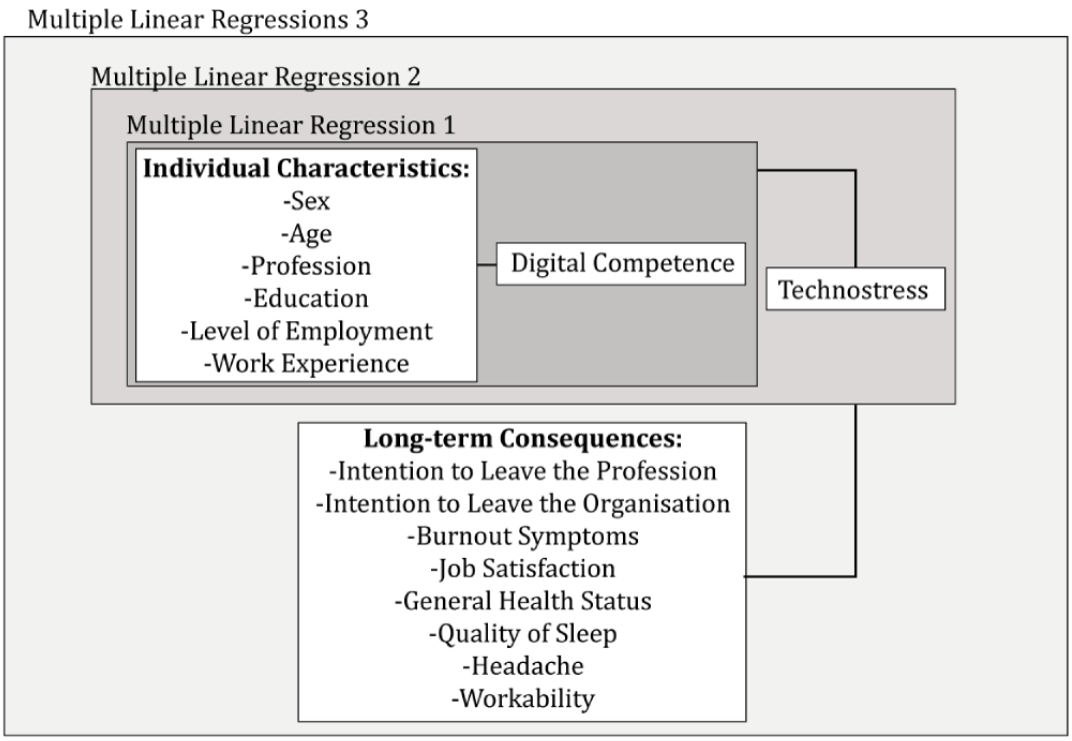
The single items of digital competence were tested for construct validity by conducting exploratory factor analysis and reliability tests. The requirements for factor analysis were met with item correlations above 0.3 and a significant Bartlett test of sphericity (χ^2_4=39.4, P<.001) and the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy with acceptable values above 0.6 (KMO=0.81). A scree plot was used to test for loadings on one factor. The reliability test for the 5 developed items on digital competence revealed satisfactory internal consistency (Cronbach α=.87; Multimedia Appendix 1).

STRAIN Questionnaire

The outcome variables (Figure 1) for long-term consequences stem from the STRAIN questionnaire [23,31], which comprises well-known, valid, and reliable scales such as the Copenhagen Psychosocial Questionnaire (COPSOQ) [32], the self-rated general health status [33], the Nurses' Early Exit study questionnaire [34], the von Korff questionnaire [35], and the

workability index [36]. The scores from the COPSOQ, the Nurses' Early Exit study questionnaire, the von Korff questionnaire, and the general health status ranged from a value of 0 (*do not agree at all*) to 100 (*fully agree*) or from 0 (*worst imaginable health state*) to 100 (*best imaginable health state*) for the general health status and from 0 (*no influence*) to 100 (*could no longer perform activity*) for the von Korff questionnaire. The COPSOQ scale scores were included if at least half of the items had no missing values [37]. The total score of the workability index questionnaire ranged from 7 (*minimum working capacity*) to 49 (*maximum working capacity*).

Figure 1. Scales used for the multiple linear regression models.



Data Analysis

The analysis was conducted using R version 3.6.1 [38] and included descriptive statistics for technostress and digital competence. Multiple linear regression models were calculated using the MASS package [39]. The predictor and outcome variables were chosen to cover the dimensions of the DSM [24]. The model describes the correlation between technostress, inhibitors of technostress, and consequences of technostress. Furthermore, individual characteristics (eg, age, education, and sex) were added to the model, as they have been identified as relevant predictors elsewhere [10]. To answer the research questions, multiple linear regressions were conducted (1) with digital competence as the outcome and individual characteristics as predictors; (2) with technostress as the outcome and individual characteristics and digital competence as predictors; and (3) with long-term consequences as outcome variables and technostress, digital competence, and individual characteristics as predictors (Figure 1). For each of the following long-term consequences, a separate multiple linear regression was calculated: intention to leave the organization [23], intention to leave the profession [23], burnout symptoms [32], job satisfaction [32], general health status [33], quality of sleep [34], headache [35] and workability [36].

To minimize the effect of internal dropouts, missing data were filled in based on multiple imputation expecting data to be missing completely at random, using the MICE package [40]. To test for multicollinearity, the variance inflation factor was computed (1.06-1.70), which is regarded as acceptable to proceed if variables show values less than 3 [41]. The assumption of heteroskedasticity was tested using the Breusch-Pagan test. This was met for multiple linear regressions. Therefore, SEs, *P* values, and CIs were bootstrapped (r=999, bias corrected and accelerated, 95% CI). A stepwise model selection was conducted for the multiple linear regressions based on the Akaike information criterion [42].

Ethical Considerations

The local Swiss ethical board confirmed that the study did not warrant a full ethical application and did not fall under the Swiss Federal Act on research involving human beings (Req-2020-00179). The participants were professionals and could take responsibility for their own participation. They received written information before the start of the study regarding the subject, aim, and voluntary nature of their participation. Filling in the questionnaire was counted as informed participation. The data were gathered anonymously and could not be traced back to individual participants.

Results

In total, 493 health professionals participated in the study, corresponding to a response rate of 27.9% (493/1767). Among the participants, 60% (296/493) were nurses, 12.3% (61/493)

were psychologists, 11.1% (55/493) were social workers, 8.7% (43/493) were physicians, and 7.7% (38/493) were medical-therapeutic professionals. The mean age of the participants was 41 (SD 12.33) years, and the majority were female (349/493, 71%). For technostress, health professionals reported a moderate mean score of 1.30 (SD 0.55). Nursing staff (mean 1.41, SD 0.54) and physicians (mean 1.41, SD 0.54) had the highest scores among the professions included, followed by medical-therapeutic professionals (mean 1.23, SD 0.60), social workers (mean 1.15, SD 0.57), and psychologists (mean 0.95, SD 0.40). Health professionals rated their digital competence high, with a mean score of 2.82 (SD 0.76): social workers were found to have the highest score (mean 3.18, SD 0.57), followed by medical-therapeutic professionals (mean 2.90, SD 0.84), psychologists (mean 2.89, SD 0.73), physicians (mean 2.82, SD 0.66), and nurses (mean 2.71, SD 0.78).

Technostress

Table 1 summarizes the results of the multiple linear regression, with technostress as the outcome variable. The regression model was shown to be significant $F_{5,487}$=19.81 (P<.001) and explained 20% of the variance (R^2). Being a physician (β=.22; P=.03) or a nurse (β=.17; P=.02) was shown to have an increasing association with technostress, compared with being a social worker (intercept), whereas being a psychologist was negatively associated with technostress (β=−0.23; P=.01). Digital competence was also negatively associated with technostress (β=−0.20; P<.001). This means that an increase in digital competence of 1 point results in a decrease in technostress by −0.20 points of the mean score.

Table 1. Multiple linear regression with technostress as the outcome [observations N=493; technostress: 0 (no technostress) to 4 (high technostress)].

Coefficient	β	SE	*t* value (*df*)	*P* value	95% CI
Intercept	1.63	0.15	10.86 (487)	<.001	1.62 to 1.64
Age	.004	0.002	2.21 (1)	.03[a]	0.004 to 0.004
Physicians	.22	0.10	2.22 (1)	.03[a]	0.22 to 0.23
Psychologists	−.23	0.09	−2.53 (1)	.01[a]	−0.24 to −0.23
Nurses	.17	0.07	2.30 (1)	.02[a]	0.16 to 0.17
Digital competence	−.20	0.03	−6.71 (1)	<.001	−0.21 to −0.20

[a]With bootstrap.

Digital Competence

The multiple linear regression with digital competence as the outcome was shown to be significant $F_{6,486}$=10.47 (P<.001) and explained 13% of the variance (R^2). Being male was shown to be positively but not significantly associated with digital competence (β=.11; P=.15). In addition, the level of employment was positively associated with digital competence (β=.006; P<.001). Age was negatively associated with digital competence (β=−0.014; P<.001), meaning that digital competence decreased marginally with increasing age (Table 2).

Table 2. Multiple linear regression with digital competence as outcome [observations N=493; digital competence: 0 (no digital competence) to 4 (high digital competence)].

Coefficient	β	SE	*t* value (*df*)	*P* value	95% CI
Intercept	3.25	0.21	15.52 (486)	<.001	3.24 to 3.26
Sex (male)	.11	0.08	1.45 (1)	.15[a]	0.10 to 0.11
Age	−.014	0.003	−5.29 (1)	<.001	−0.01 to −0.01
Level of employment	.006	0.002	3.21 (1)	<.001	0.006 to 0.006
Physicians	−.46	0.15	−3.11 (1)	<.001	−0.47 to −0.45
Psychologists	−.26	0.13	−1.92 (1)	.06[a]	−0.26 to −0.25
Nurse	−.48	0.11	−4.55 (1)	<.001	−0.49 to −0.48

[a]With bootstrap.

Long-Term Consequences

The results of the multiple regression models with long-term consequences as the outcome variables are shown in Multimedia Appendices 2 and 3. The models indicate that the independent variables predict the outcome *burnout symptoms* as best (R^2=0.16, $F_{10,482}$=9.28; P<.001), followed by *intention to leave the organization* (R^2=0.15, $F_{13,485}$=6.37; P<.001) and *job satisfaction* (R^2=0.15, $F_{12,480}$=5.28; P<.001). *General health status* turned out to have the lowest explanatory power with the included predictor variables (R^2=0.06, $F_{3,489}$=9.88; P<.001).

In all models, technostress was significantly associated with outcome variables. The highest impact was found for *burnout*

symptoms, with an increase of 10.32 (*P*<.001) associated with an increase in technostress of 1 point. Technostress was also positively associated with *headache* (β=6.58; *P*<.001) and the outcomes *intention to leave the profession* (β=4.53; *P*=.02) and *intention to leave the organization* (β=4.53; *P*<.001). Moreover, technostress was negatively associated with *job satisfaction* (β=−6.08; *P*<.001), *general health status* (β=−4.47; *P*<.001), *quality of sleep* (β=−5.87; *P*<.001), and *workability* (β= −1.40; *P*<.001).

The predictor variable, digital competence, was included in 6 of the 8 models. The effect of digital competence was lower than that of technostress. Digital competence was positively associated with *quality of sleep* (β=4.19; *P*<.001), *job satisfaction* (β=2.26; *P*=.02), and *workability* (β=.79; *P*=.002). When interpreting the results, attention must be paid to the possible scores of the outcome variables. Thus, an increase in digital competence of 1 point leads to an increase in workability of 0.79, whereby workability can range from 7 to 49. An increase of 1 point in digital competence leads to an increase of 2.26 points in job satisfaction on a possible range of 0 to 100.

Discussion

Principal Findings

Health professionals in psychiatry rate their technostress as moderate, and their digital competence as high. Higher digital competence was also significantly associated with lower technostress. Individual characteristics differ in their relevance to the models. The age of health professionals is significantly associated with technostress and digital competence. Older health care professionals appear to experience higher technostress and perceive themselves as having lower digital competence. Physicians and nurses appear in the models to have higher technostress and lower competence compared with the other professions surveyed. Being a nurse was shown to have the highest estimates across all outcomes.

To answer the question of the association between technostress and long-term outcomes of health professionals, it should be noted that technostress has a nonnegligible impact on long-term consequences, such as burnout symptoms, job satisfaction, and headache. Thus, technostress has a measurable association with the mental and physical health of health professionals. In addition, technostress promotes the intention to leave the organization or the profession.

Comparison With Prior Work

The significant association of digital competence with technostress is in line with another study in which *computer self-efficacy* (ie, digital competence) is described as an antecedent of technostress [10]. This association highlights the potential of enhanced digital competence to reduce technostress. However, the β values in the technostress model were equally high for the professions, which could mean that health professionals need to interact with digital technologies to varying degrees at work.

Interestingly, physicians and nurses who are known to have higher technostress [16] and thought to have more interaction with digital technologies than other health professionals were shown to have lower digital competence. This is in contrast with the findings of Kuek and Hakkennes [43], who found that health professionals with high-frequency digital technology use also showed higher digital competence. However, they argued that the organization in which the study took place was digitized more than organizations in comparable studies. One reason for the reported lower digital competence in this study could be past experience with digital technologies rather than a lack of knowledge and skills. Past experiences could have been negative because of a lack of *suitable rooms or technical equipment and failing support systems* [28]. Furthermore, it raises the question of whether health professionals who have experienced fewer negative interactions rate their digital competence higher because of the absence of digital technologies at work. These results are somewhat at odds with the results of other studies in which people who have little contact with digital technologies show higher levels of technostress because they lack opportunities to adapt and develop their own skills in using them [24]. This phenomenon could be explained by the Dunning-Kruger paradigm for this study. Studies “repeatedly show that people with little expertise [in the specific field] often grossly overestimate how much they know and how well they perform” [44]. However, this study does not provide any insights into the extent of interactions of health professionals with digital technologies.

Furthermore, lower digital competence (ie, computer proficiency) has been found to be a barrier to successful implementation of electronic health records in psychiatric hospitals [11]. This would imply that Swiss psychiatric hospitals have a good precondition for the successful implementation of digital technologies, as the digital competence of health professionals was rated high. However, being an active user of electronic health records was one of the inclusion criteria for the study, which means that participants self-rated their digital competence by having sufficient experience of interaction with digital technologies. According to Staggers et al [45], there are 4 different levels of digital competence for nurses. They propose that experienced nurses (level 2) are “highly skilled in using information management and computer technology skills” [45]. This expands the understanding of the core competences necessary for consideration as an experienced professional and places a requirement on educational organizations and psychiatric hospitals to support health professionals in fulfilling this aim. Recent findings also highlight the importance of leaders investing in technical support for their employees, such as “receiving low support in learning and using digital tools” [46], which is expected to contribute to enhanced digital competence [28].

Concerning gender, there was no strong evidence as to whether males or females were more affected by technostress. However, the model for digital competence indicated that being male was slightly but not significantly associated with digital competence (*P*=.15). One reason for this result could be that the clear majority of participants were female (71%), which could have led to an underestimation of the potential difference between the sexes. Regarding the technical support described earlier, females seem to compensate for their lower digital competence by relying on the organization’s helpdesk, whereas males tend

to exchange expertise [47]. This implies that health organizations might want to invest in a low-threshold helpdesk and train health professionals with an affinity for digital technologies to become peer supporters.

Evidence for the effects of individual characteristics is inconsistent, particularly with respect to age and sex [10]. This study contributes to the discussion by indicating that age is a relevant predictor of both technostress and digital competence. In terms of digital competence, the results of this study appear to confirm that younger health care professionals perceive themselves as having higher digital competency [48]. However, recent findings, albeit nonspecific to the health care setting, indicate that females tend to be more affected by technostress [49]. In this respect, a possible effect of sex should be considered in future studies that focus on health care professionals. If it turns out that women are more affected by technostress in the health care system, the intended measures must take this possible precondition into consideration.

In terms of the association between technostress and its long-term consequences, other findings from other sectors underline that higher technostress leads to higher intention to leave the profession or organization and lower job satisfaction [50]. Furthermore, additional influencing factors in health care appear to have a more important impact on long-term consequences for health professionals, such as work-private life conflict or quantitative demands at work [23,51]. However, some aspects of private life conflicts are incorporated into the technostress scale used. One of the themes of technostress is *techno-invasion*, which measures the self-perceived aspect that one can be reached at any time. Also, the theme *invasion of private life* is part of the technostress scale, assessing the feeling that one's private life is affected by digital technologies at work. Although these aspects are included in the technostress scale, the findings in this study do not reach the explained variance of the study indicated above. Therefore, it seems that digital technologies do not currently play a vital role in the context of private life conflicts among health professionals in psychiatric hospitals.

In view of the fact that the Swiss health care system is still only partly digitized in terms of international comparison [52] and that psychiatry is not expected to lead the way in digitization, these findings seem logical. However, with a future increase of digitization in psychiatric hospitals [53], the topic's relevance is expected to rise. For example, a recent study described the empowerment and enslavement paradox of digital technologies for surgeons [54]. The study highlights the issue that with an increase in possibilities because of digital technologies, the danger of misuse increases, which negatively impacts the outcomes of health professionals and patients. The implication for psychiatric hospitals is, therefore, that technostress is not a major issue at the moment. However, psychiatric hospitals are encouraged to invest in monitoring the digital competence of their health professionals, especially along with the implementation of digital technologies, and offer suitable training to their employees. Furthermore, decision makers should involve health professionals in the development and implementation of digital technologies, as involvement has been identified as crucial for positive experiences with digital technologies, increasing motivation toward innovations and dismantling prejudices [10]. Health professionals must recognize that they are going to face digitization at their workplace. However, because many health professionals have a rather reserved attitude toward digital technologies at work, decision makers should approach this process thoughtfully.

Strengths and Limitations

This study contributes to the emerging topic of technostress among health professionals in a psychiatric setting. It provides first insights into the association of digital competence with technostress and the association of the two with long-term consequences. This study enriches the discussion on the potential influence of individual characteristics, such as age, sex, profession, and education. Furthermore, a digital competence scale with satisfactory properties was developed and evaluated in this study. This scale is made available to the community for use in further research (Multimedia Appendix 1).

However, this study had several limitations. First, convenience sampling was performed. Of the 12 psychiatric hospitals invited, only 3 agreed to participate. It cannot be excluded that psychiatric hospitals whose staff generally experience lower technostress agreed to participate because they were more sensitized to the topic. In addition, the sample did not reflect the typical distribution of health professionals in Swiss psychiatric hospitals. In this study, physicians were underrepresented (9%), compared with the usual proportion of 17% [55]. This might be because physicians are increasingly reluctant to participate in surveys for reasons such as information overload, survey fatigue, or privacy concerns [56]. In addition, a response rate of 27.9% (493/1767) is considered low but rather common for web-based surveys with health professionals [57,58]. Unfortunately, forecasts indicate even lower average response rates soon [59]. Furthermore, participants could decide to use either a paper or web-based questionnaire. The comparability of paper and web-based questionnaires has been discussed in the literature. Psychological factors, such as mood state or fatigue during the inquiry, can have an impact on responses and can be influenced by *environmental stimuli or distractions* [60]. Especially in health care organizations in which the number of computers on the wards is limited and no quiet place is available to withdraw, this could have had a deleterious effect on responses. In addition, one organization opted exclusively for web-based inquiry. Staff members who feel highly stressed by digital technologies could have been excluded by this decision because they did not want to use the computer unnecessarily for longer than was required by their work. Moreover, no causal conclusions can be drawn, as this study used cross-sectional data. These implications must be considered when interpreting the results.

Conclusions

Health professionals in Swiss psychiatric hospitals experience moderate technostress at work. They rated their digital competence as high. It might be that health professionals with little interaction with digital technologies at work overestimate their digital competence. Therefore, to generate reliable results on this hypothesis in the future, the degree of digitization of the organization and the degree of contact with digital technologies

at the individual level must be additionally assessed. In this context, research should evaluate whether self-rated digital competence corresponds to an objective assessment of digital competence at work, which would contribute to further development of the measurement tool for digital competence.

Technostress has been shown to have a relevant association with long-term consequences for staff, especially those with burnout symptoms. Further digitization in psychiatric hospitals is expected to have an increasing impact on the technostress experienced. Additional digital competence will be needed as an inhibitor of technostress for health professionals to sustainably cope with technostress and, thus, lower the risk of long-term consequences.

Health professionals and professionals in educational organizations do not yet recognize the need for future digital competences. Health and educational organizations are responsible for the adequate preparation of future health professionals; however, this should include training aimed at digital competence.

Psychiatric hospitals can draw conclusions based on these results. As digital competence significantly reduced technostress, further in-house education to promote digital competence should be established. Furthermore, the duties of younger health professionals could be extended to support older health professionals in managing digital technologies at work. Mutual support is demonstrably conducive to acquiring new competences and strengthening the sense of community in the team. However, this presupposes that such a duty is appropriately appreciated and remunerated.

Psychiatric hospitals in Switzerland are still in their early days in terms of the impact of digital technologies on health professionals. The necessary digital competences will emerge as the digitization process progresses. Researchers must continue to monitor this development and generate recommendations for measures to reduce technostress and develop suitable educational content from intervention studies.

Acknowledgments

The authors would like to thank the hospitals and the health professionals for their participation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Questionnaire Digital Competence.
[DOCX File , 45 KB - mental_v8i11e31408_app1.docx]

Multimedia Appendix 2
Multiple linear regression models with long-term consequences as outcomes part 1 (observations N=493).
[DOCX File , 49 KB - mental_v8i11e31408_app2.docx]

Multimedia Appendix 3
Multiple linear regression models with long-term consequences as outcomes part 2 (observations N=493).
[DOCX File , 48 KB - mental_v8i11e31408_app3.docx]

References

1. Hariman K, Ventriglio A, Bhugra D. The future of digital psychiatry. Curr Psychiatry Rep 2019 Aug 13;21(9):88. [doi: 10.1007/s11920-019-1074-4] [Medline: 31410728]
2. Gratzer D, Torous J, Lam RW, Patten SB, Kutcher S, Chan S, et al. Our digital moment: innovations and opportunities in digital mental health care. Can J Psychiatry 2021 Jan 30;66(1):5-8 [FREE Full text] [doi: 10.1177/0706743720937833] [Medline: 32603188]
3. Was ist das EPD? eHealth Suisse. 2021. URL: https://www.patientendossier.ch/de/bevoelkerung/kurz-erklaert [accessed 2021-02-15]
4. Keesara S, Jonas A, Schulman K. Covid-19 and health care's digital revolution. N Engl J Med 2020 Jun 04;382(23):e82 [FREE Full text] [doi: 10.1056/nejmp2005835]
5. Ćosić K, Popović S, Šarlija M, Kesedžić I. Impact of human disasters and COVID-19 pandemic on mental health: potential of digital psychiatry. Psychiatr Danub 2020 Apr 15;32(1):25-31 [FREE Full text] [doi: 10.24869/psyd.2020.25] [Medline: 32303026]
6. Greysen SR, Chretien KC, Kind T, Young A, Gross CP. Physician violations of online professionalism and disciplinary actions: a national survey of state medical boards. J Am Med Assoc 2012 Mar 21;307(11):1141-1142. [doi: 10.1001/jama.2012.330] [Medline: 22436951]

7. Melnick ER, Dyrbye LN, Sinsky CA, Trockel M, West CP, Nedelec L, et al. The association between perceived electronic health record usability and professional burnout among US physicians. Mayo Clin Proc 2020 Mar;95(3):476-487 [FREE Full text] [doi: 10.1016/j.mayocp.2019.09.024] [Medline: 31735343]
8. Momenipur A, Pennathur PR. Balancing documentation and direct patient care activities: a study of a mature electronic health record system. Int J Ind Ergon 2019 Jul;72:338-346. [doi: 10.1016/j.ergon.2019.06.012] [Medline: 32201437]
9. Melnick ER, Harry E, Sinsky CA, Dyrbye LN, Wang H, Trockel MT, et al. Perceived electronic health record usability as a predictor of task load and burnout among US physicians: mediation analysis. J Med Internet Res 2020 Dec 22;22(12):e23382 [FREE Full text] [doi: 10.2196/23382] [Medline: 33289493]
10. La Torre G, Esposito A, Sciarra I, Chiappetta M. Definition, symptoms and risk of techno-stress: a systematic review. Int Arch Occup Environ Health 2019 Jan 8;92(1):13-35. [doi: 10.1007/s00420-018-1352-1] [Medline: 30196317]
11. Jung SY, Hwang H, Lee K, Lee D, Yoo S, Lim K, et al. User perspectives on barriers and facilitators to the implementation of electronic health records in behavioral hospitals: qualitative study. JMIR Form Res 2021 Apr 08;5(4):e18764 [FREE Full text] [doi: 10.2196/18764] [Medline: 33830061]
12. Pickersgill M. Digitising psychiatry? Sociotechnical expectations, performative nominalism and biomedical virtue in (digital) psychiatric praxis. Sociol Health Illn 2019 Oct 02;41 Suppl 1(S1):16-30 [FREE Full text] [doi: 10.1111/1467-9566.12811] [Medline: 30175439]
13. Bucci S, Schwannauer M, Berry N. The digital revolution and its impact on mental health care. Psychol Psychother 2019 Jun 28;92(2):277-297. [doi: 10.1111/papt.12222] [Medline: 30924316]
14. Wang K, Shu Q, Tu Q. Technostress under different organizational environments: an empirical investigation. Comput Human Behav 2008 Sep;24(6):3002-3013. [doi: 10.1016/j.chb.2008.05.007]
15. Brod C. Technostress: The Human Cost Of The Computer Revolut. Boston, MA: Addison Wesley Publishing Company; 1984:1-242.
16. Golz C, Peter KA, Zwakhalen SM, Hahn S. Technostress among health professionals - a multilevel model and group comparisons between settings and professions. Inform Health Soc Care 2021 Jun 02;46(2):136-147. [doi: 10.1080/17538157.2021.1872579] [Medline: 33646891]
17. Califf C, Sarker S, Sarker S, Fitzgerald C. The bright and dark sides of technostress: an empirical study of healthcare workers. In: Proceedings of the International Conference on Information Systems - Exploring the Information Frontier, ICIS 2015. 2015 Presented at: International Conference on Information Systems - Exploring the Information Frontier, ICIS 2015; December 13-16, 2015; Fort Worth, Texas, USA URL: https://www.researchgate.net/publication/309285357_The_Bright_and_Dark_Sides_of_Technostress_An_Empirical_Study_of_Healthcare_Workers
18. Babbott S, Manwell LB, Brown R, Montague E, Williams E, Schwartz M, et al. Electronic medical records and physician stress in primary care: results from the MEMO Study. J Am Med Inform Assoc 2014 Feb 01;21(e1):100-106 [FREE Full text] [doi: 10.1136/amiajnl-2013-001875] [Medline: 24005796]
19. Friedberg MW, Chen PG, Van Busum KR, Aunon F, Pham C, Caloyeras J, et al. Factors affecting physician professional satisfaction and their implications for patient care, health systems, and health policy. Rand Health Q 2014;3(4):1 [FREE Full text] [Medline: 28083306]
20. Gimpel H, Lanzl J, Manner-Romberg T, Nüske N. Digitaler stress in Deutschland. In: Eine Befragung von Erwärbstätigen zu Belastung und Beanspruchung durch Arbeit mit digitalen Technologien. Düsseldorf, Germany: Hans Böckler Stiftung; 2018:1-57.
21. Chiappetta M. The Technostress: definition, symptoms and risk prevention. Sens Sci 2017;4(1):358361. [doi: 10.14616/sands-2017-1-358361]
22. Berg-Beckhoff G, Nielsen G, Larsen E. Use of information communication technology and stress, burnout, and mental health in older, middle-aged, and younger workers - results from a systematic review. Int J Occup Environ Health 2017 Apr 20;23(2):160-171 [FREE Full text] [doi: 10.1080/10773525.2018.1436015] [Medline: 29460697]
23. Peter KA, Hahn S, Schols JM, Halfens RJ. Work-related stress among health professionals in Swiss acute care and rehabilitation hospitals - A cross-sectional study. J Clin Nurs 2020 Aug 16;29(15-16):3064-3081. [doi: 10.1111/jocn.15340] [Medline: 32447796]
24. Gimpel H, Lanzl J, Regal C. Gesund Digital Arbeiten?! Eine Studie zu Digitalem Stress in Deutschland. Augsburg: Fraunhofer-Institut für Angewandte Informationstechnik FIT. Bundesanstalt für Arbeitsschutz und Arbeitsmedizin, MF/M-Bayreuth; 2019.
25. Hunter K, McGonigle D, Hebda T, Sipes C, Hill T, Lamblin J. TIGER-Based Assessment of Nursing Informatics Competencies (TANIC). In: New Contributions in Information Systems and Technologies. Cham: Springer; 2015:171-177.
26. Collins S, Yen P, Phillips A, Kennedy M. Nursing informatics competency assessment for the nurse leader: the delphi study. J Nurs Adm 2017 Apr;47(4):212-218. [doi: 10.1097/NNA.0000000000000467] [Medline: 28333789]
27. Yen P, Phillips A, Kennedy M, Collins S. Nursing informatics competency assessment for the nurse leader: instrument refinement, validation, and psychometric analysis. J Nurs Adm 2017 May;47(5):271-277. [doi: 10.1097/NNA.0000000000000478] [Medline: 28422933]
28. Konttila J, Siira H, Kyngäs H, Lahtinen M, Elo S, Kääriäinen M, et al. Healthcare professionals' competence in digitalisation: a systematic review. J Clin Nurs 2019 Mar;28(5-6):745-761. [doi: 10.1111/jocn.14710] [Medline: 30376199]

29. Virtanen L, Kaihlanen A, Laukka E, Gluschkoff K, Heponiemi T. Behavior change techniques to promote healthcare professionals' eHealth competency: a systematic review of interventions. Int J Med Inform 2021 May;149:104432 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104432] [Medline: 33684712]
30. Ayyagari R, Grover V, M R RAJESH. Technostress: technological antecedents and implications. MIS Q 2011;35(4):831. [doi: 10.2307/41409963]
31. Golz C, Peter K, Hahn S. Cognitive Pretesting and pretest of the STRAIN questionnaire to elaborate work-related stress of health care staff in Switzerland. Int J Health Prof 2018;5(1):109-120. [doi: 10.2478/ijhp-2018-0011]
32. Nübling M, Stößel U, Hasselhorn H, Michaelis M, Hofmann F. Measuring psychological stress and strain at work - Evaluation of the COPSOQ Questionnaire in Germany. Psychosoc Med 2006 Oct 18;3:1-14 [FREE Full text] [Medline: 19742072]
33. Janssen MF, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. Qual Life Res 2013 Sep 25;22(7):1717-1727 [FREE Full text] [doi: 10.1007/s11136-012-0322-4] [Medline: 23184421]
34. Hasselhorn H, Müller B, Tackenberg P, Kümmerling A, Simon M. Berufsausstieg bei Pflegepersonal: Arbeitsbedingungen und Beabsichtigter Berufsausstieg bei Pflegepersonal in Deutschland und Europa. Wirtschaftsverlag NW: Verlag für Neue Wissenschaften; 2005.
35. Von Korff M, Ormel J, Keefe FJ, Dworkin SF. Grading the severity of chronic pain. Pain 1992;50(2):133-149. [doi: 10.1016/0304-3959(92)90154-4]
36. Tuomi K, Ilmarinen J, Jahkola A, Katajarinne L, Tulkki A. Work Ability. 2 Ed. Helsinki: Finnish Institute of Occupational Health; 1998.
37. Nübling M, Lincke H, Wahl-Wachendorf A, Jurkschat R, Panter W. Psychosocial strain at work, work-related state of health and the health behaviour of occupational physicians. ASUI 2014 Jul 04;2014(07):512-521. [doi: 10.17147/asui.2014-07-04-02]
38. RStudio: integrated development for R. RStudio, Inc, Boston, MA. URL: http://www.rstudio.com/ [accessed 2021-10-16]
39. Ripley B, Venables B, Bates D. Package 'MASS'. Cran r. 2013. URL: https://cran.r-project.org/web/packages/MASS/MASS.pdf [accessed 2021-10-16]
40. Böwing-Schmalenbrock M, Jurczok A. Multiple Imputation in der Praxis. Ein Sozialwissenschaftliches Anwendungsbeispiel. Potsdam: Universität Potsdam; 2012.
41. Tabachnick B, Fidell L. Using Multivariate Statistics, Vol 6. London, U.K: Pearson; 2012:1-1032.
42. Venables W, Ripley B. Modern Applied Statistics with S. Fourth Ed. New York: Springer; 2002.
43. Kuek A, Hakkennes S. Healthcare staff digital literacy levels and their attitudes towards information systems. Health Informatics J 2020 Mar 15;26(1):592-612 [FREE Full text] [doi: 10.1177/1460458219839613] [Medline: 30983476]
44. Schlösser T, Dunning D, Johnson KL, Kruger J. How unaware are the unskilled? Empirical tests of the “signal extraction” counterexplanation for the Dunning–Kruger effect in self-evaluation of performance. J Econ Psychol 2013 Dec;39:85-100. [doi: 10.1016/j.joep.2013.07.004]
45. Staggers N, Gassert CA, Curran C. A Delphi study to determine informatics competencies for nurses at four levels of practice. Nurs Res 2002;51(6):383-390. [doi: 10.1097/00006199-200211000-00006] [Medline: 12464758]
46. Bregenzer A, Jimenez P. Risk factors and leadership in a digitalized working world and their effects on employees' stress and resources: web-based questionnaire study. J Med Internet Res 2021 Mar 12;23(3):e24906 [FREE Full text] [doi: 10.2196/24906] [Medline: 33709933]
47. Sasidharan S. Technostress in the workplace: a social network perspective. Inform Technol People 2021 Jun 24:Ahead-of-print. [doi: 10.1108/itp-09-2020-0649]
48. Kleib M, Nagle L. Factors associated with Canadian nurses' informatics competency. Comput Inform Nurs 2018 Aug;36(8):406-415. [doi: 10.1097/CIN.0000000000000434] [Medline: 29596068]
49. La Torre G, De Leonardis V, Chiappetta M. Technostress: how does it affect the productivity and life of an individual? Results of an observational study. Public Health 2020 Dec;189:60-65. [doi: 10.1016/j.puhe.2020.09.013] [Medline: 33166856]
50. Boyer-Davis S. Technostress: an antecedent of job turnover intention in the accounting profession. J Busin Account 2019;12(1):49-63 [FREE Full text]
51. Hämmig O. Explaining burnout and the intention to leave the profession among health professionals - a cross-sectional study in a hospital setting in Switzerland. BMC Health Serv Res 2018 Oct 19;18(1):785. [doi: 10.1186/s12913-018-3556-1] [Medline: 30340485]
52. Thiel R, Deimel L, Schmidtmann D. #SmartHealthSystems. Digitalisierungsstrategien im Internationalen Vergleich. Gütersloh: Bertelsmann Stiftung; 2018.
53. Sabin JE, Harland JC. Professional ethics for digital age psychiatry: boundaries, privacy, and communication. Curr Psychiatry Rep 2017 Sep 20;19(9):55. [doi: 10.1007/s11920-017-0815-5] [Medline: 28726059]
54. Cohen TN, Jain M, Gewertz BL. Personal communication devices among surgeons-exploring the empowerment/enslavement paradox. JAMA Surg 2020 Dec 23;156(4):302. [doi: 10.1001/jamasurg.2020.5627] [Medline: 33355614]
55. Kennzahlen der Schweizer Spitäler. Bundesamt für Gesundheit, Bern. 2019. URL: https://www.bag.admin.ch/bag/de/home/zahlen-und-statistiken/zahlen-fakten-zu-spitaelern/kennzahlen-der-schweizer-spitaeler.html [accessed 2021-10-18]

56. Taylor T, Scott A. Do physicians prefer to complete online or mail surveys? Findings from a national longitudinal survey. Eval Health Prof 2019 Mar 01;42(1):41-70. [doi: 10.1177/0163278718807744] [Medline: 30384770]
57. Nulty DD. The adequacy of response rates to online and paper surveys: what can be done? Assess Eval High Edu 2008 Jun;33(3):301-314. [doi: 10.1080/02602930701293231]
58. Cho YI, Johnson TP, Vangeest JB. Enhancing surveys of health care professionals: a meta-analysis of techniques to improve response. Eval Health Prof 2013 Sep;36(3):382-407. [doi: 10.1177/0163278713496425] [Medline: 23975761]
59. Stedman RC, Connelly NA, Heberlein TA, Decker DJ, Allred SB. The end of the (Research) world as we know it? Understanding and coping with declining response rates to mail surveys. Society Natur Resour 2019 Mar 29;32(10):1139-1154. [doi: 10.1080/08941920.2019.1587127]
60. McCoy S, Marks P, Carr C, Mbarika V. Electronic versus paper surveys: Analysis of potential psychometric biases. In: Proceedings of the 37th Annual Hawaii International Conference on System Sciences. 2004 Presented at: 37th Annual Hawaii International Conference on System Sciences; Jan. 5-8, 2004; Big Island, HI, USA. [doi: 10.1109/hicss.2004.1265634]

Abbreviations

COPSOQ: Copenhagen Psychosocial Questionnaire
KMO: Kaiser-Meyer-Olkin
STRAIN: Work-Related Stress Among Health Professionals in Switzerland

Edited by J Torous; submitted 21.06.21; peer-reviewed by S Pesälä, E Predescu; comments to author 29.08.21; revised version received 13.09.21; accepted 16.09.21; published 04.11.21.

Please cite as:
Golz C, Peter KA, Müller TJ, Mutschler J, Zwakhalen SMG, Hahn S
Technostress and Digital Competence Among Health Professionals in Swiss Psychiatric Hospitals: Cross-sectional Study
JMIR Ment Health 2021;8(11):e31408
URL: https://mental.jmir.org/2021/11/e31408
doi:10.2196/31408
PMID:34734840

Original Paper

An Online Assessment to Evaluate the Role of Cognitive Biases and Emotion Regulation Strategies for Mental Health During the COVID-19 Lockdown of 2020: Structural Equation Modeling Study

Ivan Blanco[1,2], PhD; Teresa Boemo[1], MSc; Alvaro Sanchez-Lopez[1], PhD

[1]Department of Clinical Psychology, School of Psychology, Complutense University of Madrid, Pozuelo de Alarcon, Spain
[2]Cardenal Cisneros University Centre, Alcala de Henares, Spain

Corresponding Author:
Ivan Blanco, PhD
Department of Clinical Psychology
School of Psychology
Complutense University of Madrid
Campus de Somosaguas
Pozuelo de Alarcon, 28223
Spain
Phone: 34 650692547
Email: ivan.blanco.martinez@ucm.es

Abstract

Background: Extant research supports causal roles of cognitive biases in stress regulation under experimental conditions. However, their contribution to psychological adjustment in the face of ecological major stressors has been largely unstudied.

Objective: We developed a novel online method for the ecological examination of attention and interpretation biases during major stress (ie, the COVID-19 lockdown in March/April 2020) and tested their relations with the use of emotion regulation strategies (ie, reappraisal and rumination) to account for individual differences in psychological adjustment to major COVID-19–related stressors (ie, low depression and anxiety, and high well-being and resilience).

Methods: Participants completed an online protocol evaluating the psychological impact of COVID-19–related stressors and the use of emotion regulation strategies in response to them, during the initial weeks of the lockdown of March/April 2020. They also completed a new online cognitive task designed to remotely assess attention and interpretation biases for negative information. The psychometric properties of the online cognitive bias assessments were very good, supporting their feasibility for ecological evaluation.

Results: Structural equation models showed that negative interpretation bias was a direct predictor of worst psychological adjustment (higher depression and anxiety, and lower well-being and resilience; χ^2_9=7.57; root mean square error of approximation=0.000). Further, rumination mediated the influence of interpretation bias in anxiety (*P*=.045; 95% CI 0.03-3.25) and resilience (*P*=.001; 95% CI −6.34 to −1.65), whereas reappraisal acted as a mediator of the influence of both attention (*P*=.047; 95% CI −38.71 to −0.16) and interpretation biases (*P*=.04; 95% CI −5.25 to −0.12) in well-being.

Conclusions: This research highlights the relevance of individual processes of attention and interpretation during periods of adversity and identifies modifiable protective factors that can be targeted through online interventions.

(JMIR Ment Health 2021;8(11):e30961) doi:10.2196/30961

KEYWORDS

COVID-19; emotion regulation; cognitive biases; psychological adjustment; resilience

Introduction

The occurrence of major stressors (eg, serious illnesses, loss of beloved ones, job loss, and economic difficulties) has a deep psychological impact on individuals in terms of both increased depression and anxiety symptoms [1,2], and reduced well-being [3]. Extant empirically supported “diathesis-stress” models [4] highlight how such a psychological impact would be the result of life stressors, particularly in individuals who have pre-existing vulnerabilities. Among those vulnerabilities, cognitive models

have posited the relevant role of individual differences in cognitive processes of attention and interpretation [5,6]. These processes are thought to be on the basis of dysfunctional emotion and stress regulation [7], and are key mechanisms in the onset and maintenance of affective psychopathology in response to stress [5].

Experimental psychopathology research largely supports these assumptions. Stress-related disorders, such as depression and anxiety, have been consistently associated with a marked tendency to process (ie, attend and interpret) emotional information in a negative manner in laboratory studies. For instance, while eye-tracking studies have shown that higher psychological well-being levels are associated with attentional biases toward positive information [8,9], this type of research also shows that depressed individuals are characterized by sustained attention and difficulties disengaging from negative information [8,10], as well as reduced attention toward positive information [11,12]. Furthermore, a biased tendency to interpret ambiguous scenarios in a negative manner has been consistently observed in experimental studies in both depressed [13] and anxious individuals [14].

Conversely, cognitive models posit that attention and interpretation biases would contribute to stress-related psychopathology through their contribution to dysfunctional stress and emotion regulation [15]. This claim has also been experimentally supported. For instance, it has been found that, after negative mood induction, a participant who spent more time attending to positive emotional information (ie, happy faces) recovered faster from induced transient negative moods, whereas sustained attention to negative emotional information predicted impaired stress recovery [16,17]. Importantly, individual differences in the habitual use of emotion regulation strategies are related to the modulation of these forms of affective processing, contributing to maladaptive stress regulation. The habitual use of reappraisal, a strategy typically associated with enhanced stress recovery [18], has been found to modulate attention directed to negative information [19]. Further, the momentary use of reappraisal has been found to predict higher positive interpretation biases to solve ambiguities [20]. In contrast, rumination (ie, passively and repetitively focusing on the symptoms and consequences of distress [21]) hinders the ability to recover from stress [22]. Rumination has been found to interfere with adaptive attention processing, being related to both attention biases toward negative information [23] and negative interpretation biases [24].

In summary, laboratory studies have consistently supported relations between cognitive biases and processes of emotion dysregulation to account for stress-related psychopathology and reduced psychological well-being. Yet, the ecological manifestation of these cognitive biases, as they unfold during the occurrence of real-life major stressors, still remains largely unstudied. This step is crucial to understand how these processes may act as mechanisms of vulnerability and/or resilience to the onset and/or maintenance of psychological impairments in the face of major stressful experiences. This study aimed to provide an initial examination of the interplays among cognitive biases, emotion regulation processes, and outcomes of psychological adaptation to major stress, introducing a novel online method that allows for remote ecological assessment of attention and interpretation biases during daily life functioning. The method was based on a computerized paradigm that allows the online assessment (and intervention) of both attention and interpretation biases during the processing of emotional information [17]. It comprises a modified version of the scrambled sentence task (SST) [25], where participants are asked to create (interpret) self-referent statements using 5 out of 6 presented words (eg, “the future looks very dismal” or “the future looks very bright”) derived from unambiguous items (eg, “looks the future bright very dismal”), where eye tracking–based techniques are used to monitor the time attending to negative and positive information (eg, “dismal” vs “bright”). Using this method to manipulate attention and interpretation biases under experimental conditions, it has been shown that emotional biases in attention and interpretation are causally involved in the spontaneous use of rumination and the ability to use reappraisal in response to laboratory-based negative situations [17]. However, as highlighted above, less is known about the relations between attention and interpretation biases and emotion regulation strategies when people are faced with major stressors. This study integrated an online evaluation of these mechanisms during the occurrence of a global major stressor (the beginning of the COVID-19 pandemic in early 2020) and, specifically, during the restrictive lockdown implemented to face the pandemic at the end of March to the beginning of April of that year.

The COVID-19 pandemic had a dramatic impact on not only public health and socioeconomic status [26] but also citizens’ psychological functioning. It is well-established that pandemic situations are related to increased levels of stress and have a large impact on the prevalence of psychopathologies, such as anxiety and depression, in the general population [27,28]. Until date, the available data with regard to the psychological impact of the COVID-19 pandemic are in this line. Despite the heterogeneity and methodological issues that initial research in the context of urgency had to face [29], extant literature has consistently shown a significant reduction in well-being and an increase in the rates of mental health problems in the general population as a result of the pandemic. For instance, a previous study [30] evaluated a representative sample of 7236 Chinese participants and found a significant increase in the overall prevalence of depressive symptoms (20.1%), anxiety symptoms (35.1%), and poor sleep quality (18.2%) during the beginning of the COVID-19 pandemic. Studies in other geographical areas obtained similar results, reporting increased rates of anxiety and depression in the general population due to the pandemic [28,31]. In Spain, one of the countries more strongly affected by the COVID-19 pandemic during the first half of 2020, studies assessing nationally representative samples found that the rates of clinical depression and anxiety were 22.1% and 19.6%, respectively, in that period [32]. Additionally, these studies found a significant reduction in well-being associated with social (eg, loneliness) and mental health factors (eg, anxiety reactivity) derived as a result of the COVID-19 pandemic. These data highlight the urgent need to understand the underlying factors that have a potential role in reducing the psychological impact of major stressors, such as those derived from the COVID-19 situation.

In the first attempt, some studies assessed self-reported indicators of resilience and their contributions to psychological adjustment during the pandemic. Using equational structural models, it was shown that optimism and positive beliefs about the world might facilitate posttraumatic growth. On the contrary, suspiciousness and intolerance to uncertainty were related to posttraumatic stress symptoms during the pandemic [33]. Further research showed that self-reported positive reappraisal style (ie, the ability to take perspective and to reinterpret situations) was the strongest factor related to the ability to face adversities derived from COVID-19 [34]. Taken together, these findings underlined the relevance of individual differences in cognitive processing (eg, optimism and/or positive beliefs about the world) and adaptive emotion regulation processes (eg, positive reappraisal style) to facilitate psychological adjustment when facing the stress derived from the COVID-19 pandemic. With this study, we aimed to establish whether ecological online assessments of cognitive biases would relate to maladaptive processes of emotion regulation and, ultimately, to psychological adaptation in the face of corona-related major stressors.

To the best of our knowledge, no study has ecologically assessed the relations between these factors and mental health outcomes in the context of major stressors derived from the COVID-19 pandemic. Therefore, the aim of this study was to analyze the daily life role of cognitive biases (i.e., attention and interpretation biases) in emotion regulation and symptom development when facing major stressors. More specifically, the main aim of the study was to test, using structural equation modeling, the predictive role of ecological cognitive biases to emotional information (attention and interpretation biases) and emotion regulation strategies (use of reappraisal and brooding rumination in response to stress during the initial weeks of the pandemic) in psychological maladjustment to major stress (ie, higher depression and anxiety, and lower well-being and resilience). Cognitive biases were monitored through an online test that was completed by participants during the initial weeks of the restrictive lockdown experienced in Spain as a result of the COVID-19 pandemic. In line with previous research supporting the interrelation between cognitive biases and emotion regulation strategies to account for stress regulation, we first hypothesized that cognitive biases (ie, attention and interpretation biases) would have a direct effect on psychological adjustment. Moreover, beside these direct effects, we hypothesized that the use of emotion regulation strategies (ie, rumination or reappraisal) would act as mediators in the pathways between cognitive biases (ie, attention and interpretation biases) and psychological adjustment to stress. We specifically expected that negative cognitive biases would enhance the use of rumination, leading to worse psychological adjustment to stress (ie, higher depression and anxiety, and less well-being and resilience). Conversely, we hypothesized that negative cognitive biases would hinder the use of reappraisal as a strategy to facilitate psychological adjustment (ie, less depression and anxiety, and higher well-being and resilience).

Methods

Participants

A total of 100 participants voluntarily completed an online survey regarding their psychological functioning during the COVID-19 lockdown in Spain, during the period between the end of March and the beginning of April 2020 (3/4 weeks following the beginning of a very restrictive lockdown to prevent the expansion of COVID-19 in this country). Immediately after completing the online survey, all participants were invited to complete an online attention and interpretation experimental task through a custom-built Android smartphone app. Twenty participants were excluded owing to technical issues with the online test (10/100, 10% of the sample) or dropouts (10/100, 10% of the sample). Therefore, the final sample with completed measures of attention and interpretation biases during the lockdown included 80 participants (female: 62/80, 78%), and the mean age was 27.7 years (SD 11.3 years). The study was conducted in accordance with the Declaration of Helsinki 2013, and it was approved by the ethical committee of the Faculty of Psychology at the Complutense University of Madrid (reference 2019/20-028).

General Procedure

Participants were recruited via extensive advertising on social media and social networks. First, all participants completed an online survey administered via Qualtrics Software [35] (see Multimedia Appendix 1 for the Checklist for Reporting Results of Internet E-Surveys [CHERRIES]). This survey comprised an informed consent form, and a series of sociodemographic and self-reported psychological measures (see below). Immediately afterwards, participants were invited (via email) to install and complete on their phones an adaptation of the SST [25] designed for online remote assessment of attention and interpretation biases during daily life functioning. This was done through a novel smartphone app, adapting the computerized version of the SST for online assessment [17].

Materials

Self-reported Measures

Depressive symptoms were assessed through the Center for Epidemiological Studies on Depression-8 scale [36]. Participants reported how often they had experienced depression-related symptomatology during the last week on a 4-point Likert scale (ranging from 0 [none or almost none of the time] to 3 [all or almost all of the time]). Higher values represent the presence of depression symptoms, whereas lower values represent the absence of depressive symptomatology. The reliability in our study was good (α=.84).

Participants' anxiety symptoms were assessed through the Generalized Anxiety Disorder-7 scale [37]. It has 7 items and uses a 4-point Likert scale (from 0 [not at all sure] to 3 [nearly every day]), where general anxiety-related symptoms (irritability, worry, etc) are assessed with reference to the last 2 weeks. Higher values represent the presence of anxiety symptoms, whereas lower values represent the absence of anxious symptomatology. In this study, the internal consistency was good (α=.85).

Participants' psychological well-being was assessed using the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) [38]. It has 14 items and uses a 5-point Likert scale (from 1 [none of the time] to 5 [all of the time]) to measure a broad range of factors of psychological well-being, including emotional aspects, cognitive dimensions, interpersonal relationships, and positive functioning. Higher values represent higher levels of well-being. In this study, the internal consistency of the scale was very good (α=.92).

Finally, participants' resilience was measured using the Brief Resilience Scale [39]. This scale has 6 items and uses a Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). It conceptualizes resilience as the ability to bounce back from adversity or stress. In this study, questions were framed in relation to specific abilities to deal with the experience of COVID-19 stressors and were framed with reference to the last week (ie, during the lockdown period). Higher values represent a better ability to deal with the situation. In the present sample, its internal consistency was good (α=.82).

The use of emotion regulation strategies during the lockdown was evaluated. The use of rumination as an emotion regulation strategy since the beginning of the COVID lockdown was assessed through the brooding rumination subscale from the Ruminative Response Scale [40]. It comprises 5 items and uses a Likert scale (from 1 [almost never] to 4 [almost always]). Furthermore, the use of reappraisal as an emotion regulation strategy since the beginning of the COVID lockdown was evaluated through the reappraisal subscale from the Emotion Regulation Questionnaire [41]. This scale comprises 4 items and uses a Likert scale (from 1 [totally agree] to 5 [totally disagree]). Higher values represent a marked tendency to use rumination or reappraisal. Both scales showed adequate internal consistency in our study (α=.76 and α=.79, respectively).

To capture the influence of the lockdown situation on participants' emotion regulation and psychological adjustment, all measures were framed with reference to the 2 weeks before the assessment.

Online Attention and Interpretation Bias Task

Attention and interpretation biases were assessed using an online variant of the SST [25], adapted from the computerized procedure for online attention and interpretation bias assessment that has been previously validated [42]. A total of 15 scrambled sentences with 6 words (eg, "looks the future bright very dismal") were presented to the participants. The number of trials was established based on previous extensive piloting of sufficient required SST trials to obtain reliable cognitive bias indices related to stress vulnerability and depression status (Martín-Romero, unpublished data, July 2021). Participants were instructed to mentally unscramble the sentences, as fast as possible, using only 5 out of the 6 words, to create a grammatically correct and meaningful sentence. These sentences could only be unscrambled with a negative or a positive meaning (eg, "the future looks very dismal" or "the future looks very bright"). Participants were instructed to unscramble the words into the valid sentence that first came to their mind. To control for the influence of word positioning, emotional words (ie, positive or negative) were always displayed in the second and fifth positions. Additionally, these positions were counterbalanced, with positive and negative words similarly allocated in the second and fifth positions across trials.

The task was completed on participants' smartphones. Each trial started with a fixation cross in the left position of the screen to promote natural left-to-right reading patterns. Participants were asked to press the cross with their finger to start the trial. Immediately after, a reading phase started, where participants had to read and mentally unscramble the words in a limited time of 14 seconds. Using a moving window procedure, the 6 words were hidden in individual boxes. In order to read them, participants had to move their finger throughout a scroll bar below the boxes to unhide the corresponding word. Once participants moved their finger from one word to another, the previous words were hidden again. During this reading phase, the position of the finger on the screen was monitored, allowing to compute the time spent (in milliseconds) reading (attending to) each word of the scrambled sentence, and thus, the proportion of total time reading negative over positive words could be assessed (ie, negative attention bias). After the time limit, or when participants decided (pressing a "Ready" button), the final response phase began (Figure 1).

In the response phase, all words were unhidden. With a time limit of 7 seconds, participants had to create a meaningful sentence by pressing, as fast as possible and in the appropriate order, the corresponding chosen series of 5 words. If participants made any mistake during the construction of the sentence, they could modify it by unselecting the wrong word and selecting a new one (Figure 2). Once the 5 words were selected, participants pressed the "Ready" button at the bottom of the screen and started a new trial. The system recorded responses for each trial to compute the interpretation bias index (see below).

Figure 1. Example of the reading phase.

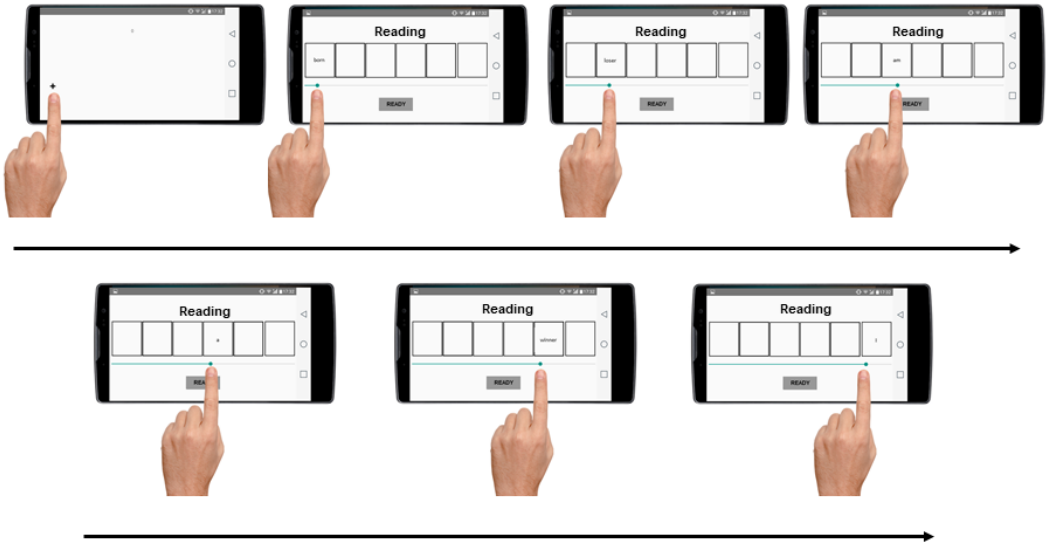
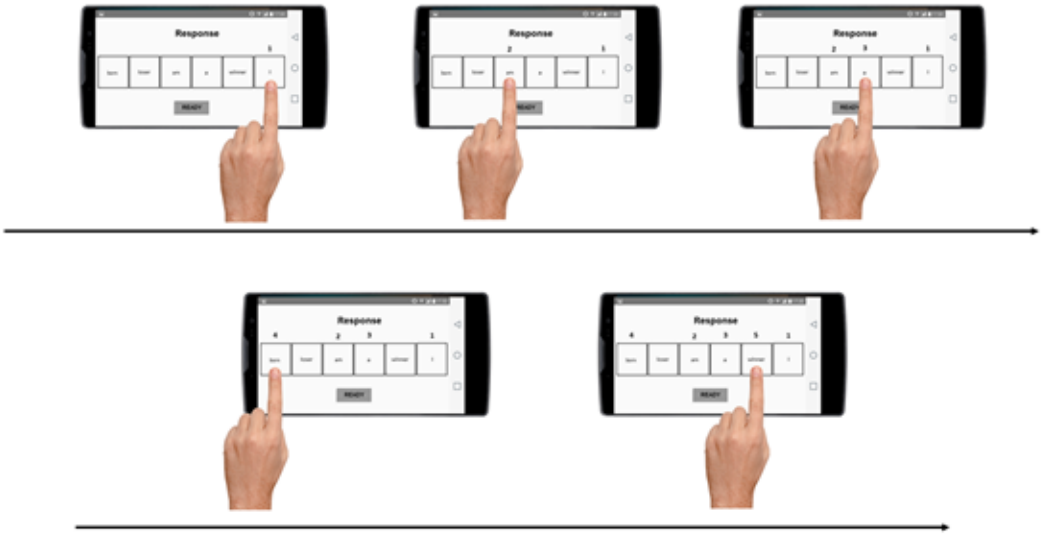


Figure 2. Example of the response phase.



Attention and Interpretation Bias Indices

The task was designed to allow for online assessment of the total time attending to negative over positive words during the reading phase, as well as the proportion of negative over positive interpretations made during the response phase. The program registered the total time (in milliseconds) that participants spent reading (attending to) negative and positive words. We analyzed the reliability of each measure. Reliability analysis showed very good reliability for both measures of total time attending to positive (α=.87) and total time attending to negative (α=.90) stimuli. Following previous studies [42], an attention bias index was computed by dividing the total time attending to negative words by the total time attending to both emotional (ie, positive and negative) words. Values above 0.5 are indicative of an attention bias toward negative information, whereas values below 0.5 are indicative of an attention bias toward positive information. The program also computed the number of positive and negative grammatically correct sentences that were unscrambled by each participant during the response phase. An interpretation bias index was computed by dividing the number of negative sentences by the total number of unscrambled sentences (ie, positive and negative). Split-half reliability analysis showed good reliability for this index (r=0.75; ρ=0.86). As with the attention bias index, values above 0.5 indicate a negative interpretation bias, whereas values below 0.5 indicate a positive interpretation bias.

Data Analysis Plan

Once we established the good psychometric properties of the cognitive bias measures, in terms of their reliability for ecological online attention and interpretation bias indexing, we conducted the main analyses in the study.

Demographics, COVID-19–Related Variables, and Psychological Measures

We conducted descriptive analyses of demographics and psychological measures of participants, including gender, age, civil status, and education level, as well as computed the mean

(SD) levels of self-report measures and online cognitive bias assessments.

Relations Between Cognitive Biases and Emotion Regulation Strategies With Psychological Adjustment Indices

We conducted a series of Pearson bivariate correlations to analyze the relations between the attention and interpretation biases and emotion regulation strategy measures with psychological adjustment indices.

Structural Equation Models

We tested an equation structure model including those variables that were significantly correlated. Thus, we tested a model where attention and interpretation biases act as exogenous variables, all of which predicted psychological adjustment (ie, depression, anxiety, well-being, and resilience) directly and also indirectly through the use of emotion regulation strategies (ie, use of rumination or reappraisal), which would act as mediators. Moreover, we tested the reverse model where psychological adjustment variables were introduced as predictors, emotion regulation strategies as mediators, and cognitive bias indexes as outcome variables. The estimation of the standardized parameters of the model followed the full information maximum likelihood (FIML) estimation method. To test the adjustment of our model, we used the following standard criteria [43]: (1) χ^2, a nonsignificant value indicates a perfect fit; (2) χ^2/df, a value lower than 2 indicates a good fit; (3) comparative fit index and Tucker-Lewis index, a value ≥0.95 indicates a good fit; (4) root mean square error of approximation, a value ≤0.05 indicates a good fit; (5) standardized root mean square, a smaller value indicates a better fit between the observed data and the tested model; and (6) Akaike information criterion, a lower value indicates the preference for selecting a model when compared to another model. Moreover, we used the Mardia coefficient for assessing multivariate normality (a value ≤5 indicates the possibility to assume multivariate normality) [44]. Finally, the hypothesized mediation pathways within the model (ie, cognitive bias → emotion regulation strategy → psychological adjustment outcome) were tested via the estimation of indirect effects within the final model. All the structural equation models were tested using AMOS v18.0 (SPSS Inc). A *P* value <.05 was used to determine statistical significance in all analyses.

Results

Demographics, COVID-19–Related Variables, and Psychological Measures

Descriptive data of demographics and psychological measures of the participants in this study are shown in Table 1.

Table 1. Descriptive data of demographics and psychological measures.

Variable	Value (N=80)
Gender: female, n (%)	62 (78)
Age (years), mean (SD)	27.7 (11.3)
Civil status, n (%)	
Single	34 (43)
Married	34 (43)
In a relationship	7 (9)
Divorced/widower	5 (6)
Educational level, n (%)	
Without studies	0 (0)
Primary school	0 (0)
High school	43 (54)
University graduate	37 (46)
Negative interpretation bias, mean (SD)	0.28 (0.23)
Negative attention bias, mean (SD)	0.51 (0.03)
Rumination level, mean (SD)	11.54 (3.63)
Reappraisal level, mean (SD)	12.45 (3.37)
Depression level, mean (SD)	7.06 (2.81)
Anxiety level, mean (SD)	2.60 (3.04)
Well-being level, mean (SD)	48.25 (8.08)
Resilience level, mean (SD)	18.44 (4.65)

Relations Between Cognitive Biases and Emotion Regulation Strategies With Psychological Adjustment Indices

Depression and anxiety were significantly positively related to the use of rumination (r=0.398 and r=0.450, respectively) and negative interpretation biases (r=0.619 and r=0.488, respectively) during the lockdown. Resilience and well-being were also significantly but negatively related to the use of rumination (r=−0.575 and r=−0.502, respectively) and negative interpretation biases (r=−0.536 and r=−0.574, respectively) during the lockdown, and significantly positively related to the use of reappraisal during the lockdown (r=0.374 and r=0.330, respectively). Moreover, all these psychological adjustment variables (ie, levels of depression, anxiety, resilience, and well-being) were significantly related among each other (all P<.001).

With regard to cognitive biases and the use of emotion regulation strategies during the lockdown, higher use of rumination was significantly associated with lower use of reappraisal (r=−0.293) and with higher levels of negative interpretation biases (r=0.543). In the case of the use of reappraisal, it was negatively related to both negative attention and interpretation biases (r=−0.224 and r=−0.275, respectively) (see Multimedia Appendix 2 for all correlation results).

Structural Equation Models

The Mardia coefficient yielded a value of 2.15, which is far below the critical value (±5), assuming multivariate normality in our data [44]. Based on the previous bivariate correlation analysis and following the predictions from current cognitive models [15,45], we tested an equation model where psychological adjustment variables (ie, depression, anxiety, resilience, and well-being) were predicted by cognitive biases directly and/or indirectly through the use of emotion regulation strategies. All the goodness-of-fit indices are shown in Table 2.

Table 2. Goodness-of-fit indices for the tested models.

Model	Chi-square (*df*)	*P* value	χ^2/df	CFI[a]	TLI[b]	RMSEA[c] (90% CI)	SRMR[d]	AIC[e]
Model 1[f]	62.4 (15)	<.001	4.16	0.79	0.61	0.20 (0.14-0.25)	0.1061	120.1
Model 2[g]	137.2 (15)	<.001	9.14	0.46	−0.01	0.321 (0.27-0.37)	0.2878	195.1
Model 1R[h]	7.6 (9)	.58	0.84	1	1.02	0.000 (0.00-0.11)	0.0554	77.56

[a]CFI: comparative fit index.

[b]TLI: Tucker-Lewis index.

[c]RMSEA: root mean square error of approximation.

[d]SRMR: standardized root mean square.

[e]AIC: Akaike information criterion.

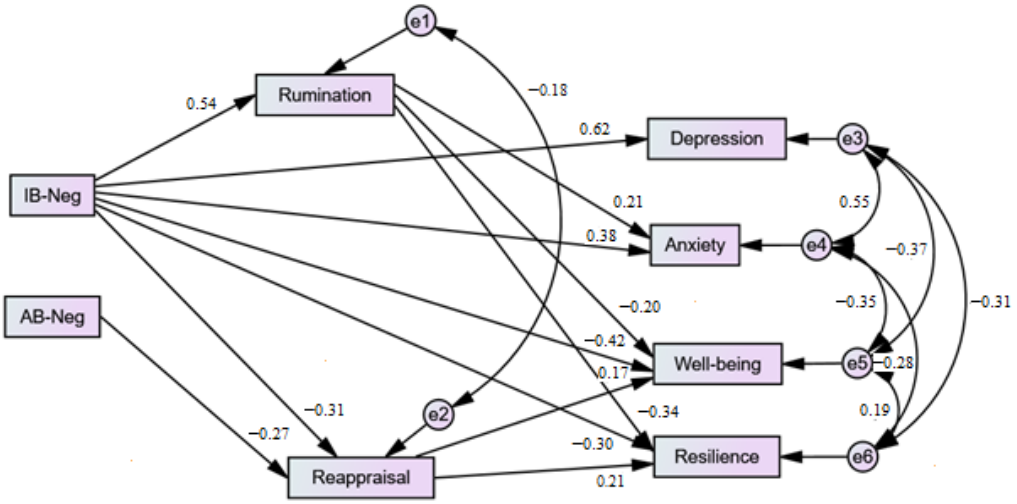
[f]Model 1: initial model.

[g]Model 2: alternative model.

[h]Model 1R: initial model respecified.

As shown in Table 2, the goodness-of-fit indices were better for our hypothesized model (Model 1) than for the alternative reverse model (Model 2). However, since the fit of our initial model (Model 1) was poor, respecification was carried out following Wald and Lagrange multiplier tests [46]. All paths with nonsignificant P values were removed consecutively. Only the path *rumination* to *depression* was removed. No additional paths were included in the model (Figure 3). The final respecified model (Model 1R) showed very good fit in all of the indices (Table 2).

Figure 3. The respecified model (Model 1R) with standardized regression weights. AB-Neg: negative attention bias; IB-Neg: negative interpretation bias.



Finally, indirect effects were tested using a bias-corrected bootstrap estimation (2000 bootstrap samples with 95% CI). As shown in Table 3, significant indirect effects were found between negative interpretation biases and anxiety and resilience via rumination (*P*=.045 and *P*=.001, respectively). Additionally, via the reappraisal path, indirect effects between negative cognitive biases (ie, both attention and interpretation biases) and well-being were statistically significant (*P*=.047 and *P*=.04, respectively).

Table 3. Bootstrap mediational analysis.

Variable	Indirect effects (95% CI)		SE	*P* value
	Lower	Upper		
Indirect effect via rumination				
Interpretation bias → anxiety	0.032	3.246	0.819	.045
Interpretation bias → well-being	−8.736	0.492	2.344	.08
Interpretation bias → resilience	−6.341	−1.647	1.192	.001
Indirect effect via reappraisal				
Interpretation bias → well-being	−5.251	−0.117	1.226	.04
Interpretation bias → resilience	−3.792	0.079	0.981	.08
Attention bias → well-being	−38.714	−0.159	8.904	.047
Attention bias → resilience	−24.647	0.411	6.241	.07

Discussion

The main aim of this study was to assess the predictive role of cognitive biases and emotion regulation strategies on different indices of psychological adjustment to a major stressor, the COVID-19 lockdown (namely, lower depression and anxiety, and higher well-being and resilience in the face of experienced corona stress). Using structural equation modeling, we analyzed how ecological online assessments of cognitive biases (ie, attention and interpretation biases remotely measured through a novel app-based system integrating the SST) were directly related to the outcomes of psychological adjustment to corona stress and/or indirectly related to them through the use of emotion regulation strategies during the lockdown (ie, rumination and reappraisal in response to experienced negative events). Our results highlight the central role of negative interpretation bias as a vulnerability factor during the lockdown period, accounting for significant variance in all indicators of psychological adjustment to major stress (namely, depression and anxiety, and well-being and resilience in the face of experienced corona stress), above and beyond attention bias and the use of emotion regulation strategies (ie, rumination and reappraisal). We also found mediation effects of the use of strategies (rumination and reappraisal) between negative cognitive biases and psychological adjustment outcomes.

These results show the direct effect of cognitive biases on psychological adjustment to the COVID-19 lockdown, and support our initial hypotheses. It is worth noting that the sentences used in our online SST paradigm (remotely measured through a novel app-based system) were related to different central cognitive schemas for psychopathology and well-being (such as self-concept, world, and future beliefs). It seems that the interpretation of those sentences in a negative manner (eg, “the future looks very dismal”), in contrast to the interpretation

in a positive manner (eg, "the future looks very bright"), emerged as an important risk factor that enhanced the impact of the major stressful situation on psychological functions (ie, increasing depression and anxiety levels) and reduced positive functioning variables, such as psychological well-being and resilient responses to corona stress. This result is consistent with previous research showing that individual differences in positive beliefs about the world were one of the major predictors of posttraumatic growth in the face of corona stress [32]. In contrast, attention biases to negative versus positive information did not have any direct effect on psychological adjustment. These findings suggest that the role of biased attention as a direct correlate of psychological functioning (ie, depression and anxiety levels or well-being and resilience) may be limited. Multiple studies support the idea that attention biases might exert indirect influences in psychological functioning through their influence on elaborative processes such as interpretation bias [47-49]. However, we did not find any statistical relation between attention and interpretation biases. Therefore, it is plausible that, despite the high reliability of the attention bias index, the task used for assessment was not able to fully capture the actual attentional processes in the present sample. Moreover, it might be plausible that the negative interpretation bias index introduced in our model accounted for all the variance in psychological outcomes explained by the negative attention bias index. In fact, previous research has also shown that attention bias indices did not demonstrate significant relevance to directly account for psychological symptoms when other related elaborative cognitive processes, such as memory biases, were modeled together [50]. Thus, the results found regarding attention biases in this study should be considered cautiously.

As previously mentioned, we also analyzed the mediational role that the use of emotion regulation strategies during the COVID-19 lockdown played in the interplay between negative cognitive biases and consequent psychological adjustment outcomes. Analysis showed that the use of rumination emerged as a significant mediator between interpretation bias and anxiety and resilience. Regarding the use of reappraisal, our results showed that while reappraisal partially mediated the relation between negative interpretation bias and well-being, it totally mediated the association of negative attention bias with well-being. These findings are in line with current theories with regard to the major role that cognitive processes play on emotion regulation [15,45]. Our findings indicate the relevance of interpretation biases as a particularly central mechanism to hinder or buffer the impact of adverse situations, such as those derived from the COVID-19 emergency and the resulting lockdown period during March/April 2020. This is in line with former empirical evidence in the context of COVID-19. For instance, a positive appraisal style was found to be the major contributor for resilience during the COVID-19 pandemic [34]. However, our data go beyond previous studies using self-reported measures of these processes and suggest that direct ecological assessments of negative interpretation biases, as they manifest during daily functioning, might reduce the ability to use positive reappraisal, hindering psychological adjustment.

Taken together, our results support the idea that individual differences in the way reality is perceived and interpreted (ie, the construction of self-relevant meanings from ongoing experiences) may be central to increase (or reduce) the psychological impact of ongoing adversities (such as the one experienced during the COVID-19 pandemic). In times of major stress and uncertainty, as the period under study, difficulties in accessing standard in-person resources of psychological assistance may emerge. Conversely, applied work to intervene in these biases could be efficiently integrated into remote online interventions. This includes novel online protocols to directly train positive interpretation biases, with consistent results for changes in emotion regulation and clinical outcomes [51], as well as cognitive tools designed to actively train attention operations involved in interpretation bias change and adaptive emotion regulation [42]. Therefore, future research is warranted to adapt these promising tools for easy access online implementations that can facilitate stress regulation in daily life and positive psychological functioning during the occurrence of major adversities.

It is worth noting the strengths and limitations of this study. As for the strengths of the study, to our knowledge, this is the first study that has ecologically assessed cognitive biases of affective processing during the occurrence of a major stressor, such as the COVID-19 lockdown of early 2020. Furthermore, the adaptation of a previously validated paradigm to remotely assess attention and interpretation biases [42] increases the ecological validity of the present results in terms of the indices of attention and interpretation bias performance. Furthermore, the reliabilities of these cognitive bias measures were very good, supporting their feasibility for use in online remote assessments during the occurrence of major stressors. Moreover, the study was conducted in Spain, which was one of the countries more dramatically hit by the COVID-19 situation at the time of the study, with data being collected during a very restrictive lockdown. Given all these conditions, we were able to test purported mechanisms of psychological (mal)adjustment to major stress with considerable ecological validity.

With regard to limitations, our sample was relatively small. Yet, our current findings were consistent across different forms of psychological adjustment to major stress. This supports the relevance of these findings and informs about the potential of further investigating these models in more representative samples to fully determine the role of cognitive affective processes in buffering the impact of major stressors. Furthermore, the mobile app developed to remotely assess cognitive biases could only be adapted to work on Android smartphones, limiting the number of screened participants that could be included in the study, and thus, partially restricting the representativeness of the sample. Future research is warranted to adapt this new tool for other operating systems, which will allow access to bigger samples and to replicate these initial findings in representative samples under different related conditions of major stress and adversity. Furthermore, cognitive theories have pointed out that attention and interpretation biases interplay with other cognitive biases such as memory biases [47]. However, in this study, only attention and interpretation were assessed. Future studies should also consider developing ecological online assessments of memory biases to analyze their specific roles in accounting for emotion regulation and

psychological functioning when facing major stressors in daily life.

In summary, our study presents a novel approach that allows the analysis of the interplay of cognitive biases, emotion regulation strategies, and psychological adjustment when facing major stressors, which are assessed in naturalistic settings.

Acknowledgments

This work was supported by a grant from the Program of Attraction of Scientific Talent of Community of Madrid (2017-T1/SOC-5359) and a grant from the Spanish Ministry of Science (PGC2018-095723-A-I00) awarded to ASL, and Complutense University of Madrid Predoctoral Fellowships CT63/19-CT64/19 awarded to TB. The funding source had no role in the design of this study and will not have any role during its execution, the analyses, the interpretation of the data, or the decision to submit the results.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Checklist for Reporting Results of Internet E-Surveys (CHERRIES).
[DOCX File , 18 KB - mental_v8i11e30961_app1.docx]

Multimedia Appendix 2
Bivariate correlations among cognitive biases, use of emotion regulation strategies, and psychological variables (depression, anxiety, resilience, and well-being).
[DOCX File , 15 KB - mental_v8i11e30961_app2.docx]

References

1. Liu RT, Alloy LB. Stress generation in depression: A systematic review of the empirical literature and recommendations for future study. Clin Psychol Rev 2010 Jul;30(5):582-593 [FREE Full text] [doi: 10.1016/j.cpr.2010.04.010] [Medline: 20478648]
2. Uliaszek AA, Zinbarg RE, Mineka S, Craske MG, Griffith JW, Sutton JM, et al. A longitudinal examination of stress generation in depressive and anxiety disorders. J Abnorm Psychol 2012 Feb;121(1):4-15. [doi: 10.1037/a0025835] [Medline: 22004114]
3. Marum G, Clench-Aas J, Nes RB, Raanaas RK. The relationship between negative life events, psychological distress and life satisfaction: a population-based study. Qual Life Res 2014 Mar 13;23(2):601-611. [doi: 10.1007/s11136-013-0512-8] [Medline: 24026629]
4. Ingram RE, Luxton DD. Chapter 2: Vulnerability-Stress Models. In: Hankin BL, Abela JRZ, editors. Development of Psychopathology: A Vulnerability-Stress Perspective. Thousand Oaks, CA: SAGE Publications Inc; 2005:32-46.
5. Beck AT, Bredemeier K. A unified model of depression. Clinical Psychological Science 2016 Jun 21;4(4):596-619. [doi: 10.1177/2167702616628523]
6. Armstrong T, Olatunji BO. Eye tracking of attention in the affective disorders: a meta-analytic review and synthesis. Clin Psychol Rev 2012 Dec;32(8):704-723 [FREE Full text] [doi: 10.1016/j.cpr.2012.09.004] [Medline: 23059623]
7. Joormann J, Vanderlind WM. Emotion regulation in depression. Clinical Psychological Science 2014 Jun 30;2(4):402-421. [doi: 10.1177/2167702614536163]
8. Johnson SL, LeMoult J, Vanderlind WM, Joormann J. Mood Disorders: Biological Bases. In: Blaney PH, Krueger RF, Millon T, editors. Oxford Textbook of Psychopathology. Oxford, United Kingdom: Oxford University Press; 2015:222-256.
9. Sanchez A, Vazquez C. Looking at the eyes of happiness: Positive emotions mediate the influence of life satisfaction on attention to happy faces. The Journal of Positive Psychology 2014 Apr 22;9(5):435-448. [doi: 10.1080/17439760.2014.910827]
10. Blanco I, Vazquez C. Integrative well-being leads our attentional system: an eye-tracking study. J Happiness Stud 2020 Apr 06;22(2):787-801. [doi: 10.1007/s10902-020-00251-7]
11. Duque A, Vázquez C. Double attention bias for positive and negative emotional faces in clinical depression: evidence from an eye-tracking study. J Behav Ther Exp Psychiatry 2015 Mar;46:107-114. [doi: 10.1016/j.jbtep.2014.09.005] [Medline: 25305417]
12. Blanco I, Poyato N, Nieto I, Boemo T, Pascual T, Roca P, et al. Attentional biases in dysphoria when happy and sad faces are simultaneously presented. J Behav Ther Exp Psychiatry 2019 Dec;65:101499. [doi: 10.1016/j.jbtep.2019.101499] [Medline: 31352298]

13. Peckham AD, McHugh RK, Otto MW. A meta-analysis of the magnitude of biased attention in depression. Depress Anxiety 2010 Dec;27(12):1135-1142. [doi: 10.1002/da.20755] [Medline: 21049527]
14. Everaert J, Podina IR, Koster EH. A comprehensive meta-analysis of interpretation biases in depression. Clin Psychol Rev 2017 Dec;58:33-48. [doi: 10.1016/j.cpr.2017.09.005] [Medline: 28974339]
15. Mathews A, Mackintosh B, Fulcher EP. Cognitive biases in anxiety and attention to threat. Trends in Cognitive Sciences 1997 Dec;1(9):340-345. [doi: 10.1016/s1364-6613(97)01092-9]
16. Sanchez A, Vazquez C, Gomez D, Joormann J. Gaze-fixation to happy faces predicts mood repair after a negative mood induction. Emotion 2014 Feb;14(1):85-94. [doi: 10.1037/a0034500] [Medline: 24188064]
17. Sanchez A, Vazquez C, Marker C, LeMoult J, Joormann J. Attentional disengagement predicts stress recovery in depression: an eye-tracking study. J Abnorm Psychol 2013 May;122(2):303-313. [doi: 10.1037/a0031529] [Medline: 23421524]
18. Jamieson JP, Mendes WB, Nock MK. Improving acute stress responses. Curr Dir Psychol Sci 2013 Feb 01;22(1):51-56. [doi: 10.1177/0963721412461500]
19. Kim SA, Kim H, Kim SH. Reappraisal modulates attentional bias to angry faces. Front Psychol 2016;7:1841 [FREE Full text] [doi: 10.3389/fpsyg.2016.01841] [Medline: 27920749]
20. Gordon NS, Chesney SA, Reiter K. Thinking positively: Optimism and emotion regulation predict interpretation of ambiguous information. Cogent Psychology 2016 Jun 20;3(1):1195068. [doi: 10.1080/23311908.2016.1195068]
21. Treynor W, Gonzalez R, Nolen-Hoeksema S. Rumination reconsidered: A psychometric analysis. Cognitive Therapy and Research 2003;27(3):247-259. [doi: 10.1023/A:1023910315561]
22. Donaldson C, Lam D, Mathews A. Rumination and attention in major depression. Behav Res Ther 2007 Nov;45(11):2664-2678. [doi: 10.1016/j.brat.2007.07.002] [Medline: 17692819]
23. LeMoult J, Arditte KA, D'Avanzato C, Joormann J. State rumination: associations with emotional stress reactivity and attention biases. J Exp Psychopathol 2013;4(5):471-484 [FREE Full text] [doi: 10.5127/jep.029112] [Medline: 25431652]
24. Mor N, Hertel P, Ngo TA, Shachar T, Redak S. Interpretation bias characterizes trait rumination. J Behav Ther Exp Psychiatry 2014 Mar;45(1):67-73. [doi: 10.1016/j.jbtep.2013.08.002] [Medline: 24001990]
25. Wenzlaff RM, Bates DE. Unmasking a cognitive vulnerability to depression:How lapses in mental control reveal depressive thinking. Journal of Personality and Social Psychology 1998;75(6):1559-1571. [doi: 10.1037/0022-3514.75.6.1559]
26. Nicola M, Alsafi Z, Sohrabi C, Kerwan A, Al-Jabir A, Iosifidis C, et al. The socio-economic implications of the coronavirus pandemic (COVID-19): A review. Int J Surg 2020 Jun;78:185-193 [FREE Full text] [doi: 10.1016/j.ijsu.2020.04.018] [Medline: 32305533]
27. Reissman D, Watson P, Klomp R, Tanielian T, Prior S. Pandemic influenza preparedness: adaptive responses to an evolving challenge. Journal of Homeland Security and Emergency Management 2006;3(2):1-24. [doi: 10.2202/1547-7355.1233]
28. Lee SA, Jobe MC, Mathis AA. Mental health characteristics associated with dysfunctional coronavirus anxiety. Psychol. Med 2020 Apr 16;51(8):1403-1404. [doi: 10.1017/s003329172000121x]
29. Nieto I, Navas JF, Vázquez C. The quality of research on mental health related to the COVID-19 pandemic: A note of caution after a systematic review. Brain Behav Immun Health 2020 Aug;7:100123 [FREE Full text] [doi: 10.1016/j.bbih.2020.100123] [Medline: 32835299]
30. Huang Y, Zhao N. Mental health burden for the public affected by the COVID-19 outbreak in China: Who will be the high-risk group? Psychol Health Med 2021 Jan;26(1):23-34. [doi: 10.1080/13548506.2020.1754438] [Medline: 32286091]
31. Shevlin M, McBride O, Murphy J, Miller JG, Hartman TK, Levita L, et al. Anxiety, depression, traumatic stress and COVID-19-related anxiety in the UK general population during the COVID-19 pandemic. BJPsych Open 2020 Oct 19;6(6):e125 [FREE Full text] [doi: 10.1192/bjo.2020.109] [Medline: 33070797]
32. Valiente C, Contreras A, Peinado V, Trucharte A, Martínez AP, Vázquez C. Psychological adjustment in Spain during the COVID-19 pandemic: positive and negative mental health outcomes in the general population. Span. J. Psychol 2021 Feb 08;24:7. [doi: 10.1017/sjp.2021.7]
33. Vazquez C, Valiente C, García FE, Contreras A, Peinado V, Trucharte A, et al. Post-traumatic growth and stress-related responses during the COVID-19 pandemic in a national representative sample: the role of positive core beliefs about the world and others. J Happiness Stud 2021 Jan 11:1-21 [FREE Full text] [doi: 10.1007/s10902-020-00352-3] [Medline: 33456320]
34. Veer IM, Riepenhausen A, Zerban M, Wackerhagen C, Puhlmann LMC, Engen H, et al. Psycho-social factors associated with mental resilience in the Corona lockdown. Transl Psychiatry 2021 Jan 21;11(1):67 [FREE Full text] [doi: 10.1038/s41398-020-01150-4] [Medline: 33479211]
35. Qualtrics. URL: https://www.qualtrics.com/ [accessed 2021-10-05]
36. Turvey CL, Wallace RB, Herzog R. A revised CES-D measure of depressive symptoms and a DSM-based measure of major depressive episodes in the elderly. Int Psychogeriatr 1999 Jun;11(2):139-148. [doi: 10.1017/s1041610299005694] [Medline: 11475428]
37. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]

38. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007 Nov 27;5:63 [FREE Full text] [doi: 10.1186/1477-7525-5-63] [Medline: 18042300]
39. Smith BW, Dalen J, Wiggins K, Tooley E, Christopher P, Bernard J. The brief resilience scale: assessing the ability to bounce back. Int J Behav Med 2008;15(3):194-200. [doi: 10.1080/10705500802222972] [Medline: 18696313]
40. Nolen-Hoeksema S, Morrow J. A prospective study of depression and posttraumatic stress symptoms after a natural disaster: The 1989 Loma Prieta earthquake. Journal of Personality and Social Psychology 1991;61(1):115-121. [doi: 10.1037/0022-3514.61.1.115]
41. Gross JJ, John OP. Individual differences in two emotion regulation processes: implications for affect, relationships, and well-being. J Pers Soc Psychol 2003 Aug;85(2):348-362. [doi: 10.1037/0022-3514.85.2.348] [Medline: 12916575]
42. Sanchez-Lopez A, De Raedt R, van Put J, Koster EH. A novel process-based approach to improve resilience: Effects of computerized mouse-based (gaze)contingent attention training (MCAT) on reappraisal and rumination. Behav Res Ther 2019 Jul;118:110-120. [doi: 10.1016/j.brat.2019.04.005] [Medline: 31048096]
43. Weston R, Gore PA. A brief guide to structural equation modeling. The Counseling Psychologist 2016 Jun 30;34(5):719-751. [doi: 10.1177/0011000006286345]
44. Bentler PM. EQS 6 Structural Equations Program Manual. Encino, CA: Multivariate Software, Inc; 2006.
45. Gross JJ. Emotion regulation: Conceptual and empirical foundations. In: Handbook of emotion regulation. New York, NY: The Guilford Press; 2014:3-20.
46. Kline RB. Principles and Practice of Structural Equation Modeling. New York, NY: The Guilford Press; 2015.
47. Everaert J, Duyck W, Koster EHW. Attention, interpretation, and memory biases in subclinical depression: a proof-of-principle test of the combined cognitive biases hypothesis. Emotion 2014 Apr;14(2):331-340. [doi: 10.1037/a0035250] [Medline: 24512247]
48. Everaert J, Grahek I, Duyck W, Buelens J, Van den Bergh N, Koster EHW. Mapping the interplay among cognitive biases, emotion regulation, and depressive symptoms. Cogn Emot 2017 Jun;31(4):726-735. [doi: 10.1080/02699931.2016.1144561] [Medline: 26878897]
49. Sanchez A, Everaert J, De Putter LM, Mueller SC, Koster EH. Life is … great! Emotional attention during instructed and uninstructed ambiguity resolution in relation to depressive symptoms. Biol Psychol 2015 Jul;109:67-72. [doi: 10.1016/j.biopsycho.2015.04.007] [Medline: 25939559]
50. Marchetti I, Everaert J, Dainer-Best J, Loeys T, Beevers CG, Koster EH. Specificity and overlap of attention and memory biases in depression. J Affect Disord 2018 Jan 01;225:404-412. [doi: 10.1016/j.jad.2017.08.037] [Medline: 28850855]
51. Hirsch CR, Krahé C, Whyte J, Bridge L, Loizou S, Norton S, et al. Effects of modifying interpretation bias on transdiagnostic repetitive negative thinking. J Consult Clin Psychol 2020 Mar;88(3):226-239. [doi: 10.1037/ccp0000455] [Medline: 32068424]

Abbreviations

SST: Scrambled Sentence Task

Edited by J Torous; submitted 04.06.21; peer-reviewed by J Vrijsen, K Kishore; comments to author 01.07.21; revised version received 22.07.21; accepted 01.09.21; published 02.11.21.

Please cite as:
Blanco I, Boemo T, Sanchez-Lopez A
An Online Assessment to Evaluate the Role of Cognitive Biases and Emotion Regulation Strategies for Mental Health During the COVID-19 Lockdown of 2020: Structural Equation Modeling Study
JMIR Ment Health 2021;8(11):e30961
URL: https://mental.jmir.org/2021/11/e30961
doi:10.2196/30961
PMID:34517337

Original Paper

A Digital Human for Delivering a Remote Loneliness and Stress Intervention to At-Risk Younger and Older Adults During the COVID-19 Pandemic: Randomized Pilot Trial

Kate Loveys[1], MSc; Mark Sagar[2,3], PhD; Isabella Pickering[1], BSc; Elizabeth Broadbent[1], PhD

[1]Department of Psychological Medicine, The University of Auckland, Auckland, New Zealand

[2]Auckland Bioengineering Institute, The University of Auckland, Auckland, New Zealand

[3]Soul Machines Ltd, Auckland, New Zealand

Corresponding Author:
Elizabeth Broadbent, PhD
Department of Psychological Medicine
The University of Auckland
Building 507, Level 3
22-30 Park Avenue, Grafton
Auckland, 1023
New Zealand
Phone: 64 9 923 0003
Email: e.broadbent@auckland.ac.nz

Abstract

Background: Loneliness is a growing public health issue that has been exacerbated in vulnerable groups during the COVID-19 pandemic. Computer agents are capable of delivering psychological therapies through the internet; however, there is limited research on their acceptability to date.

Objective: The objectives of this study were to evaluate (1) the feasibility and acceptability of a remote loneliness and stress intervention with digital human delivery to at-risk adults and (2) the feasibility of the study methods in preparation for a randomized controlled trial.

Methods: A parallel randomized pilot trial with a mixed design was conducted. Participants were adults aged 18 to 69 years with an underlying medical condition or aged 70 years or older with a Mini-Mental State Examination score of >24 (ie, at greater risk of developing severe COVID-19). Participants took part from their place of residence (independent living retirement village, 20; community dwelling, 7; nursing home, 3). Participants were randomly allocated to the intervention or waitlist control group that received the intervention 1 week later. The intervention involved completing cognitive behavioral and positive psychology exercises with a digital human facilitator on a website for at least 15 minutes per day over 1 week. The exercises targeted loneliness, stress, and psychological well-being. Feasibility was evaluated using dropout rates and behavioral observation data. Acceptability was evaluated from behavioral engagement data, the Friendship Questionnaire (adapted), self-report items, and qualitative questions. Psychological measures were administered to evaluate the feasibility of the trial methods and included the UCLA Loneliness Scale, the 4-item Perceived Stress Scale, a 1-item COVID-19 distress measure, the Flourishing Scale, and the Scale of Positive and Negative Experiences.

Results: The study recruited 30 participants (15 per group). Participants were 22 older adults and 8 younger adults with a health condition. Six participants dropped out of the study. Thus, the data of 24 participants were analyzed (intervention group, 12; waitlist group, 12). The digital human intervention and trial methods were generally found to be feasible and acceptable in younger and older adults living independently, based on intervention completion, and behavioral, qualitative, and some self-report data. The intervention and trial methods were less feasible to nursing home residents who required caregiver assistance. Acceptability could be improved with additional content, tailoring to the population, and changes to the digital human's design.

Conclusions: Digital humans are a promising and novel technological solution for providing at-risk adults with access to remote psychological support during the COVID-19 pandemic. Research should further examine design techniques to improve their acceptability in this application and investigate intervention effectiveness in a randomized controlled trial.

Trial Registration: Australia New Zealand Clinical Trials Registry ACTRN12620000786998; https://www.anzctr.org.au/Trial/Registration/TrialReview.aspx?id=380113

(JMIR Ment Health 2021;8(11):e31586) doi:10.2196/31586

KEYWORDS

COVID-19; loneliness; stress; well-being; eHealth; digital human; conversational agent; older adults; chronic illness

Introduction

This study investigated the feasibility and acceptability of a digital human (DH) that delivered a psychological intervention to mitigate the effects of social restrictions on loneliness, stress, and well-being in vulnerable populations during the COVID-19 pandemic. The results will inform the design of a randomized controlled trial (RCT) to evaluate intervention effectiveness. To provide a rationale and context for the study, the introduction describes the effects of the COVID-19 pandemic on loneliness, the importance of treating loneliness, and previous work on robot and conversational agent (CA) interventions for loneliness.

Many countries have adopted socially restrictive public health measures over recent months to slow the spread of the COVID-19 pandemic, including the United Kingdom, Canada, the European Union, Japan, and Australia [1]. Precautions have included bans on mass gatherings, closure of schools and businesses, mandatory work from home conditions, and limits on social activities [2]. New Zealand adopted some of the strictest lockdown rules globally, which included 2-m physical distancing between people and staying at home, except for essential trips to a supermarket or pharmacy, or to seek medical care, with restrictions gradually eased as appropriate [3]. Individuals who were at risk of developing a severe illness should they contract COVID-19 were advised to take additional precautions to social distance and isolate [4]. This included older adults over the age of 70 years (who are at greater risk of dying from COVID-19) [5] and younger adults with an underlying medical condition who may be immunocompromised [6].

While these precautions can help protect vulnerable populations, there are mental health implications of strict social distancing, including increased loneliness [7]. Older adults and adults with underlying health conditions were already at greater risk of loneliness prepandemic [8,9], and these restrictions have exacerbated this risk. Interventions to reduce loneliness are especially important for this group given the long-term implications for health as described below.

Loneliness is a subjective psychological state in which a person perceives a mismatch between their actual and desired social relations [10]. While brief feelings of loneliness can serve as an adaptive motivator to seek social interaction, chronic loneliness has negative effects on physical and mental health outcomes [11,12]. Loneliness is associated with feelings of stress [13], which activate the body's "fight or flight" response. The sympathetic nervous system becomes activated, and over a prolonged period, it creates negative downstream effects on the cardiovascular, neuroendocrine, and immune systems [14]. As a result, chronic loneliness has been associated with increased risks of morbidity (eg, coronary heart disease, high blood pressure, and stroke) [11] and mortality [15]. Loneliness can be improved through psychological interventions that target the following 4 key areas: changing maladaptive social cognitions, increasing social support, increasing opportunities for social interaction, and improving social skills [16].

Loneliness interventions can be delivered in-person or remotely through technology, and both have been shown to be effective [16], including for older adults [17]. In-person loneliness interventions have included individual psychotherapy involving social cognitive training as part of cognitive behavioral therapy [18], mindfulness-based therapies [19], and social support groups [20]. However, remote interventions may be more suitable for at-risk individuals in isolation as a result of the pandemic. Remote therapies for loneliness have included internet-based cognitive behavioral therapy [21] and internet skills training to access online support [22]. However, research has shown that engagement with technology-based interventions is often lower outside of a clinical trial context [23,24].

Artificial agents may be a particularly engaging way to provide psychological support to people during a pandemic. People have been shown to feel a sense of social presence with artificial agents, which can improve technology engagement [25,26]. Social robots are artificial agents with embodiment in a physical hardware form that are capable of social interaction and are programmed to autonomously interact with their physical environment [27]. CAs are artificial agents that include a dialogue system, and may or may not include a digital embodiment or face [28]. Under the umbrella term of CAs fall chatbots, embodied CAs, voice assistants, and DHs, among others. CAs may be more feasible for providing remote support than robots because they are less expensive and more scalable [29], as they can be accessed through websites or software applications on devices that many patients already own (eg, smartphones and computers).

A recent scoping review on robot-facilitated loneliness interventions found evidence supporting their use with older adults [30]. For example, Paro, a companion robot in the form of a fluffy baby harp seal, alleviated feelings of loneliness in older adults in nursing homes by providing direct companionship in a manner akin to a pet [31]. Other robots include Giraff (a telepresence robot that connects users and their families over video call [32]), MARIO (which includes a My Memories function where users can show photographs to others as a conversation starter [33]), and SYMPARTNER (which reminds people of their upcoming social engagements [34]). Social robots may also improve loneliness in younger adults [35]. Robots have been shown to be effective at delivering other kinds of psychological interventions, such as positive psychology interventions for well-being [36].

Research looking at the clinical effectiveness of CAs in health care is relatively limited, and a more robust methodology is required [28]. However, a study found that daily conversations with an animal-like embodied CA over the course of a hospital stay significantly improved loneliness in older adults [37].

Another study found that daily interactions over 1 week with a human-like embodied CA that used a proactive communication strategy improved loneliness and happiness in older adults [38].

CAs also show promise for delivering psychological therapies to improve stress and well-being; outcomes that may be worsened by chronic loneliness. Vivibot, a Facebook messenger chatbot that delivered positive psychology exercises over 4 weeks, was found to be acceptable and effective for reducing anxiety in young adults with a chronic health condition [39]. Other research has found that a Facebook messenger chatbot that delivered cognitive behavioral therapy exercises, such as mindfulness and gratitude activities, improved stress and well-being [40].

DHs are a new type of CA that use artificial intelligence to build social and emotional engagement with users [41], which could help to reduce loneliness. DHs differ from other CAs in that they are modeled off real people using Hollywood light room technology and computer-generated imagery (CGI) animation techniques [42]. This provides DHs with a very life-like appearance. In addition, DHs include a complex cognitive architecture modeled off humans and involve a digital brain with virtual neurotransmitters to influence behavior [43]. For example, while in "high oxytocin mode," DHs show attachment and separation distress toward users, which can help to build a bond [44]. DHs use live neural networks while interacting with people to classify their emotional state, and respond to people using a combination of speech, facial behaviors, and body gestures. DHs may be a particularly promising technology to deliver a remote loneliness intervention given their engaging social abilities and scalability; all that users require to access one is a computer and an internet connection. However, as DHs are a relatively new technology, it is unknown whether they are a feasible and acceptable way to deliver a remote loneliness intervention.

This study aimed to investigate whether a DH was a feasible and acceptable method of delivering a remote loneliness and stress intervention to high-risk adults during the COVID-19 pandemic. In addition, this study evaluated the feasibility of the study methods in advance of a future definitive RCT. It was hypothesized that a DH would be a feasible and acceptable method of intervention delivery, and that the study methods would be feasible. The results will inform the design of an RCT to investigate the effectiveness of the DH intervention.

Methods

Trial Registration

This trial was reported in keeping with the CONSORT (Consolidated Standards of Reporting Trials) 2010 statement extension for randomized pilot and feasibility trials [45]. Ethics approval was obtained from the University of Auckland Human Participants Ethics Committee on July 06, 2020 (approval number: 024752). The trial was prospectively registered with the Australia New Zealand Clinical Trials Registry on August 04, 2020 (registration number: ACTRN12620000786998).

Trial Design

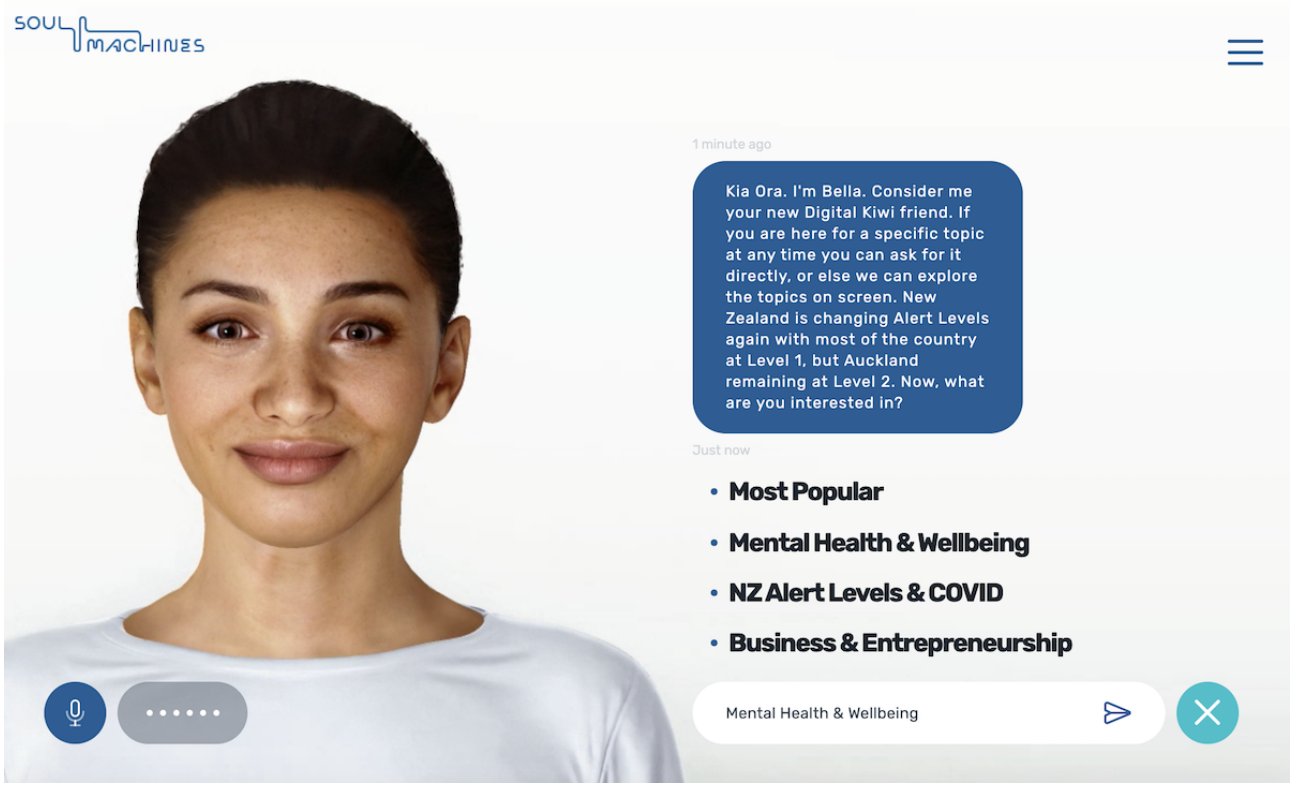
A randomized pilot trial was conducted involving a parallel mixed design with a waitlist control condition (1:1 allocation ratio). The primary outcomes were feasibility and acceptability, and the secondary outcomes were rapport with the DH, loneliness, stress, COVID-19 distress, positive and negative experiences, and psychological well-being. No major changes were made to the methods after commencing the trial.

DH Intervention

The DH facilitator ("Bella") was developed by Soul Machines Ltd (Auckland, New Zealand) (Figure 1). Bella was autonomously animated and presented on a website that participants accessed from their personal computer, tablet, or smartphone. Bella was modeled to be a young adult female of mixed race (Māori and New Zealand European). She was synthesized from the visual features of several human models (ie, not modeled off a singular person). Bella was presented in front of a white background in a portrait view of her head and shoulders. Her appearance, background, and proximity to the screen remained consistent throughout the study.

Bella autonomously responded to participants' language using a finite state conversation engine with preprogrammed responses. Bella was programmed to have some autonomous variation in her language for phrases that would not affect her intervention delivery (eg, she varied her greetings each day). Bella spoke using a computer-generated female voice with an Australian accent ("Wavenet C – female" by Google). Participants could communicate with Bella in 1 of the following 3 ways: (1) speech, (2) typing, and (3) clicking on-screen buttons (where present). Bella always responded to participants in speech; however, if participants opened the messenger window to type, they could see a typed version of Bella's speech as well (see Figure 1 for an example). If Bella did not understand a participant's language, she would say, "I'm sorry, I didn't understand. Could you please repeat or reword your statement?" or similar. If she did not understand after a couple of attempts, she would redirect the participant back to her main menu.

Figure 1. The digital human's user interface when using the messenger function.



Bella engaged in human-like facial and body gestures as she spoke, including blinking, maintaining eye gaze, raising her eyebrows, and moving her head and shoulders. She showed emotional expressions on her face as she spoke to portray joy and concern, which were preprogrammed and triggered by phrases she spoke using text-to-speech emotional markup language. This involved a process of manually tagging language in her script to elicit particular facial emotions each time Bella spoke the phrase. Bella's facial expressions were autonomously generated in real-time using visual computing and neurobehavioral modeling techniques (described in greater detail in previous reports [41-43,46]). Bella had a virtual nervous system that contained virtual neurotransmitters and live neural networks to process emotional data and inform her responses; however, these capabilities were not used in this study in order to maintain experimental control.

Bella was designed to deliver several relationship building strategies derived from psychology [47] and human-computer interaction research [48]. These included engaging in shared activities with the user, mutual self-disclosure, showing empathy, expressing the value of the friendship, and being nonjudgmental. These relationship building strategies were incorporated into Bella's language at various points in the interaction.

Participants were informed that Bella continuously collected speech and video data in order to communicate (eg, to hear speech and to make eye contact). These data were not recorded, saved, or analyzed by the researchers. Bella's data collection and use processes are in keeping with the European Union General Data Protection Regulation (GDPR) [49,50].

DH Intervention Content

Participants were asked to prioritize visiting the mental health and well-being content that Bella offered as part of their daily website visit. This content included evidence-based exercises to improve loneliness, stress, and psychological well-being, as described below.

The Expressing Kindness Challenge

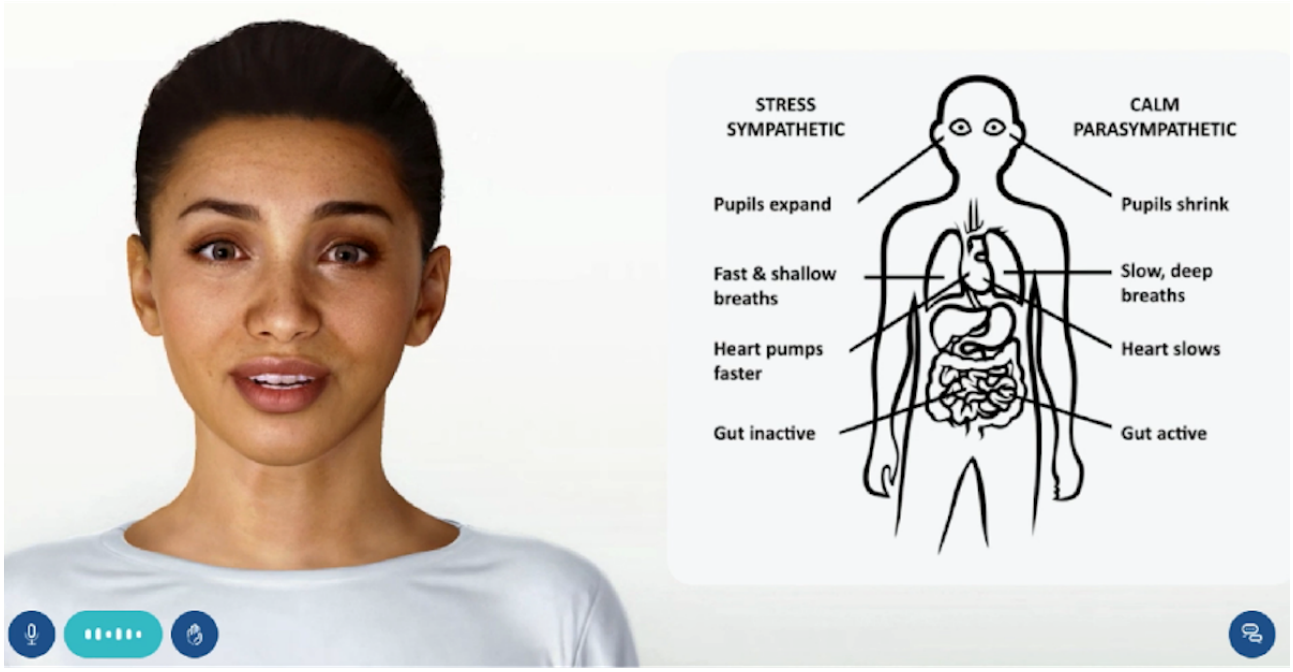
Three challenges were delivered over 3 days and included evidence-based strategies to improve loneliness and psychological well-being. The first 2 challenges were (1) to make contact with an old friend, relative, or someone the participant had not been in touch with for a while and (2) to contact someone to let them know something that the participant appreciated about them. These tasks aimed to increase opportunities for social interaction, strengthen social support, and improve social skills. The third challenge asked the participant to make a list of 3 things that they were grateful for, as a positive psychology exercise. Each of the challenges was accompanied by examples to help the participant generate ideas (eg, on day 2, Bella told participants something that she appreciated about them). At the end of the module, participants were reminded to continue practicing kindness toward others and themselves.

The Brain and Stress Module

This module provided psychoeducation about stress and stress awareness through verbal explanations and diagrams over 1 visit (Figure 2). It covered how stress affects the body and symptoms that are associated with the stress response. The module encouraged participants to reflect on the sources of stress in their lives, and informed participants of behavioral strategies for stress management. These included educating

participants about a deep breathing exercise that they could practice, and linking participants to the Headspace website [51], where participants could access audio recordings of deep breathing and meditation exercises. At the end, participants were encouraged to visit the mental health tips, which are described in further detail below.

Figure 2. The digital human interface during psychoeducation as part of the brain and stress module.



Mental Health Tips

Six modules each focused on a separate psychological well-being tip. The tips encouraged social connection, exercise, acknowledging feelings, being mindful of anxiety-provoking news media consumption, doing activities that elicit positive emotions, and trying out behaviors from a self-care guide.

Other Conversation Modules

Participants were able to talk with Bella about a range of other topics beyond mental health and well-being. This included information about the COVID-19 pandemic (eg, New Zealand's alert levels, details about the virus, symptoms and prevention, and New Zealand's Healthline and health support resources), and information regarding business and entrepreneurship (eg, remote work and business support organizations).

Participants

Thirty participants were recruited. Participants were adults who were at greater risk of developing severe illness if they contracted COVID-19, and as a result, they were asked by the local New Zealand Government to self-isolate to a greater degree during the pandemic. They included (1) older adults aged 70 years or older and (2) adults aged 18 to 69 years who had at least one underlying medical condition that increased the risk of contracting severe COVID-19. The underlying medical condition could have included a serious respiratory disease (such as a chronic lung disease or moderate to severe asthma), a serious heart condition, an immunocompromised condition (such as cancer treatment, smoking-related illness, bone marrow or organ transplantation, hematologic neoplasms, immune deficiency, uncontrolled HIV or AIDS, and prolonged use of corticosteroids and/or other immune-weakening medications such as disease-modifying antirheumatic drugs), a BMI of 40 or higher, diabetes, chronic kidney disease, dialysis, liver disease, and/or pregnancy at the third trimester stage. Participants were required to have English fluency, and access to a computer and internet connection at home. Participants who were 70 years or older were required to achieve a score of 25 or higher on the Mini-Mental State Examination (MMSE). Potential participants were excluded if they received a score of 24 or lower on the MMSE, which would indicate cognitive decline to a moderate or greater degree. Participants aged 70 years or older were not excluded on the basis of whether or not they had an underlying health condition, as their age placed them at a higher risk of developing severe COVID-19.

Twenty-two older adult participants (aged 70 years or older) were recruited from 5 Summerset retirement village sites around the greater Auckland area. Recruitment methods involved presentations to residents about the research, email flyers, and caregiver word of mouth. Residents approached the research team if they were interested in participating. Eligibility screening involving the MMSE and an informed consent procedure (for those who were eligible) were conducted in-person at the retirement village with a member of the research team.

Eight younger adult participants (aged 18-69 years with an underlying medical condition) were recruited from a flyer posted to a staff email list at the University of Auckland, in addition to targeted Facebook advertising, word of mouth, and a Summerset retirement village presentation. Younger adults interested in taking part completed an eligibility screen and informed consent procedure online via a survey website (Qualtrics), except 1 participant who was recruited from a retirement village presentation. This participant completed an eligibility screen and informed consent procedure in-person.

A recruitment target of 30 participants was set, as a minimum of 12 participants per group is recommended for a feasibility study due to precision about the mean and variance [52], and to allow for 20% attrition. Recruitment took place between November 11, 2020, and March 04, 2021, with a 3-week break from late December. Recruitment stopped once the quota of 30 participants had been reached.

Data were collected from online questionnaires using Qualtrics, which participants completed from their place of residence. For older adult participants, this may have included completion from a Summerset retirement village independent living villa or apartment, or from the nursing home facility. For younger adult participants, participation took place online from their place of residence in the community or a Summerset care home facility. Data collection took place between November 16, 2020, and March 11, 2021. All participants in the study were provided with a NZ $30 (US $21.50) shopping voucher for their involvement in the research.

Randomization

Participants were randomly allocated to an intervention or waitlist control group by a member of the research team (EB) (1:1 allocation ratio). Simple randomization was performed using a computerized sequence generation software called Research Randomizer. Allocations were concealed in sealed opaque envelopes from the researchers who enrolled participants (KL and IP) until after participants were enrolled and allocated an ID code. At this point, the researcher was deblinded to assign participants to conditions and provide participants with the appropriate instructions. Participants were deblinded after their assignment to conditions.

Procedure

Once enrolled, participants were contacted over email with instructions for proceeding in the trial. For nursing home residents, their caregiver was copied in the email communications and facilitated the participant's involvement in the study.

All participants completed an online baseline questionnaire on day 1 of their participation. Then, participants in the intervention group completed a DH training session with a member of the research team for 30 minutes. For all older adults (plus 1 younger adult participant), this took place in-person at their retirement village or nursing home facility. For 7 of 8 younger adults, this took place either in-person at the University of Auckland Clinical Research Centre or online over Zoom video conferencing software (Zoom Video Communications), depending on the lockdown conditions.

All participants received the same technology training, which involved learning how to interact with Bella and completing "day 1" of their intervention week with the researcher present to answer questions. The researcher ensured that the software worked on each participant's computer. Participants were provided with written instructions and pictures of the user interface that summarized the training session content. For sessions over Zoom videoconferencing, the screen share feature was used and participants received a PDF copy of the interaction instructions. Three participants were trained over Zoom, and 23 participants were trained in-person.

Participants were asked to interact with Bella for at least 15 minutes per day over 1 week. Participants visited Bella's website independently from their place of residence. The daily 15 minutes could include time spent interacting with Bella and doing therapy activities (eg, a deep breathing exercise). They were asked to prioritize completing the mental health and well-being modules before visiting other content. Participants interacted with Bella at their chosen time of day. Participants were sent a daily text reminder to engage in the intervention and were informed that they could text back to receive technical support.

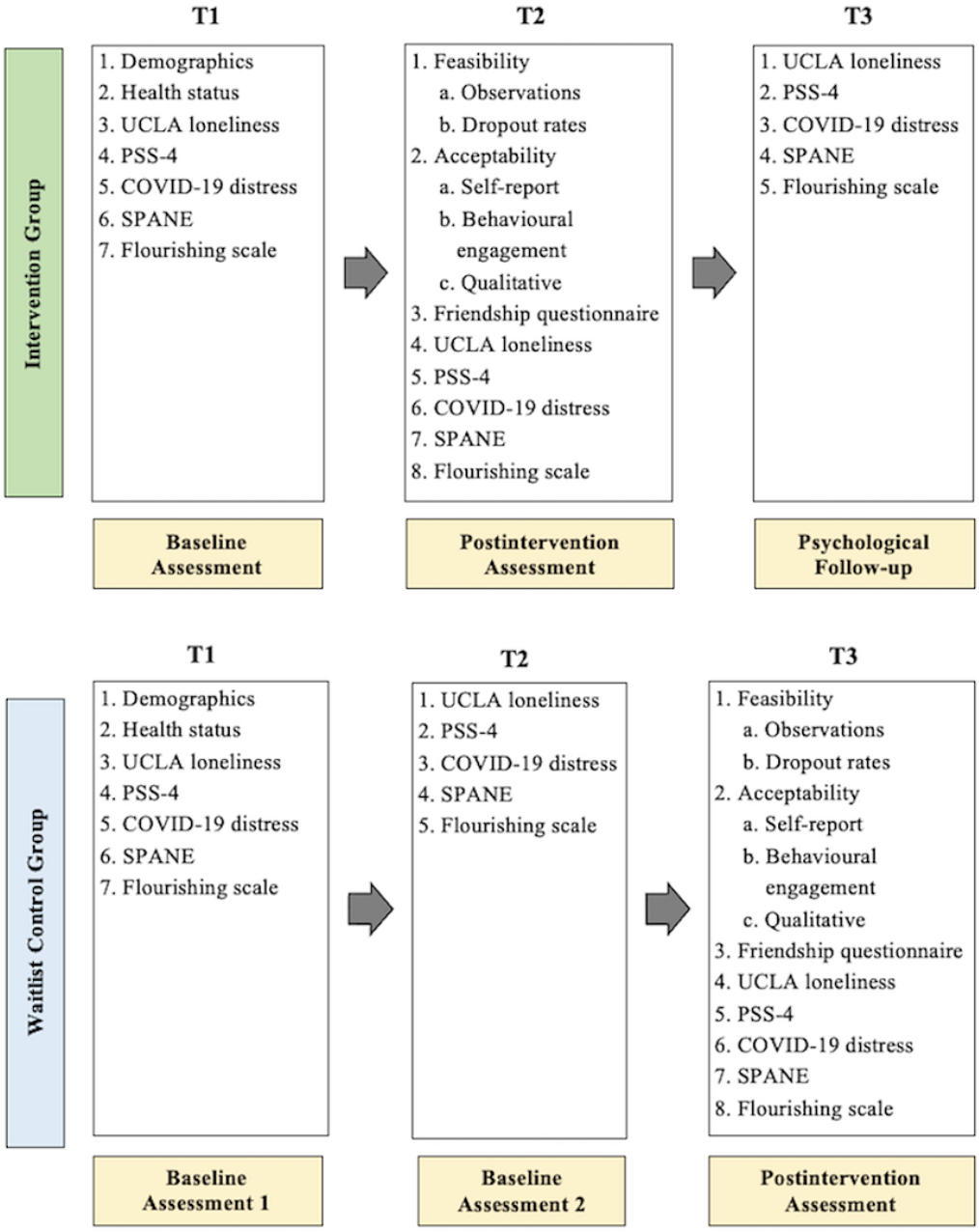
On day 8, the intervention group finished their intervention week and completed an online postintervention questionnaire. One week later, on day 15, intervention group participants filled out an online follow-up questionnaire.

For participants in the waitlist control group, the order of the procedure was slightly different. Participants in the waitlist group completed an online baseline questionnaire on day 1 and then waited for 1 week. On day 8, waitlist participants completed a second online questionnaire, completed the technology training session, and began their intervention week. On day 15, at the end of their intervention week, waitlist participants completed the postintervention questionnaire.

Measures

Figure 3 depicts the time points at which each measure was administered. Questionnaires were administered online using Qualtrics, a secure survey website.

Figure 3. Time points for assessments. PSS-4: Perceived Stress Scale 4 items; SPANE: Scale of Positive and Negative Experiences.



Feasibility Measures

Feasibility of the DH Intervention

Observations were made by a member of the research team (KL) about how the intervention was used (eg, independently or with the aid of a caregiver), along with dropout rates and reasons. Observations were also made regarding the feasibility of the technology training methods for younger and older adults, and nursing home residents. Instances were recorded where participants refused to receive training through a particular delivery method (eg, video calling).

Feasibility of the Study Methods

Observations were recorded during recruitment and data collection by a member of the research team (KL). Observations pertained to the success rate of different recruitment strategies for younger and older adult participants, and challenges associated with data collection from the online forms that participants completed independently.

Acceptability Measures

Acceptability of the DH

Bella's acceptability was measured using quantitative self-report items and open-ended qualitative questions designed for the study. Behavioral engagement data were also collected. The acceptability measures are outlined in further detail below.

Self-Report Items

Participants were asked to rate whether (1) they felt Bella was helpful for promoting resilience and psychological well-being,

(2) they felt Bella was helpful for improving feelings of loneliness, and (3) they would be willing to use Bella again in the future, using a 7-point scale with response anchors (1, "definitely no" to 7, "definitely yes").

Qualitative Responses

Participants provided written responses to the following open-ended questions: *What did you like most about Bella?* and *How do you think Bella could be improved?* These questions were intended to provide an overall indication of Bella's acceptability and to identify aspects of the technology that could be improved.

Behavioral Engagement

Behavioral engagement with Bella over 1 week was evaluated by retrospective self-report. Participants reported on which days of the week they visited Bella and estimated approximately how long they used Bella each day in minutes.

Acceptability of the Intervention Content

The acceptability of each psychological intervention module was evaluated separately in the postintervention questionnaire. Participants rated how much they liked the brain and stress module, Headspace (if they visited), and the expressing kindness challenge (including each of its 3 activities) on a 7-point scale with response anchors (1, "not at all" to 7, "very much"). Participants rated how beneficial they found the expressing kindness challenge for well-being, and how well they felt the brain and stress module improved their understanding of the stress response on a 7-point scale (1, "not at all" to 7, "very much"). Participants who visited Headspace were asked whether they felt that Headspace was a helpful resource to link to with a dichotomous yes/no response option. The participants were also asked the following qualitative question: *Were there any particular topics that you would have liked to talk about with Bella, which were not available?* Participants provided written responses. Self-reported behavioral engagement data were collected on whether participants visited each module and whether they did the activity that the module asked of them.

Rapport With the DH

Rapport with Bella was measured using the 20-item Friendship Questionnaire developed by Johanson et al [53], with items adapted to suit a DH. It is comprised of items taken from multiple friendship scales, including the McGill Friendship Questionnaire, the McGill Friendship Questionnaire Functions scale, the Interactant Satisfaction Survey, and the Acquaintance Description Form-F2 [53]. Participants indicated their agreement with each item using a 5-point Likert scale from 1 ("strongly disagree") to 5 ("strongly agree"). Responses were summed to derive a total score from 20 to 100, where a higher score indicated greater rapport. The friendship questionnaire has been shown to have good internal consistency reliability when used to evaluate a social robot in a New Zealand adult sample (α=.94) [53]. The scale showed good internal consistency reliability in this study sample when adapted for use with a DH (α=.95). The adapted scale has been included in Multimedia Appendix 1.

Loneliness

Loneliness was measured using the 20-item UCLA Loneliness Scale (Version 3) [54]. Participants rated how often they felt the way described in each statement using a 4-point scale. Responses could range from 1 ("never") to 4 ("always"). Items were reverse coded where appropriate, and responses were summed to derive a total score from 20 to 80. A higher score indicated greater perceived loneliness. This scale was developed with language to improve readability and has demonstrated acceptable psychometric properties with older adults. This includes good internal consistency reliability (α=.89), discriminant validity with social support, and construct validity [54].

Psychological Stress

Perceived stress was measured using the 4-item Perceived Stress Scale (PSS-4) [55], which evaluated the degree of stress participants felt over the past week using a 5-point scale (0, "never" to 4, "very often"). Items 2 and 3 were reverse coded, and all responses were summed to form a total score from 0 to 16. A higher score indicated greater perceived stress. Although the psychometric properties of the PSS-10 and PSS-14 have been shown to be superior, the PSS-4 was chosen as it is a shorter measure of perceived stress that reduces participant burden and has adequate internal consistency reliability [56].

COVID-19 Distress

Worry about contracting COVID-19 was measured using a 1-item scale [57]. The scale evaluated participants' degree of worry over the past week on a 4-point scale as follows: 0, "I do not worry about getting COVID-19;" 1, "I occasionally worry about getting COVID-19;" 2, "I spend much of my time worrying about getting COVID-19;" and 3, "I spend most of my time worrying about getting COVID-19."

Positive and Negative Affect

The Scale of Positive and Negative Experiences (SPANE) has two 6-item subscales that measure positive emotions (SPANE-P) and negative emotions (SPANE-N) [58]. The subscales measured the extent to which positive or negative emotions were experienced over the past week using a 5-point scale (1, "very rarely or never" to 5, "very often or always"). For each subscale, responses were summed, and a total score was derived ranging from 6 to 30. A higher score indicated stronger positive or negative affect, depending on the subscale. Affect balance scores (SPANE-B) were calculated, which indicate the participant's balance of positive and negative affect from −24 to 24, where positive scores indicate more positive than negative affect during the period. The scale has good internal consistency (SPANE-B: α=.89; SPANE-P: α=.87; SPANE-N: α=.81) and convergent validity [58].

Psychological Well-Being

Psychological well-being was measured using the 8-item Flourishing Scale [58]. Participants were asked to rate their perceived success across items pertaining to different aspects of psychological well-being, including purpose, relationships, self-esteem, and optimism, using a 7-point Likert scale (1, "strong disagreement" to 7, "strong agreement"). Responses

were summed to derive a total well-being score between 8 and 56. Higher scores indicated greater well-being. The Flourishing Scale has been shown to have good psychometric properties including convergent validity and discriminant validity. It has also been shown to have good reliability and validity in a nationally representative New Zealand sample [59].

Data Analysis

Quantitative Data

Data were analyzed using SPSS software (version 27; IBM Corp). Missing data were addressed by imputing the mean score of the participant's other responses to the scale at the timepoint. For 1-item scales, where it was not possible to impute a score or where the participant did not complete a full scale, the participant's data were excluded from analysis of the relevant variable.

Baseline demographic and psychological variables were calculated for the overall sample, and compared between groups using chi-square tests and independent samples *t* tests. Average acceptability and rapport scores were calculated for the overall sample, and independent samples *t* tests were conducted to compare group means. A series of mixed factorial analyses of variance (ANOVA) were conducted to evaluate the main and interaction effects of condition and time on psychological outcomes. Data were checked for violations of test assumptions. Greenhouse-Geisser–adjusted values were reported for data where sphericity assumptions were violated (COVID-19 distress, SPANE-P, and SPANE-B). Exploratory pair-wise comparisons with Bonferroni corrections were conducted as follow-up analyses for significant or trending effects.

Qualitative Data

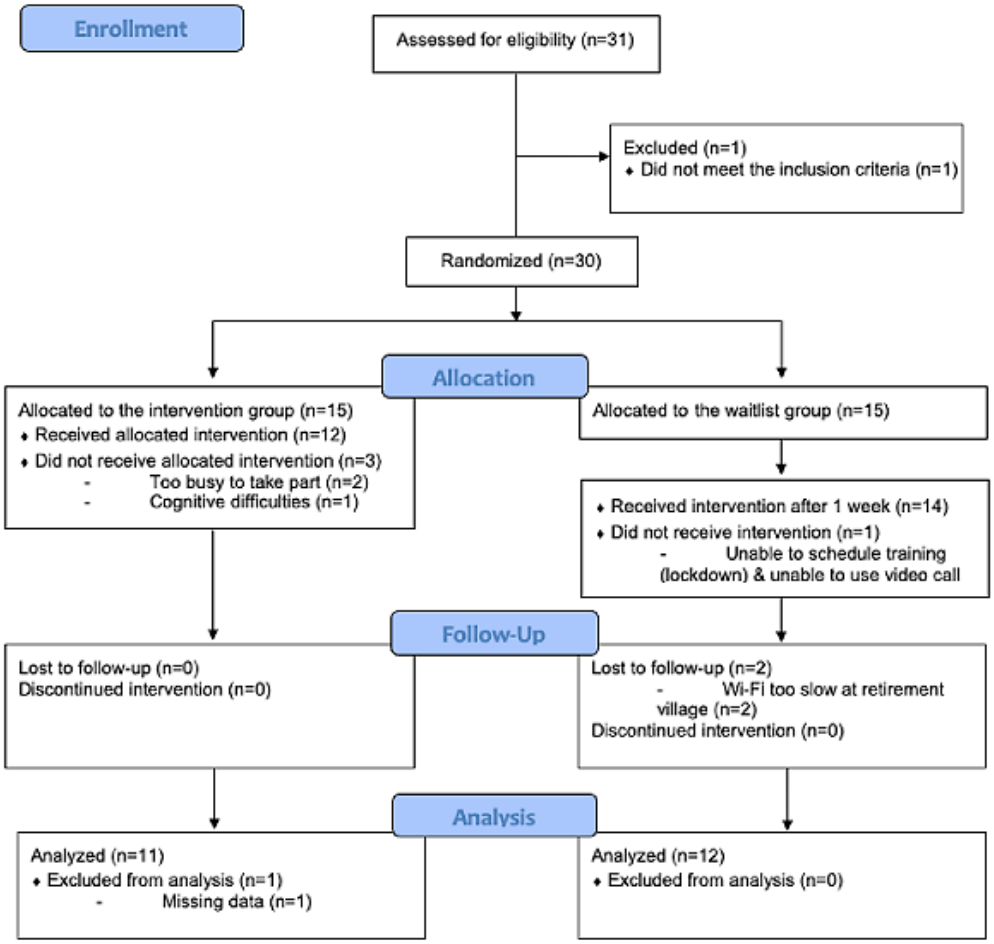
Written responses to 3 open-ended questions were analyzed using reflexive thematic analysis [60], which is theoretically flexible and suitable for analyzing the content of language data. One member of the research team (KL) conducted the analysis in keeping with recommendations by Braun & Clarke [60], using the following steps: (1) familiarization with the data, (2) coding, (3) generating initial themes, (4) reviewing themes, (5) defining and naming themes, and (6) writing results. An inductive approach was taken whereby coding and theme development were informed by the content of the data. As part of the theme development in stages 3, 4, and 5, themes and subthemes were checked against the original data set and each other to ensure that they were internally coherent (ie, organized around a clear central concept), consistent, and distinctive. Themes and subthemes were split or combined during the review process (stage 4) to improve specificity. All coded data for each theme and subtheme were collated to assist with result writeup. Data were combined across groups as both received the same intervention.

Results

Participants

Participants were predominantly female (24/30, 80%) and Caucasian (22/30, 73%), and mainly had high school or less education (14/30, 47%). Half of the sample (15/30, 50%) reported an underlying medical condition. Participants reported low levels of loneliness (mean 37.79, SD 9.90) and stress at baseline (mean 3.86, SD 2.88). Participant characteristics at baseline are reported in Multimedia Appendix 2. A CONSORT diagram depicts participant flow through the study in Figure 4.

Figure 4. CONSORT (Consolidated Standards of Reporting Trials) diagram of participant flow.



Feasibility of the DH Intervention

DH Training Method

Older adults required technology training to be completed in-person at their retirement village with a member of the research team (KL or IP). This method worked well as it avoided any discomfort with using the video calling software. One older adult was offered training over the Zoom video calling software during the lockdown period and refused as she was not able to use the software. Only 1 of 22 older adult participants did not have a webcam as part of their computer, which was uncovered during the technology training (the DH website requires a webcam). To solve this, a webcam was borrowed from the retirement village reception and installed by a member of the research team (IP) during the training session.

Younger adult participants were generally able to be trained either in-person from a clinic room at the university (outside of the lockdown period) or online over Zoom (during the lockdown period). Video calling did not appear to impact the effectiveness of the training. Technical support requests were low for younger adults during the study, irrespective of how their training was delivered. One younger adult participant who was a nursing home resident required in-person technology training.

Dropout

Six participants withdrew from the study (all older adults). The reasons for withdrawal were as follows: (1) the Wi-Fi speed at the retirement village location was too slow for Bella to load properly (n=2); (2) cognitive health difficulties interfered with understanding study instructions (n=1); (3) the participant was too busy to take part after enrollment (n=2); and (4) technology training could not be scheduled (n=1).

Intervention Completion

Twenty-four participants completed the intervention, 22 of whom completed it independently after training. Two participants (1 younger adult and 1 older adult) who were both nursing home residents required caregiver assistance to access the website and interact with Bella.

Feasibility of Study Methods

Feasibility of Recruitment Methods

The majority of older adult participants were recruited through information sessions held at retirement villages (21/22, 96%). One older adult participant was recruited through an email flyer sent by a village staff member to residents.

For younger adults, the most effective recruitment method was by advertisement to a university junk email list (5/8, 63%). Facebook advertising and word of mouth each resulted in

recruitment of 1 participant, and 1 participant was recruited from an information session at a nursing home.

Feasibility of Assessment Delivery Methods

Most participants completed assessments online using Qualtrics without significant issue. Two participants reported instances where they were unsure whether their responses had been submitted. Nursing home residents were unable to complete assessments independently on a website and required caregiver assistance.

Acceptability of the DH

Overall, participants reported that Bella was somewhat helpful for promoting resilience and psychological well-being (mean score 4.39 out of 7, SD 1.83) and for improving any feelings of loneliness (mean score 4.09 out of 7, SD 1.76), as responses on average were above the mid-point. Participants were somewhat willing to use Bella again in the future (mean score 4.09 out of 7, SD 1.98). Younger and older adults rated Bella similarly across the acceptability items, and no significant differences were found.

On average, participants interacted with Bella 6 out of 7 days (mean 6.23, SD 1.19). Participants interacted with Bella for approximately 20 minutes per day (mean 20.20, SD 13.95); 5 minutes longer than the 15 minutes per day requested by the researchers. The average total interaction time with Bella over 1 week was 128 minutes (mean 128.33, SD 102.77). There were no significant differences between younger and older adults in engagement behavior.

Participants identified several strengths and limitations of Bella through responses to the following 2 written open-ended questions: *What did you like most about Bella?* and *How do you think Bella could be improved?* Themes, subthemes, and representative quotes are presented below in Tables 1 and 2. Definitions of themes are presented in Multimedia Appendix 3. Overall, participants liked aspects of Bella's appearance, speech, and interpersonal skills; the informational support Bella provided; the user experience; and the interaction with a new technology. Aspects of Bella that participants felt could be improved were the human likeness of her interaction behaviors and voice, and aspects of the conversation design (eg, more personalization and conversation topics). Some participants reported that they felt Bella would be improved with gradual advances in the underlying technology (eg, natural language understanding). Other participants reported that they would have preferred to interact with a real human, and some participants did not request any improvements.

Table 1. Themes, subthemes, and representative quotes describing what participants liked most about Bella.

Themes and subthemes	Representative quotes
Bella's appearance	
Facial expressions	*Her friendly smile.* [Participant ID 106]
Human-like	*I liked the 'human' aspects of her.* [Participant ID 124]
Attractive face	*She is attractive looking.* [Participant ID 112]
Similarity to user	*I felt in some ways visually represented by Bella.* [Participant ID 124]
Bella's speech	
Gentle voice	*Quite relaxing. Liked the soft voice. You can hear compassion in her voice.* [Participant ID 115]
Self-disclosure	*I really appreciated how the conversation would be 'softened' by more personal statements from her.* [Participant ID 124]
Clear language	*Clear speaking.* [Participant ID 118]
Bella's interpersonal skills	
Companionship	*That she was there.* [Participant ID 117]
Nonjudgmental	*Feel like you can tell her just about anything and she wouldn't be shocked. Like talking to a priest in confession.* [Participant ID 115]
Friendly personality	*Friendly and likeable.* [Participant ID 119]
Validating	*Nice being told what you're feeling is normal.* [Participant ID 115]
Calm personality	*Her calmness.* [Participant ID 125]
Informational support	
Quality resources	*A good selection of resources.* [Participant ID 126]
Accessible delivery	*Accessibility, most people would find her approachable.* [Participant ID 103]
User experience	
Interaction modalities	*The direct interaction.* [Participant ID 130]
User controls interaction	*A good medium that allowed me to have plenty of control.* [Participant ID 124]
Easy to use	*That it was easy to use.* [Participant ID 127]
Overall experience	*Enjoyed the 'experience' of Bella and certainly a helpful person if you were lonely.* [Participant ID 111]
Novel technology	
Something different	*Something to do with somebody to talk to me. She was different.* [Participant ID 104]
A new technology	*New technology is always intriguing, and I had heard of Bella before.* [Participant ID 105]

Table 2. Themes, subthemes, and representative quotes describing what participants felt could be improved about Bella.

Themes and subthemes	Representative quotes
Interaction behaviors	
More human-like movements	*The movement of Bella is still quite robotic and her eyes cannot really focus, which makes her sometimes not seem very engaged in the interaction.* [Participant ID 125]
Ability to touch	*A sad thing is you can't touch her. Make a doll out of her. A nice cuddly soft doll.* [Participant ID 115]
More positive emotional expression	*Smiles.* [Participant ID 118]
Conversation design	
Extra conversation topics	*A wider range of subjects.* [Participant ID 101]
Greater interactivity	*More interaction by talking to her, rather than just a yes or no.* [Participant ID 107]
More personalized responses	*Answers need to be more individualized (e.g., welcoming the participant by name and building on each day's responses).* [Participant ID 112]
Regularly update information	*Information wasn't as up to date (eg, COVID levels).* [Participant ID 129]
Avoid human-like backstories	*I didn't like the comments she made such as 'I contacted my friend today.' I found it weird that she was pretending to be real. I would have preferred if it was just accepted as an interface that had a good selection of resources that you could navigate in an interesting dynamic way.* [Participant ID 126]
Address user by name	*By addressing each person by their name, that way we could feel in the moment.* [Participant ID 116]
Incorporate user's life experience	*Most older people (I am 82) have many years of life experience and perhaps some way could be found to take life experience into consideration.* [Participant ID 112]
Robotic speech	
A more human-like voice	*Perhaps maybe not sound so robotic? It might be hard to achieve but all the inflections and warmth that someone would have in their tone and delivery was missing and I think that's what would have made Bella more engaging for me.* [Participant ID 127]
Formal speech delivery	*Improve pronunciation. Use correct English (nope and yeah are not acceptable).* [Participant ID 105]
Technology advances	
Natural language understanding	*I think given advances in technology this will happen anyway. Found the response from her using the audio didn't always work so found it easier to type the responses to her.* [Participant ID 102]
General technology advances	*I guess technology will advance and make changes but pretty amazing now.* [Participant ID 111]
Preference for a real human	*Personally, I believe talking to a real person is far more desirable.* [Participant ID 106]
No changes	*I accept it for what she is. Saying that we are not all the same, she is different, she is what she is.* [Participant ID 104]

Acceptability of Intervention Content

The Expressing Kindness Challenge

Of 24 participants, 22 visited the expressing kindness challenge. Overall, the expressing kindness challenge was liked by participants (mean score 5.50 out of 7, SD 1.34), as were the 3 daily challenges of which it was comprised. Participants reported liking reaching out to a friend (mean score 5.95 out of 7, SD 1.13), telling a friend what they appreciate about them (mean score 5.77 out of 7, SD 1.09), and making a gratitude list (mean score 5.71 out of 7, SD 1.14). Participants reported that the expressing kindness challenge felt beneficial for their well-being (mean score 5.00 out of 7, SD 1.95). There were no significant differences between younger and older adults in terms of how much they reported to like the expressing kindness challenge (mean score 5.57, SD 1.39 vs mean score 5.47, SD 1.36; t_{20}=−0.17; P=.87), its activities (all P>.27), or how beneficial the module felt for well-being (mean score 5.57, SD 1.13 vs mean score 4.73, SD 2.22; t_{20}=−0.94; P=.36).

The majority of participants (13/24, 59%) visited all 3 tasks of the expressing kindness challenge. Two participants (9%) visited only 2 tasks, and 7 participants (32%) visited only 1 task. Most participants (15/24, 68%) completed the expressing kindness challenge on 3 consecutive days. One participant completed the challenge in 1 day, and 6 participants (27%) completed the challenge in other ways (eg, spread over a week).

Most participants who visited the expressing kindness challenge attempted the activities. All 20 participants who visited day 1 completed the activity (ie, reaching out to a friend). Of 16 participants who visited day 2, 15 completed the activity (ie, telling a friend what they appreciate about them). All 14 participants who visited day 3 did the activity (ie, make a gratitude list).

The Brain and Stress Module

Twenty-one participants visited the brain and stress module. On average, participants reported that they liked the brain and stress module (mean score 5.52 out of 7, SD 1.25), and that it improved their understanding of the stress response (mean score 4.90 out of 7, SD 1.61). There were no significant differences in how much younger and older adults liked the brain and stress module (mean score 5.71, SD 1.38 vs mean score 5.43, SD 1.22; t_{19}=−0.48; *P*=.63) or how helpful they found the module for improving their understanding of stress (mean score 5.43, SD 1.13 vs mean score 4.64, SD 1.78; t_{19}=−1.06; *P*=.30).

Of 21 participants who visited the brain and stress module, 18 (86%) reported looking at the mental health tips section afterwards to learn about stress management and mental well-being. Additionally, 17 participants (81%) visited Headspace, which is a meditation website that the DH linked to at the end of the brain and stress module [51]. Of these participants, 6 (35%) tried a deep breathing meditation from Headspace. On average, participants reported liking the meditation exercise that they tried (mean score 5.33 out of 7, SD 1.21). There was no significant difference in how much younger and older adults liked the meditation exercise (mean score 5.00, SD 1.41 vs mean score 6.00, SD 0.00; t_4=0.94; *P*=.40). Moreover, 13 participants (77%) agreed that Headspace was a helpful resource for Bella to share.

Other Conversation Modules

Participants visited an average of 9.39 (SD 5.23) other modules beyond the expressing kindness challenge and the brain and stress module (ie, the mental health modules that the researchers asked them to complete in particular). There were no significant differences in how many additional modules younger and older adults visited (mean 8.50, SD 6.37 vs mean 9.87, SD 4.69; t_{21}=0.59; *P*=.56).

Module Visit Behavior

Multimedia Appendix 4 depicts how many participants visited each module. The most popular modules were brain and stress, expressing kindness challenge day 1, move your body, do things that bring joy, watch what you consume, and self-care guide. The least popular module was COVID-19: healthline and resources.

Requests for Conversation Topics

Seventeen participants responded to the question *Were there any particular topics that you would have liked to talk about with Bella which were not available?* Six participants reported no additional topics, and 11 participants described topics pertaining to physical health, mental health, entertainment, New Zealand, and other areas, as outlined in Multimedia Appendix 5. Representative quotes are not presented as participants tended to list topics.

Rapport With the DH

Overall, participants reported a reasonable degree of rapport with Bella (mean score 66.92 out of 100, SD 12.63). There was no significant difference in the amount of rapport reported by younger and older adults (mean 68.13, SD 14.86 vs mean 66.31, SD 12.45; t_{22}=−0.30; *P*=.77).

Loneliness

There was no significant main effect of time ($F_{2,40}$=0.87; *P*=.43; η_p^2=0.04) or condition on perceived loneliness ($F_{1,20}$=0.87; *P*=.36; η_p^2=0.04). There was no significant interaction effect between time and condition on perceived loneliness ($F_{2,40}$=0.01; *P*=.99; η_p^2=0.00).

Stress

There was a significant main effect of condition on perceived stress ($F_{1,20}$=6.58; *P*=.02; η_p^2=0.25). The intervention group reported significantly lower stress overall (mean 2.30, SE 0.77) compared to the waitlist control group (mean 5.09, SE 0.77) (Multimedia Appendix 6). Exploratory pair-wise comparisons revealed that the intervention group reported significantly lower stress compared to the waitlist group at baseline (mean 2.36, SE 0.77 vs mean 5.46, SE 0.77; $F_{1,20}$=8.13; *P*=.01; η_p^2=0.29) and at T2 (mean 2.36, SE 0.82 vs mean 5.09, SE 0.82; $F_{1,20}$=5.47; *P*=.03; η_p^2=0.22). There was no significant main effect of time ($F_{2,40}$=0.35; *P*=.71; η_p^2=0.02) or interaction effect between time and condition on perceived stress ($F_{2,40}$=0.13; *P*=.88; η_p^2=0.01).

COVID-19 Distress

There was no significant main effect of time ($F_{1.47,29.44}$=0.12; *P*=.83; η_p^2=0.01) or condition on COVID-19 distress ($F_{1,20}$=0.03; *P*=.41; η_p^2=0.00). There was no significant interaction effect between time and condition on COVID-19 distress ($F_{1.47,29.44}$=0.83; *P*=.41; η_p^2=0.04).

Positive and Negative Affect

There was no significant main effect of time ($F_{1.44,28.89}$=0.93; *P*=.38; η_p^2=0.04) or condition on the degree of positive affect reported ($F_{1,20}$=0.45; *P*=.51; η_p^2=0.02). There was no significant interaction effect between time and condition on positive affect ($F_{1.44,28.89}$=0.26; *P*=.70; η_p^2=0.01).

There was no significant main effect of time ($F_{2,40}$=1.51; *P*=.23; η_p^2=0.07) or condition on the degree of negative affect reported ($F_{1,20}$=2.50; *P*=.13; η_p^2=0.11). There was no significant interaction effect between time and condition on negative affect ($F_{2,40}$=1.78; *P*=.18; η_p^2=0.08).

There was no significant main effect of time ($F_{1.50,40}$=1.03; *P*=.35; η_p^2=0.05) or condition on the balance of positive and negative affect reported ($F_{1,20}$=1.28; *P*=.27; η_p^2=0.06). There was no significant interaction effect between time and condition on the balance of positive and negative affect ($F_{1.50,40}$=0.89; *P*=.39; η_p^2=0.04).

Psychological Well-Being

There was a trend toward a significant main effect of condition on psychological well-being ($F_{1,20}$=3.44; P=.08; η_p^2=0.15). The intervention group reported greater well-being overall (mean 49.00, SE 1.80) compared to the waitlist group (mean 44.27, SE 1.80) (Multimedia Appendix 7). Exploratory pair-wise comparisons revealed a trend toward the intervention group reporting greater well-being compared to the waitlist group at baseline (mean 49.27, SE 1.99 vs mean 43.91, SE 1.99; $F_{1,20}$=3.64; P=.07; η_p^2=0.15) and at T3 only (mean 49.46, SE 2.03 vs mean 43.82, SE 2.03; $F_{1,20}$=3.84; P=.06; η_p^2=0.16). There was no significant main effect of time ($F_{2,40}$=0.01; P=.99; η_p^2=0.00) or interaction effect between time and condition on psychological well-being ($F_{2,40}$=1.29; P=.29; η_p^2=0.06).

Discussion

Contextualization

Technology has come to play an important role in combatting the COVID-19 pandemic. Artificial intelligence technologies have been rapidly deployed to assist in diagnosing COVID-19 cases and forecasting epidemic development, contact tracing, aiding in drug and vaccine discovery research, and predicting patient outcomes such as disease severity, length of hospital stay, and mortality risk [61,62]. This study proposes that DHs may be an additional technology to aid in health care during the COVID-19 pandemic by providing remote psychological support to people at risk of developing more severe illness. Indeed, other studies have found that digital psychological interventions have been effective during the pandemic (eg, mHealth apps) [63,64].

This study found that a DH was a feasible and acceptable way to deliver a remote loneliness and stress intervention to at-risk older adults living independently and to younger adults with a chronic health condition based on behavioral, qualitative, and some self-report data. The intervention was less feasible for nursing home residents who required caregiver assistance to participate, which may have increased caregiver burden.

Prior to the pandemic, evidence had been building in support of the effectiveness, feasibility, and acceptability of CAs, including embodied agents, at delivering remote psychology interventions and assessments [65,66]. However, their actual adoption in health care settings was low [65], and their efficacy varied depending on the intervention they delivered [66]. Some CAs have technological limitations such as issues with speech recognition, which will be improved as technology advances, but until that stage, these limitations may negatively impact usage intentions [67]. DHs are a new type of CA with an engaging hyperrealistic appearance and neural network-driven behaviors that, prior to this study, had not been evaluated for providing remote loneliness interventions to older adults or adults with chronic health conditions. This study achieved positive results that align with prior CA research showing good acceptability at delivering loneliness interventions to older people [38,68,69], and psychological support for well-being [70] and anxiety in adults with chronic health conditions [39].

A challenge of evaluating the effectiveness of CAs in psychology applications is the large heterogeneity of outcome measures, psychological interventions, and technology features across the literature, which makes comparisons difficult, alongside a shortage of RCTs [48,65,66]. This study was conducted in preparation for a larger RCT to investigate intervention effectiveness. In this pilot RCT, the trial methods were found to be feasible, and they support conducting a future RCT. Exploratory analyses of the psychological variables did not reveal any significant effects. However, this is not unexpected as the pilot trial was not powered to detect any significant group differences in psychological outcomes. Furthermore, it is likely that a 1-week intervention is not long enough to see effects on general loneliness.

Questions remain around how to optimally design CAs for health care applications [66,71]. Some research suggests that greater personalization of CAs (eg, through feedback, daily health reports, and recommendations) may improve acceptability and user engagement [72]. Indeed, some participants from this study reported that they would have liked more personalized responses from Bella. Other research has found that a variety of verbal and nonverbal relational behaviors may contribute to better relationships and usage intention with embodied agents [48,71]. However, there may be interaction effects between relational behaviors, user characteristics, and use context [48]. Participants in this study requested more relational behaviors, such as increased positive emotional expression, addressing the user by name, and incorporating the user's life experience, among others. Incorporating these changes to Bella's design may help to boost her acceptability scores, alongside gradual developments in her underlying technology. Indeed, other research has argued the importance of a co-design process with users and stakeholders to increase acceptability and encourage successful implementation, and future research should adopt this process [73,74].

Strengths and Limitations

This study investigated a novel application of DH technology and adopted a pilot RCT design to inform the methodology of future trials. However, there were several methodological limitations. A sample bias may have occurred whereby participants who volunteered may have been more digitally literate or comfortable with using novel technologies. Moreover, the sample predominantly included Caucasian women; therefore, it is unclear how well the results would generalize to a more diverse population. Even though randomization was conducted, there were significant group differences at baseline in stress, and a larger sample likely would have eliminated these differences. Changes in and out of lockdown conditions in Auckland during the data collection period could have affected the psychological results and degree of engagement in the study. Moreover, there was no control for the psychological follow-up data of our waitlist group, and this should be addressed in a future trial. It is also unclear what the level of engagement with the DH would be outside of a clinical trial context. Research has shown that engagement with eHealth interventions is often lower than what is observed in trials [23,24].

Future Research

The results suggest several directions for future research. A fully powered RCT should investigate the effects of the DH intervention on loneliness and stress. This trial could address the methodological limitations of our study. An active control condition (eg, a chatbot and a website) could be used to provide stronger evidence of effectiveness and reduce the chance that outcome improvements are due to confounding variables (eg, passage of time and researcher attention). The length of follow-up for psychological measures should be extended, along with the length of the intervention. Many loneliness interventions take place over 4 to 6 months with weekly sessions that take an hour or more [16]. The intervention content could be expanded with evidence-based techniques, such as cognitive behavioral and mindfulness exercises to reduce maladaptive social cognition, which have been shown to be the most effective techniques for reducing loneliness in a meta-analysis [16,19,75]. Additionally, the conversation topics that participants requested (eg, physical health support and entertainment) and their feedback should be incorporated to increase acceptability. Other methodological changes for a future RCT could include using other recruitment strategies to achieve a more diverse sample that is more representative of the general population. A future trial could also change the eligibility criteria to require a moderate or high loneliness score. Individuals with higher loneliness at baseline may find the activities more beneficial for well-being and may have more room to improve their loneliness scores. Intervention effectiveness could be investigated separately in younger and older adults. This would allow for tailoring of the intervention content to the age group (with age-appropriate activities, examples, and conversation topics), as well as adapting the DH's design to be more similar to the user population (eg, older adults could interact with an older DH). Moreover, separate trials would allow for more streamlined processes for recruitment and technology training. Lastly, future research could examine DHs in other therapeutic applications and in more diverse patient populations. More research is also needed to discern how DHs in psychology applications should be designed to maximize acceptability and engagement.

Conclusion

Bella, a DH, was found to be a feasible and acceptable way to deliver a remote loneliness intervention to at-risk adults facing social restrictions during the COVID-19 pandemic, based on behavioral, qualitative, and some self-report data. The results support conducting a larger and longer RCT to investigate intervention effectiveness, and indicate that several changes should be made to the technology, intervention content, and trial design. DHs are a novel technological solution that may provide remote psychological support to socially restricted at-risk groups during pandemics. Research should examine the use of DHs in other health care applications with diverse patient populations.

Acknowledgments

We sincerely thank Summerset retirement village and the staff for supporting the research. Orquidea Tamayo Mortera, Steve Williamson, Bobby Gerneke, Claire-Marie Blair, Mike Wharepouri, Jo Clark, Abba Samman, Joseph Lascano, Lianne Gray, and Kate Chastukhina played an integral part in the success of our recruitment of residents and their involvement in the research throughout. We also wish to acknowledge the engineering team at Soul Machines Ltd for their support with developing the digital human. In particular, we would like to thank Erica Lloyd, Hilary O'Connor, Zara Jillings, Keegan Swenson, and Nitin Bhandari. We also wish to thank Roisin Kelly for contributing illustrations for the brain and stress module. We further thank Mikaela Law and Gabrielle Sebaratnam for their assistance with conducting eligibility screens and enrolling participants at Summerset retirement village. Soul Machines Ltd supported the research by building the digital human.

Conflicts of Interest

MS is the chief executive officer of Soul Machines Ltd (a New Zealand artificial intelligence company), which supports KL with a PhD stipend and contracts EB for consultancy work.

Multimedia Appendix 1
The Friendship Questionnaire (Johanson et al, 2020 [<xref ref-type="bibr" rid="20ref53">53</xref>]) adapted to the digital human.
[DOCX File , 15 KB - mental_v8i11e31586_app1.docx]

Multimedia Appendix 2
Participant characteristics at baseline.
[DOCX File , 19 KB - mental_v8i11e31586_app2.docx]

Multimedia Appendix 3
Definitions of themes in response to the qualitative questions.
[DOCX File , 14 KB - mental_v8i11e31586_app3.docx]

Multimedia Appendix 4
Number of participants who visited each module.
[DOCX File , 14 KB - mental_v8i11e31586_app4.docx]

Multimedia Appendix 5
Conversation topics that participants would like to talk about with Bella.
[DOCX File , 13 KB - mental_v8i11e31586_app5.docx]

Multimedia Appendix 6
Perceived stress (mean scores) between groups over the 3 time points.
[DOCX File , 72 KB - mental_v8i11e31586_app6.docx]

Multimedia Appendix 7
Psychological well-being (mean scores) between groups across the 3 time points.
[DOCX File , 73 KB - mental_v8i11e31586_app7.docx]

Multimedia Appendix 8
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 952 KB - mental_v8i11e31586_app8.pdf]

References

1. Piovani D, Christodoulou MN, Hadjidemetriou A, Pantavou K, Zaza P, Bagos PG, et al. Effect of early application of social distancing interventions on COVID-19 mortality over the first pandemic wave: An analysis of longitudinal data from 37 countries. J Infect 2021 Jan;82(1):133-142 [FREE Full text] [doi: 10.1016/j.jinf.2020.11.033] [Medline: 33275956]
2. Considerations for public health and social measures in the workplace in the context of COVID-19. World Health Organization. 2020. URL: https://www.who.int/publications-detail/considerations-for-public-health-and-social-measures-in-the-workplace-in-the-context-of-covid-19 [accessed 2021-06-15]
3. COVID-19 Alert System. New Zealand Government. URL: https://covid19.govt.nz/alert-system/covid-19-alert-system/ [accessed 2021-06-15]
4. People at higher risk of severe illness from COVID-19. New Zealand Government. URL: https://covid19.govt.nz/individuals-and-households/health-and-wellbeing/at-risk-people/ [accessed 2020-08-10]
5. Mills J, Kaye K, Mody L. COVID-19 in older adults: clinical, psychosocial, and public health considerations. JCI Insight 2020 May 21;5(10):e139292 [FREE Full text] [doi: 10.1172/jci.insight.139292] [Medline: 32302293]
6. Jordan RE, Adab P, Cheng KK. Covid-19: risk factors for severe disease and death. BMJ 2020 Mar 26;368:m1198. [doi: 10.1136/bmj.m1198] [Medline: 32217618]
7. Bu F, Steptoe A, Fancourt D. Who is lonely in lockdown? Cross-cohort analyses of predictors of loneliness before and during the COVID-19 pandemic. Public Health 2020 Sep;186:31-34 [FREE Full text] [doi: 10.1016/j.puhe.2020.06.036] [Medline: 32768621]
8. Victor CR, Yang K. The prevalence of loneliness among adults: a case study of the United Kingdom. J Psychol 2012;146(1-2):85-104. [doi: 10.1080/00223980.2011.613875] [Medline: 22303614]
9. Deckx L, van den Akker M, Buntinx F. Risk factors for loneliness in patients with cancer: a systematic literature review and meta-analysis. Eur J Oncol Nurs 2014 Oct;18(5):466-477. [doi: 10.1016/j.ejon.2014.05.002] [Medline: 24993076]
10. Peplau LA, Perlman D. Perspectives on Loneliness. In: Peplau LA, Perlman D, editors. Loneliness: A Sourcebook of Current Theory, Research and Therapy. New York, NY: Wiley; Mar 1984:203.
11. Valtorta NK, Kanaan M, Gilbody S, Ronzi S, Hanratty B. Loneliness and social isolation as risk factors for coronary heart disease and stroke: systematic review and meta-analysis of longitudinal observational studies. Heart 2016 Jul 01;102(13):1009-1016 [FREE Full text] [doi: 10.1136/heartjnl-2015-308790] [Medline: 27091846]
12. Hawkley LC, Cacioppo JT. Loneliness matters: a theoretical and empirical review of consequences and mechanisms. Ann Behav Med 2010 Oct;40(2):218-227 [FREE Full text] [doi: 10.1007/s12160-010-9210-8] [Medline: 20652462]
13. Weiss R. Loneliness: The Experience of Emotional and Social Isolation. Cambridge, MA: MIT Press; 1973:1-236.
14. Hawkley LC, Cacioppo JT. Loneliness and pathways to disease. Brain, Behavior, and Immunity 2003 Feb;17(1):98-105. [doi: 10.1016/s0889-1591(02)00073-9]
15. 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]
16. Masi CM, Chen H, Hawkley LC, Cacioppo JT. A meta-analysis of interventions to reduce loneliness. Pers Soc Psychol Rev 2011 Aug 17;15(3):219-266 [FREE Full text] [doi: 10.1177/1088868310377394] [Medline: 20716644]

17. Poscia A, Stojanovic J, La Milia DI, Duplaga M, Grysztar M, Moscato U, et al. Interventions targeting loneliness and social isolation among the older people: An update systematic review. Exp Gerontol 2018 Feb;102:133-144. [doi: 10.1016/j.exger.2017.11.017] [Medline: 29199121]
18. Chiang K, Chu H, Chang H, Chung M, Chen C, Chiou H, et al. The effects of reminiscence therapy on psychological well-being, depression, and loneliness among the institutionalized aged. Int J Geriatr Psychiatry 2010 Apr;25(4):380-388. [doi: 10.1002/gps.2350] [Medline: 19697299]
19. Creswell JD, Irwin MR, Burklund LJ, Lieberman MD, Arevalo JM, Ma J, et al. Mindfulness-Based Stress Reduction training reduces loneliness and pro-inflammatory gene expression in older adults: a small randomized controlled trial. Brain Behav Immun 2012 Oct;26(7):1095-1101 [FREE Full text] [doi: 10.1016/j.bbi.2012.07.006] [Medline: 22820409]
20. Collins CC, Benedict J. Evaluation of a community-based health promotion program for the elderly: lessons from Seniors CAN. Am J Health Promot 2006 Jan 29;21(1):45-48. [doi: 10.1177/089011710602100108] [Medline: 16977912]
21. Käll A, Jägholm S, Hesser H, Andersson F, Mathaldi A, Norkvist BT, et al. Internet-Based Cognitive Behavior Therapy for Loneliness: A Pilot Randomized Controlled Trial. Behav Ther 2020 Jan;51(1):54-68. [doi: 10.1016/j.beth.2019.05.001] [Medline: 32005340]
22. Jones RB, Ashurst EJ, Atkey J, Duffy B. Older people going online: its value and before-after evaluation of volunteer support. J Med Internet Res 2015 May 18;17(5):e122 [FREE Full text] [doi: 10.2196/jmir.3943] [Medline: 25986724]
23. Rathbone AL, Prescott J. The Use of Mobile Apps and SMS Messaging as Physical and Mental Health Interventions: Systematic Review. J Med Internet Res 2017 Aug 24;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]
24. 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]
25. Heerink M, Kröse B, Evers V, Wielinga B. Influence of Social Presence on Acceptance of an Assistive Social Robot and Screen Agent by Elderly Users. Advanced Robotics 2012 Apr 02;23(14):1909-1923. [doi: 10.1163/016918609x12518783330289]
26. Schuetzler R, Grimes G, Giboney J. An Investigation of Conversational Agent Relevance, Presence, and Engagement. In: AMCIS 2018 Proceedings. 2018 Presented at: AMCIS; 2018; New Orleans, LA.
27. Hegel F, Muhl C, Wrede B, Hielscher-Fastabend M, Sagerer G. Understanding social robots. 2009 Presented at: Second International Conferences on Advances in Computer-Human Interactions; February 1-7, 2009; Cancun, Mexico p. 169-174. [doi: 10.1109/achi.2009.51]
28. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
29. Hung L, Liu C, Woldum E, Au-Yeung A, Berndt A, Wallsworth C, et al. The benefits of and barriers to using a social robot PARO in care settings: a scoping review. BMC Geriatr 2019 Aug 23;19(1):232-210 [FREE Full text] [doi: 10.1186/s12877-019-1244-6] [Medline: 31443636]
30. Gasteiger N, Loveys K, Law M, Broadbent E. Friends from the Future: A Scoping Review of Research into Robots and Computer Agents to Combat Loneliness in Older People. CIA 2021 May;16:941-971. [doi: 10.2147/cia.s282709]
31. Robinson H, Macdonald B, Kerse N, Broadbent E. The psychosocial effects of a companion robot: a randomized controlled trial. J Am Med Dir Assoc 2013 Sep;14(9):661-667. [doi: 10.1016/j.jamda.2013.02.007] [Medline: 23545466]
32. Niemelä M, van Aerschot L, Tammela A, Aaltonen I, Lammi H. Towards Ethical Guidelines of Using Telepresence Robots in Residential Care. Int J of Soc Robotics 2019 Feb 22;13(3):431-439. [doi: 10.1007/s12369-019-00529-8]
33. Barrett E, Burke M, Whelan S, Santorelli A, Oliveira BL, Cavallo F, et al. Evaluation of a Companion Robot for Individuals With Dementia: Quantitative Findings of the MARIO Project in an Irish Residential Care Setting. J Gerontol Nurs 2019 Jul 01;45(7):36-45. [doi: 10.3928/00989134-20190531-01] [Medline: 31237660]
34. Gross H, Scheidig A, Müller S, Schütz B, Fricke C, Meyer S. Living with a Mobile Companion Robot in your Own Apartment - Final Implementation and Results of a 20-Weeks Field Study with 20 Seniors. 2019 Presented at: International Conference on Robotics and Automation (ICRA); May 20-24, 2019; Montreal, QC, Canada p. 2253-2259. [doi: 10.1109/icra.2019.8793693]
35. Jeong K, Sung J, Lee H, Kim A, Kim H, Park C, et al. Fribo: A Social Networking Robot for Increasing Social Connectedness through Sharing Daily Home Activities from Living Noise Data. In: HRI '18: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction. 2018 Presented at: ACM/IEEE International Conference on Human-Robot Interaction; March 5-8, 2018; Chicago, IL p. 114-122. [doi: 10.1145/3171221.3171254]
36. Jeong S, Alghowinem S, Aymerich-Franch L, Arias K, Lapedriza A, Picard R, et al. A Robotic Positive Psychology Coach to Improve College Students' Wellbeing. 2020 Presented at: 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN); August 31-September 4, 2020; Naples, Italy p. 187-194. [doi: 10.1109/ro-man47096.2020.9223588]
37. Bott N, Wexler S, Drury L, Pollak C, Wang V, Scher K, et al. A Protocol-Driven, Bedside Digital Conversational Agent to Support Nurse Teams and Mitigate Risks of Hospitalization in Older Adults: Case Control Pre-Post Study. J Med Internet Res 2019 Oct 17;21(10):e13440 [FREE Full text] [doi: 10.2196/13440] [Medline: 31625949]

38. Ring L, Shi L, Totzke K, Bickmore T. Social support agents for older adults: longitudinal affective computing in the home. J Multimodal User Interfaces 2014 Jun 18;9(1):79-88. [doi: 10.1007/s12193-014-0157-0]
39. Greer S, Ramo D, Chang Y, Fu M, Moskowitz J, Haritatos J. Use of the Chatbot "Vivibot" to Deliver Positive Psychology Skills and Promote Well-Being Among Young People After Cancer Treatment: Randomized Controlled Feasibility Trial. JMIR Mhealth Uhealth 2019 Oct 31;7(10):e15018 [FREE Full text] [doi: 10.2196/15018] [Medline: 31674920]
40. Williams R. Development and Pilot Trial of a Chatbot as a Digital Wellbeing Intervention to Reduce Stress in Tertiary Students. Auckland, New Zealand: University of Auckland; 2020:1-158.
41. Sagar M, Bullivant D, Robertson P, Efimov O, Jawed K, Kalarot R, et al. A neurobehavioural framework for autonomous animation of virtual human faces. In: SA '14: SIGGRAPH Asia 2014 Autonomous Virtual Humans and Social Robot for Telepresence. 2014 Presented at: SIGGRAPH Asia 2014 Autonomous Virtual Humans and Social Robot for Telepresence; December 3-6, 2014; Shenzhen, China. [doi: 10.1145/2668956.2668960]
42. Delivering on the Promise of AI. Soul Machines. 202. URL: https://www.soulmachines.com/resources/whitepapers/download-digital-humans-whitepaper/ [accessed 2021-06-15]
43. Sagar M, Seymour M, Henderson A. Creating connection with autonomous facial animation. Commun. ACM 2016 Dec 01;59(12):82-91. [doi: 10.1145/2950041]
44. Loveys K, Fricchione G, Kolappa K, Sagar M, Broadbent E. Reducing Patient Loneliness With Artificial Agents: Design Insights From Evolutionary Neuropsychiatry. J Med Internet Res 2019 Jul 08;21(7):e13664 [FREE Full text] [doi: 10.2196/13664] [Medline: 31287067]
45. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, PAFS consensus group. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ 2016 Oct 24;355:i5239 [FREE Full text] [doi: 10.1136/bmj.i5239] [Medline: 27777223]
46. Sagar M. Making simulated faces come alive. 2016 Presented at: 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI); March 7-10, 2016; Christchurch, New Zealand. [doi: 10.1109/hri.2016.7451726]
47. Parks MR, Floyd K. Meanings for Closeness and Intimacy in Friendship. Journal of Social and Personal Relationships 2016 Jun 30;13(1):85-107. [doi: 10.1177/0265407596131005]
48. Loveys K, Sebaratnam G, Sagar M, Broadbent E. The Effect of Design Features on Relationship Quality with Embodied Conversational Agents: A Systematic Review. Int J of Soc Robotics 2020 Sep 24;12(6):1293-1312. [doi: 10.1007/s12369-020-00680-7]
49. Ethics Policy. Soul Machines. URL: https://www.soulmachines.com/wp-content/uploads/Ethics_Policy_1.0-1-1.pdf [accessed 2021-06-15]
50. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). European Union. 2016. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679 [accessed 2021-06-15]
51. headspace. URL: https://www.headspace.com [accessed 2021-06-15]
52. Julious SA. Sample size of 12 per group rule of thumb for a pilot study. Pharmaceut. Statist 2005 Oct;4(4):287-291. [doi: 10.1002/pst.185]
53. Johanson D, Ahn H, Sutherland C, Brown B, MacDonald BA, Lim JY, et al. Smiling and use of first-name by a healthcare receptionist robot: Effects on user perceptions, attitudes, and behaviours. Paladyn, Journal of Behavioral Robotics 2020;11(1):40-51. [doi: 10.1515/pjbr-2020-0008]
54. Russell DW. UCLA Loneliness Scale (Version 3): reliability, validity, and factor structure. J Pers Assess 1996 Feb;66(1):20-40. [doi: 10.1207/s15327752jpa6601_2] [Medline: 8576833]
55. Cohen S. Perceived stress in a probability sample of the United States. In: Spacapan S, Oskamp S, editors. The Social Psychology of Health. New York, NY: Sage Publications Inc; 1988:31-67.
56. Lee EH. 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]
57. Jia R, Ayling K, Chalder T, Massey A, Broadbent E, Coupland C, et al. Mental health in the UK during the COVID-19 pandemic: cross-sectional analyses from a community cohort study. BMJ Open 2020 Sep 15;10(9):e040620 [FREE Full text] [doi: 10.1136/bmjopen-2020-040620] [Medline: 32933965]
58. 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]
59. Hone L, Jarden A, Schofield G. Psychometric Properties of the Flourishing Scale in a New Zealand Sample. Soc Indic Res 2013 Nov 6;119(2):1031-1045. [doi: 10.1007/s11205-013-0501-x]
60. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
61. Abd-Alrazaq A, Alajlani M, Alhuwail D, Schneider J, Al-Kuwari S, Shah Z, et al. Artificial Intelligence in the Fight Against COVID-19: Scoping Review. J Med Internet Res 2020 Dec 15;22(12):e20756 [FREE Full text] [doi: 10.2196/20756] [Medline: 33284779]

62. Golinelli D, Boetto E, Carullo G, Nuzzolese AG, Landini MP, Fantini MP. Adoption of Digital Technologies in Health Care During the COVID-19 Pandemic: Systematic Review of Early Scientific Literature. J Med Internet Res 2020 Nov 06;22(11):e22280 [FREE Full text] [doi: 10.2196/22280] [Medline: 33079693]
63. Sun S, Lin D, Goldberg S, Shen Z, Chen P, Qiao S, et al. A mindfulness-based mobile health (mHealth) intervention among psychologically distressed university students in quarantine during the COVID-19 pandemic: A randomized controlled trial. J Couns Psychol 2021 Jul 15:1-47. [doi: 10.1037/cou0000568] [Medline: 34264696]
64. Rauschenberg C, Schick A, Hirjak D, Seidler A, Paetzold I, Apfelbacher C, et al. Evidence Synthesis of Digital Interventions to Mitigate the Negative Impact of the COVID-19 Pandemic on Public Mental Health: Rapid Meta-review. J Med Internet Res 2021 Mar 10;23(3):e23365 [FREE Full text] [doi: 10.2196/23365] [Medline: 33606657]
65. Provoost S, Lau HM, Ruwaard J, Riper H. Embodied Conversational Agents in Clinical Psychology: A Scoping Review. J Med Internet Res 2017 May 09;19(5):e151 [FREE Full text] [doi: 10.2196/jmir.6553] [Medline: 28487267]
66. Chattopadhyay D, Ma T, Sharifi H, Martyn-Nemeth P. Computer-Controlled Virtual Humans in Patient-Facing Systems: Systematic Review and Meta-Analysis. J Med Internet Res 2020 Jul 30;22(7):e18839 [FREE Full text] [doi: 10.2196/18839] [Medline: 32729837]
67. Stara V, Vera B, Bolliger D, Rossi L, Felici E, Di Rosa M, et al. Usability and Acceptance of the Embodied Conversational Agent Anne by People With Dementia and Their Caregivers: Exploratory Study in Home Environment Settings. JMIR Mhealth Uhealth 2021 Jun 25;9(6):e25891 [FREE Full text] [doi: 10.2196/25891] [Medline: 34170256]
68. Chi N, Sparks O, Lin S, Lazar A, Thompson HJ, Demiris G. Pilot testing a digital pet avatar for older adults. Geriatr Nurs 2017;38(6):542-547 [FREE Full text] [doi: 10.1016/j.gerinurse.2017.04.002] [Medline: 28479082]
69. Vardoulakis L, Ring L, Barry B, Sidner C, Bickmore T. Designing Relational Agents as Long Term Social Companions for Older Adults. In: Nakano Y, Neff M, Paiva A, Walker M, editors. Intelligent Virtual Agents. IVA 2012. Lecture Notes in Computer Science, vol 7502. Berlin, Heidelberg: Springer; 2012:289-302.
70. Easton K, Potter S, Bec R, Bennion M, Christensen H, Grindell C, et al. A Virtual Agent to Support Individuals Living With Physical and Mental Comorbidities: Co-Design and Acceptability Testing. J Med Internet Res 2019 May 30;21(5):e12996 [FREE Full text] [doi: 10.2196/12996] [Medline: 31148545]
71. ter Stal S, Kramer LL, Tabak M, op den Akker H, Hermens H. Design Features of Embodied Conversational Agents in eHealth: a Literature Review. International Journal of Human-Computer Studies 2020 Jun;138:102409. [doi: 10.1016/j.ijhcs.2020.102409]
72. Kocaballi AB, Berkovsky S, Quiroz JC, Laranjo L, Tong HL, Rezazadegan D, et al. The Personalization of Conversational Agents in Health Care: Systematic Review. J Med Internet Res 2019 Nov 07;21(11):e15360 [FREE Full text] [doi: 10.2196/15360] [Medline: 31697237]
73. Kramer LL, Ter Stal S, Mulder BC, de Vet E, van Velsen L. Developing Embodied Conversational Agents for Coaching People in a Healthy Lifestyle: Scoping Review. J Med Internet Res 2020 Feb 06;22(2):e14058 [FREE Full text] [doi: 10.2196/14058] [Medline: 32022693]
74. Santini S, Galassi F, Kropf J, Stara V. A Digital Coach Promoting Healthy Aging among Older Adults in Transition to Retirement: Results from a Qualitative Study in Italy. Sustainability 2020 Sep 09;12(18):7400. [doi: 10.3390/su12187400]
75. Cacioppo S, Grippo AJ, London S, Goossens L, Cacioppo JT. Loneliness: clinical import and interventions. Perspect Psychol Sci 2015 Mar;10(2):238-249 [FREE Full text] [doi: 10.1177/1745691615570616] [Medline: 25866548]

Abbreviations

CA: conversational agent
DH: digital human
MMSE: Mini-Mental State Examination
PSS: Perceived Stress Scale
RCT: randomized controlled trial
SPANE: Scale of Positive and Negative Experiences

Edited by J Torous; submitted 26.06.21; peer-reviewed by V Stara, J Andrews; comments to author 25.08.21; revised version received 22.09.21; accepted 24.09.21; published 08.11.21.

Please cite as:
Loveys K, Sagar M, Pickering I, Broadbent E
A Digital Human for Delivering a Remote Loneliness and Stress Intervention to At-Risk Younger and Older Adults During the COVID-19 Pandemic: Randomized Pilot Trial
JMIR Ment Health 2021;8(11):e31586
URL: https://mental.jmir.org/2021/11/e31586
doi:10.2196/31586
PMID:34596572

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org