JMIR Mental Health

Journal Impact Factor (JIF) (2023): 4.8
Volume 9 (2022), Issue 8 ISSN 2368-7959 Editor in Chief: John Torous, MD, MBI

Contents

Original Papers

Reviews

Original Paper

The Effect of Mental Health App Customization on Depressive Symptoms in College Students: Randomized Controlled Trial

Stephanie G Six[1], MSc; Kaileigh A Byrne[1], PhD; Heba Aly[2], MSc; Maggie W Harris[1], BA

[1]Department of Psychology, Clemson University, Clemson, SC, United States

[2]Department of Computer Science, Clemson University, Clemson, SC, United States

Corresponding Author:
Kaileigh A Byrne, PhD
Department of Psychology
Clemson University
418 Brackett Hall
Clemson University
Clemson, SC, 29634-0001
United States
Phone: 1 864 656 3935
Email: kaileib@clemson.edu

Abstract

Background: Mental health apps have shown promise in improving mental health symptoms, including depressive symptoms. However, limited research has been aimed at understanding how specific app features and designs can optimize the therapeutic benefits and adherence to such mental health apps.

Objective: The primary purpose of this study is to investigate the effect of avatar customization on depressive symptoms and adherence to use a novel cognitive behavioral therapy (CBT)–based mental health app. The secondary aim is to examine whether specific app features, including journaling, mood tracking, and reminders, affect the usability of the mental health app.

Methods: College students were recruited from a university study recruitment pool website and via flyer advertisements throughout campus. A total of 94 participants completed a randomized controlled trial in which they were randomized to either customization or no customization version of the app. Customization involved personalizing a virtual avatar and a travel vehicle to one's own preferences and use of one's name throughout the app. Participants completed a 14-day trial using a novel CBT-based mental health app called AirHeart. Self-report scores for depressive symptoms, anxiety, and stress were measured at baseline and after the intervention. Postintervention survey measures also included usability and avatar identification questionnaires.

Results: Of the 94 enrolled participants, 83 (88%) completed the intervention and postintervention assessments. AirHeart app use significantly reduced symptoms of depression (P=.006) from baseline to the end of the 2-week intervention period for all participants, regardless of the customization condition. However, no differences in depressive symptoms (P=.17) or adherence (P=.80) were observed between the customization (39/83, 47%) and no customization (44/83, 53%) conditions. The frequency of journaling, usefulness of mood tracking, and helpfulness of reminders were not associated with changes in depressive symptoms or adherence (P>.05). Exploratory analyses showed that there were 3 moderate positive correlations between avatar identification and depressive symptoms (identification: r=−0.312, P=.02; connection: r=−0.305, P=.02; and lack of relatability: r=0.338, P=.01).

Conclusions: These results indicate that CBT mental health apps, such as AirHeart, have the potential to reduce depressive symptoms over a short intervention period. The randomized controlled trial results demonstrated that customization of app features, such as avatars, does not further reduce depressive symptoms over and above the CBT modules and standard app features, including journal, reminders, and mood tracking. However, further research elucidating the relationship between virtual avatar identification and mental health systems is needed as society becomes increasingly more digitized. These findings have potential implications for improving the optimization of mental health app designs.

Trial Registration: Open Science Framework t28gm; https://osf.io/t28gm

(JMIR Ment Health 2022;9(8):e39516) doi:10.2196/39516

KEYWORDS

depression; mental health apps; customization; personalization; cognitive behavioral therapy; avatars; mobile phone

Introduction

Background

Ranked as the 6th most expensive health condition, depression costs the United States approximately 326.2 billion dollars in treatment and workplace costs in 2018 [1]. Depression is highly prevalent across all age groups, genders, and racial groups, causing anhedonia, irritability, extreme sadness, and other emotional and physical symptoms [2]. This rise in spending parallels the rate of diagnosis of major depressive disorder, which has increased 7-fold over the past 5 years [3]. Certain therapeutic techniques, including cognitive behavioral therapy (CBT), have been implemented to alleviate depressive symptoms.

CBT can be efficacious in reducing depressive symptoms and improving the quality of life in both clinical [4-6] and nonclinical populations [7] of multiple age groups [8,9]. In addition to depression, CBT has been successful in treating symptoms of anxiety and stress [10-15]. Depression is a complex mental health condition that is often comorbid with anxiety and exacerbated by stress [16-18]. Therefore, reducing the symptoms of depression may further aid in alleviating the negative manifestations present in other comorbid or associative disorders.

CBT has been shown to be effective in reducing symptoms of depression not only in in-person therapy environments but also in mobile apps [19,20]. The combination of CBT and mobile technology has burgeoned in the last 10 years, with an estimated 10,000 to 20,000 mental health apps currently existing in the Apple App Store and Google Play Store [21]. Despite its prevalence, there is a dearth of research investigating the interaction between mobile CBT and technological features, such as mobile journaling, reminders, mood tracking, and customization, on mental health symptom reduction. Certain features may complement, augment, or detract from CBT delivery and its effectiveness in symptom reduction. Thus, this study sought to experimentally address this gap in knowledge by investigating how specific app features, including the use of customized avatars, would affect depressive symptoms and adherence.

Computerized CBT (cCBT) is a web-based form of CBT that is accessed through a computer, smartphone, or tablet [22]. Many cCBT mental health apps, such as Space from Depression and MoodGYM, use a time line similar to brief CBT, which is typically 4 to 8 sessions or modules [23-25]. This type of CBT has shown effectiveness in reducing depressive symptoms in both clinical [26,27] and subclinical [28,29] depressive populations of varying ages, although the results from 2 meta-analyses suggest that cCBT mental health apps may be more effective for subclinical than for clinical levels of depression [30,31]. One caveat of this type of therapy is its low adherence rate [32,33]. A proposed explanation for the low adherence rates includes individuals not progressing as quickly as expected, leading to the conclusion that treatment is not effective [34]. In addition, participants also reported quitting if they had negative expectations about their treatment outcomes. Individuals who do not believe that their treatment will be effective often perceive it as inconvenient and quit trying another therapy technique or spending their time elsewhere [35]. Overall, adherence remains a problem in a variety of cCBT programs, but various elements and tools can be used to encourage adherence.

Researchers from different disciplines have presented multiple suggestions to counteract the low adherence rates. One notable suggestion is gamification, which can be defined as the implementation of game elements, such as challenges, rewards, badges, or levels, into a system [36,37]. Gamification has emerged as one of the most widely used solutions for increasing adherence [38,39]. Although it may not provide any additional benefits in reducing depressive symptoms when coupled with therapeutic techniques such as cCBT [40], it has been shown to increase adherence in therapy trials for a variety of mental health disorders, such as depression and anxiety [41,42]. Some specific technological elements, including journaling, mood tracking, and reminders, have been shown to effectively increase engagement *and* aid in the reduction of depressive symptoms in observational studies [43-46]. However, few studies have experimentally investigated the effect of these elements on adherence and mental health symptoms within mental health apps.

One technological feature that has been largely overlooked in mental health app research is customization. In the technology domain, customization is the process of changing a product or interface to make it more personalized to an individual's preferences or needs. Customization permeates through mobile technology, such as the Apple iPhone, which allows the user to create custom alarms, reminders, or ringtones. More specifically, customization within mobile apps can include the creation of a self-representative avatar. An avatar is a virtual representation of a genuine user, where the user can alter various features, such as hairstyle, clothes, skin color, and facial features. Users may have a strong preference for programs that include customizable avatars. For example, a qualitative study conducted focus groups and interviews with adolescents exhibiting depressive symptoms to investigate the usability of a cCBT fantasy game with avatars (SPARX) [47]. This study found that participants enjoyed the option of personalizing their characters, because they could easily relate to the personalized characters [47]. On the basis of prior research, this study used avatar customization to increase app adherence and identification with the avatar.

To the best of our knowledge, only one study has experimentally examined the connection between customization and mobile mental health interventions [48]. Participants were randomized to a condition in which they either created their avatar or were assigned a random avatar that they could not personalize. Baseline anxiety levels were assessed, and participants completed either an attention bias modification training or a no training control activity. Ultimately, the study found that the participants' ability to customize their avatar increased their resilience to the induction of negative moods, their identification with the avatar, their engagement, and the efficacy of therapeutic training [48]. This study was among the first to directly investigate the relationship between customization and anxiety, as well as how avatar customization affects identification, engagement, and efficacy. However, it is unclear whether

customizable avatars can affect *depressive* symptoms. Consequently, this study sought to fill this gap in prior research by investigating the efficacy of avatar customization within a CBT mental health app for depressive symptoms.

This Study and Hypotheses

Overview

Overall, prior research suggests that customization has the potential to influence mental health symptoms, such as anxiety, and increase adherence within a mental health app. When combined with CBT, customization can increase adherence by augmenting self-representation and escapism. The addition of reminders and mood tracking may further increase adherence by promoting app engagement and use of therapeutic tools. The primary aim (aim 1) of this study was to examine the effect of customization on depressive symptoms and adherence to a mental health app, controlling for depression diagnosis. We controlled for depression diagnosis, because prior research has shown that cCBT mental health apps may be more effective for those with subclinical compared with clinical levels of depression [30,31]. The secondary aim (aim 2) was to explore how the perceived usability of the mental health app and its internal features (mood tracking, reminders, and journaling) influenced depressive symptoms and adherence. To examine these aims, this study used a 2-arm randomized controlled trial (1:1 allocation ratio) to compare a mental health app with customization (the intervention) and without customization (the active control) on adherence and depressive symptoms. The specific hypotheses of this study are outlined as follows.

Primary Hypotheses for Aim 1

The primary hypotheses for aim 1 include the following:

- H1: It is hypothesized that the ability to customize a personal avatar will further reduce depressive symptoms compared with the active control.
- H2: It is hypothesized that the level of adherence, measured through the completion of the log-in questionnaires, will be higher in the customization condition than in the active control condition.

Primary Hypotheses for Aim 2

The primary hypotheses for aim 2 include the following:

- H3: It is hypothesized that as the number of journal entries increases, symptoms of depression will decrease, and adherence level will increase.
- H4: It is hypothesized that higher usefulness scores for the reminders will lead to a higher level of adherence and decreased depressive symptoms.
- H5: It is hypothesized that greater use and understanding of mood patterns will lead to a higher level of adherence and decreased depressive symptoms.

Exploratory Aims and Hypotheses

It is possible that individuals with depressive symptoms may differ in their identification with their avatar; however, there is limited prior research aimed at examining this possible relationship. Thus, as an exploratory analysis, we examined whether depressive symptoms were associated with differences in identification with one's avatar. In addition to investigating the effect of CBT-based mental health app features on depression symptoms, this study also examined the effect of customization on anxiety and stress symptoms. Customization is expected to reduce anxiety and stress.

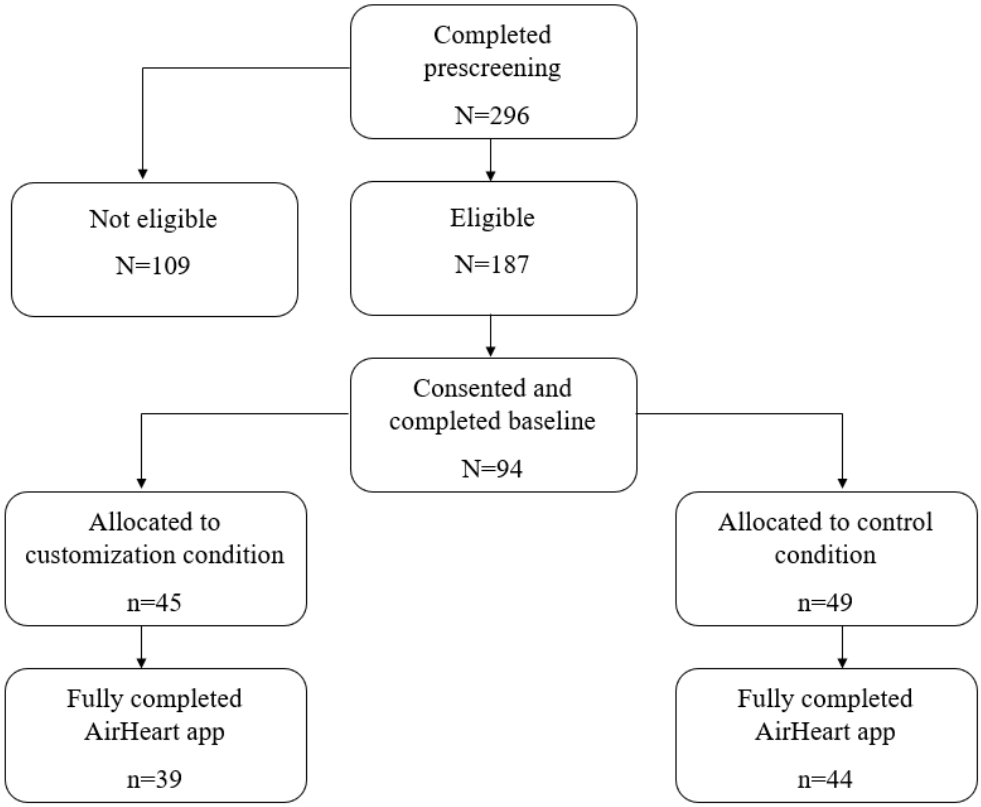
Methods

Participants

The target population for this study was college students enrolled full time. To address H1, an a priori power analysis (*F*-test, repeated measures ANOVA, and within-between interaction) was performed using the G*Power 3.1 (Universität Kiel). The analysis sought to determine the number of participants necessary to maintain a power level of 80% to detect a possible effect at the *P* value of .05 level with 2 groups and 2 measurement time points. A meta-analysis was conducted to determine whether various gamification elements improved the reduction of depressive symptoms in different mental health apps [40]. Cohen *f* for this experiment (*f*=0.16) was calculated from the Hedges *g* (*g*=0.32) provided in the meta-analysis, because both projects investigated mental health apps for depression. According to this analysis, a sample of 80 participants would be needed to have 80% power to detect an effect. Multimedia Appendix 1 presents the log of this power analysis.

To recruit a sample representative of the depressive population, 296 participants were screened before beginning the study. Participants were recruited through SONA Systems software, a cloud-based participant recruitment pool or flyer advertisements. The recruitment flyer and web-based information said only that "beta testers" were needed, and compensation would be provided. Participants received a course or extra credit for their classes, if applicable, and a US $20 Amazon gift card if they fully completed the study. Despite the large prescreening sample size, many individuals (n=109) were ineligible or declined to complete the study (n=94). A total of 94 students from Clemson University completed the prescreening and preassessment, and 88% (83/94) of students completed all 3 mandatory requirements (the prescreening, preassessment, and postassessment questionnaires). Participants were randomly assigned by the computer database to either the customization intervention (39/83, 47%; mean 20.462, SD 2.437) or the active control group (44/83, 53%; mean 20.978, SD 2.633). The recruitment, screening, and study period lasted from February 2022 to May 2022, when the semester was concluded. Figure 1 shows visualization of the participant flow diagram. Multimedia Appendix 2 presents the inclusion and exclusion criteria.

Figure 1. Participant flow diagram.



Materials

AirHeart App

AirHeart is a CBT mental health app designed exclusively for this study to aid in reducing depressive symptoms. This app immerses participants into a world of discovery as they travel in a hot-air balloon to the 7 wonders of the modern world. Each stop along their journey provides new depressive symptom management techniques in the form of 7 cognitive behavioral training modules, which encourage the implementation of new cognitive strategies and offer new healthier behaviors. The scripts for each module are presented in Multimedia Appendix 3. AirHeart includes other features, such as log-in questionnaires, mood tracking, journaling, and reminders. Further details of these features are provided in Multimedia Appendix 4.

Experimental Manipulation: Customization and Personalization

A total of 2 different elements within the intervention group were customized: an avatar and a hot-air balloon. One of the first steps of the AirHeart tutorial, led by both the app and a research assistant introducing the participants to the app, was avatar customization. This customization prompted the participants to create an avatar to embody themselves by tailoring the avatar's skin, eye, hair color, and clothes. The participants also customized their avatar's hot-air balloon, specifically tailoring the color of the balloon. Multimedia Appendix 5 provides customization instructions, avatar examples, and hot-air balloon examples.

In addition to the customization of the avatar and hot-air balloon, the app asked for the participant's name to personalize the journal and mood tracking chart to the user. Use of the participant's name could be observed on the cover of the journal and as a header on the mood tracker page.

Active Control Group: No Customization or Personalization

The control version of AirHeart did not include avatar or hot-air balloon customization. The control avatar was designed as a gray, gender-neutral person with no specific features, and the hot-air balloon was gray colored. This version of AirHeart asked the participant's name but did not use it in the journal or mood tracker. The avatar for the active control condition is presented in Multimedia Appendix 6.

Study Design

This study used a 2 (app condition: customization vs no customization)×2 (time: baseline vs 14-day postintervention period) mixed design, controlling for depression diagnosis (coded as major depression disorder [MDD]: yes vs no). The app condition followed a between-subjects design, time followed a within-subjects design, and depression diagnosis represented a covariate. Participants completed a baseline training and setup session along with baseline questionnaires via a face-to-face assessment; thereafter, participants completed the intervention and postintervention questionnaires on the web.

Measures: Mental Health Symptoms

All questionnaires were based on self-reported data from participants during the past 2 weeks.

Depressive Symptom Questionnaire

The Patient Health Questionnaire (PHQ; PHQ-8) was administered at baseline and after the intervention, and the PHQ-4 was administered as part of the app design. Compared with the PHQ-9, the PHQ-8 omitted questions regarding self-harm and suicidal thoughts. The AirHeart app was not specifically designed to alleviate suicidal ideologies, and it falls out of the scope of this study. The PHQ-8 is an 8-item questionnaire that assesses the frequency of depressive symptoms over the past 2 weeks on a Likert scale, ranging from 0 (*not at all*) to 3 (*nearly every day)*. The PHQ-8 indicates symptoms of minimal depression (score of 0-4) to severe depression (score of 17-24) [49]. The PHQ-4 is a 4-item, shortened version of the PHQ-8.

Anxiety Symptom Questionnaire

The Generalized Anxiety Disorder-7 (GAD-7) is a 7-item questionnaire measuring anxiety over the past 2 weeks, ranging from minimal anxiety (0-4) to severe anxiety (15-21). Participants responded to the 7 questions on a 4-point Likert scale, ranging from 0 (*not at all*) to 3 (*nearly every day*). This questionnaire asks about the frequency of nervousness, worry, relaxation, irritability, restlessness, and fear [50].

Stress Questionnaire

The Perceived Stress Scale (PSS; PSS-10) was administered at baseline and after the intervention, and the shortened PSS-4 was administered as part of the app design. The PSS is a 10-item questionnaire measure of stress over the past month, with scores ranging from 0 to 40, with 0 to 13 indicating low stress and 27 to 40 indicating high stress [51]. This questionnaire, which is a shortened version of the PSS-14, presents scenarios to the user and asks them to rank their responses on a 5-point Likert scale, ranging from 0 (*never*) to 4 (*very often*).

Positive and Negative Affect Scale—Short Form

The Positive and Negative Affect Scale—Short Form is a 20-item questionnaire investigating participants' positive and negative emotions. This questionnaire contains 2 scales, one for positive affect and the other for negative affect, with both scores cumulatively ranging from 20 to 100 or 10 to 50 for each scale. Some examples of the adjectives on the positive affect scale used to describe participants' current feelings include "interested," "strong," or "excited," whereas the negative affect scale includes adjectives such as "distressed," "hostile," or "scared" [52].

Measures: Usability and Adherence

Demographic Questionnaire

Various demographic factors, such as the participant's gender and age, were recorded during both the pre- and postintervention surveys. In addition, during the preintervention period, questions regarding the participants' prior mental health app use and clinical diagnosis were included.

Usability Questionnaire

The System Usability Scale (SUS) is a 10-item questionnaire measuring the usability of a system or product using a 5-item Likert scale, ranging from "strongly agree" to "strongly disagree." The scale ranges from 0 to 100, with any score of ≥68 being average and any score <68 is below average [53,54].

In addition to the SUS, the postassessment questionnaire also collected self-report data regarding efficacy, convenience, and usefulness of the reminders, mood tracker, journal, positive feedback, and storytelling elements and whether participants would continue to use the app if it was available for longer than 14 days.

Adherence Data

Adherence data were collected each day the participants logged in to the app. The number of log-ins along with the days when the participants used the app were compiled.

Procedure

Ethics Approval

Approval was first received from the Clemson University Institutional Review Board (IRB2021-0879). In addition, this study was preregistered on the open science framework [55].

Screening

Before participants began the baseline session using the AirHeart app, they completed the PHQ-8 screening. To be eligible for the study, participants received a score of ≥5, indicating at least mild symptoms of depression. The participants had access to the PHQ-8 via a Qualtrics survey. If a participant received a score of ≤4, they were thanked for their time but were not provided with the passcode to join the full AirHeart study.

Baseline Session

At the initial in-person baseline session, participants provided written informed consent and completed the PHQ-8, GAD-7, PSS-10, demographic survey, and filler questions on campus laboratory computers via the Qualtrics web platform. The informed consent document included privacy statements (eg, participants' email would be retained) and campus mental health hotline information. Participants were then randomly assigned to either the intervention or the active control condition. Participants were granted access to their version of the app on their mobile phones. The participants were blinded to their condition and were thus unaware of the experimental manipulation. Blinding of the researcher to the condition was impossible because the researchers were required to assist participants in the download procedure of the AirHeart app. A physical step-by-step guide was given to the participants, who were instructed to follow along with the guide and were encouraged to ask questions along the way.

To begin, the participants opened the app, which asked them to register for an account, after which they met their avatars. In the intervention condition, participants customized their avatar to look like themselves, and the active control was simply given the gray avatar. Next, participants completed the 16-item log-in questionnaire consisting of the PHQ-4, PSS-4, and 8 positive affect questions from the Positive and Negative Affect Scale—Short Form and their first journal entry. After finishing the prerequisites for the CBT modules, the home page was shown, and participants were encouraged to start exploring their first wonder (and thus the first CBT module). Once the first

CBT module, "wonder," was finished, participants were prompted to set up reminders on their phone. They were then informed that US $20 compensation would be given if they completed the 7 modules, 7 journal entries, 7 log-in questionnaires, and a postintervention survey. Once the participants completed the steps on the guide in the app, they exited the laboratory.

Intervention Period

After the initial session, participants completed 6 more modules over the following 2-week period that were nearly identical to the app experience in the baseline session. Multimedia Appendix 7 provides a detailed outline of the intervention period.

End-of-Intervention Assessment

A survey was emailed to the participants 1 day after the completion of the 2-week intervention period. This survey contained the original 3 questionnaires, the PHQ-8, GAD-7, and PSS-10 as well as a usability scale (SUS) and questions regarding the relationship of the participant to the avatar, identification, and usability elements of the app features.

Data Analysis

To test aim 1, we performed a 2 (intervention type: customization vs no customization control condition)×2 (time: baseline vs postintervention) mixed effects analysis of covariance, controlling for depression diagnosis. This analysis was conducted to identify the effect of customization on depressive symptoms over time (H1). We also conducted a 2-tailed independent sample *t* test for the adherence outcome, the number of log-ins (H2). All analyses were performed using an intent-to-treat approach.

To test aim 2, a correlation analysis was performed between depressive symptom scores and journal entries (H3). Multiple linear regressions were performed to assess the effect of reminder usefulness (H4) and mood tracking understanding and usefulness (H5) on adherence and depressive symptoms.

Results

Participant Characteristics

A total of 296 participants completed the prescreening, and 187 (63.2%) qualified for the full AirHeart study, but only 94 consented and completed the preassessment questionnaire. A total of 83 participants also completed the postassessment questionnaire ($mean_{age}$ 20.771, SD_{age} 2.539 years); 33% (13/39) reported a diagnosis of MDD in the intervention ($mean_{age}$ 20.487, SD_{age} 2.516 years), and 25% (11/44) reported MDD in the control condition ($mean_{age}$ 21.091, SD_{age} 2.701 years). In addition, 28% (11/39) of participants reported prior use of a mental health app in the intervention condition, and 23% (10/44) of the participants reported prior use of a mental health app in the control condition. Further information regarding baseline participant characteristics is presented in Table 1.

Table 1. Baseline participant characteristics overall and by condition.

Variables	Overall sample (N=83)	Customization condition (n=39)	No customization condition (n=44)	Significance level (*P* value)
Age (years), mean (SD)	20.77 (2.54)	20.46 (2.44)	21.05 (2.62)	.30
Gender[a]				.37
Female, n (%)	60 (72)	28 (72)	32 (73)	
Male, n (%)	19 (23)	8 (21)	11 (25)	
Nonbinary, n (%)	4 (5)	3 (7)	1 (2)	
Prior app use (yes), n (%)	21 (25)	11 (28)	10 (23)	.57
Major depression disorder diagnosis (yes), n (%)	24 (29)	13 (33)	11 (25)	.41
Depression scores, mean (SD)	9.39 (4.99)	9.00 (4.82)	9.73 (5.16)	.51
Anxiety scores, mean (SD)	8.60 (4.52)	7.62 (4.05)	9.48 (4.78)	.06
Stress scores, mean (SD)	22.39 (3.91)	22.56 (3.73)	22.23 (4.10)	.70

[a]$F_{1,81}$=0.82.

Descriptive Information for System Usability

The SUS reached an average of 54.349 (SD 18.293; range 25-85) on a scale from 0 to 100, which was below the average of 68. This below average score indicates that the AirHeart app is below the average usability point, suggesting the necessity of an update to make the app potentially less complex, cumbersome, and more intuitive; however, the large SD indicates a wide range of differing opinions.

Aim 1 Analyses

Effect of Customization Versus No Customization on Depressive Symptoms

In accordance with the first hypothesis (H1), a 2 (time: baseline and 14-day postassessment period) ×2 (condition: customization or control) mixed analysis of covariance, controlling for depression diagnosis (MDD: *yes vs no*), was conducted to investigate whether the customization of a virtual avatar would further reduce depressive symptoms over time. A significant

main effect of time was observed ($F_{1,79}$=8.044; P=.006; η_p^2=0.092); however, no other effects demonstrated significant differences: time×condition ($F_{1,79}$=1.965; P=.17; η_p^2=0.024); time×diagnosis ($F_{1,79}$=2.575; P=.11; η_p^2=0.032); time×condition×diagnosis ($F_{1,79}$=1.269; P=.26; η_p^2=0.016); condition×diagnosis ($F_{1,79}$=.026; P=.87; η_p^2<0.001). Thus, the app was effective in reducing depressive symptoms across all participants, but customization had no significant effect. Thus, H1 was not supported.

Effect of Customization Versus No Customization on App Adherence

The independent sample t test results showed no significant difference between the customization and no customization active control conditions on AirHeart app number of log-ins (P=.95). These results do not support H2. The descriptive statistics for the log-in questionnaire, modules, and journal entries are presented in Table 2.

Table 2. Descriptives for log-in questionnaires, journal entries, and modules completed overall and by condition.

Variables	Overall sample (N=83), mean (SD)	Customization condition (n=39), mean (SD)	No customization condition (n=44), mean (SD)	Significance level (P value)
Log-in questionnaire	7.49 (2.42)	0.51 (2.16)	0.48 (2.65)	.95
Journal entries	6.28 (2.45)	6.13 (1.77)	6.41 (2.94)	.61
Completed modules	6.75 (0.84)	6.92 (0.35)	6.59 (1.09)	.07

Aim 2 Analyses

Relationship Between Depressive Symptoms and App Features

Journal entry frequency was not significantly associated with changes in depressive symptoms (postintervention minus baseline depressive symptom levels: r=−0.076; P=.49).

A multiple linear regression with the 6 reminder statements predicting changes in depressive symptoms did not reach statistical significance ($F_{6,68}$=1.305; P=.27), nor were any of the individual 6 statements associated with depressive symptoms (P=.08-.88). Table 3 presents the regression results.

A similar regression with the 4 mood tracking and usefulness questions (Table 4) predicting changes in depressive symptoms also failed to reach significance for the overall model ($F_{4,68}$=0.339; P=.85) and individual items (P=.32-.99). These results failed to support H3 to H5 for the depressive symptom outcome measure.

Table 3. Results of the multiple linear regression between reminder variables and change in depressive symptoms[a].

Reminder variables	ß	t test (df)	Significance level (P value)
The reminders helped me to remember to complete my modules	−.057	−0.382 (68)	.70
The reminders were annoying	−.338	−1.792 (68)	.08
The reminders were inconvenient	.283	1.206 (68)	.23
I would turn off the reminders if I could	.296	1.377 (68)	.17
The reminders helped improve the quality of the app	.200	1.214 (68)	.23
I got excited when I saw the reminders	.021	0.151 (68)	.88

[a]Changes in depressive symptoms were defined as postintervention scores minus baseline depressive symptom scores.

Table 4. Results of the multiple linear regression between mood tracking variables and change in depressive symptoms[a].

Mood tracking variables	ß	t test (df)	Significance level (P value)
The mood tracking helped me understand my pattern of moods	−.159	−0.997 (68)	.32
I liked being able to track my mood and symptoms	.105	0.584 (68)	.56
I did not use the mood tracking	−.002	−0.016 (68)	.99
The mood tracking made me want to use the app	−.067	−0.448 (68)	.66

[a]Changes in depressive symptoms were defined as postintervention scores minus baseline depressive symptom scores.

Relationship Between Adherence and App Features

In terms of adherence, journal entry frequency was positively associated with adherence (r=0.638; P<.001). In addition, the overall model for multiple linear regression with the 6 reminder statements predicting the number of log-in questionnaires did not reach statistical significance ($F_{6,68}$=0.390; P=.88). None of the individual reminder statement predictors were significant (Table 5) in predicting usefulness on their own; therefore, H4 was not supported.

In addition to journaling and reminders, questions regarding the use and understanding of mood trackers were investigated through multiple linear regression. The overall model did not reach significance ($F_{4,68}$=2.115; *P*=.09), nor did any individual factor reach the level of significance. Therefore, H5 is not supported. Table 6 presents the regression results.

Table 5. Results of the multiple linear regression between reminder variables and the number of log-in questionnaires completed.

Reminder variables	ß	*t* test (*df*)	Significance level (*P* value)
The reminders helped me to remember to complete my modules	.097	0.629 (68)	.53
The reminders were annoying	.115	0.587 (68)	.56
The reminders were inconvenient	−.016	−0.065 (68)	.95
I would turn off the reminders if I could	−.049	−0.219 (68)	.83
The reminders helped improve the quality of the app	−.114	−0.660 (68)	.51
I got excited when I saw the reminders	.156	1.089 (68)	.28

Table 6. Results of the multiple linear regression between mood tracking variables and the number of log-in questionnaires completed.

Mood tracking variables	ß	*t* test (*df*)	Significance level (*P* value)
The mood tracking helped me understand my pattern of moods	.230	1.516 (68)	.13
I liked being able to track my mood and symptoms	−.077	−0.455 (68)	.65
I did not use the mood tracking	−.238	−1.814 (68)	.07
The mood tracking made me want to use the app	−.084	−0.588 (68)	.56

Exploratory Analyses

Effect of Customization Versus No Customization on Anxiety and Stress Symptoms

A 2 (time: baseline and 14-day postassessment period)×2 (condition: customization or control) mixed ANOVA was conducted to investigate whether the customization of a virtual avatar would further reduce anxiety symptoms over time. The effects of time (*P*=.08), condition (*P*=.13), and time×condition interaction (*P*=.28) were all nonsignificant. Customization did not influence anxiety symptoms, and the AirHeart app did not significantly reduce anxiety symptoms from the baseline.

Mixed ANOVA results for stress indicated that stress levels significantly declined from baseline to after the intervention ($F_{1,79}$=11.438; *P*=.001; η_p^2=0.126), but customization did not influence stress symptoms (time×condition: *P*=.29). The main effect of the condition (*P*=.21) was nonsignificant.

Relationship Between Depressive Symptoms and Avatar Identification

Bivariate correlations were conducted to investigate the relationship between depressive symptoms and identification with one's avatar. Significant associations were observed between depressive symptom scores and the statements "I identified with my avatar" (*r*=−0.312; *P*=.02), "I felt a connection with my avatar" (*r*=−0.305; *P*=.02), and "my avatar was not like me" (*r*=0.338; *P*=.01); however, no other statement reached significance. The results for the depressive symptoms are presented in Table 7. Multimedia Appendix 8 presents the results for anxiety and stress symptoms.

Table 7. Correlations between Patient Health Questionnaire-8 scores after 14 days and the Avatar Identification Questionnaire.

Avatar Identification Questionnaire	Pearson correlation	Significance level (*P* value)
Identified with avatar	−0.312[a]	.02
Connection with avatar	−0.305[a]	.02
Avatar was not like me	0.338[a]	.01
Avatar is more accomplished	0.174	.20
I like my avatar	−0.169	.21
Avatar made AirHeart more enjoyable	−0.217	.11
Avatar made me want to use AirHeart	−0.244	.07
Avatar helped during modules	−0.189	.16

[a]Correlation is significant at the .05 level (2-tailed).

Discussion

Principal Findings

This study tested the effectiveness of customization within a novel CBT-based mental health app (AirHeart) on depressive symptoms and app adherence. Customization focused on virtual avatar self-representations, vehicle representations, and the use of an individual's name throughout the app. The results indicated that, on average, depressive symptoms decreased over time in all participants. However, customization and personalization of app features did not lead to a further reduction in depressive symptoms (H1) or adherence (H2) compared with the control group. Therefore, although depressive symptoms declined overall, customization and personalization did not exert a significant benefit on symptom reduction or adherence. Instead, the core features of cCBT implemented within an app appear to be independently effective.

The finding that depressive symptoms declined from baseline after the 14-day app intervention period is consistent with randomized controlled trials evaluating cCBT mental health app efficacy [20,23,56-59]. More specifically, this finding aligns with other studies showing that brief CBT-based mental health apps, which consist of a short 4- to 8-module intervention, can mitigate symptoms of depression [23]. However, this study did not include a no app control group. Although the results showed a significant reduction in symptoms from baseline to after the intervention, the primary goal of the study was to compare customization within 2 variants of a CBT-based mental health app rather than to compare the effectiveness of the AirHeart app to a control group.

In addition to depressive symptoms, we explored the effect of AirHeart and AirHeart feature customization on anxiety and stress symptoms. The results indicated that self-reported stress levels decreased from baseline to after the intervention, but there were no significant changes in anxiety symptoms. Similar to the results for depressive symptoms, customization did not significantly affect stress or anxiety levels. The AirHeart cCBT modules targeted strategies for alleviating depressive symptoms but were not designed for anxiety or stress symptoms. Thus, the reduction in stress levels may represent an added benefit of the cCBT mental health app.

The second aim of this study was to examine whether specific app features, including journaling, mood tracking, and reminders, were related to depressive symptoms and adherence. Journal entry frequency was linked to adherence but not to depressive symptoms. Users who journaled more frequently had higher adherence levels. However, given the correlative nature of this result, the directionality of this relationship is inconclusive. It is possible that individuals who used the app more frequently were also incidentally journaling more. Furthermore, in contrast to our hypotheses, the results showed that the likability and usefulness of mood tracking and the helpfulness of reminders did not impact depressive symptoms or adherence. This finding was surprising, given that prior qualitative, mixed methods, and review studies have shown evidence that journaling, mood tracking, and reminder features within mental health apps for depression can effectively increase engagement and aid in the reduction of depressive symptoms [44-46]. We speculate that the nature of CBT module delivery—exploring depressive symptom management strategies by navigating through the wonders of the world—may have been more engaging than the other features that we assessed. Future research is needed to understand the role of these specific app features in mental health symptoms and adherence.

Despite the prevalence of virtual avatars in apps, video games, and virtual meeting platforms, there is an exceptionally limited work characterizing individual difference factors that influence the connection with a self-representative virtual avatar. As an exploratory analysis, we examined the relationship between virtual avatar identification and depressive symptoms. An interesting finding of this study was that individuals with higher levels of depressive symptoms did not identify or connect with their virtual avatars. This finding was observed across both the customization and no customization control conditions. Thus, this negative relationship between depressive symptoms and avatar identification was observed, regardless of the level of customization and personalization within the app. This finding diverges from other research, suggesting that participants would form an attachment and identify with their virtual avatar [48,60-62]. Although this finding may not support our hypothesis, it suggests that individuals with higher depression symptoms may have more difficulty in identifying with virtual self-representative avatars, regardless of aesthetics or similarity to themselves. This finding may be explained by specific symptoms of depression. In particular, depressed individuals often experience increased levels of self-loathing [2], which could reduce their feelings of positive connection with a self-representative avatar. In other words, if an individual does not like themselves, it is reasonable to expect that they would not like a virtual representation of themselves. Alternatively, it is possible that the lack of certain avatar customization options within the AirHeart app that are present within other apps may have reduced identification.

With highly customizable avatars common in popular video games, such as The Sims, Skyrim, and Animal Crossing, as well as on mobile devices, such as Bitmojis, both unconscious requirements and conscious judgments for not providing specific hairstyles, skin color, accessories, or clothing may reduce the identity of one's avatar. Although several studies have examined player expectations for in-game behavior, little research has been documented regarding players' expectations for the game avatar creation system [63,64], and in particular, how mental health conditions, such as depression, may influence avatar use and identification.

Limitations

This study has certain limitations. First, technical difficulties within the AirHeart app impacted the experiences of a few participants. Participants who at times lacked a steady Wi-Fi connection reported having difficulty accessing the log-in questionnaires at the correct time. Although all participants who reached out with difficulties responded within a 24-hour period either by the researcher (SGS) or developer (HA), this could have impacted the usability and effectiveness of the AirHeart

app. Furthermore, this limitation may have affected the SUS scores reported by the participants.

In addition, owing to a technical error with the Qualtrics website along with user error, one questionnaire regarding the participants' opinions about the features within the AirHeart app was not saved to the Qualtrics survey; thus, participants did not originally complete this questionnaire in their postassessment period. Although a total of 16 participants completed the questionnaire as a separate survey, there was a 19% (16/83) participant loss for this questionnaire due to this error. This loss of data impacted H3 and H4, which stated that the number of journal entries and usefulness of reminders would both be associated with heightened adherence and diminished depressive symptoms.

Finally, customization options were limited to skin, eye, hair color, and clothing choice. Participants were not able to customize their facial expressions, facial features (eg, facial hair and piercings), or personal features such as tattoos. In addition, customization lacked some level of religious and accessibility inclusivity, including the lack of wheelchair, hijab, or yamaka options. Individuals may feel that such features are a strong reflection of their identity, and without such options, they may feel detached from their avatar.

Future Directions

The AirHeart mental health app provides a foundation for future mental health app development and informs work on avatar development. Future studies could replicate this design with improved graphics and more customization options for both the avatar and hot-air balloons. Animation could be added as a way of increasing the potential connectivity with one's avatar. More research is needed on the overall connection with human avatars, specifically on whether people identify more with avatars that look like them or potentially look different, whether through age, gender, or even species. Future studies are needed to characterize the guidelines and requirements for high-level avatar identification. In addition, ways to increase such avatar identification for individuals with mental health conditions, such as depression, are needed. Finally, future work is needed to identify whether these findings can be generalized to a broader, nonstudent population.

Conclusions

Numerous mental health apps use avatars, customization, and gamification; however, the therapeutic benefits of these features appear limited [40]. The findings of this study may provide insights into the usability and functionality of certain mental health app features to further refine app development. This study provides empirical evidence that customization features within cCBT mental health do not provide significant benefits for symptom reduction or adherence over and above the CBT intervention itself. Identifying specific app features that improve app effectiveness is required to optimize the design of mental health apps and improve users' mental health symptoms.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Log of the study power analysis.
[DOCX File , 50 KB - mental_v9i8e39516_app1.docx]

Multimedia Appendix 2
Inclusion and exclusion criteria.
[DOCX File , 13 KB - mental_v9i8e39516_app2.docx]

Multimedia Appendix 3
AirHeart app cognitive behavioral therapy module scripts.
[DOCX File , 237 KB - mental_v9i8e39516_app3.docx]

Multimedia Appendix 4
AirHeart app features.
[DOCX File , 636 KB - mental_v9i8e39516_app4.docx]

Multimedia Appendix 5
Customization instructions, customized avatar examples, and customized hot-air balloon examples.
[DOCX File , 230 KB - mental_v9i8e39516_app5.docx]

Multimedia Appendix 6
Avatar for the no customization active control condition.
[DOCX File , 145 KB - mental_v9i8e39516_app6.docx]

Multimedia Appendix 7
Detailed procedure of the 14-day app intervention.
[DOCX File , 13 KB - mental_v9i8e39516_app7.docx]

Multimedia Appendix 8
Results of the association between anxiety, stress, and avatar identification.
[DOCX File , 14 KB - mental_v9i8e39516_app8.docx]

Multimedia Appendix 9
CONSORT-eHEALTH checklist (version 1.6.1).
[PDF File (Adobe PDF File), 17974 KB - mental_v9i8e39516_app9.pdf]

References

1. Greenberg PE, Fournier AA, Sisitsky T, Simes M, Berman R, Koenigsberg SH, et al. The economic burden of adults with Major Depressive Disorder in the United States (2010 and 2018). Pharmacoeconomics 2021 Jun;39(6):653-665 [FREE Full text] [doi: 10.1007/s40273-021-01019-4] [Medline: 33950419]
2. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5). 5th edition. Washington, DC, USA: American Psychiatric Association; 2013.
3. Bueno-Notivol J, Gracia-García P, Olaya B, Lasheras I, López-Antón R, Santabárbara J. Prevalence of depression during the COVID-19 outbreak: a meta-analysis of community-based studies. Int J Clin Health Psychol 2021;21(1):100196 [FREE Full text] [doi: 10.1016/j.ijchp.2020.07.007] [Medline: 32904715]
4. Chaves C, Lopez-Gomez I, Hervas G, Vazquez C. A comparative study on the efficacy of a positive psychology intervention and a cognitive behavioral therapy for clinical depression. Cogn Ther Res 2017;41(3):417-433. [doi: 10.1007/s10608-016-9778-9]
5. Dobkin RD, Menza M, Allen LA, Gara MA, Mark MH, Tiu J, et al. Cognitive-behavioral therapy for depression in Parkinson's disease: a randomized, controlled trial. Am J Psychiatry 2011 Oct;168(10):1066-1074 [FREE Full text] [doi: 10.1176/appi.ajp.2011.10111669] [Medline: 21676990]
6. Grosse Holtforth M, Krieger T, Zimmermann J, Altenstein-Yamanaka D, Dörig N, Meisch L, et al. A randomized-controlled trial of cognitive-behavioral therapy for depression with integrated techniques from emotion-focused and exposure therapies. Psychother Res 2019 Jan;29(1):30-44. [doi: 10.1080/10503307.2017.1397796] [Medline: 29130400]
7. Tandon SD, Leis JA, Mendelson T, Perry DF, Kemp K. Six-month outcomes from a randomized controlled trial to prevent perinatal depression in low-income home visiting clients. Matern Child Health J 2014 May;18(4):873-881 [FREE Full text] [doi: 10.1007/s10995-013-1313-y] [Medline: 23793487]
8. Gould RL, Coulson MC, Howard RJ. Cognitive behavioral therapy for depression in older people: a meta-analysis and meta-regression of randomized controlled trials. J Am Geriatr Soc 2012 Oct;60(10):1817-1830. [doi: 10.1111/j.1532-5415.2012.04166.x] [Medline: 23003115]
9. Oud M, de Winter L, Vermeulen-Smit E, Bodden D, Nauta M, Stone L, et al. Effectiveness of CBT for children and adolescents with depression: a systematic review and meta-regression analysis. Eur Psychiatry 2019 Apr;57:33-45. [doi: 10.1016/j.eurpsy.2018.12.008] [Medline: 30658278]
10. Chiu AW, Langer DA, McLeod BD, Har K, Drahota A, Galla BM, et al. Effectiveness of modular CBT for child anxiety in elementary schools. Sch Psychol Q 2013 Jun;28(2):141-153 [FREE Full text] [doi: 10.1037/spq0000017] [Medline: 23750860]
11. Klein B, Mitchell J, Gilson K, Shandley K, Austin D, Kiropoulos L, et al. A therapist-assisted Internet-based CBT intervention for posttraumatic stress disorder: preliminary results. Cogn Behav Ther 2009;38(2):121-131. [doi: 10.1080/16506070902803483] [Medline: 20183691]
12. Querstret D, Cropley M, Kruger P, Heron R. Assessing the effect of a Cognitive Behaviour Therapy (CBT)-based workshop on work-related rumination, fatigue, and sleep. Eur J Work Organ Psychol 2016;25(1):50-67. [doi: 10.1080/1359432x.2015.1015516]
13. Simon N, McGillivray L, Roberts NP, Barawi K, Lewis CE, Bisson JI. Acceptability of Internet-based cognitive behavioural therapy (i-CBT) for post-traumatic stress disorder (PTSD): a systematic review. Eur J Psychotraumatol 2019 Aug 27;10(1):1646092 [FREE Full text] [doi: 10.1080/20008198.2019.1646092] [Medline: 31497259]
14. Sportel BE, de Hullu E, de Jong PJ, Nauta MH. Cognitive bias modification versus CBT in reducing adolescent social anxiety: a randomized controlled trial. PLoS One 2013 May 14;8(5):e64355 [FREE Full text] [doi: 10.1371/journal.pone.0064355] [Medline: 23691203]
15. Watkins E, Scott J, Wingrove J, Rimes K, Bathurst N, Steiner H, et al. Rumination-focused cognitive behaviour therapy for residual depression: a case series. Behav Res Ther 2007 Sep;45(9):2144-2154. [doi: 10.1016/j.brat.2006.09.018] [Medline: 17367751]

16. Abramson LY, Metalsky GI, Alloy LB. Hopelessness depression: a theory-based subtype of depression. Psychol Rev 1989;96(2):358-372. [doi: 10.1037/0033-295x.96.2.358]
17. Sartorius N, Ustün TB, Lecrubier Y, Wittchen HU. Depression comorbid with anxiety: results from the WHO study on psychological disorders in primary health care. Br J Psychiatry Suppl 1996 Jun(30):38-43. [Medline: 8864147]
18. Zetsche U, D'Avanzato C, Joormann J. Depression and rumination: relation to components of inhibition. Cogn Emot 2012;26(4):758-767. [doi: 10.1080/02699931.2011.613919] [Medline: 21970297]
19. 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]
20. 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]
21. Clay RA. Mental health apps are gaining traction. Monitor on Psychology. American Psychological Association. 2021 Jan 1. URL: https://www.apa.org/monitor/2021/01/trends-mental-health-apps [accessed 2021-11-05]
22. Donovan CL, Poole C, Boyes N, Redgate J, March S. Australian mental health worker attitudes towards cCBT: what is the role of knowledge? Are there differences? Can we change them? Internet Interv 2015 Nov;2(4):372-381. [doi: 10.1016/j.invent.2015.09.001]
23. 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]
24. Richards D, Murphy T, Viganó N, Timulak L, Doherty G, Sharry J, et al. Acceptability, satisfaction and perceived efficacy of " Space from Depression" an Internet-delivered treatment for depression. Internet Interv 2016 Sep;5:12-22 [FREE Full text] [doi: 10.1016/j.invent.2016.06.007] [Medline: 30135802]
25. A Provider's Guide to Brief CBT | South Central MIRECC. US Department of Veterans Affairs. 2013. URL: https://www.mirecc.va.gov/visn16/guide-to-brief-cbt-manual.asp [accessed 2021-10-20]
26. Cavanagh K, Shapiro DA, Van Den Berg S, Swain S, Barkham M, Proudfoot J. The effectiveness of computerized cognitive behavioural therapy in routine care. Br J Clin Psychol 2006 Nov;45(Pt 4):499-514. [doi: 10.1348/014466505X84782] [Medline: 17076960]
27. Smith P, Scott R, Eshkevari E, Jatta F, Leigh E, Harris V, et al. Computerised CBT for depressed adolescents: randomised controlled trial. Behav Res Ther 2015 Oct;73:104-110. [doi: 10.1016/j.brat.2015.07.009] [Medline: 26301756]
28. 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]
29. Spek V, Nyklícek I, Smits N, Cuijpers P, Riper H, Keyzer J, et al. Internet-based cognitive behavioural therapy for subthreshold depression in people over 50 years old: a randomized controlled clinical trial. Psychol Med 2007 Dec;37(12):1797-1806. [doi: 10.1017/S0033291707000542] [Medline: 17466110]
30. 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 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
31. Heber E, Ebert DD, Lehr D, Cuijpers P, Berking M, Nobis S, et al. The benefit of Web- and computer-based interventions for stress: a systematic review and meta-analysis. J Med Internet Res 2017 Feb 17;19(2):e32 [FREE Full text] [doi: 10.2196/jmir.5774] [Medline: 28213341]
32. Kaltenthaler E, Sutcliffe P, Parry G, Beverley C, Rees A, Ferriter M. The acceptability to patients of computerized cognitive behaviour therapy for depression: a systematic review. Psychol Med 2008 Nov;38(11):1521-1530. [doi: 10.1017/S0033291707002607] [Medline: 18205964]
33. So M, Yamaguchi S, Hashimoto S, Sado M, Furukawa TA, McCrone P. Is computerised CBT really helpful for adult depression?-A meta-analytic re-evaluation of CCBT for adult depression in terms of clinical implementation and methodological validity. BMC Psychiatry 2013 Apr 15;13:113 [FREE Full text] [doi: 10.1186/1471-244X-13-113] [Medline: 23587347]
34. Schindler A, Hiller W, Witthöft M. What predicts outcome, response, and drop-out in CBT of depressive adults? A naturalistic study. Behav Cogn Psychother 2013 May;41(3):365-370. [doi: 10.1017/S1352465812001063] [Medline: 23211066]
35. Rost T, Stein J, Löbner M, Kersting A, Luck-Sikorski C, Riedel-Heller SG. User acceptance of computerized cognitive behavioral therapy for depression: systematic review. J Med Internet Res 2017 Sep 13;19(9):e309 [FREE Full text] [doi: 10.2196/jmir.7662] [Medline: 28903893]
36. Hamari J, Koivisto J, Sarsa H. Does gamification work? -- A literature review of empirical studies on gamification. In: Proceedings of the 47th Hawaii International Conference on System Sciences. 2014 Presented at: HICSS '14; January 6-9, 2014; Waikoloa, HI, USA p. 3025-3034. [doi: 10.1109/hicss.2014.377]
37. Seaborn K, Fels DI. Gamification in theory and action: a survey. Int J Human Comput Stud 2015 Feb;74:14-31. [doi: 10.1016/j.ijhcs.2014.09.006]

38. 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]
39. Abdul Rahim MI, Thomas RH. Gamification of medication adherence in epilepsy. Seizure 2017 Nov;52:11-14 [FREE Full text] [doi: 10.1016/j.seizure.2017.09.008] [Medline: 28934624]
40. 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 Nov 29;8(11):e32199 [FREE Full text] [doi: 10.2196/32199] [Medline: 34847058]
41. Dias LP, Barbosa JL, Vianna HD. Gamification and serious games in depression care: a systematic mapping study. Telemat Inform 2018 Apr;35(1):213-224. [doi: 10.1016/j.tele.2017.11.002]
42. 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 Sep 2;15(9):e0237220 [FREE Full text] [doi: 10.1371/journal.pone.0237220] [Medline: 32877425]
43. Asbury ET, Casey J, Desai K. Family eJournal: benefits of online guided group journaling for women. J Public Mental Health 2018 Sep 17;17(3):135-141. [doi: 10.1108/jpmh-01-2018-0008]
44. Grist R, Porter J, Stallard P. Acceptability, use, and safety of a mobile phone app (BlueIce) for young people who self-harm: qualitative study of service users' experience. JMIR Ment Health 2018 Mar 23;5(1):e16 [FREE Full text] [doi: 10.2196/mental.8779] [Medline: 29475823]
45. Stawarz K, Preist C, Tallon D, Wiles N, Coyle D. User experience of cognitive behavioral therapy apps for depression: an analysis of app functionality and user reviews. J Med Internet Res 2018 Jun 06;20(6):e10120 [FREE Full text] [doi: 10.2196/10120] [Medline: 29875087]
46. Zhang R, Nicholas J, Knapp AA, Graham AK, Gray E, Kwasny MJ, et al. Clinically meaningful use of mental health apps and its effects on depression: mixed methods study. J Med Internet Res 2019 Dec 20;21(12):e15644 [FREE Full text] [doi: 10.2196/15644] [Medline: 31859682]
47. Cheek C, Bridgman H, Fleming T, Cummings E, Ellis L, Lucassen MF, et al. Views of young people in rural Australia on SPARX, a fantasy world developed for New Zealand youth with depression. JMIR Serious Games 2014 Mar 18;2(1):e3 [FREE Full text] [doi: 10.2196/games.3183] [Medline: 25659116]
48. Birk MV, Mandryk RL. Improving the efficacy of cognitive training for digital mental health interventions through avatar customization: crowdsourced quasi-experimental study. J Med Internet Res 2019 Jan 08;21(1):e10133 [FREE Full text] [doi: 10.2196/10133] [Medline: 30622095]
49. 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]
50. Spitzer RL, Kroenke K, Williams JB, 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]
51. Reis RS, Hino AA, Añez CR. Perceived stress scale: reliability and validity study in Brazil. J Health Psychol 2010 Jan;15(1):107-114. [doi: 10.1177/1359105309346343] [Medline: 20064889]
52. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988 Jun;54(6):1063-1070. [doi: 10.1037//0022-3514.54.6.1063] [Medline: 3397865]
53. Bangor A, Kortum PT, Miller JT. An empirical evaluation of the system usability scale. Int J Human Comput Interact 2008 Jul 30;24(6):574-594. [doi: 10.1080/10447310802205776]
54. Lewis JR. The system usability scale: past, present, and future. Int J Human Comput Interact 2018 Mar 30;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
55. Six S, Byrne KA. Investigating the efficacy of mental health app features on depressive symptoms and adherence. Open Source Framework. Charlottesville, VA, USA: Center for Open Science; 2021 May 27. URL: https://osf.io/7wv9s/ [accessed 2022-04-26]
56. 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]
57. 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]
58. 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]
59. 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]
60. Birk MV, Mandryk RL, Atkins C. The motivational push of games: the interplay of intrinsic motivation and external rewards in games for training. In: Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play. 2016 Presented at: CHI PLAY '16; October 16-19, 2016; Austin, TX, USA p. 291-303. [doi: 10.1145/2967934.2968091]

61. Li D, Liau A, Khoo A. Examining the influence of actual-ideal self-discrepancies, depression, and escapism, on pathological gaming among massively multiplayer online adolescent gamers. Cyberpsychol Behav Soc Netw 2011 Sep;14(9):535-539. [doi: 10.1089/cyber.2010.0463] [Medline: 21332374]
62. Yee N, Bailenson J. The Proteus effect: the effect of transformed self-representation on behavior. Human Comm Res 2007 Jul;33(3):271-290. [doi: 10.1111/j.1468-2958.2007.00299.x]
63. Lin H, Wang H. Avatar creation in virtual worlds: behaviors and motivations. Comput Human Behav 2014 May;34:213-218. [doi: 10.1016/j.chb.2013.10.005]
64. Triberti S, Durosini I, Aschieri F, Villani D, Riva G. Changing avatars, changing selves? The influence of social and contextual expectations on digital rendition of identity. Cyberpsychol Behav Soc Netw 2017 Aug;20(8):501-507. [doi: 10.1089/cyber.2016.0424] [Medline: 28806125]

Abbreviations

CBT: cognitive behavioral therapy
cCBT: computerized cognitive behavioral therapy
GAD-7: Generalized Anxiety Disorder-7
MDD: major depression disorder
PHQ: Patient Health Questionnaire
PSS: Perceived Stress Scale
SUS: System Usability Scale

Edited by J Torous, G Eysenbach; submitted 06.06.22; peer-reviewed by J Frommel, S Chen; comments to author 27.06.22; revised version received 05.07.22; accepted 18.07.22; published 09.08.22.

Please cite as:
Six SG, Byrne KA, Aly H, Harris MW
The Effect of Mental Health App Customization on Depressive Symptoms in College Students: Randomized Controlled Trial
JMIR Ment Health 2022;9(8):e39516
URL: https://mental.jmir.org/2022/8/e39516
doi:10.2196/39516
PMID:35943788

Original Paper

Using Voice Biomarkers to Classify Suicide Risk in Adult Telehealth Callers: Retrospective Observational Study

Ravi Iyer[1], BSc, MSc; Maja Nedeljkovic[1], PhD; Denny Meyer[1], PhD

Centre for Mental Health, Swinburne University of Technology, Hawthorn, Australia

Corresponding Author:
Ravi Iyer, BSc, MSc
Centre for Mental Health
Swinburne University of Technology
34 Wakefield Street
Hawthorn, 3122
Australia
Phone: 61 454565575
Email: raviiyer@swin.edu.au

Abstract

Background: Artificial intelligence has the potential to innovate current practices used to detect the imminent risk of suicide and to address shortcomings in traditional assessment methods.

Objective: In this paper, we sought to automatically classify short segments (40 milliseconds) of speech according to low versus imminent risk of suicide in a large number (n=281) of telephone calls made to 2 telehealth counselling services in Australia.

Methods: A total of 281 help line telephone call recordings sourced from On The Line, Australia (n=266, 94.7%) and 000 Emergency services, Canberra (n=15, 5.3%) were included in this study. Imminent risk of suicide was coded for when callers affirmed intent, plan, and the availability of means; level of risk was assessed by the responding counsellor and reassessed by a team of clinical researchers using the Columbia Suicide Severity Rating Scale (=5/6). Low risk of suicide was coded for in an absence of intent, plan, and means and via Columbia suicide Severity Scale Ratings (=1/2). Preprocessing involved normalization and pre-emphasis of voice signals, while voice biometrics were extracted using the statistical language r. Candidate predictors were identified using Lasso regression. Each voice biomarker was assessed as a predictor of suicide risk using a generalized additive mixed effects model with splines to account for nonlinearity. Finally, a component-wise gradient boosting model was used to classify each call recording based on precoded suicide risk ratings.

Results: A total of 77 imminent-risk calls were compared with 204 low-risk calls. Moreover, 36 voice biomarkers were extracted from each speech frame. Caller sex was a significant moderating factor (β=–.84, 95% CI –0.85, –0.84; t=6.59, P<.001). Candidate biomarkers were reduced to 11 primary markers, with distinct models developed for men and women. Using leave-one-out cross-validation, ensuring that the speech frames of no single caller featured in both training and test data sets simultaneously, an area under the precision or recall curve of 0.985 was achieved (95% CI 0.97, 1.0). The gamboost classification model correctly classified 469,332/470,032 (99.85%) speech frames.

Conclusions: This study demonstrates an objective, efficient, and economical assessment of imminent suicide risk in an ecologically valid setting with potential applications to real-time assessment and response.

Trial Registration: Australian New Zealand Clinical Trials Registry ACTRN12622000486729; https://www.anzctr.org.au/ACTRN12622000486729.aspx

(JMIR Ment Health 2022;9(8):e39807) doi:10.2196/39807

KEYWORDS

voice biometrics; suicide prevention; machine learning; telehealth; suicide; telehealth; risk prediction; prediction model; voice biomarker; mental health

Introduction

Suicide remains the 4th leading cause of death among 15- to 45-year-olds internationally [1]. However, traditional risk factor–based assessment has failed to identify suicide risk in a targeted and timely manner [2]. There has been historically a poor understanding of which risk factors contribute most to identifying an escalation in suicide risk [2]. This has led to calls for alternative approaches to evaluation, coupled with more powerful means of analysis [3].

Suicide risk assessment using voice biomarkers holds significant promise. Several candidate voice biomarkers have been identified that discriminate accurately between low and high risk of suicide, including timing and prosody–based features [4]. When combined with high-powered forms of statistical analysis (eg, machine learning), voice biomarkers offer an objective, unobtrusive, and economically feasible approach for this purpose.

In a promising study by Pestian and colleagues [5], an accuracy of 85% was obtained using a range of voice biomarkers that classified 379 calls according to the 3 categories of high risk of suicide, mentally ill without prior history of suicidal ideation, and healthy controls. However, the classification was obtained using support vector machines, a powerful machine learning approach that can analyze nonlinear data, but for which post hoc interpretability is unavailable [6]. Thus, with a support vector machine, it is difficult to understand which voice biomarkers are important and which are not. These important considerations will be addressed in our new study.

Souririajan and colleagues [7] replicated the Pestian and colleagues [5] study with 94 US veterans, meeting the criteria for Gulf War syndrome measured at months 0, 1, 2, 3, 6, and 12 after recruitment. A range of voice biomarkers provided only modest discrimination between low and high risk of suicide (area under the receiver operating characteristic curve=0.64) [7]. However, male Gulf War veterans, who formed the majority of the sample (80.0%), are at lower risk of suicide than the general population [8], and reliance upon item 9 of the Patient Health Questionnaire alone (“Thoughts that you’d be better off dead, or thoughts of hurting yourself in some way?”) is associated with higher rates for false positives when compared with the more comprehensive Columbia Suicide Severity Rating Scale used by Pestian and colleagues [5].

Attempts have been made by a number of international jurisdictions to codify a hierarchy of patient suicide risk and appropriate response. Following the UK guidelines, Victoria, Australia has developed the Statewide Mental Health Triage Scale [9]. Seven levels of risk are defined, with “current actions endangering self” afforded the highest level of risk, followed by very high risk of imminent harm, high risk, moderate risk, low risk, referral required, and advice or information provision at the lowest level of risk. The second highest category, very high risk, specifies acute suicidal ideation accompanied by clear plan and means. Neither Pestian and colleagues [5] nor Sourirajan [7] clearly indicated which level of risk was being targeted in their studies. In this new study, we are targeting low risk and below compared with very high risk of imminent harm.

Furthermore, the studies by Pestian and colleagues [5] and Sourirajan [7] lack translation into real-world settings. Based on their studies, high levels of accuracy seem plausible only when participants are recruited from inpatient services [10,11], interviewed under lab conditions, and risk of suicide is assumed to remain static over time. To extend the generalizability of these findings, participants need to be recruited from ecologically valid settings and assessed when elevation in suicide risk occurs.

The help line services we partnered with in this study represent ecologically valid settings. Help line services have played an important role in early detection and response to suicide risk in the community since the early 1950s [12]. In recent years, help line services have witnessed a significant increase in the volume of suicide-related presentations resulting from the COVID-19 pandemic [12,13]. Help line services support important avenues of suicide detection and prevention by providing equitability of access, promotion of disclosures and trust, and supplementation of traditional forms of health care [14]. However, suicide assessment via telehealth is challenged by the absence of nonverbal cues, by time limitations, and by the reticence of some callers to verbally express suicidal intent [15].

Where there is reasonable suspicion that the caller may have taken actions to endanger themselves, emergency management protocols can be triggered. This typically involves dispatch of police and ambulance to perform a welfare check. However, scarce emergency resources can also be dispatched when the caller is not at imminent risk, thus potentially diverting life-saving services from other emergencies. Alternatively, there is the threat of a serious risk of harm when imminent risk of suicide is not detected and therefore not responded to. These unfortunate high-stake scenarios can result in high-pressure work environments that can adversely affect service providers and the individual affected, making it critically important that assessments of imminent suicide risk are as close to 100% correct as possible.

Artificial intelligence has the potential to detect risk of suicide in an accurate, efficient, and timely manner. Although there is initial evidence for the efficacy of such an approach [4], current evidence lacks application to real-world ecologies and real-time assessment, both of which are essential if these insights are to move beyond the lab. Thus, we aimed to use artificial intelligence approaches to automatically classify in real time a large sample of telephone counselling calls made to Australian suicide-prevention help line services using voice biomarkers. By classifying counselling help line calls to a very high level of accuracy, we aim to demonstrate a viable support to existing help line infrastructure that can be employed in real time.

Methods

Multimedia Appendix 1 illustrates the analysis workflow.

Call Recordings

A total of 532 telephone call recordings were sourced for this retrospective observational study. Of these, 77 (14.5%) featured imminent risk of suicide, while 204 (38.3%) featured low risk of suicide. Participants were callers of Suicide Call-Back Service

(a national help line service coordinated by On The Line, Australia) and 000 Emergency Services, Canberra wishing to discuss themes relevant to suicide risk. On The Line call recordings (n=517, 97.2%) were randomly sampled from July 1, 2019, to June 30, 2021, stratified by organizationally determined suicide risk level and disclosed sex of caller. In the case of 000 calls (n=15, 2.8%), call recordings were randomly sampled over the same time to reflect callers exhibiting imminent risk of suicide, necessitating emergency services' response. Moderate-risk calls (236/517, 45.6%; Columbia Suicide Severity Risk ratings= 3 or 4) were removed from further analysis as they were not relevant to the aims of the study.

Ethics Approval

No contact information for callers was possible, and a waiver of consent was granted by the Swinburne University Human Research Ethics Committee (reference number: 2021-4340). This study is reported in accordance with the CONSORT (Consolidated Standards of Reporting Trials) checklist [16]. A CONSORT attrition flowchart is provided in Multimedia Appendix 2. This study was registered with the Australian and New Zealand Clinical Trials Registry (ACTRN12622000486729) [17].

Preprocessing of Calls

All calls were recorded in monochannel 8-kHz, 32-bit float format. Preprocessing involved transformation to 16-bit pulse-code modulation format, normalization, and pre-emphasis, which attenuated low signals and emphasized higher frequency signals to clarify the degree of audibility. This was important to reduce the effect of background noise. Listwise removal of silent frames (1,283,286/1,752,618, 73.2% speech frames) was performed prior to the following analyses.

Selection of Low Versus Imminent Risk of Suicide Calls for Analysis

A multigated approach informed the designation of suicide risk level. Imminent risk of suicide was confirmed via affirmative responses (by the caller) to the following 3 screening questions: "are you having thoughts of suicide?", "do you have a plan?", and "are the means available?", in compliance with triage guidelines [11]. The level of risk was then reassessed at the conclusion of each call by the responding counsellor using an organizationally developed framework; 6-point Likert-style scale (0-1=low; 2-3=medium; and 4-6=high). Responding counsellors also made clinical notes (eg, presentation of important content), which were inspected to ensure good correspondence to the assigned levels of suicide risk.

The level of risk for each call was then reassessed by a team of associate researchers (n=6), blinded to the initial rating. The associate researchers were psychologists either provisionally or fully registered with the Psychology Board of Australia, who had substantial prior experience working with suicidal presentations in telehealth settings. A random sample of calls (n=100) was provided to each researcher for reassessment using the Columbia Suicide Severity Rating Scale, a validated measure of suicide risk when used by clinicians [18] and administered via telephone [19]. Interrater reliability (kappa) of the Suicide Severity Rating Scale among the team of 6 associate researchers was 0.92 for a random selection of 12 recordings. The researchers were also asked to annotate segments of each recording using appropriate audio software (Audacity, Version 2.4.2; Audacity Team). Annotated segments of each recording were to be free from the counsellors' voice as much as possible. Each annotated segment of speech was also described using mental status examination language.

Derivation of Voice Biomarkers for Identifying Imminent-Risk Calls

Each annotated sound segment was divided into 50% overlapping 40 milliseconds Blackman-filter windowed frames [20]. The frame size ensured an adequate level of magnification of important characteristics at the center of the frame, while the degree of overlap ensured that the tails of each window did not remove valuable information. The modelling of risk occurred via 36 different voice biomarkers. Voice biomarkers are defined at both the 40 milliseconds speech frame and segment levels in the generalized additive mixed effects regression model described below.

Reduction of Voice Biomarkers Using Penalized Lasso Regression

Penalized Lasso regression [21] was performed in the first instance to reduce the number of possible predictors to only those with a strong relationship with suicide risk. However, this model assumes linear relationships between predictors and response (conveyed via a logit link function) and ignores gender effects and correlations among segments across a single call. Thus, this model was used primarily to reduce the set of predictors that informed subsequent analyses.

Validation via Mixed Effects Generalized Linear Regression

A 3-level model best reflected the approach to data collection. This model was used to confirm the significance of the reduced predictor set and test for significant moderation by caller sex, while allowing for the correlation between speech frames within each call. Model variables are summarized in Multimedia Appendix 3.

Splines were applied to each biomarker to account for nonlinearity [22]. Random intercepts at level 3 accounted for differences between individual calls, and a binomial model with logit link was used to identify imminent risk speech frames in terms of the level 1 and 2 voice biomarkers.

Without a comparable prior study, a power analysis for the final classification algorithm was not feasible. However, Pestian and colleagues [5] were able to achieve levels of classification accuracy of 85% with 371 recordings (level 2) and ~15 voice biomarkers (level 1). With a more precise classification model (component-wise gradient boosting) and a mixture of level 1 and level 2 predictors, we anticipated that a smaller sample size would suffice for this new study.

Classification of Calls Using a Gradient Boosting Classification Model

Although powerful, support vector machines, as used by Pestian and colleagues [5], have notable disadvantages. Computation time is prohibitive when the data set is large (eg, >100,000 observations) and the choice of kernel, which allows the algorithm to choose a path of demarcation between groups while minimizing misclassification error, can be difficult. This is especially true when there is little to guide the choice of kernel, which is the case in analysis problems concerning voice biometrics. Finally, the mathematical complexity of support vector machines reduces the transparency of classification decision-making.

In comparison to support vector machines, gradient boosting is a computationally simpler approach that addresses many of the aforementioned problems. However, in its base implementation, it assumes linearity among the predictors. This can be remedied with an alternative implementation. Component-wise gradient boosting can analyze nonlinear data by first estimating a generalized additive mixed model with splines added, and then applying each model component (individual predictors and random components) to achieve the best reduction in classification error (eg, see Hofner [23] for a detailed overview). It is an approach that also allows for sex-moderated effects for all biomarkers.

Leave-one-out cross validation was used to test the classification accuracy of the gamboost model and to prevent information leakage occurring if data from one participant was used in both the training and test data sets. Thus, n–1 callers were used to train the model, leaving the nested speech frames of a single caller as the test case, ensuring independence of data between training and test data sets. Classification probabilities were derived for each speech frame (40 milliseconds) within each hold-out caller. Frame level classification probabilities were summarized by the mean classification probability for each hold-out caller.

The Youden J index was used to derive the ideal cut point that maximized upon both sensitivity and specificity of the classification accuracy across all hold-out callers in relation to binary precoded suicide risk level. A total of 1000 bootstrap samples were estimated, and the mean of these estimated samples was used as the ideal cut point. This approach minimizes sample-specific bias and possible overestimation of diagnostic utility, as discussed in Thiele and Herschfeld [24], and is an approach used by other authors, such as Hentschel [25].

Overfitting is suggested when there is a drop in classification accuracy between training and validation classification accuracy, suggesting that the algorithm has *memorized* the basis for classification and applies these insights poorly to new data. Classification accuracy was determined via accuracy measures including area under the receiver operating characteristic and area under the precision-recall curve. Plain language descriptions of all voice biometrics are contained in Multimedia Appendix 4.

Results

Select caller demographics are summarized in Multimedia Appendix 5. The sample comprised 77/281 (27.4%) callers at imminent risk of suicide and 204/281 (72.6%) at low risk of suicide (n=87, 40% male callers and n=194, 70% female callers). Voice biomarkers were derived and analyzed for each of the 470,032 forty-millisecond speech frames. Median number of annotated segments per recording was 13.0 (SD 14.58), and median length of each segment was 118.50 (SD 120.19) milliseconds.

Reduction of Voice Biomarkers Using Penalized Lasso Regression

Penalized Lasso regression was performed to reduce the number of voice biomarkers used to predict imminent suicide. A total of 36 initial voice biomarkers were reduced to 12. The significant predictors are summarized in Multimedia Appendix 6.

Validation via Mixed Effects Generalized Linear Regression

A generalized additive mixed model was employed to validate the 12 predictors chosen by the Lasso regression. The model explained 12.0% (adjusted R^2) of the variance in risk level at the segment level (N=3070 annotated segments).

Table 1 summarizes the significance of spline coefficients in the generalized additive mixed model. Sex of caller was a significant moderator. The effective degrees of freedom indicate the degree of nonlinearity for each voice biomarker, with higher effective degrees of freedom indicating a greater degree of nonlinearity, and effective degrees of freedom close to 1 indicating linearity. Figure 1 illustrates the relationship between each voice biomarkers and the probability of imminent suicide, separately for male and female callers. For example, the plot of root mean squared amplitude suggests that both male and female callers speak with less signal strength (conceptually analogous to speaking in hushed tones) when at imminent risk of suicide. Conversely, increases in spectral slope were observed in both male and female callers as the level of risk of suicide increased, suggesting an increase in physiological effort when experiencing increasing suicidal stress.

These results confirmed that 11 (92%) of the 12 voice biomarkers were significant predictors of imminent suicide risk. First formant frequency, which proved nonsignificant for both male and female callers, was not included in the subsequent Component-wise gradient boosting model.

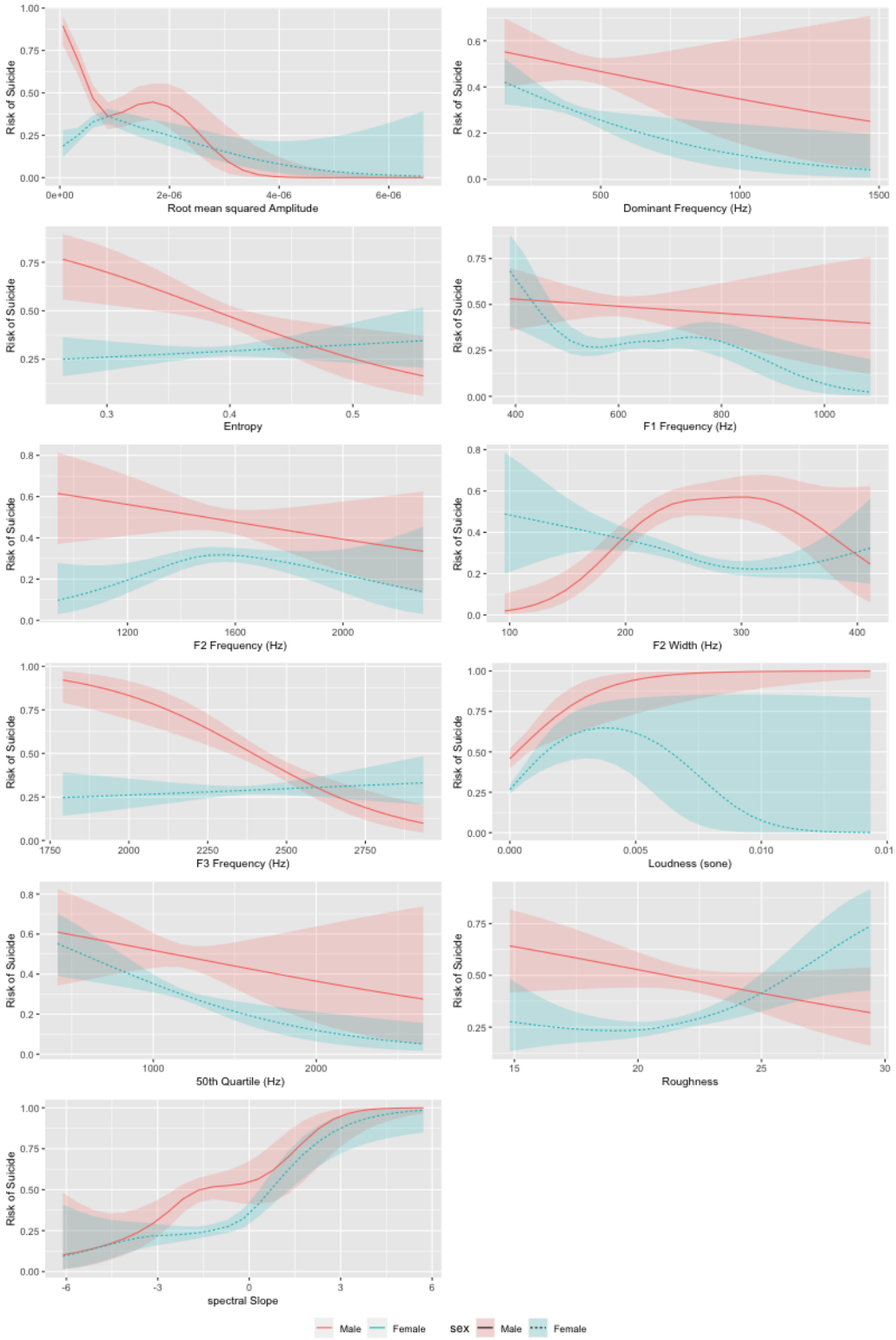
Table 1. Voice biomarker significance in the prediction of suicide risk: generalized additive mixed model[a] (adjusted R2=0.12; N=3070).

Voice biomarker significance	β	SE	95% CI	EDF[b]	*F*-test	*P* value
Male						
Root mean squared amplitude (dB)	–19.02	6.91	(–19.26, –18.77)	4.23	10.99	<.001
Dominant frequency (Hz)	–1.10	1.10	(–1.14, –1.07)	1.00	1.00	.32
Entropy	–2.21	0.80	(–2.24, –2.18)	1.00	7.56	.006
$Formant_1$ frequency (Hz)	–0.42	0.89	(–0.45, –0.39)	1.00	0.23	.63
$Formant_1$ width (Hz)	–0.60	1.28	(–0.65, –0.56)	2.46	3.05	.06
$Formant_2$ frequency (Hz)	–0.88	0.84	(–0.91, –0.85)	1.00	1.11	.29
$Formant_2$ width (Hz)	2.09	0.86	(2.06, 2.12)	3.57	10.16	<.001
$Formant_3$ frequency (Hz)	–3.61	0.76	(–3.64, –3.59)	1.00	22.68	<.001
Loudness	8.57	2.79	(8.47, 8.67)	1.00	9.52	<.001
50th quartile (Hz)	–1.14	1.27	(–1.19, –1.10)	1.03	0.82	.37
Roughness	–1.05	0.71	(–1.07, –1.02)	1.00	2.21	.12
Spectral slope	7.60	2.09	(7.53, 7.67)	4.10	4.76	<.001
Female						
Root mean squared amplitude (dB)	–3.67	2.28	(–3.75, –3.59)	3.74	7.11	<.001
Dominant frequency (Hz)	–2.41	0.90	(–2.44, –2.37)	1.02	6.63	.01
Entropy	0.36	0.49	(0.34, 0.38)	1.00	0.54	.47
$Formant_1$ frequency (Hz)	–3.09	1.10	(–3.13, –3.05)	4.34	2.71	.02
$Formant_1$ width (Hz)	3.28	1.35	(3.23, 3.32)	3.44	2.00	.07
$Formant_2$ frequency (Hz)	0.17	0.76	(0.14, 0.19)	2.79	4.18	.02
$Formant_2$ width (Hz)	–0.74	0.58	(–0.76, –0.72)	2.90	5.52	<.001
$Formant_3$ frequency (Hz)	0.32	0.50	(0.30, 0.34)	1.01	0.39	.53
Loudness	–5.97	4.55	(–6.13, –5.81)	1.91	8.34	<.001
50th quartile (Hz)	–2.53	0.74	(–2.56, –2.50)	1.03	10.56	<.001
Roughness	1.62	0.62	(1.60, 1.64)	2.58	4.65	.005
Spectral slope	4.99	1.20	(4.94, 5.03)	4.10	10.53	<.001

[a]Male versus female: β=–.84; SE 0.002, 95% CI (–0.85, –0.84), *t*=6.59 (2 tailed); *P*<.001.

[b]EDF: effective degrees of freedom.

Figure 1. Plots of generalized additive mixed effects model predictors; nonlinear relationship between voice biomarkers and the risk of suicide with 95% CIs.



Classification of Calls Using a Gradient Boosting Classification Model

Component-wise gradient boosting was used to classify each speech frame in terms of low and imminent risk of suicide. Leave-one-out cross-validation was used to test the classification accuracy of the gamboost model. A Youden J index value of 0.51 optimized upon both sensitivity and specificity in the classification of imminent risk. We correctly classified 469,332/470,032 (99.85%) speech frames (area under the receiver operating characteristic=1.0, 95% CI 1.0-1.0; area under the precision-recall curve=0.989, 95% CI 0.97-1.00).

While all low–suicide-risk speech frames were correctly classified, 700 (0.53%) of the 132,741 imminent risk frames were misclassified as low. This corresponded with the speech frames of a single caller in the 000 Emergency Services,

Canberra sample. Upon closer inspection, this caller presented in an intoxicated manner, having ingested a "large amount of sleeping tablets." Mental status examination annotations made by the reviewing team of psychologists indicated this caller spoke with slow-to-normal rate of speech and flat-to-neutral affect, and was responsive to all questions asked, a presentation similar to many low–suicide-risk callers.

Discussion

The development of a timely and accurate form of suicide risk assessment remains a significant challenge, especially if implemented in a real-time capacity as required by suicide-prevention help lines. In this study, we sought to automatically classify short segments of speech obtained from 2 suicide-prevention telehealth services in Australia, according to low and imminent risk of suicide using supervised machine learning approaches. We successfully classified 469,332/470,032 (99.85%) speech frames, with only a small number (700/132,741, 0.53%) of high-risk speech frames misclassified.

Our study compares favorably with the findings of Pestian and colleagues [5], who successfully classified 322 (85.0%) of 379 participant recordings discriminating between low and high risk of suicide, using support vector machines. However, our study differs in a number of important ways from the aforementioned study. Rather than classify suicide risk at the holistic recording level, we instead classified risk at the 40 milliseconds frame level. This allowed us to expand upon the size of the data set upon which the classification algorithm could be trained and validated, allowing for a more nuanced assessment of each voice biomarker. This approach also demonstrates that only a short segment of a call is required for suicide risk classification, suggesting that the algorithm can be used for triage purposes based on only a short exchange (eg, an exchange with a triaging chatbot).

Second, we achieved a greater level of transparency and refinement than was afforded by the support vector machine in the study by Pestian and colleagues [5]. Figure 1 illustrates the exact nature of the relationship between the 11 voice biomarkers and the level of risk of suicide for men and women and suggests several discernible nuances in the ways male and female callers might speak when experiencing suicidal stress.

Finally, we reduced the numbers of possible predictors (via Lasso regression) to ensure that only the most statistically relevant biomarkers were included in later models. Our choice of generalized additive mixed model validated the use of all but one of the voice biomarkers selected by the Lasso regression.

However, there are limitations in our approach. We did not include callers of minority status, such as members of the LGBTIQ+ (lesbian, gay, bisexual, transgender, intersex, queer, and other people of diverse sexuality, gender, or bodily characteristics) communities and callers of non–English speaking backgrounds. These community members may offer valuable information that can further enhance the diversity of input, practical outcomes, and nuance of our analyses overall.

We were also not always successful in classifying all calls in the imminent suicide risk category. The similarity between the presentations of the single misclassified imminent risk caller and other low suicide risk callers suggests a possible subsample of callers who may be in the midst of a suicide attempt that is not being recognized by our classification approach. This is of concern given this presentation is most at need of timely emergency support. This suggests a possible role for other forms of classification such as natural language processing of speech-to-text translation, which might reduce similar misclassifications in the future.

There were also notable strengths in our approach. Our industry partnerships with On The Line, Australia and the Australian Federal Police, Canberra ensured that we could trial this novel technology in ecologically valid settings, where call quality is often degraded and background noise evident. This contrasts with the clinical settings within which the majority of studies have thus far been conducted. Our multigated approach to the assignment of suicide risk to each call ensured the establishment of a credible ground truth that was pivotal in accurately training the classification algorithm. Our choice of advanced statistical modelling has ensured a robust account of error variance in estimating the probability of imminent suicide. A final strength of this study is the visualization of imminent risk of suicide in terms of voice biomarkers allowing for nonlinearity.

This study has taken evidence from 25 years of pilot research and extended it to real world scenarios involving present-moment intent to suicide. However, it would be beneficial to control for caller age. We did not control for age in any of our analyses; however, given the well-documented age-related changes in vocal quality, this should feature in subsequent analyses and would boost an account of variance achieved by future mixed effects modelling. Finally, these compelling findings suggest possible implementation within a suicide-prevention telehealth service as an avenue worthy of further exploration.

Authors' Contributions

RI was responsible for conceptualization, data curation, investigation, methodology, and writing—original draft. MN was responsible for supervision and writing—review and editing. DM was responsible for methodology, supervision, and writing—review and editing. RI and DM have accessed and verified the underlying data reported in this manuscript. All authors had full access to the data reported in this manuscript. All authors accept full responsibility for the submission of this manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Analysis workflow.
[PNG File , 589 KB - mental_v9i8e39807_app1.png]

Multimedia Appendix 2
CONSORT (Consolidated Standards of Reporting Trials) attrition flowchart.
[PNG File , 341 KB - mental_v9i8e39807_app2.png]

Multimedia Appendix 3
Summary generalized additive mixed effect model predictors.
[DOCX File , 14 KB - mental_v9i8e39807_app3.docx]

Multimedia Appendix 4
Definitions of voice biomarkers terminology.
[DOCX File , 15 KB - mental_v9i8e39807_app4.docx]

Multimedia Appendix 5
Summary caller characteristics.
[DOCX File , 15 KB - mental_v9i8e39807_app5.docx]

Multimedia Appendix 6
Summary of significant level 1 and level 2 Lasso regression predictors.
[DOCX File , 15 KB - mental_v9i8e39807_app6.docx]

References

1. Suicide in the world: global health estimates. World Health Organization. 2019. URL: https://apps.who.int/iris/bitstream/handle/10665/326948/WHO-MSD-MER-19.3-eng.pdf [accessed 2020-11-17]
2. 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]
3. Čukić M, López V, Pavón J. Classification of Depression Through Resting-State Electroencephalogram as a Novel Practice in Psychiatry: Review. J Med Internet Res 2020 Nov 03;22(11):e19548 [FREE Full text] [doi: 10.2196/19548] [Medline: 33141088]
4. Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. Speech Communication 2015 Jul;71:10-49. [doi: 10.1016/j.specom.2015.03.004]
5. Pestian JP, Sorter M, Connolly B, Bretonnel Cohen K, McCullumsmith C, Gee JT, et al. A Machine Learning Approach to Identifying the Thought Markers of Suicidal Subjects: A Prospective Multicenter Trial. Suicide Life Threat Behav 2017 Feb;47(1):112-121. [doi: 10.1111/sltb.12312] [Medline: 27813129]
6. Došilović FK, Brčić M, Hlupić N. Explainable artificial intelligence: A survey. 2018 Presented at: 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO); July 2, 2018; Opatija, Croatia. [doi: 10.23919/MIPRO.2018.8400040]
7. Sourirajan V, Belouali A, Dutton MA, Reinhard M, Pathak J. A Machine Learning Approach to Detect Suicidal Ideation in US Veterans Based on Acoustic and Linguistic Features of Speech. arXiv 2020 Sep 14:Preprint. [doi: 10.1186/s13040-021-00245-y] [Medline: 33531048]
8. Barth SK, Kang HK, Bullman T. All-Cause Mortality Among US Veterans of the Persian Gulf War: 13-Year Follow-up. Public Health Rep 2016 Nov;131(6):822-830 [FREE Full text] [doi: 10.1177/0033354916676278] [Medline: 28123229]
9. Fisk M, Livingstone A, Pit SW. Telehealth in the Context of COVID-19: Changing Perspectives in Australia, the United Kingdom, and the United States. J Med Internet Res 2020 Jun 09;22(6):e19264 [FREE Full text] [doi: 10.2196/19264] [Medline: 32463377]
10. Hashim NW, Wilkes M, Salomon R, Meggs J, France DJ. Evaluation of Voice Acoustics as Predictors of Clinical Depression Scores. J Voice 2017 Mar;31(2):256.e1-256.e6. [doi: 10.1016/j.jvoice.2016.06.006] [Medline: 27473933]
11. Venek V, Scherer S, Morency L, Rizzo AS, Pestian J. Adolescent Suicidal Risk Assessment in Clinician-Patient Interaction. IEEE Trans. Affective Comput 2017 Apr 1;8(2):204-215. [doi: 10.1109/taffc.2016.2518665]
12. Annual report 2020-21. Lifline. 2021. URL: https://www.lifeline.org.au/media/vwop50aj/lifeline-annual-report-2021-v2.pdf [accessed 2022-02-21]
13. Brülhart M, Klotzbücher V, Lalive R, Reich SK. Mental health concerns during the COVID-19 pandemic as revealed by helpline calls. Nature 2021 Dec;600(7887):121-126. [doi: 10.1038/s41586-021-04099-6] [Medline: 34789873]

14. 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. [doi: 10.2196/13388] [Medline: 32432554]
15. Hunt T, Wilson CJ, Woodward A, Caputi P, Wilson I. Intervention among Suicidal Men: Future Directions for Telephone Crisis Support Research. Front Public Health 2018;6:1 [FREE Full text] [doi: 10.3389/fpubh.2018.00001] [Medline: 29404319]
16. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
17. Iyer R, Nedeljkovic M, Meyer D. ANZCTR study protocol. Australian New Zealand Clinical Trials Registry. URL: https://www.anzctr.org.au/ACTRN12622000486729.aspx [accessed 2022-03-02]
18. Barzilay S, Yaseen ZS, Hawes M, Kopeykina I, Ardalan F, Rosenfield P, et al. Determinants and Predictive Value of Clinician Assessment of Short-Term Suicide Risk. Suicide Life Threat Behav 2019 Apr 17;49(2):614-626. [doi: 10.1111/sltb.12462] [Medline: 29665120]
19. Katz I, Barry CN, Cooper SA, Kasprow WJ, Hoff RA. Use of the Columbia-Suicide Severity Rating Scale (C-SSRS) in a large sample of Veterans receiving mental health services in the Veterans Health Administration. Suicide Life Threat Behav 2020 Feb;50(1):111-121. [doi: 10.1111/sltb.12584] [Medline: 31441952]
20. Zhang L, Qu Y, Jin B, Jing L, Gao Z, Liang Z. An Intelligent Mobile-Enabled System for Diagnosing Parkinson Disease: Development and Validation of a Speech Impairment Detection System. JMIR Med Inform 2020 Sep 16;8(9):e18689 [FREE Full text] [doi: 10.2196/18689] [Medline: 32936086]
21. Desboulets L. A Review on Variable Selection in Regression Analysis. Econometrics 2018 Nov 23;6(4):45. [doi: 10.3390/econometrics6040045]
22. Coretta S. Modelling electroglottographic data with wavegrams and generalised additive mixed models. OSF Preprints 2019 Aug 29:Preprint. [doi: 10.31219/osf.io/m623d]
23. Hofner B, Mayr A, Robinzonov N, Schmid M. Model-based boosting in R: a hands-on tutorial using the R package mboost. Comput Stat 2012 Dec 22;29(1-2):3-35. [doi: 10.1007/s00180-012-0382-5]
24. Thiele C, Hirschfeld G. cutpointr: improved estimation and validation of optimal cutpoints in R. J. Stat. Soft 2021;98(11):1-27. [doi: 10.18637/jss.v098.i11]
25. Hentschel AE, Beijert IJ, Bosschieter J, Kauer PC, Vis AN, Lissenberg-Witte BI, et al. Bladder cancer detection in urine using DNA methylation markers: a technical and prospective preclinical validation. Clin Epigenetics 2022 Feb 05;14(1):19 [FREE Full text] [doi: 10.1186/s13148-022-01240-8] [Medline: 35123558]

Abbreviations

CONSORT: Consolidated Standards of Reporting Trials
LGBTIQ+: lesbian, gay, bisexual, transgender, intersex, queer, and other people of diverse sexuality, gender, or bodily characteristics

Edited by J Torous; submitted 24.05.22; peer-reviewed by K Zhang, D Helic; comments to author 14.06.22; revised version received 17.06.22; accepted 22.07.22; published 15.08.22.

Please cite as:
Iyer R, Nedeljkovic M, Meyer D
Using Voice Biomarkers to Classify Suicide Risk in Adult Telehealth Callers: Retrospective Observational Study
JMIR Ment Health 2022;9(8):e39807
URL: https://mental.jmir.org/2022/8/e39807
doi:10.2196/39807
PMID:35969444

Original Paper

A Game-Based School Program for Mental Health Literacy and Stigma on Depression (Moving Stories): Cluster Randomized Controlled Trial

Anouk Tuijnman[1,2], MSc; Marloes Kleinjan[2,3], MSc, PhD; Merlijn Olthof[1], MSc; Evert Hoogendoorn[4], MA; Isabela Granic[5], MSc, PhD; Rutger CME Engels[6], MSc, PhD

[1]Behavioural Science Institute, Radboud University Nijmegen, Nijmegen, Netherlands

[2]Trimbos-institute, Utrecht, Netherlands

[3]Department of Interdisciplinary Social Sciences, Youth Studies, Utrecht University, Utrecht, Netherlands

[4]IJsfontein, Amsterdam, Netherlands

[5]Faculty of Social Sciences, McMaster University, Hamilton, ON, Canada

[6]Erasmus School of Social and Behavioural Sciences, Erasmus University, Rotterdam, Netherlands

Corresponding Author:
Anouk Tuijnman, MSc
Behavioural Science Institute
Radboud University Nijmegen
Postbus 9104
Nijmegen, 6500 HE
Netherlands
Phone: 31 243610082
Email: anouk.tuijnman@ru.nl

Abstract

Background: Depressive symptoms are highly prevalent among adolescents in Western countries. However, although treatment for depressive symptoms is available, many adolescents do not seek help when they need it. Important barriers to help-seeking among adolescents include low mental health literacy and high stigma. Therefore, we have developed a game-based school program, *Moving Stories*, which combines mental health literacy training for depression with contact with someone with lived experience both in the digital and nondigital world.

Objective: The aim of this study is to conduct a first test of the effectiveness of the newly developed game-based program, *Moving Stories*, using a cluster randomized controlled trial.

Methods: A total of 185 adolescents participated, divided over 10 classes from 4 schools. Half of the classes were randomly selected to follow the *Moving Stories* program, whereas the other half were in the control group, where no intervention was provided. The adolescents filled out digital questionnaires at 4 time points, with questions on mental health literacy, stigma, depressive symptoms, and the program itself (before the program, after the program, 3-month follow-up, and 6-month follow-up). Using R (R Foundation for Statistical Computing), we ran linear mixed-effects models for all continuous outcome variables and generalized linear mixed-effects models for all binary outcome variables.

Results: Compared with the control group, participants in the *Moving Stories* group improved after the program in personal stigma (b=−0.53, 95% CI −1.02 to −0.03; $t_{179.16}$=−2.08; P=.04). Effects on personal stigma lasted over time (3-month follow-up: b=−0.57, 95% CI −1.11 to −0.03; $t_{174.39}$=−2.07; P=.04). Most adolescents in the *Moving Stories* group participated in the introduction (97/99, 98%) and contact session (93/99, 94%), played the game for 4 or 5 days (83/99, 83%), and indicated that they would recommend the game to their peers (90/98, 92%).

Conclusions: The results of this study show the potential of *Moving Stories* as a stigma reduction program. With changes in the program to improve its effects on mental health literacy, *Moving Stories* could be implemented in schools to improve help-seeking in adolescents and reduce the negative consequences and burden of depressive symptoms.

Trial Registration: Dutch Trial Register NTR7033; https://trialsearch.who.int/Trial2.aspx?TrialID=NTR7033

International Registered Report Identifier (IRRID): RR2-10.2196/11255

(JMIR Ment Health 2022;9(8):e26615) doi:10.2196/26615

KEYWORDS

depression; help-seeking behavior; helping behavior; health literacy; stigma; video games; adolescence; secondary schools; mental health; digital health

Introduction

Background

Depressive symptoms are highly prevalent among adolescents in Western countries [1-3] and increase the risk of developing a depressive disorder later in life [4,5]. Both depressive symptoms and a depressive disorder have many negative consequences, including social, academic, and physical problems [6,7]. In extreme cases, a depressive disorder can lead to suicidal thoughts and suicide attempts [2,4]. Suicide is one of the main causes of death among youths worldwide [2]. Owing to these detrimental consequences, it is vital that adolescents seek help if they experience depressive symptoms. However, although treatment for depressive symptoms is available [8-12], many adolescents do not seek help when they need it [13-15]. Moreover, the longer it takes to receive help, the worse the response to treatment is [16].

Important barriers to help-seeking among adolescents are low mental health literacy and high stigma [13,17]. Mental health literacy comprises "knowledge and beliefs about mental disorders which aid their recognition, management or prevention" [18], whereas stigma "exists when elements of labeling, stereotyping, separation, status loss, and discrimination occur together in a power situation that allows them" [19]. Social support, on the other hand, increases help-seeking [13]. In light of the high prevalence and negative consequences of depressive symptoms, it is pivotal to lower barriers to help-seeking. Therefore, we developed a game-based school program, called *Moving Stories*, that targets mental health literacy and stigma on depression [20]. The aim of this study is to examine the effectiveness of this program.

Mental Health Literacy and Stigma

Mental health literacy is a broad concept that not only concerns knowledge of mental health disorders but also refers to connected actions, both by those who need help themselves and by those close to them [18,21]. More specifically, Jorm [21] defined the main components of mental health literacy as "(a) knowledge of how to prevent mental disorders, (b) recognition of when a disorder is developing, (c) knowledge of help-seeking options and treatments available, (d) knowledge of effective self-help strategies for milder problems, and (e) first aid skills to support others who are developing a mental disorder or are in a mental health crisis." The *Moving Stories* program targets recognition (component b), knowledge of help-seeking options and treatments (component c), and first aid skills (component e), specifically for depression in youth [20]. The goal of the *Moving Stories* program is to improve help-seeking. As component (a) of mental health literacy concerns the prevention of mental disorders and component (d) concerns knowledge of self-help strategies, we decided not to focus on these 2 components.

In adolescents, recognizing mental health disorders such as depression has been linked to choosing appropriate help [22] and increased help-seeking recommendations to peers in need [22,23]. However, most adolescents have trouble identifying depressive symptoms [24]. Adolescents also have limited knowledge of help-seeking options and available treatments. Often, they do not know how and where to seek help [25]. Moreover, adolescents prefer to seek help from people they know [23,26-28] and believe this help to be the most beneficial [29]. Unfortunately, seeking help solely from peers is inadvisable as peers are usually not able to provide the help that is warranted [21,30]. Therefore, improving symptom recognition, knowledge of appropriate help and treatment, and first aid skills in adolescents is important in furthering help-seeking.

Although improving symptom recognition is a good strategy to enhance help-seeking, recognizing depressive symptoms and labeling someone as mentally ill have also been linked to stigmatizing attitudes [31]. Both personal and perceived stigmatizing attitudes have been found to hinder help-seeking [13,32-35]. Personal stigma refers to a person's "own attitudes to a mentally ill person," whereas perceived stigma refers to a person's "perceptions of the attitudes held by other people" [36]. High personal stigma in adolescents is also related to poorer first aid skills [37]. Therefore, programs that target mental health literacy should also, to a similar extent, focus on reducing stigmatizing attitudes.

Stigma and Mental Health Literacy Programs

Several studies have tested the effectiveness of programs specifically aimed at reducing stigma toward mental health. Meta-analyses have shown that antistigma programs, in general, are effective in reducing stigma [38,39]. Successful elements of antistigma programs are education regarding mental health disorders and direct or indirect contact with someone who has experienced a mental health disorder. Research on mental health literacy programs among youth is scarce. Most programs that include components targeting mental health literacy primarily focus on enhancing mental health rather than literacy [40]. Moreover, most studies on mental health literacy lack a rigorous evaluation with a randomized design, validated measures, or follow-up measurements [40-42]. There are a few exceptions, and these more rigorous studies show promising results (eg, *Headstrong* [43] and *teen Mental Health First Aid* [44-47]). Despite these promising results, currently available mental health literacy programs for youth are subject to limitations. Moreover, there has been some criticism of the way contact sessions with someone with lived experience with a mental health disorder are organized in stigma programs for youth [48,49].

Few mental health literacy programs target behavioral mental health literacy components (ie, help-seeking and first aid

behavior). Most programs aim to improve knowledge and attitudes but do not focus on changing behavior [21]. In addition, most mental health literacy programs are fairly didactic, whereas adolescents themselves prefer more interaction [40]. Studies on stigma programs in young adolescents have reported possible limitations of contact sessions, including that when effects are not found, it could be because of (1) the length of a contact session (eg, when it is just 1 hour [49]), (2) the way the person with lived experience is introduced (eg, when a surprise introduction is used, which distracts the adolescents from the provided information [48]), and (3) the age difference between the adolescents and the person with lived experience (eg, when an adult is presenting their story to adolescents [48]). We argue that a program that makes use of a video game, which is combined with a contact session with someone with lived experience, could address these limitations. In video games, adolescents can practice help-seeking and first aid behaviors in a safe but engaging environment [50]. Games are excellent learning tools as they provide immediate feedback on players' actions, encouraging them to continue and learn more [51]. Games also allow for contact with people who youths normally might not often encounter, including peers with depression. Moreover, as games can be programmed to be played over a longer period, these interactions with peers with depression can be repeated over a period of multiple days. Games are also an important part of youths' lives [52], making a game-based program relevant for young adolescents. In addition, there are already some successful examples of video games that teach adolescents health knowledge and skills [53,54] and decrease mental health stigma [55]. However, there are none that focus on both mental health literacy and stigma in youth. Finally, a contact session with someone with lived experience could add to the interactions that adolescents prefer, transfer the digital experience to the nondigital world, and further help decrease the stigma for depression. Especially when a program allows for informal discussions between adolescents and a person with lived experience, it could increase its effects on stigma [56]. For these reasons, we developed the game-based school program, *Moving Stories*, which combines mental health literacy training for depression with contact with someone with lived experience, both in the digital and nondigital world [20].

This Study

The goal of this study is to conduct a first test of the effectiveness of the newly developed game-based program *Moving Stories* using a cluster randomized controlled trial [20]. Our primary hypotheses are that adolescents who participate in the *Moving Stories* program, compared with a control group (business as usual), will have better mental health literacy and lower stigma after the program and at the 3-month follow-up. Our secondary hypotheses concern the effectiveness of the *Moving Stories* program at the 6-month follow-up on mental health literacy and stigma and at the 3- and 6-month follow-ups on first aid and help-seeking behaviors. Specifically, we expect that adolescents who participate in the *Moving Stories* program will provide more appropriate first aid if they are in contact with a peer with a mental health problem and will be more likely to seek help personally if they themselves experience mental health symptoms compared with adolescents in the control group. Finally, we have included process measures to assess the use of the *Moving Stories* program and the possible side effects, including depressive symptoms.

Methods

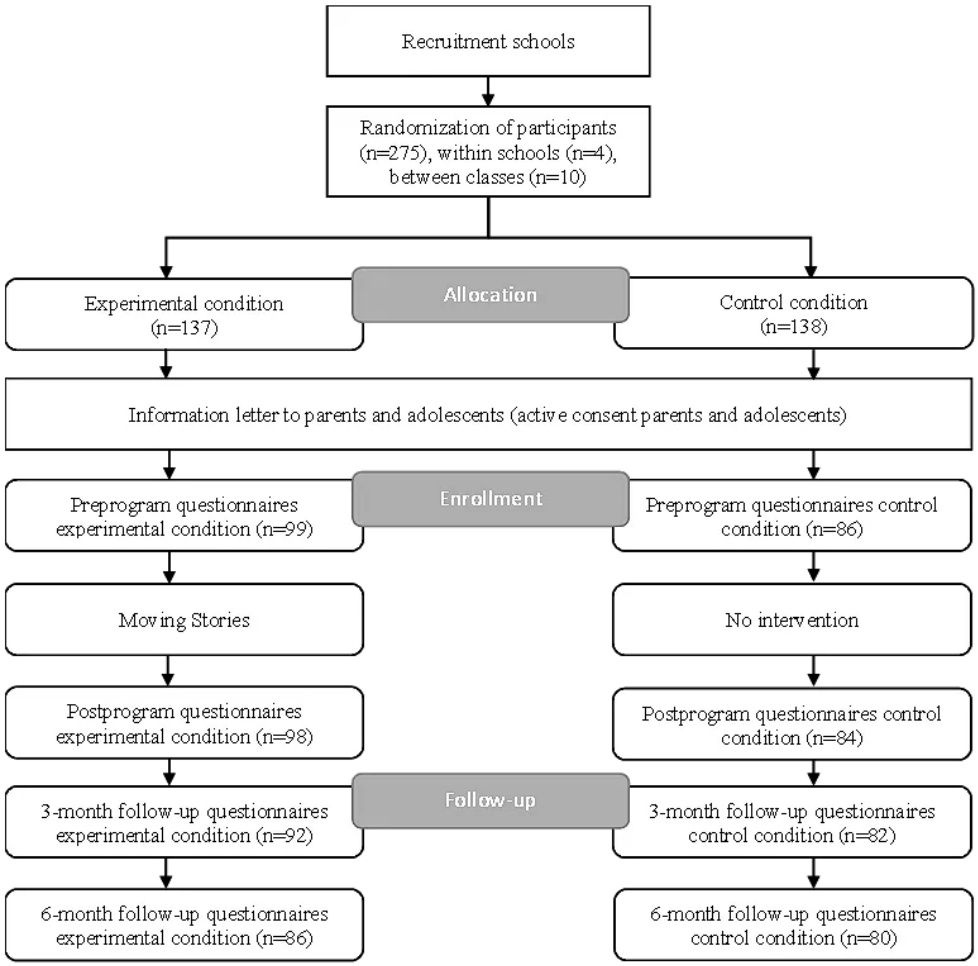
Procedure

A total of 275 adolescents from 10 classes in 4 schools were asked to participate. Active informed consent was obtained; both adolescents and their parents received an information letter and consent form. The adolescents who did not get permission from their parents to participate or refused to participate themselves were excluded from the study and the program, leaving a sample of 185 participants (185/275, 67.3%). Parents and adolescents were allowed to withdraw their consent at any time. Figure 1 shows the flowchart of the allocation, enrollment, and follow-up of participants in the study. A summary of the procedure is presented in this section. For a more detailed explanation of the procedure, refer to the study protocol [20]. The CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist (V 1.6.1) can be found in Multimedia Appendix 1.

The adolescents filled out digital questionnaires at 4 time points with a personal ID code and password (before the program, after the program, 3-month follow-up, and 6-month follow-up). Most adolescents completed the questionnaire at school. If they were not able to do so (eg, because of illness or because the school could not provide a time slot), they could complete the questionnaire at home. The adolescents who failed to fill in the questionnaire after the program or at the 3-month follow-up were allowed to participate in the subsequent time point. All dropouts were because of the adolescents forgetting or not willing to fill in the questionnaire at home, even after repeated reminders by phone or email. The adolescents received €12.50 (US $12.73) for completing at least the preprogram and 6-month follow-up questionnaires.

As *Moving Stories* was designed to motivate adolescents to talk about depression, the primary teachers of the participating classes and school care professionals (in both conditions) received an information booklet and information session in which possible actions were provided to discuss depressive symptoms or suicidal thoughts with students. They were given the opportunity to follow a web-based e-learning program on depression and suicidality in youth [57,58]. Furthermore, participants' depressive symptoms were assessed at each time point to verify that *Moving Stories* did not increase depressive symptoms or suicidal ideation. If an adolescent had clinically high depressive symptoms or suicidal ideation, both the adolescent and their parents were contacted by phone by a clinically trained member of the research team to inform them of the result and give advice on where to seek professional help. In total, 4 adolescents and their parents were contacted. They did not withdraw from participating in the program or study.

Figure 1. Flowchart of allocation, enrollment, and follow-up.



Ethics Approval

Ethical approval was provided by the ethics committee of the Faculty of Social Sciences at Radboud University (ECSW2017-2306-526). The study was registered in the Dutch Trial Register (NTR7033).

Sample Size

We expected a small to medium effect (Cohen *d*=0.40) of *Moving Stories* on mental health literacy and stigma at the 3-month follow-up based on the results of a previous study with a school program for mental health literacy [43]. Our power analysis was based on our original analytic plan [20], using Stata (version 14.2; StataCorp) [59] and assuming baseline-adjusted regression analyses (α=.05; β=.20). Our provisional estimates for the correlations between pre- and postprogram measurements and between postprogram and 3-month follow-up measurements were 0.50. A coefficient of 0.19 (estimated mean cluster size 18; estimated cluster size range 11-25 [60]) and an intracluster correlation coefficient of 0.02 [43] led to a design effect of 1.35. Considering the design effect, we calculated that we needed 3.75 classes per condition to show the expected effect, rounding up to 4 classes per condition, with 18 adolescents per class. To adjust for a *t* distribution [61], we added 1 class per condition, resulting in 5 classes (ie, 90 adolescents) per condition and a total necessary sample size of 180.

Program

Adolescents in the experimental condition participated in the *Moving Stories* program. *Moving Stories* is a game-based school program that comprises 3 parts: (1) an introduction lesson; (2) a single-player, mobile, and 3D video game [62]; and (3) a contact session with someone who has lived experience with a depressive disorder. For each class, the full program was delivered within 1 week. In the introduction lesson, adolescents were told what the video game was about (without using the term *depression*), and they were able to download and start the game with a classroom password. This password was linked to their class schedule to allow for joint playing time and shared feedback moments in the game.

The adolescents could play the game individually but in the same period as their classmates for 5 days in the morning (one of which was a weekend day). In the game, they interacted with the character Lisa (Figure 2), who showed signs of depression (without being labeled as depressive symptoms). Lisa's symptoms were based on symptoms that are most common in adolescent girls in the Netherlands who have subclinical or clinical depression. These symptoms were derived from multiple data sets from Dutch studies on depression during adolescence. The adolescents were asked to help Lisa and were able to do so by performing 5 actions each day. Some of these actions had a positive effect on their relationship with Lisa, and others had a negative effect. During the day, at set time points, the

adolescents received messages from Lisa with feedback on their actions. After 5 days of playing, the adolescents were able to see a final scene in which Lisa explained that she got help and thanked the player for their efforts in trying to help her.

Figure 2. Screenshot of the *Moving Stories* video game.

The program ended with a contact session for the whole class, which was led by trainers with lived experience with depression. The trainers told their own life stories about depression. Using this life story and the experience the adolescents had in the game, the trainer discussed 5 first aid skills that the adolescents could use when one of their friends would experience depressive feelings. A detailed description of the program is available in the study protocol [20]. The adolescents in the control condition did not participate in any part of the *Moving Stories* program. They only filled in the questionnaires, and their teachers received information on depression and suicide as described in the *Procedure* section.

Measures

Overview

We deviated from our study protocol by calculating McDonald ω instead of Cronbach α to assess the reliability of the measures. Many scholars have argued that in most psychological research, the assumptions made by Cronbach α are violated (eg, τ equivalence, normality, and a unidimensional scale), and that ω is a better alternative that can deal with these violations [63-66]. We used the package psych in R (R Foundation for Statistical Computing) to calculate the ω total [67]. We did not test the external validity of the various outcome measures. These concepts have been studied in multiple studies on mental health literacy and are related to behavioral outcomes. For example, in a study by Yap and Jorm [68], the researchers found that mental health first aid intentions predicted helping actions 2 years later.

Outcome Measures

Overview

Mental health literacy regarding depression was measured by (1) symptom recognition, (2) first aid intentions, (3) knowledge of first aid, (4) first aid confidence, (5) beliefs about help, and (6) help-seeking intentions. Depression stigma was measured by (1) personal stigma, (2) perceived stigma, and (3) social distance. A full description of all the measures is given in the study protocol [20].

Symptom Recognition

Symptom recognition was assessed using 3 vignettes with gender-matched descriptions of adolescents aged 15 years with depression, social anxiety, and psychosis [20,30]. Recognition of depression was defined by labeling the person in the depression vignette as depressed, whereas overestimation of depression was defined by labeling the person in the social anxiety or psychosis vignette as depressed. Responses to the vignettes were coded by undergraduate students (2 per vignette), and incongruencies were discussed. For all the following measures (unless otherwise mentioned), the vignette of the person with depression was used as an example.

First Aid Confidence

Confidence in providing first aid was measured by asking how confident the participant would be to help the person in the vignette if they were a friend [47].

First Aid Intentions and Skills

To measure general first aid intentions, participants were asked how much they agreed with the following statement: “If [name of the person in the vignette] was a friend, I would help

him/her." Specific first aid skills were assessed by asking whether the participant would perform each action on a list of 12 helpful and harmful first aid actions [44]. The scores for the harmful actions were reverse-scored. The helpful and harmful action scales have both shown acceptable to good reliability in a previous study [44]. However, in this sample, reliability was low for the 2 separate scales. We added up the helpful and reverse-scored harmful items to create a total score for first aid skills, with higher scores representing better first aid skills. The value of ω was between 0.71 and 0.82 for all 4 time points.

Beliefs About Help

Beliefs about help were assessed by asking whether a specific person would make the person's situation in the vignette "better"; "not better", "not worse"; or "worse." A score for beliefs about appropriate help was calculated by adding up the number of selected adult sources (ie, parent, other relative, psychologist or social worker, phone helpline, general practitioner, teacher, school welfare coordinator or school counselor, and religious leader) deemed to be helpful (better) [47]. The value of ω was between 0.67 and 0.74 for all 4 time points.

Help-Seeking Intentions

Help-seeking intentions were measured using the General Help-Seeking Questionnaire, which has good internal consistency and excellent test–retest reliability [69]. Average scores for the following 3 scales were calculated: (1) general help-seeking intentions (all categories; ω between 0.78 and 0.82), (2) help-seeking intentions toward informal sources (categories 1-4; ω between 0.70 and 0.74), and (3) help-seeking intentions toward formal sources (categories 5-10; ω between 0.84 and 0.85). Higher scores indicated higher intentions.

Stigma

Stigma was measured with 3 scales. Both personal and perceived stigma was measured using the Dutch Depression Stigma Scale [70,71], whereas social distance was measured with the 5 items from the Social Distance Scale for youths [71]. The original personal and perceived stigma scales have both shown acceptable to good internal consistency [72], whereas the Social Distance Scale has shown excellent internal consistency [44]. In this sample, ω was between 0.62 and 0.67 for personal stigma, between 0.74 and 0.83 for perceived stigma, and between 0.84 and 0.87 for social distance. Higher scores for personal and perceived stigma indicated higher stigma, whereas higher scores for social distance indicated lower stigma.

First Aid Behavior

First aid behavior was measured by asking whether the participant has had contact with someone who has experienced a problem similar to the one seen in the vignette ever (before the program) or within the past 3 months (at 3- and 6-month follow-ups). If the participant answered *yes* or *maybe*, they were then asked whether they offered the other person their help. If so, or if they were unsure, they were asked what they did out of the 12 actions and the open-ended option mentioned in the *First Aid Intentions and Skills* section [44].

Help-Seeking Behavior

Help-seeking behavior was assessed by asking whether the participants themselves had experienced a problem similar to the situation in the vignette. If they responded with *yes* or *not sure*, they were asked whether someone had helped them with this problem in the past 3 months and who this person was. Multiple answers could be provided.

Distal Measure

Depressive symptoms were assessed with the Children's Depression Inventory [73,74]. The Children's Depression Inventory comprises 27 items, each with 3 statements to choose from. The participants were asked to pick the statement that best described how they felt in the previous 2 weeks.

Process Measures

Contamination Check

To check for possible contamination effects because of the within-school randomization, participants in the control group were asked at the 6-month follow-up whether they had heard of the game *Moving Stories* and, if so, whether they had played it. If they indicated that they had played it, they were asked on what platform they had done so.

Evaluation of Moving Stories

After the program, participants in the intervention group were asked to evaluate the program *Moving Stories* with 7 items. To distinguish between the different components of the program and the study (ie, game, evaluation session, and research), the participants were also asked which of the components they would recommend to a friend if they had the opportunity to participate in the study over the following year.

Statistical Analyses

Data preparation and descriptive analyses were performed using SPSS Statistics (version 25; IBM Corp) [75]. All other analyses were conducted using R (version 3.5.1) [76]. We used multilevel analyses in R instead of regression analyses according to the intention-to-treat principle in Mplus [77], as described in the study protocol. Multilevel analyses in R were chosen for 2 reasons: (1) because our data were nested within participants and classes and therefore the analyses needed to account for this nested structure and (2) because multilevel analyses do not delete participants listwise, nor are imputations of data needed [78,79]. We ran linear mixed-effects models for all continuous outcome variables from preprogram measurements to postprogram measurements, 3-month follow-up, or 6-month follow-up and generalized linear mixed-effects models for all binary outcome variables using the R lme4 package [80]. We included random intercepts per participant and per class, as well as a random slope for condition per class. Therefore, our random effects structure modeled (1) baseline differences between participants, (2) baseline differences between classes, and (3) differences in the effectiveness of *Moving Stories* between classes. By including random effects, we knew that the differences between the experimental group and the control group would not be biased by differences between the classes. Therefore, the random effects structure benefited the interpretation in this 2-group design. When random effect

variances were estimated as 0 in one of our models, we simplified the random effect structure of the model by first omitting random correlations and, if necessary, the random intercept for class. As recommended for confirmatory hypothesis testing, the random slope for our predictor of interest (ie, condition) was never omitted [81]. Time was set as a factor, and all categorical predictors were coded using sum-to-zero contrasts. All continuous predictors were standardized into *Z* scores. The significance of the coefficients was tested using *t* tests (2-tailed), with correction of the df based on the Satterthwaite approximation. A chi-square test was used to test if the condition had an effect on whether adolescents had provided help to someone in their network with a mental health problem. We reported the estimate, 95% CI, and 2-tailed *t* test results with $P<.05$ for the linear mixed-effects models and the odds ratio, 95% CI, and *Z* test results with $P<.05$ for the generalized linear mixed-effects models. The other results can be found in Multimedia Appendix 2 and the materials at the Open Science Framework [82].

Open Science

All materials have been made publicly available via Open Science Framework [82]. The design and analysis plans for the experiments were published before finishing data collection at *JMIR Research Protocols* [20]. Deviations from the original analysis plans are described in this manuscript or in the materials at Open Science Framework.

Results

Participants

A total of 185 adolescents participated, with an age range of 12 to 15 years (mean 13.43, SD 0.67 years). Of the 185 participants, 101 (54.6%) were boys, and 179 (96.8%) were born in the Netherlands. The educational levels of the classes were either only preparatory secondary vocational education or preparatory secondary vocational education including higher general secondary education (low to middle level; 72/185, 38.9%) and preuniversity education (highest level; 113/185, 61.1%). Approximately 91.9% (170/185) of the adolescents indicated that they played video games and did so for an average of 8.48 (SD 7.26) hours per week.

Dropout at postprogram measurements (3/185, 1.6%) and the 3-month follow-up was low (11/185, 5.9%). At the 6-month follow-up, approximately 89.7% (166/185) of the adolescents finished the questionnaire (10% dropout). There were no significant differences in study condition, gender, country of birth, and gaming behavior and frequency among adolescents who dropped out at the 6-month follow-up and adolescents who finished the final questionnaire. However, the adolescents who did not finish the final questionnaire were significantly older ($t_{183}=2.47$; $P=.01$) and had higher depressive symptoms ($t_{183}=2.63$; $P=.009$) and a lower educational level ($\chi^2_1=18.3$; $P<.001$) than the adolescents who finished the final questionnaire.

Baseline Differences

Differences were found between the experimental and control groups in terms of gender ($\chi^2_1=10.5$; $P=.001$), whether the adolescents generally played video games ($\chi^2_1=7.4$; $P=.007$), and number of hours playing video games ($t_{181.98}=2.93$; $P=.004$). In the experimental group, 66% (65/99) of the participants were male, whereas in the control group, 42% (36/86) were male. Significantly more adolescents in the experimental group indicated that they played video games, and they did so more hours a week than those in the control group. We included gaming frequency (standardized) and gender as covariates in all models. Tables 1 and 2 show the descriptives of the outcomes for both conditions. At preprogram measurements, the experimental group had less adequate first aid skills than the control group ($t_{182.35}=-2.09$; $P=.04$) and more social distance toward peers with depression ($t_{183}=-2.04$; $P=.04$). Moreover, the experimental group deemed more adults to be helpful than the control group ($t_{183}=2.07$; $P=.04$).

Table 1. Percentages of adolescents who mentioned depression at the 3 vignettes (N=185).

Depression mentioned in vignette per group	Preprogram measurement		Postprogram measurement		3-month follow-up		6-month follow-up	
	n (%)	N	n (%)	N	n (%)	N	n (%)	N
Depression vignette								
Moving Stories	62 (63)	99	90 (92)[a]	98	70 (76)[a]	92	64 (74)	86
Control	57 (66)	86	75 (89)[a]	84	71 (87)[a]	82	68 (85)	80
Social anxiety vignette								
Moving Stories	0 (0)	99	2 (2)[a]	98	1 (1)[a]	92	1 (1)	86
Control	1 (1)	86	1 (1)[a]	84	0 (0)[a]	82	1 (1)	80
Psychosis vignette								
Moving Stories	13 (13)	99	34 (35)[a]	98	29 (32)[a]	92	21 (24)	86
Control	9 (10)	86	13 (15)[a]	84	17 (21)[a]	82	20 (25)	80

[a]Primary outcomes.

Table 2. Means and SDs of the outcomes of the continuous variables.

Outcome and group	Preprogram measurement, mean (SD)	Postprogram measurement, mean (SD)	3-month follow-up, mean (SD)	6-month follow-up, mean (SD)
First aid confidence				
Moving Stories	3.60 (0.89)	3.48 (1.02)[a]	3.47 (1.03)[a]	3.29 (1.10)
Control	3.48 (1.03)	3.49 (1.02)[a]	3.46 (1.08)[a]	3.55 (1.01)
General first aid intentions				
Moving Stories	4.45 (1.02)	4.27 (0.97)[a]	4.25 (0.88)[a]	4.22 (1.01)
Control	4.42 (0.93)	4.25 (1.02)[a]	4.23 (1.01)[a]	4.40 (0.92)
Specific first aid skills				
Moving Stories	43.12 (4.61)	43.00 (4.76)[a]	42.55 (4.74)[a]	42.57 (4.47)
Control	44.41 (3.77)	43.18 (4.87)[a]	43.87 (4.63)[a]	43.59 (4.38)
Beliefs about help				
Moving Stories	3.03 (1.63)	2.97 (1.79)[a]	2.45 (1.63)[a]	2.51 (1.70)
Control	2.56 (1.45)	2.36 (1.56)[a]	2.29 (1.51)[a]	2.39 (1.56)
Help-seeking intentions (total)				
Moving Stories	4.06 (0.98)	4.18 (1.02)[a]	3.98 (0.91)[a]	3.94 (0.83)
Control	3.87 (0.90)	3.86 (0.81)[a]	3.85 (0.75)[a]	3.85 (0.75)
Help-seeking intentions (informal)				
Moving Stories	4.83 (1.15)	4.86 (1.22)[a]	4.85 (1.07)[a]	4.85 (1.07)
Control	4.79 (1.14)	4.67 (1.08)[a]	4.61 (1.20)[a]	4.72 (1.08)
Help-seeking intentions (formal)				
Moving Stories	3.21 (1.23)	3.42 (1.34)[a]	3.08 (1.28)[a]	2.99 (1.23)
Control	2.93 (1.06)	3.00 (1.00)[a]	2.89 (0.94)[a]	2.93 (1.13)
Personal stigma				
Moving Stories	18.56 (3.83)	17.74 (3.36)[a]	17.14 (4.10)[a]	16.81 (4.06)
Control	17.57 (3.70)	17.87 (3.58)[a]	17.59 (3.55)[a]	16.79 (3.47)
Perceived stigma				
Moving Stories	22.71 (4.14)	22.24 (4.81)[a]	22.73 (4.90)[a]	22.62 (4.84)
Control	21.97 (4.48)	22.61 (4.70)[a]	22.01 (4.86)[a]	22.08 (4.28)
Social distance (stigma)				
Moving Stories	14.27 (2.78)	14.56 (2.82)[a]	13.87 (3.03)[a]	14.19 (2.97)
Control	15.13 (2.91)	14.54 (2.87)[a]	14.88 (2.99)[a]	14.58 (3.06)

[a]Primary outcomes.

Outcome Measures

Symptom Recognition

No effects of condition on recognition of depressive symptoms in the depression vignette or on overestimation of depressive symptoms in the psychosis vignette were found from preprogram to postprogram measurements and 3- and 6-month follow-ups. For the social anxiety vignette, variance was close to 0 (approximately no adolescents indicated they thought the vignette was about depression); therefore, we did not conduct any analyses on the social anxiety vignette.

First Aid Confidence

No effects of condition were found at the postprogram measurement and 3-month follow-up for first aid confidence. Participation in the experimental group predicted changes over time in first aid confidence from the preprogram measurement

to the 6-month follow-up compared with the control group (b=−0.20, 95% CI −0.35 to −0.05; $t_{169.28}$=−2.61; P=.01). Contrary to our secondary hypothesis, the experimental group had a decrease in first aid confidence from the preprogram measurement to the 6-month follow-up (t_{86}=3.74; P<.001), whereas this effect was not found in the control group.

First Aid Intentions

Participation in the experimental group did not significantly predict changes over time in intentions for general or specific first aid skills.

Beliefs About Help

Participation in the experimental group did not significantly predict changes over time in beliefs about the helpfulness of adults.

Help-Seeking Intentions

Participation in the experimental group did not significantly predict changes over time in help-seeking intentions toward informal, formal, or all help sources.

Stigma

Participation in the experimental group, compared with the control group, was related to changes over time in personal stigma from pre- to postprogram measurements (b=−0.53, 95% CI −1.02 to −0.03; $t_{179.16}$=−2.08; P=.04) and from preprogram measurements to the 3-month follow-up (b=−0.57, 95% CI −1.11 to −0.03; $t_{174.39}$=−2.07; P=.04). At the postprogram measurements and 3-month follow-up, in line with our primary hypotheses, personal stigma decreased significantly in the experimental group (postprogram measurement: t_{97}=2.40, P=.02; 3-month follow-up: t_{91}=3.81, P<.001), whereas this was not the case for the control group. No effects of condition on perceived stigma and social distance were found at postprogram measurement and the 3- and 6-month follow-ups, and the effects on personal stigma did not last at the 6-month follow-up.

First Aid Behavior

There were no significant differences at the 3- and 6-month follow-ups between the experimental and control groups on whether adolescents had provided help to someone in their network with a mental health problem. At preprogram measurements, of the 71 participants who were in contact with someone in their network with a mental health problem, 66 (93%) indicated having offered help. At the 3-month follow-up, this was 89% (31/35 participants), and at the 6-month follow-up, it was 67% (24/36 participants).

Help-Seeking Behavior

There were no significant differences at the 3- and 6-month follow-ups between the experimental and control conditions on whether adolescents had a personal experience with a mental health problem. At preprogram measurements, 17.3% (32/185) of the participants indicated whether they ever had a mental health problem (or did not know for sure). At the 3-month follow-up, 10.9% (19/174) of the participants indicated having had a mental health problem in the previous 3 months (or did not know for sure), whereas at the 6-month follow-up, this was 5.4% (9/166) of the participants. None of the adolescents at the 3- and 6-month follow-ups mentioned that they had not been helped.

Distal Measure

Participation in the experimental group was not related to changes in depressive symptoms over time.

Process Measures

Contamination Check

In the control group, at the 6-month follow-up, 3% (2/80) of the adolescents indicated that they had heard of *Moving Stories*. None of them indicated they had played the game.

User Statistics

Most of the adolescents in the experimental group participated in the introduction and contact sessions (97/99, 98% and 93/99, 94% of participants, respectively). During the playing days, 49% (49/99) of the adolescents played the game for 5 days, 34% (34/99) played it for 4 days, 14% (14/99) played it for 3 days, and 3% (3/99) played it for ≤2 days.

Evaluation of Moving Stories

The adolescents in the experimental group rated *Moving Stories* positively at the postprogram measurements. When asked whether they would recommend elements of the study and program to a friend, 92% (90/98) of the adolescents indicated they would recommend the game, 66% (65/98) indicated they would recommend the contact session, and 79% (77/98) indicated they would recommend the research element. Of the 98 adolescents, 75 (77%) understood the game, 67 (68%) thought they had learned something from it, and 74 (76%) were interested in how Lisa would be doing in 6 months. Fewer adolescents rated the contact session positively (44/98, 45%); however, 53% (52/98) of the adolescents indicated that they had learned something from it. Approximately 40% (39/98) indicated that they would play the game a second time to get a better score. In terms of stigmatizing attitudes toward Lisa, 23% (23/98) of the adolescents indicated that they would not hang out with Lisa in daily life, whereas the others were neutral or positive toward her.

Discussion

Principal Findings

The goal of this cluster randomized controlled trial was to test the effects of the newly developed game-based school program *Moving Stories* [20] on mental health literacy and stigma regarding depression in adolescents.

We found that participation in the *Moving Stories* program reduced personal stigma over time compared with participation in the control condition. These results are in line with previous research showing that contact with someone with lived experience can reduce stigma [39]. Interacting with a video game character with depressive symptoms combined with meeting someone with lived experience in the *Moving Stories* program seemed to be similarly effective, at least in terms of immediate effects. However, improvements in personal stigma

did not continue beyond 3 months, and the effects could be a result of regression to the mean [83]. These results suggest that repeated conversations about mental health in the classroom or increased interactions with people with mental health disorders (either virtually or face-to-face) may be needed to achieve longer-lasting improvements. Furthermore, implementing booster sessions with the game may be beneficial as well [84-86].

Contrary to our secondary hypothesis, we found that adolescents in the *Moving Stories* group were less confident about providing help at the 6-month follow-up compared with adolescents in the control group; however, there were no group differences at the 3-month follow-up or after the program. The adolescents in the *Moving Stories* group may have felt less confident after 6 months about remembering and applying the helpful skills taught in the *Moving Stories* program, whereas the adolescents in the control group had no knowledge of whether specific skills were helpful or harmful. Indeed, we did not find any effects on first aid skills. It could be that, for confidence to increase, a strong and longer-lasting knowledge improvement would be needed.

No group differences were found for symptom recognition, first aid intentions, beliefs about the helpfulness of adults in general, help-seeking intentions in general, or toward informal and formal sources specifically. We did not find group differences in perceived stigma, social distance, or providing and seeking help. The null results in our study are not in line with previous work, as other mental health literacy programs have been found to be effective in improving mental health literacy, specifically in terms of knowledge [43,44,47]. An explanation could be that our study was conducted on a population of younger adolescents. Most successful mental health literacy programs have been tested on older adolescents [43,44], and a study with younger adolescents found weaker effects [45]. Younger adolescents may be less developmentally equipped or lack experience with depressed peers to relate to the topic [45]. Another reason for not finding significant effects could be that the content of the *Moving Stories* program was not sufficient to improve mental health literacy. Other mental health literacy programs have incorporated didactic sessions in which facts about depression and first aid skills are explained. The *Moving Stories* program did not include didactic sessions but was focused on learning by doing [50]. It is possible that the experiences in the video game did not sufficiently lead to improved knowledge. As the contact session took approximately 50 minutes and focused on sharing experiences and translating the story of Lisa to the real world, providing information about depression was not prioritized. Future iterations of *Moving Stories* could combine the learning by doing in the video game with the didactic sessions from other programs to improve its effectiveness. This might be especially beneficial for a younger, less-experienced group of adolescents. Finally, it could also be that we did not have sufficient power to show effects in a Dutch context. We based our power analysis on an Australian study; however, the contexts of the 2 countries might have differed too much. Future studies in a Dutch context with *Moving Stories*, or any other mental health literacy program, should start with a lower expected effect size.

Finally, the process evaluation showed that *Moving Stories* was well-received by the adolescents, and adherence to the full program was high. This is as important as the stigma results, as both digital and nondigital mental health programs often suffer from engagement problems [87]. If we want a school program to be successful outside the research context, we need adolescents to be intrinsically motivated to participate in the program. Although the *Moving Stories* program does not have the same entertainment goals as the video games adolescents play today, adolescents in our study were still engaged by the game and interested in the content of both the game and the contact session.

Strengths and Limitations

This study used a rigorous design to test the effects of a new mental health literacy program. Most studies on the effects of health literacy programs do not include a control group, and very few studies test the effects over a longer period [40-42]. Another strength of this study is the assessment of ≥1 element of mental health literacy [21]. Many studies on mental health literacy focus solely on knowledge of depression facts [88]. In this study, we also assessed knowledge of available help, first aid skills, and help-seeking and first aid intentions.

A limitation of our study was the difference between dropouts and completers at the 6-month follow-up. Age and education level differences in attrition were mostly because of a class that received the questionnaire right before the Christmas break and therefore received fewer reminders. Compared with the other classes, adolescents in this class had a higher mean age and lower educational level. However, this particular class did not explain the attrition differences found in depressive symptoms. Adolescents with depressive symptoms may have been less comfortable with answering questions about their mental state and therefore did not fill in subsequent questionnaires [89-91]. In addition, all participants were made aware that in the exceptional case that the data showed that the adolescent was *feeling very bad*, they themselves and their parents would be contacted by the research team. To avoid such a call, the adolescents could have chosen not to fill out later questionnaires.

A second limitation was the relatively low reliability of some of the outcome measures (ie, beliefs about help and personal stigma). These outcomes should be interpreted with caution. In general, low psychometric quality is found for mental health literacy measures [88] and, to enable comparisons, we included questionnaires that were used in studies on one of the most rigorously evaluated mental health literacy programs (*teen Mental Health First Aid* [44]). Although the use of vignettes had the benefit of not relying on labels to assess mental health first aid, a disadvantage was that we did not know what the effects of repeated use of the vignettes were. However, our study did show a variety in responses to the vignettes over time, meaning that the participants did not simply repeat their answers. More work is needed to develop measures of good quality that help us assess the effects of mental health literacy programs more effectively [88].

Finally, we were unable to explore the behavioral effects of the *Moving Stories* program because of small subsamples and differences at baseline in the quality of the provided help

between the 2 conditions. If we are interested in studying whether adolescents show appropriate and effective behavioral responses when coming into contact with a friend with depressive symptoms after participation in a mental health literacy program, long-term follow-ups with larger samples are needed to allow for the occurrence of such an event. Moreover, assessing reaching out for help next to assessing receiving help would provide even more useful information.

Future Directions

On the basis of the results of this study, we have several recommendations for future directions of the *Moving Stories* program specifically and mental health literacy programs in general. One recommendation is to compare the effects of a mental health literacy program between younger and older adolescents. The original idea of *Moving Stories* was to teach adolescents first aid skills and help-seeking behavior before most mental health problems occurred. However, it could be that a mental health literacy program is more effective when students have experienced mental health problems themselves or in their social environment, which is more likely when they are older. Another recommendation is to tailor mental health programs to the preferences of the users. This becomes a more feasible option when digital tools are used and might improve their effectiveness [87]. For example, for *Moving Stories*, we could create a male or nonbinary version of Lisa and increase the number of first aid and dialog options in the game. Then, the game could be tailored to the preferences of a user but also to their knowledge and experience level. Finally, although many programs, both for mental health literacy and stigma, include the presence of a person with lived experience, the organization of these sessions on a large scale has proven to be difficult. Other options could be to include video material of people with lived experience while a school mental health professional leads a discussion. An added benefit is that the students participating in these discussion sessions are immediately in contact with someone who can provide help if they need it.

Conclusions

The results of this clustered randomized controlled trial seem to show that the *Moving Stories* program could be effective in reducing personal stigma regarding depression in adolescents, although the effects were small. Moreover, the program was well-received and well-adhered to, something that can be difficult to achieve with young adolescents. In its current form, the *Moving Stories* program had no effect on mental health literacy components, perceived stigma, or social distance. Our results suggest that a larger sample might be needed to show the effects on a Dutch adolescent population. Nevertheless, the stigma results and positive evaluations may be a good stepping stone for further iterations of the program. Combining the program with didactic elements from other programs and scheduling regular follow-up conversations in the classroom might improve these results. Moreover, tailoring the video game to each individual player's preferences and skill level could further increase its effects. With better results and after showing its effectiveness in studies with larger sample sizes, the program could be implemented in schools to improve help-seeking in adolescents and reduce the negative consequences and burden of depressive symptoms.

Acknowledgments

The development of *Moving Stories* and this study were funded by the Stimuleringsfonds and the Dutch Research Council Creative Industries (grant 314-99-115).

Authors' Contributions

AT was the primary researcher in this study and was supervised by MK, IG, and RCMEE. MO provided statistical support. AT has also contributed to the development of *Moving Stories*. EH was the lead designer in the development of *Moving Stories*. IG and RCMEE had advisory roles in the development of *Moving Stories*. All authors contributed to the writing of the manuscript.

Conflicts of Interest

Moving Stories was developed by IJsfontein and the Trimbos Institute in collaboration with the Behavioural Science Institute of the Radboud University Nijmegen and 113 Zelfmoordpreventie.

Multimedia Appendix 1
CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist (V 1.6.1).
[PDF File (Adobe PDF File), 9856 KB - mental_v9i8e26615_app1.pdf]

Multimedia Appendix 2
Test results linear mixed effects models.
[DOCX File , 24 KB - mental_v9i8e26615_app2.docx]

References

1. Balázs J, Miklósi M, Keresztény A, Hoven CW, Carli V, Wasserman C, et al. Adolescent subthreshold-depression and anxiety: psychopathology, functional impairment and increased suicide risk. J Child Psychol Psychiatry 2013 Jun;54(6):670-677. [doi: 10.1111/jcpp.12016] [Medline: 23330982]
2. Depression and other common mental disorders: global health estimates. World Health Organization. 2017. URL: https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf [accessed 2021-11-30]
3. Kleinjan M, Pieper I, Stevens G, van de Klundert N, Rombouts M, Boer M, et al. Geluk onder Druk? Onderzoek naar het mentaal welbevinden van jongeren in Nederland. Den Haag: UNICEF Netherlands. 2020. URL: https://www.trimbos.nl/aanbod/webwinkel/product/af1785-geluk-onder-druk [accessed 2021-11-30]
4. Fergusson DM, Horwood LJ, Ridder EM, Beautrais AL. Subthreshold depression in adolescence and mental health outcomes in adulthood. Arch Gen Psychiatry 2005 Jan;62(1):66-72. [doi: 10.1001/archpsyc.62.1.66] [Medline: 15630074]
5. Pine DS, Cohen E, Cohen P, Brook J. Adolescent depressive symptoms as predictors of adult depression: moodiness or mood disorder? Am J Psychiatry 1999 Jan;156(1):133-135. [doi: 10.1176/ajp.156.1.133] [Medline: 9892310]
6. Jaycox LH, Stein BD, Paddock S, Miles JN, Chandra A, Meredith LS, et al. Impact of teen depression on academic, social, and physical functioning. Pediatrics 2009 Oct;124(4):596-605. [doi: 10.1542/peds.2008-3348] [Medline: 19736259]
7. Fletcher JM. Adolescent depression and educational attainment: results using sibling fixed effects. Health Econ 2010 Jul;19(7):855-871. [doi: 10.1002/hec.1526] [Medline: 19582699]
8. Chi X, Bo A, Liu T, Zhang P, Chi I. Effects of mindfulness-based stress reduction on depression in adolescents and young adults: a systematic review and meta-analysis. Front Psychol 2018;9:1034. [doi: 10.3389/fpsyg.2018.01034] [Medline: 29977221]
9. Cuijpers P, van Straten A, Warmerdam L. Behavioral activation treatments of depression: a meta-analysis. Clin Psychol Rev 2007 Apr;27(3):318-326. [doi: 10.1016/j.cpr.2006.11.001] [Medline: 17184887]
10. Duffy F, Sharpe H, Schwannauer M. Review: the effectiveness of interpersonal psychotherapy for adolescents with depression - a systematic review and meta-analysis. Child Adolesc Ment Health 2019 Nov;24(4):307-317. [doi: 10.1111/camh.12342] [Medline: 32677350]
11. Ebert DD, Zarski A, Christensen H, Stikkelbroek Y, Cuijpers P, Berking M, et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a meta-analysis of randomized controlled outcome trials. PLoS One 2015;10(3):e0119895 [FREE Full text] [doi: 10.1371/journal.pone.0119895] [Medline: 25786025]
12. Oud M, de Winter L, Vermeulen-Smit E, Bodden D, Nauta M, Stone L, et al. Effectiveness of CBT for children and adolescents with depression: a systematic review and meta-regression analysis. Eur Psychiatry 2019 Apr;57:33-45. [doi: 10.1016/j.eurpsy.2018.12.008] [Medline: 30658278]
13. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
14. Raven D, Jörg F, Visser E, Oldehinkel AJ, Schoevers RA. Time-to-treatment of mental disorders in a community sample of Dutch adolescents. A TRAILS study. Epidemiol Psychiatr Sci 2017 Apr;26(2):177-188 [FREE Full text] [doi: 10.1017/S2045796016000226] [Medline: 27075651]
15. Amone-P'Olak K, Ormel J, Oldehinkel AJ, Reijneveld SA, Verhulst FC, Burger H. Socioeconomic position predicts specialty mental health service use independent of clinical severity: the TRAILS study. J Am Acad Child Adolesc Psychiatry 2010 Jul;49(7):647-655. [doi: 10.1016/j.jaac.2010.03.007] [Medline: 20610134]
16. Ghio L, Gotelli S, Marcenaro M, Amore M, Natta W. Duration of untreated illness and outcomes in unipolar depression: a systematic review and meta-analysis. J Affect Disord 2014 Jan;152-154:45-51. [doi: 10.1016/j.jad.2013.10.002] [Medline: 24183486]
17. Rickwood DJ, Deane FP, Wilson CJ. When and how do young people seek professional help for mental health problems? Med J Aust 2007 Oct 01;187(S7):35-39. [doi: 10.5694/j.1326-5377.2007.tb01334.x] [Medline: 17908023]
18. Jorm AF, Korten AE, Jacomb PA, Christensen H, Rodgers B, Pollitt P. "Mental health literacy": a survey of the public's ability to recognise mental disorders and their beliefs about the effectiveness of treatment. Med J Aust 1997 Mar 17;166(4):182-186. [doi: 10.5694/j.1326-5377.1997.tb140071.x] [Medline: 9066546]
19. Link BG, Phelan JC. Conceptualizing stigma. Annu Rev Sociol 2001 Aug;27(1):363-385. [doi: 10.1146/annurev.soc.27.1.363]
20. Tuijnman A, Kleinjan M, Hoogendoorn E, Granic I, Engels RC. A Game-Based School Program for Mental Health Literacy and Stigma Regarding Depression (Moving Stories): Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2019 Mar 14;8(3):e11255 [FREE Full text] [doi: 10.2196/11255] [Medline: 30869652]
21. Jorm AF. Mental health literacy: empowering the community to take action for better mental health. Am Psychol 2012 Apr;67(3):231-243. [doi: 10.1037/a0025957] [Medline: 22040221]
22. Wright A, Jorm AF, Harris MG, McGorry PD. What's in a name? Is accurate recognition and labelling of mental disorders by young people associated with better help-seeking and treatment preferences? Soc Psychiatry Psychiatr Epidemiol 2007 Mar;42(3):244-250. [doi: 10.1007/s00127-006-0156-x] [Medline: 17450404]
23. Coles ME, Ravid A, Gibb B, George-Denn D, Bronstein LR, McLeod S. Adolescent mental health literacy: young people's knowledge of depression and social anxiety disorder. J Adolesc Health 2016 Jan;58(1):57-62. [doi: 10.1016/j.jadohealth.2015.09.017] [Medline: 26707229]

24. Singh S, Zaki RA, Farid ND. A systematic review of depression literacy: knowledge, help-seeking and stigmatising attitudes among adolescents. J Adolesc 2019 Jul;74:154-172. [doi: 10.1016/j.adolescence.2019.06.004] [Medline: 31216495]
25. Wang C, Cramer KM, Cheng H, Do KA. Associations between depression literacy and help-seeking behavior for mental health services among high school students. School Mental Health 2019 May 6;11(4):707-718. [doi: 10.1007/s12310-019-09325-1]
26. Raviv A, Raviv A, Vago-Gefen I, Fink AS. The personal service gap: factors affecting adolescents' willingness to seek help. J Adolesc 2009 Jun;32(3):483-499. [doi: 10.1016/j.adolescence.2008.07.004] [Medline: 18786722]
27. Slone M, Meir Y, Tarrasch R. Individual differences in referral for help for severe emotional difficulties in adolescence. Child Youth Serv Rev 2013 Nov;35(11):1854-1861. [doi: 10.1016/j.childyouth.2013.08.011]
28. Burns JR, Rapee RM. Adolescent mental health literacy: young people's knowledge of depression and help seeking. J Adolesc 2006 Apr;29(2):225-239. [doi: 10.1016/j.adolescence.2005.05.004] [Medline: 15996727]
29. Offer D, Howard KI, Schonert KA, Ostrov E. To whom do adolescents turn for help? Differences between disturbed and nondisturbed adolescents. J Am Acad Child Adolesc Psychiatry 1991 Jul;30(4):623-630. [doi: 10.1097/00004583-199107000-00015] [Medline: 1890097]
30. Jorm AF, Wright A, Morgan AJ. Beliefs about appropriate first aid for young people with mental disorders: findings from an Australian national survey of youth and parents. Early Interv Psychiatry 2007 Feb;1(1):61-70. [doi: 10.1111/j.1751-7893.2007.00012.x] [Medline: 21352109]
31. Link BG, Cullen FT, Struening E, Shrout PE, Dohrenwend BP. A modified labeling theory approach to mental disorders: an empirical assessment. Am Sociol Rev 1989 Jun;54(3):400. [doi: 10.2307/2095613]
32. Corrigan P. How stigma interferes with mental health care. Am Psychol 2004 Oct;59(7):614-625. [doi: 10.1037/0003-066X.59.7.614] [Medline: 15491256]
33. Nearchou FA, Bird N, Costello A, Duggan S, Gilroy J, Long R, et al. Personal and perceived public mental-health stigma as predictors of help-seeking intentions in adolescents. J Adolesc 2018 Jul;66:83-90. [doi: 10.1016/j.adolescence.2018.05.003] [Medline: 29800758]
34. Yap MB, Reavley NJ, Jorm AF. Associations between stigma and help-seeking intentions and beliefs: findings from an Australian national survey of young people. Psychiatry Res 2013 Dec 30;210(3):1154-1160. [doi: 10.1016/j.psychres.2013.08.029] [Medline: 24011848]
35. Yap MB, Reavley N, Jorm AF. Where would young people seek help for mental disorders and what stops them? Findings from an Australian national survey. J Affect Disord 2013 May;147(1-3):255-261. [doi: 10.1016/j.jad.2012.11.014] [Medline: 23228570]
36. Yap MB, Mackinnon A, Reavley N, Jorm AF. The measurement properties of stigmatizing attitudes towards mental disorders: results from two community surveys. Int J Methods Psychiatr Res 2014 Mar;23(1):49-61 [FREE Full text] [doi: 10.1002/mpr.1433] [Medline: 24550065]
37. Yap MB, Jorm AF. The influence of stigma on first aid actions taken by young people for mental health problems in a close friend or family member: findings from an Australian national survey of youth. J Affect Disord 2011 Nov;134(1-3):473-477. [doi: 10.1016/j.jad.2011.05.039] [Medline: 21658776]
38. Corrigan PW, Morris SB, Michaels PJ, Rafacz JD, Rüsch N. Challenging the public stigma of mental illness: a meta-analysis of outcome studies. Psychiatr Serv 2012 Oct;63(10):963-973. [doi: 10.1176/appi.ps.201100529] [Medline: 23032675]
39. Yamaguchi S, Mino Y, Uddin S. Strategies and future attempts to reduce stigmatization and increase awareness of mental health problems among young people: a narrative review of educational interventions. Psychiatry Clin Neurosci 2011 Aug;65(5):405-415 [FREE Full text] [doi: 10.1111/j.1440-1819.2011.02239.x] [Medline: 21851449]
40. Cairns K, Rossetto A. School-based mental health literacy interventions. In: Okan O, Bauer U, Levin-Zamir D, Pinheiro P, Sørensen K, editors. International Handbook of Health Literacy: Research, Practice and Policy across the Life-Span. Bristol, UK: Policy Press; 2019:1-768.
41. Kelly CM, Jorm AF, Wright A. Improving mental health literacy as a strategy to facilitate early intervention for mental disorders. Med J Aust 2007 Oct 01;187(S7):26-30. [doi: 10.5694/j.1326-5377.2007.tb01332.x] [Medline: 17908021]
42. Wei Y, Hayden JA, Kutcher S, Zygmunt A, McGrath P. The effectiveness of school mental health literacy programs to address knowledge, attitudes and help seeking among youth. Early Interv Psychiatry 2013 May;7(2):109-121. [doi: 10.1111/eip.12010] [Medline: 23343220]
43. Perry Y, Petrie K, Buckley H, Cavanagh L, Clarke D, Winslade M, et al. Effects of a classroom-based educational resource on adolescent mental health literacy: a cluster randomized controlled trial. J Adolesc 2014 Oct;37(7):1143-1151. [doi: 10.1016/j.adolescence.2014.08.001] [Medline: 25151646]
44. Hart LM, Morgan AJ, Rossetto A, Kelly CM, Mackinnon A, Jorm AF. Helping adolescents to better support their peers with a mental health problem: a cluster-randomised crossover trial of teen Mental Health First Aid. Aust N Z J Psychiatry 2018 Feb 01:4867417753552. [doi: 10.1177/0004867417753552] [Medline: 29417834]
45. Hart LM, Bond KS, Morgan AJ, Rossetto A, Cottrill FA, Kelly CM, et al. Teen Mental Health First Aid for years 7-9: a description of the program and an initial evaluation. Int J Ment Health Syst 2019;13:71 [FREE Full text] [doi: 10.1186/s13033-019-0325-4] [Medline: 31788023]

46. Hart LM, Cropper P, Morgan AJ, Kelly CM, Jorm AF. 'teen Mental Health First Aid' as a school-based intervention for improving peer support of adolescents at risk of suicide: outcomes from a cluster randomised crossover trial. Aust N Z J Psychiatry 2020 Apr;54(4):382-392. [doi: 10.1177/0004867419885450] [Medline: 31707787]
47. Hart LM, Mason RJ, Kelly CM, Cvetkovski S, Jorm AF. 'teen Mental Health First Aid': a description of the program and an initial evaluation. Int J Ment Health Syst 2016;10:3 [FREE Full text] [doi: 10.1186/s13033-016-0034-1] [Medline: 26788123]
48. Chisholm K, Patterson P, Torgerson C, Turner E, Jenkinson D, Birchwood M. Impact of contact on adolescents' mental health literacy and stigma: the SchoolSpace cluster randomised controlled trial. BMJ Open 2016 Mar 19;6(2):e009435 [FREE Full text] [doi: 10.1136/bmjopen-2015-009435] [Medline: 26895983]
49. Pinto-Foltz MD, Logsdon MC, Myers JA. Feasibility, acceptability, and initial efficacy of a knowledge-contact program to reduce mental illness stigma and improve mental health literacy in adolescents. Soc Sci Med 2011 Jun;72(12):2011-2019 [FREE Full text] [doi: 10.1016/j.socscimed.2011.04.006] [Medline: 21624729]
50. Shaffer DW, Squire KR, Halverson R, Gee JP. Video games and the future of learning. Phi Delta Kappan 2005;87(2):104-111. [doi: 10.1177/003172170508700205]
51. Granic I, Lobel A, Engels RC. The benefits of playing video games. Am Psychol 2014 Jan;69(1):66-78. [doi: 10.1037/a0034857] [Medline: 24295515]
52. Lenhart A, Kahne J, Middaugh E, Rankin MA, Evans C, Vitak J. Teens, video games, and civics: teens' gaming experiences are diverse and include significant social interaction and civic engagement. Pew Internet & American Life Project. 2008. URL: https://eric.ed.gov/?id=ED525058 [accessed 2021-11-30]
53. DeSmet A, Van Ryckeghem D, Compernolle S, Baranowski T, Thompson D, Crombez G, et al. A meta-analysis of serious digital games for healthy lifestyle promotion. Prev Med 2014 Dec;69:95-107. [doi: 10.1016/j.ypmed.2014.08.026] [Medline: 25172024]
54. Kato PM, Cole SW, Bradlyn AS, Pollock BH. A video game improves behavioral outcomes in adolescents and young adults with cancer: a randomized trial. Pediatrics 2008 Aug;122(2):305-317 [FREE Full text] [doi: 10.1542/peds.2007-3134] [Medline: 18676516]
55. Cangas AJ, Navarro N, Parra JMA, Ojeda JJ, Cangas D, Piedra JA, et al. Stigma-Stop: A Serious Game against the Stigma toward Mental Health in Educational Settings. Front Psychol 2017;8:1385 [FREE Full text] [doi: 10.3389/fpsyg.2017.01385] [Medline: 28878702]
56. Rüsch N, Angermeyer MC, Corrigan PW. Mental illness stigma: concepts, consequences, and initiatives to reduce stigma. Eur Psychiatry 2005 Dec;20(8):529-539. [doi: 10.1016/j.eurpsy.2005.04.004] [Medline: 16171984]
57. Ghoncheh R, Vos C, Koot H, Kerkhof A. Mental health online: a self- report and e-learning program for enhancing recognition, guidance and referral of suicidal adolescents. In: Mishara BL, Kerkhof AJ, editors. Suicide Prevention and New Technologies. London: Palgrave Macmillan; 2013:76-86.
58. Ghoncheh R, Gould MS, Twisk JW, Kerkhof AJ, Koot HM. Efficacy of adolescent suicide prevention e-learning modules for gatekeepers: a randomized controlled trial. JMIR Ment Health 2016 Jan 29;3(1):e8 [FREE Full text] [doi: 10.2196/mental.4614] [Medline: 26825006]
59. StataCorp. Stata Statistical Software: Release 14. College Station, TX: StataCorp LP; 2015.
60. Eldridge SM, Ashby D, Kerry S. Sample size for cluster randomized trials: effect of coefficient of variation of cluster size and analysis method. Int J Epidemiol 2006 Oct;35(5):1292-1300. [doi: 10.1093/ije/dyl129] [Medline: 16943232]
61. Hemming K, Girling AJ, Sitch AJ, Marsh J, Lilford RJ. Sample size calculations for cluster randomised controlled trials with a fixed number of clusters. BMC Med Res Methodol 2011 Jun 30;11:102 [FREE Full text] [doi: 10.1186/1471-2288-11-102] [Medline: 21718530]
62. Moving stories. IJsfontein. 2018. URL: https://youtu.be/2PRsq9GXVf4 [accessed 2021-11-30]
63. Dunn TJ, Baguley T, Brunsden V. From alpha to omega: a practical solution to the pervasive problem of internal consistency estimation. Br J Psychol 2014 Aug;105(3):399-412. [doi: 10.1111/bjop.12046] [Medline: 24844115]
64. McNeish D. Thanks coefficient alpha, we'll take it from here. Psychol Methods 2018 Sep;23(3):412-433. [doi: 10.1037/met0000144] [Medline: 28557467]
65. Hayes AF, Coutts JJ. Use omega rather than Cronbach's alpha for estimating reliability. But…. Commun Methods Meas 2020 Feb 11;14(1):1-24. [doi: 10.1080/19312458.2020.1718629]
66. Trizano-Hermosilla I, Alvarado JM. Best alternatives to Cronbach's alpha reliability in realistic conditions: congeneric and asymmetrical measurements. Front Psychol 2016;7:769 [FREE Full text] [doi: 10.3389/fpsyg.2016.00769] [Medline: 27303333]
67. Revelle W. psych: Procedures for Psychological, Psychometric, and Personality Research. Evanston, Illinois: Northwestern University; 2019.
68. Yap MB, Jorm AF. Young people's mental health first aid intentions and beliefs prospectively predict their actions: findings from an Australian National Survey of Youth. Psychiatry Res 2012 Apr 30;196(2-3):315-319. [doi: 10.1016/j.psychres.2011.10.004] [Medline: 22377574]
69. Wilson C, Deane F, Ciarrochi J, Rickwood D. Measuring help seeking intentions: properties of the general help seeking questionnaire. Can J Counsel 2005;39(1):15-28 [FREE Full text]

70. Boerema AM, Zoonen KV, Cuijpers P, Holtmaat CJ, Mokkink LB, Griffiths KM, et al. Psychometric properties of the Dutch Depression Stigma Scale (DSS) and associations with personal and perceived stigma in a depressed and community sample. PLoS One 2016;11(8):e0160740 [FREE Full text] [doi: 10.1371/journal.pone.0160740] [Medline: 27500969]
71. Jorm AF, Wright A. Influences on young people's stigmatising attitudes towards peers with mental disorders: national survey of young Australians and their parents. Br J Psychiatry 2008 Feb;192(2):144-149 [FREE Full text] [doi: 10.1192/bjp.bp.107.039404] [Medline: 18245033]
72. Griffiths KM, Christensen H, Jorm AF, Evans K, Groves C. Effect of web-based depression literacy and cognitive-behavioural therapy interventions on stigmatising attitudes to depression: randomised controlled trial. Br J Psychiatry 2004 Oct;185:342-349 [FREE Full text] [doi: 10.1192/bjp.185.4.342] [Medline: 15458995]
73. Timbremont B, Braet C, Roelofs J. Children's Depression Inventory: Handleiding. Amsterdam: Pearson; 2008.
74. Kovacs M. The Children's Depression Inventory (CDI). Psychopharmacol Bull 1985;21(4):995-998. [Medline: 4089116]
75. IBM Corp. SPSS Statistics for Windows, Version 25. Armonk, NY: IBM Corp; 2017.
76. Team R. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2018.
77. Muthén L, Muthén B. Mplus User's Guide. 8th ed. Los Angeles, CA: Muthén & Muthén; 2017.
78. Singer J, Willett J. Applied Longitudinal Data Analysis. Modeling Change and Event Occurrence. USA: Oxford University Press; 2003.
79. Gueorguieva R, Krystal JH. Move over ANOVA: progress in analyzing repeated-measures data and its reflection in papers published in the Archives of General Psychiatry. Arch Gen Psychiatry 2004 Mar;61(3):310-317. [doi: 10.1001/archpsyc.61.3.310] [Medline: 14993119]
80. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
81. Barr DJ, Levy R, Scheepers C, Tily HJ. Random effects structure for confirmatory hypothesis testing: keep it maximal. J Mem Lang 2013 Apr;68(3):255-278 [FREE Full text] [doi: 10.1016/j.jml.2012.11.001] [Medline: 24403724]
82. Tuijnman A, Kleinjan M, Olthof M, Hoogendoorn E, Granic I, Engels R. Open materials for "Effectiveness of a Game-Based School Program for Mental Health Literacy and Stigma on Depression (Moving Stories): A Cluster Randomized Controlled Trial". OSF. 2021. URL: https://osf.io/9hkep/ [accessed 2021-12-01]
83. Bland JM, Altman DG. Regression towards the mean. Br Med J 1994 Jun 04;308(6942):1499 [FREE Full text] [doi: 10.1136/bmj.308.6942.1499] [Medline: 8019287]
84. Knaak S, Modgill G, Patten SB. Key ingredients of anti-stigma programs for health care providers: a data synthesis of evaluative studies. Can J Psychiatry 2014 Oct;59(10 Suppl 1):19-26 [FREE Full text] [doi: 10.1177/070674371405901s06] [Medline: 25565698]
85. National Academies of Sciences, Engineering, and Medicine. Ending Discrimination Against People with Mental and Substance Use Disorders: The Evidence for Stigma Change. Washington (DC): National Academies Press; 2016:1-170.
86. Corrigan PW. Best practices: Strategic stigma change (SSC): five principles for social marketing campaigns to reduce stigma. Psychiatr Serv 2011 Aug;62(8):824-826. [doi: 10.1176/ps.62.8.pss6208_0824] [Medline: 21807820]
87. Scholten H, Granic I. Use of the principles of design thinking to address limitations of digital mental health interventions for youth: viewpoint. J Med Internet Res 2019 Jan 14;21(1):e11528 [FREE Full text] [doi: 10.2196/11528] [Medline: 31344671]
88. Wei Y, McGrath PJ, Hayden J, Kutcher S. Mental health literacy measures evaluating knowledge, attitudes and help-seeking: a scoping review. BMC Psychiatry 2015 Nov 17;15:291 [FREE Full text] [doi: 10.1186/s12888-015-0681-9] [Medline: 26576680]
89. Pérez RG, Ezpeleta L, Domenech JM. Features associated with the non-participation and drop out by socially-at-risk children and adolescents in mental-health epidemiological studies. Soc Psychiatry Psychiatr Epidemiol 2007 Mar;42(3):251-258. [doi: 10.1007/s00127-006-0155-y] [Medline: 17273807]
90. Anstey KJ, Hofer SM. Longitudinal designs, methods and analysis in psychiatric research. Aust N Z J Psychiatry 2004 Mar;38(3):93-104. [doi: 10.1080/j.1440-1614.2004.01343.x] [Medline: 14961925]
91. Fröjd SA, Kaltiala-Heino R, Marttunen MJ. Does problem behaviour affect attrition from a cohort study on adolescent mental health? Eur J Public Health 2011 Jun;21(3):306-310. [doi: 10.1093/eurpub/ckq078] [Medline: 20570960]

Abbreviations

CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth

Edited by J Torous; submitted 18.01.21; peer-reviewed by Y Perry, D Romer, Z Feng; comments to author 15.03.21; revised version received 31.05.21; accepted 28.07.21; published 17.08.22.

Please cite as:
Tuijnman A, Kleinjan M, Olthof M, Hoogendoorn E, Granic I, Engels RCME
A Game-Based School Program for Mental Health Literacy and Stigma on Depression (Moving Stories): Cluster Randomized Controlled Trial
JMIR Ment Health 2022;9(8):e26615
URL: https://mental.jmir.org/2022/8/e26615
doi:10.2196/26615
PMID:35976200

Original Paper

A Web-Based Application for Personalized Ecological Momentary Assessment in Psychiatric Care: User-Centered Development of the PETRA Application

Fionneke M Bos[1,2], PhD; Lino von Klipstein[2], MSc; Ando C Emerencia[3], PhD; Erwin Veermans[1], MSc; Tom Verhage[2], BSc; Evelien Snippe[2], PhD; Bennard Doornbos[4], MD, PhD; Grietje Hadders-Prins[1]; Marieke Wichers[2], PhD; Harriëtte Riese[2], PhD

[1]Rob Giel Research Center, Department of Psychiatry, University Medical Center Groningen, University of Groningen, Groningen, Netherlands

[2]Interdisciplinary Center Psychopathology and Emotion Regulation, Department of Psychiatry, University Medical Center Groningen, University of Groningen, Groningen, Netherlands

[3]Research Support, Faculty of Behavioral and Social Sciences, University of Groningen, Groningen, Netherlands

[4]Lentis Research, Groningen, Netherlands

Corresponding Author:
Fionneke M Bos, PhD
Rob Giel Research Center
Department of Psychiatry
University Medical Center Groningen, University of Groningen
Hanzeplein 1
Groningen, 9700 RB
Netherlands
Phone: 31 50 361 4860
Email: f.m.bos01@umcg.nl

Abstract

Background: Smartphone self-monitoring of mood, symptoms, and contextual factors through ecological momentary assessment (EMA) provides insights into the daily lives of people undergoing psychiatric treatment. Therefore, EMA has the potential to improve their care. To integrate EMA into treatment, a clinical tool that helps clients and clinicians create personalized EMA diaries and interpret the gathered data is needed.

Objective: This study aimed to develop a web-based application for personalized EMA in specialized psychiatric care in close collaboration with all stakeholders (ie, clients, clinicians, researchers, and software developers).

Methods: The participants were 52 clients with mood, anxiety, and psychotic disorders and 45 clinicians (psychiatrists, psychologists, and psychiatric nurses). We engaged them in interviews, focus groups, and usability sessions to determine the requirements for an EMA web application and repeatedly obtained feedback on iteratively improved high-fidelity EMA web application prototypes. We used human-centered design principles to determine important requirements for the web application and designed high-fidelity prototypes that were continuously re-evaluated and adapted.

Results: The iterative development process resulted in Personalized Treatment by Real-time Assessment (PETRA), which is a scientifically grounded web application for the integration of personalized EMA in Dutch clinical care. PETRA includes a decision aid to support clients and clinicians with constructing personalized EMA diaries, an EMA diary item repository, an SMS text message–based diary delivery system, and a feedback module for visualizing the gathered EMA data. PETRA is integrated into electronic health record systems to ensure ease of use and sustainable integration in clinical care and adheres to privacy regulations.

Conclusions: PETRA was built to fulfill the needs of clients and clinicians for a user-friendly and personalized EMA tool embedded in routine psychiatric care. PETRA is unique in this codevelopment process, its extensive but user-friendly personalization options, its integration into electronic health record systems, its transdiagnostic focus, and its strong scientific foundation in the design of EMA diaries and feedback. The clinical effectiveness of integrating personalized diaries via PETRA into care requires further research. As such, PETRA paves the way for a systematic investigation of the utility of personalized EMA for routine mental health care.

(JMIR Ment Health 2022;9(8):e36430) doi:10.2196/36430

KEYWORDS

eHealth; clinical implementation; ecological momentary assessment; experience sampling method; smartphone; mobile health; mHealth; personalized diaries; personalized psychiatry; client-tailored; cocreation; shared decision-making; mobile phone

Introduction

Background

Ecological momentary assessment (EMA; also referred to as the experience sampling method or ambulatory monitoring) [1], is an increasingly popular self-monitoring method to gain insight into daily life experiences relevant to mental health problems. With EMA, clients monitor their daily life *as it is lived* via short questionnaires (ie, diaries) on their smartphones. Several times per day, they rate their experienced symptoms, affect, thoughts, activities, and (social) contexts [2]. In research, EMA has already been applied to improve our understanding of mental disorders by examining how symptoms fluctuate in daily life, in relation to stressful and rewarding events, and in a diverse range of mental disorders [3-6]. Increasingly, researchers and scientists are recognizing the potential of EMA as a tool for improving mental health care [7-10]. Self-monitoring through EMA and visualizing the data in EMA feedback may offer clients and clinicians relevant insights into the person-specific mechanisms that contribute to clients' mental health problems [8].

EMA as a Self-monitoring Tool

In many medical disciplines, self-monitoring of, for example, daily blood pressure or glucose levels, is already a common practice and is informative for diagnosis and treatment planning [11,12]. Similarly, mental health treatment may benefit from the self-monitoring of daily symptoms, (social) contextual experiences, and treatment-related factors [13]. Indeed, for this reason, many mental health treatments such as cognitive behavioral therapy already use some form of (paper-and-pencil) self-monitoring [14], and routine outcome monitoring is well established [15]. However, EMA is both more detailed and more frequent than existing monitoring methods. Typically, EMA assesses momentary experiences ("right now, I feel cheerful"), which can be rated on a visual analog scale ranging from 0 to 100 ("not at all" to "very much," respectively). In mental health care, momentary experiences can pertain to symptoms (eg, sadness, restlessness, or hearing voices) or other factors relevant to treatment (eg, harmful or helpful thoughts and behaviors). The number of EMA assessments per day, as well as the timing and duration of the assessments, depend on the goal of the EMA diary and the experienced burden of completing the diaries. Most EMA studies thus far have used 3 to 10 assessments per day for 1 to 2 weeks [16,17], whereas some have monitored for up to 4 months [18].

Clinical Promise of EMA

Self-monitoring through EMA is suggested to improve clients' self-management by increasing their insights into their well-being. By frequently reflecting on one's symptoms and learning what types of activities or situations positively or negatively influence well-being, clients may become more in control [19]. As such, EMA self-monitoring may be an intervention in itself. In addition, EMA data can be visualized to enhance the understanding of clients' personal mechanisms that contribute to mental health problems. Such EMA feedback could be as simple as demonstrating variability in affect or showing contexts in which clients have complaints. As EMA is suggested to be unaffected by memory biases [20], such feedback could provide a more reliable overview of how clients fared between treatment sessions. This could improve shared decision-making and the therapeutic alliance (ie, the client-clinician relationship, involving an agreement on the goals and tasks of treatment, as well as their interpersonal bonds [21]) [10,19]. Therefore, EMA feedback could form the basis of a more collaborative approach to diagnosis and intervention, where clients and therapists use EMA data to decide together on the next treatment step.

Empirical Evidence for Clinical Effectiveness of EMA

The first empirical investigations into the clinical effectiveness of EMA are promising. In qualitative research, clients and clinicians describe the beneficial effects of EMA on client self-management, therapeutic alliance, and treatment effectiveness [22-25]. This is supported by a clinical trial demonstrating that adding EMA monitoring and feedback to antidepressant treatment was more effective in reducing depressive symptoms than using antidepressant medication alone [26], and it improved clients' feelings of empowerment [27], emotion differentiation [28], and behavioral change [29]. However, 2 other clinical trials did not replicate these beneficial effects in individuals with depression in routine care [30] and individuals reporting a loss of interest and pleasure (anhedonia) [31]. This discrepancy may be because in clinical trials thus far, EMA diaries have been standardized for all clients. This insufficiently relates EMA feedback to clients' treatment goals, making the EMA less effective, and may increase the burden for clients, who have to frequently answer questions that are irrelevant to their current situation [32]. Therefore, personalization of the EMA seems to be a core requirement for successful clinical implementation.

Developing Web-Based EMA Technology

Given that clients, clinicians, and researchers align in their positive evaluation of the clinical utility of EMA, the field needs a flexible, user-friendly, and evidence-based application for the use of personalized EMA diaries in mental health care. More specifically, this entails the development of a digital infrastructure that enables clients and clinicians to intuitively construct personalized EMA diaries, deliver them to clients' smartphones, and dynamically visualize the gathered EMA data.

In this paper, we describe the development of Personalized Treatment by Real-time Assessment (PETRA), a web-based application that enables the use of personalized EMA diaries in Dutch mental health care. PETRA was built into electronic

health record (EHR) systems to ensure sustainability and easy access. The web application comprises a decision aid to help clients and clinicians construct personalized EMA diaries, an EMA diary item repository, a diary delivery system, and a feedback module to visualize the gathered EMA data. PETRA was developed in close collaboration with clients and clinicians as they will be the primary users. Involving them early in the development process is crucial to ensure the uptake of PETRA.

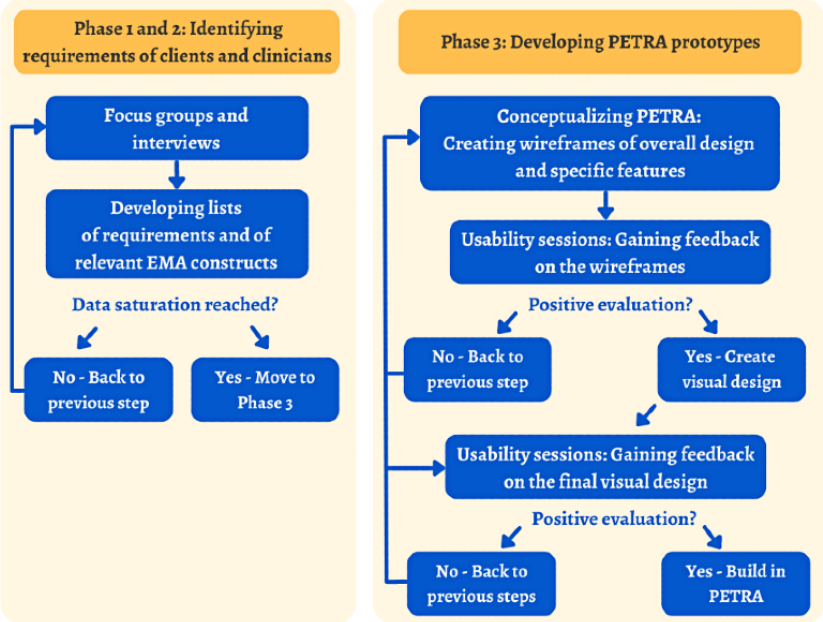
Methods

Study Design

PETRA was developed using the road map of the Center for eHealth Research (CeHRes) of the University of Twente, the Netherlands [33]. The road map comprises 5 iterative phases. In the first 2 phases of the CeHRes road map (contextual inquiry and value specification), we conducted focus groups and interviews to identify the needs and requirements of the relevant stakeholders, clients and clinicians, for a personalized diary web application [34]. The next phase (design) comprised an iterative design process in which prototypes were developed, evaluated, and adapted frequently based on feedback from clients and clinicians [35]. In this paper, we describe the results of these 3 phases and the resulting PETRA web application. The final 2 phases of the CeHRes road map (operationalization and summative evaluation) will involve implementing the web application and evaluating its use in practice.

During each of these phases, stakeholders were actively involved to ensure their perspectives are integrated into PETRA. The CeHRes road map follows human-centered design principles, attempting to understand the needs of users and actively involving them in designing solutions to meet these needs [36]. Therefore, the PETRA project team was multidisciplinary and comprised the core scientific team; a scientific programmer; the RoQua development team to ensure sustainable embedding in existing mental health care software; a user experience (UX) designer to integrate the needs of all stakeholders; and client and clinician representatives engaged in interviews, focus groups, and usability sessions. All team members critically reviewed the prototypes, which were then continuously adapted and re-evaluated (Figure 1).

Figure 1. Overview of the developmental phases of PETRA (Personalized Treatment by Real-time Assessment). EMA: ecological momentary assessment.



Setting and Participants

The PETRA web application was developed from 2016 to 2022 at the Department of Psychiatry, University Medical Center Groningen (the Netherlands), facilitated by iLab, a collaborative initiative to implement scientific innovations into clinical practice [37]. PETRA's software was built into RoQua [38], a nonprofit web-based questionnaire system for research and care. RoQua is fully integrated into the EHR systems of several mental health care facilities in the Netherlands, which are part of the Rob Giel Research Center (RGOc) [39]. The facilities participating in this collective mostly provide secondary or tertiary (specialist) mental health care services to clients presenting with a diverse range of mental disorders, such as major depression, bipolar disorder, anxiety disorders, and psychosis.

Participant Recruitment

Participants comprised a convenience sample of clients and clinicians recruited from RGOc facilities, who took part in interviews, focus groups, and usability sessions. All clients with a (self-reported) history of mental health problems (regardless of type) were eligible to participate in the study. Similarly, all clinicians working in an RGOc facility were eligible. Clinicians could be psychiatrists, psychologists, or psychiatric nurses. Participant characteristics for the first 2 phases can be found in the studies by Bos et al [22,23] but were not systematically collected in the design phase. The clients were informed of the PETRA project by the RGOc coordinator of the client representatives or their clinician. If interested, clients received a flyer and, after providing consent, were invited for an interview, focus group, or usability session. The clients were

reimbursed for their time (€10 [US $10.18] per hour). Clinicians were invited based on prior participation in ongoing diary studies or their expressed interest in PETRA during (department) presentations. Both clients and clinicians were invited to join multiple sessions to offer feedback on the adaptations that were made.

Ethics Approval

The University Medical Center Groningen Medical Ethics Committee exempted this research from a full review (reference number 201900401). Clients signed a written informed consent form to participate in the focus groups, interviews, and usability sessions.

Phases 1 and 2: Identifying the Requirements of Clients and Clinicians

Phases 1 and 2 aimed to identify perceived problems in the status quo of mental health treatment, perceived advantages and challenges of personalized electronic diaries, and core requirements for the EMA web application to be developed. In-depth qualitative interviews and focus groups were conducted with 40 clients and 27 clinicians between June 2016 and March 2018 on the perceived utility of EMA for psychiatric care and perceived important requirements for a clinical EMA tool. Focus groups with clients or clinicians were conducted in groups of 3 to 7 participants until data saturation was reached (ie, no new themes emerged). The focus groups and interviews were part of 2 qualitative studies that broadly focused on identifying applications of EMA in diverse clinical contexts and implementation requirements. For more details, we refer the reader to previous studies [22,23] (for interview guides, see Table S1 in Multimedia Appendix 1). In this paper, we specifically discuss the resulting recommendations for the development of an EMA tool on which PETRA was based. Furthermore, in this phase, we compiled a list of potentially relevant constructs to assess personalized EMA diaries. On the basis of previous EMA research and subsequent suggestions by clients and clinicians, each construct was linked to related EMA diary items (eg, “I feel anxious” for the construct “Anxiety”).

Phase 3: Development of PETRA Prototypes

Overview

On the basis of the recommendations in phases 1 and 2, we designed prototypes for the web-based personalized EMA tool PETRA. These were continuously tested and redesigned in usability sessions with clients and clinicians held from May 2019 to August 2021. Owing to the COVID-19 pandemic, all usability sessions after March 2020 were conducted via video calls (for interview guides, see Table S1 in Multimedia Appendix 1). At each session, the researcher FB and UX designer TV were present. After each session, they wrote a report with the main recommendations. The participants were invited to provide feedback on this report.

Sessions With Clinicians

On the basis of phases 1 and 2, clinicians were considered the primary users of the decision aid and feedback module: they would need to integrate PETRA into their working routines and introduce it to clients. Therefore, most of the usability sessions were held with clinicians. A total of 18 clinicians participated in ≥1 of the 31 usability sessions (5/18, 28% were men). Approximately 77% (24/31) of the sessions were held individually, and 23% (7/31) of the sessions took place in groups of 2 to 3 clinicians. Most clinicians were psychologists (10/18, 56%), followed by psychiatrists (4/18, 22%) and psychiatric nurses (4/18, 22%). Of the 18 clinicians, 7 (39%) participated once, 8 (44%) participated twice, and 3 (17%) participated thrice. During each session, clinicians were asked to describe their thoughts and expectations regarding EMA and their requirements for a diary web application. Subsequently, the clinicians were walked through a digital prototype of the PETRA web application. They were encouraged to describe what they saw on the screen and provide their initial thoughts on each page (the think-aloud method). When the clinicians offered suggestions for improvement, follow-up questions were asked until an in-depth understanding of the proposed adaptation was reached. The clinicians were also invited to explain how they would use PETRA in treatment and to suggest new features or feedback types.

Sessions With Clients

A total of 12 clients participated in ≥1 of the 4 usability sessions, which were similar to those with the clinicians. Clients were currently or had been previously in treatment for mood, psychotic, or anxiety disorders (8/12, 67% were men). Of the 12 clients, 9 (75%) participated in 1 usability session, and 3 (25%) participated in 2 sessions. The sessions took place in groups of 3 to 7 clients. One of the sessions specifically focused on designing low-fidelity (paper-and-pencil) prototypes for EMA diary feedback. The other usability sessions were used to obtain feedback on the digital high-fidelity prototypes of the PETRA web application. Feedback was sought on the workflow and its intended use in treatment.

Data Analysis

The interviews and focus groups were transcribed and analyzed thematically using the Qualitative Analysis Guide of Leuven [40]. Briefly, this approach involves the identification of important themes within qualitative data, which are then iteratively verified against the transcripts. Further details are provided elsewhere [22,23]. The resulting themes regarding the requirements of a clinical EMA tool are described in the Results section.

After each usability session, wireframes were created with the proposed changes to the PETRA web application (see Figures 2-4 for example wireframes or a slide-by-slide overview on the PETRA website [41]). A wireframe is a blueprint that explicates the flow throughout the web application. In total, 9 low-fidelity wireframes of the overarching structure and workflow of PETRA were designed, each version improving upon the previous one. The final ninth wireframe was used for the next stage, which focused on determining the types of content, visuals, and graphs depicted on the screens. This stage comprised 15 high-fidelity wireframes, which were again improved iteratively.

Figure 2. The first wireframe that was designed of the first page of the decision aid, on which participants select the goal of the EMA diary. The usability sessions indicated a need for an introduction and a more detailed progress menu. More detailed wireframes can be found here [41]. EMA: ecological momentary assessment.

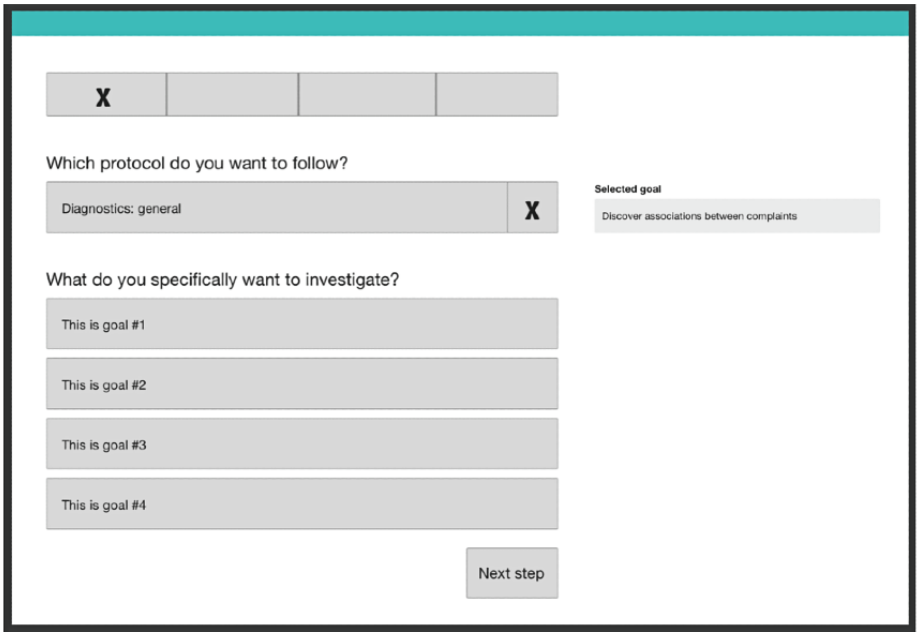


Figure 3. The ninth wireframe that was designed of the first page of the decision aid, on which participants select the goal of the EMA diary. As suggested in usability sessions, it incorporates an introduction and a more detailed progress menu. More detailed wireframes can be found here [41]. EMA: ecological momentary assessment.

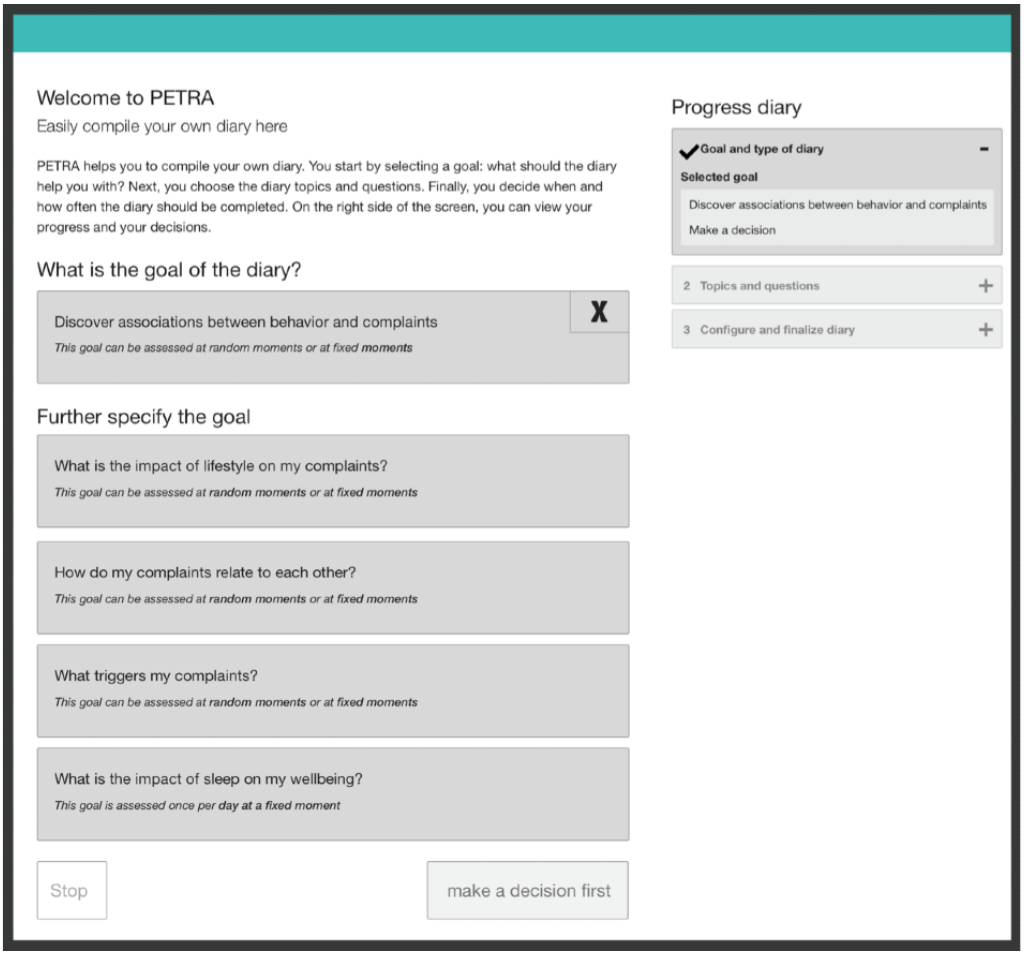
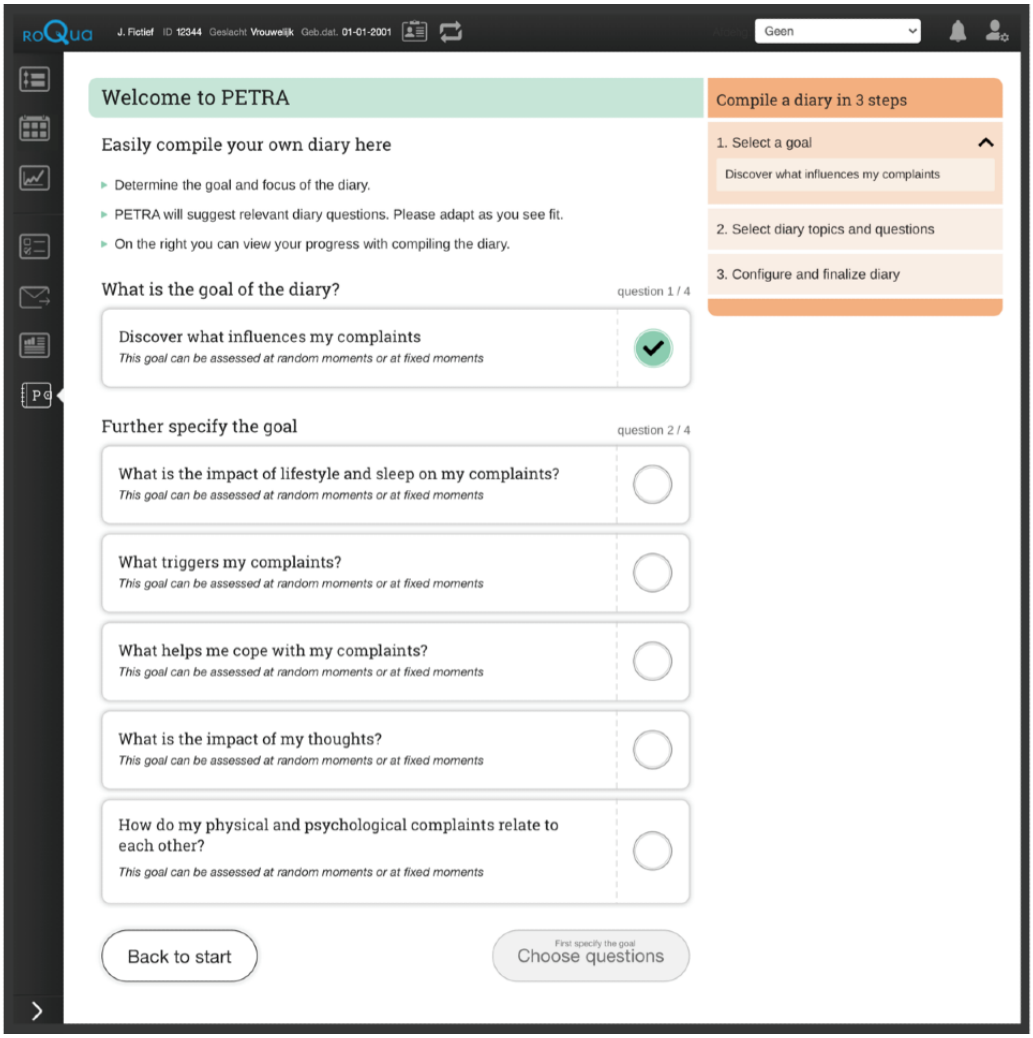


Figure 4. The final (visual) design of the first page of the decision aid, on which participants select the goal of the EMA diary. More detailed wireframes can be found here [41]. EMA: ecological momentary assessment.



Results

Phases 1 and 2: Core Requirements of PETRA

Overview

The first 2 phases of the CeHRes road map resulted in an in-depth overview of clients' and clinicians' perceptions of perceived problems in the status quo of mental health treatment, perceived advantages and challenges of personalized electronic diaries, and core requirements for the EMA web application to be developed (Table 1). The core requirements were (1) extensive possibilities for personalization, (2) resource efficiency, (3) scientifically grounded and provided by a trusted source, and (4) meeting the demands regarding privacy and safe data storage. Relevant quotes can be found in Table S2 in Multimedia Appendix 1.

Table 1. Results of phases 1 and 2 of the Center for eHealth Research road map in the development of the PETRA (Personalized Treatment by Real-time Assessment) application.

Theme	Description
Problems clients and clinicians identified in the status quo of mental health treatment	
Reliability or efficiency of current monitoring instruments	• Often paper-and-pencil based and deemed less reliable and efficient • Often focus too much on symptoms and fail to take important contextual factors and strengths into account • Usually only administered 1 to 2 times per month; missing relevant information about clients' daily lives throughout the day • Usually not person-specific enough to be directly relevant to clients
Insights	• Limited insight into overall progress and treatment outcomes • Limited insight into the effects of treatment or lifestyle adaptations • Limited insight into the frequency and severity of symptoms and when they occur • Limited insight into triggers and relapse signals • Recall of well-being in between sessions is biased by current mood or otherwise difficult to recall for clients (recall bias)
Therapeutic alliance	• Knowledge imbalance between client and clinician, weakening the therapeutic alliance • Current monitoring instruments are often not discussed in treatment, weakening the therapeutic alliance • Limited contact between client and clinician in between sessions • Client does not believe in the current treatment approach as the effects are not clearly visible
Perceived advantages of personalized electronic diaries	
Reliability or efficiency	• More reliable and efficient as assessments take place multiple times per day via smartphone and are less easily forgotten or ignored
Insight	• Offers insights into progress, treatment effects, and the flow of symptoms throughout daily life, thereby increasing client self-management • Offers insights into contextual factors and personal strengths
Therapeutic alliance	• Client and clinicians share the same information and can, therefore, collaborate more easily • Relevant diary feedback can more easily be integrated in treatment, strengthening the therapeutic alliance and clients' trust in their clinician
Perceived challenges in constructing personalized electronic diaries	
Diary construction	• What kind of clinical questions can be answered with electronic diaries? • How to formulate or select relevant and valid diary items? • How to determine the number of assessments per day, balancing client burden and the number needed to answer a clinical question? • How to determine the necessary diary schedule (eg, time-contingent or event-contingent) for a clinical question? • How to make sure the diary maximizes its advantages and minimizes disadvantages?
Diary feedback	• How to automatically analyze and visualize the diary data without the need for statistical knowledge of the clinician? • How to interpret the diary feedback in a clinical context?

Personalization

All clients and clinicians stressed the necessity for extensive personalization of the diary content, schedule, and duration. They suggested that clients and clinicians should be able to compile different EMA items based on the clients' current care needs in a flexible manner. This means that EMA diaries should be transdiagnostic and adaptable in diverse stages of treatment. The alternative, where diaries are constructed based on client diagnosis, was viewed as limiting, more burdensome, and not in line with the real-world situation where many clients present with symptoms of multiple diagnoses. Furthermore, EMA diaries should not only focus on symptoms but also on personal strengths and (social) contextual factors. This need for personalization has also been highlighted in research on monitoring in general [32,42,43].

Resource Efficiency

Both clients and clinicians emphasized that the use of EMA in treatment should be easy, time efficient, and fit with the existing workflow of clinicians. This means that constructing personalized EMA diaries and interpreting the resulting feedback should be intuitive and cost clients and clinicians limited time, effort, and resources. This corresponds with research demonstrating that eHealth is most useful if it matches client and clinician expectations [44]. Thus, the web application should be largely automated and integrated into existing digital infrastructures used in mental health care, preferably the clinician's EHR system.

Scientific Background

Clients and clinicians indicated that they found it difficult to assess the effectiveness of eHealth applications, which

corresponds to eHealth research [45,46]. Therefore, they stressed the need for a web application that is scientifically grounded so that the resulting diary and feedback are valid, trustworthy, and according to scientific standards for EMA [16]. Furthermore, as clinicians are not primarily trained in the innovative methods and statistics necessary for constructing and analyzing diary data, they indicated that the web application should help them with constructing diaries and interpreting feedback.

Privacy

The final core requirement clients and clinicians mentioned was that the gathered diary data should be safely stored, protecting the privacy of clients, according to data protection laws (eg, the General Data Protection Regulation [GDPR] law of the European Union). Given the far-reaching personalization of the diaries, the data should be treated accordingly.

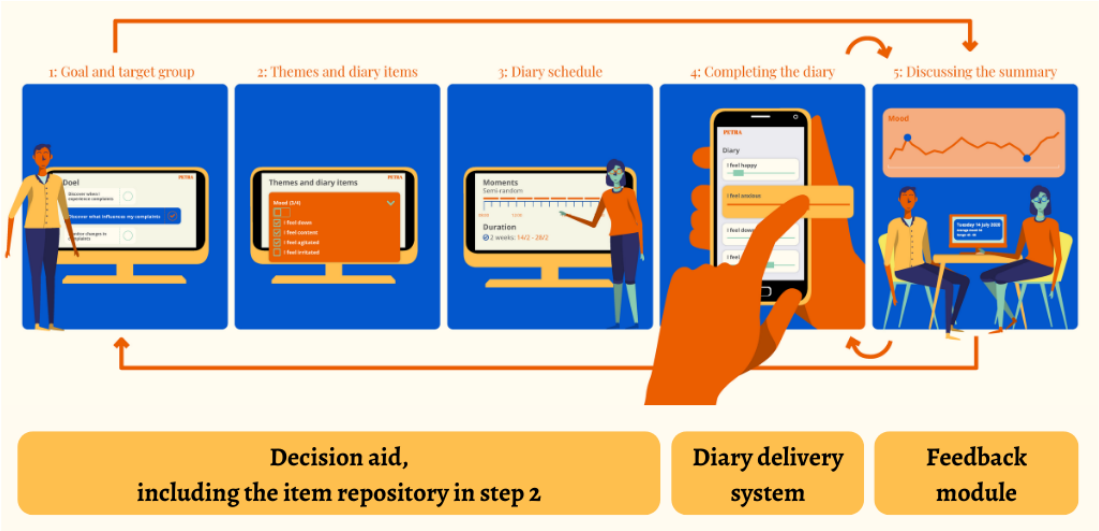
Phase 3: Design of PETRA

Overview

The core requirements for a web application for personalized EMA diaries were translated into the current design of PETRA. PETRA was developed as a web-based application, built into EHR systems via RoQua. As per the suggestion of clinicians, this means that clients and clinicians do not need additional log-in data, and clients do not need to install an app on their phones, allowing easy access for both. PETRA was developed for adult clients diagnosed with mood, anxiety, or psychotic disorders receiving treatment in specialized mental health care. Proficiency in the Dutch language and client possession of a smartphone with internet access are required to use PETRA.

PETRA comprises four main parts: (1) a decision aid, (2) an item repository of diary items, (3) an SMS text message–based diary delivery system, and (4) a feedback module. An overview of the various parts of PETRA is shown in Figure 5. Each part was iteratively evaluated and adapted based on the feedback of clients and clinicians, who mostly agreed on the design of PETRA's features. PETRA was designed to be used collaboratively: clients and clinicians together decide on the goal, content, and schedule of the EMA diary and interpret the feedback together. Clients have access to the decision aid and feedback modules via their clinician. PETRA's steps are outlined in more detail in the following sections.

Figure 5. Overview of PETRA.



Decision Aid

A decision aid was developed to assist clients and clinicians in constructing personalized EMA diaries. This decision aid is intended to be used collaboratively by clients and clinicians. The following specifications can be personalized via the decision aid: (1) the content (or diary items), (2) the number of assessments per day, (3) the diary schedule (timing of the assessments), and (4) the length of the EMA monitoring period. These specifications depend on the intended goal of the diary, anticipated burden for the client, and the client's main symptom profile.

PETRA's decision aid was developed based on the results of phases 1 and 2 and guidelines for designing EMA studies [16,47]. First, we developed a list of clinical goals for which personalized EMA may be insightful. Such goals include, for example, gaining insights into the effects of lifestyle on well-being or monitoring the effects of medication. Each goal is linked to potentially relevant EMA items, a preferred number of assessments, a preferred diary schedule, and a preferred diary period (Tables S3-S5 in Multimedia Appendix 1). This matching process was based on clinical EMA research demonstrating the requirements for different types of research questions. For example, if gaining insights into contextual factors is the primary goal, the best practice in EMA research is to opt for a semirandom diary schedule, in which the assessments are randomly generated within multiple predefined time blocks. This is generally considered the most ecologically valid way of gaining an overview of someone's activities and events during the day [48].

In addition to the goal, clients and clinicians select the symptom profile they wish to focus on. We developed a list of prevalent symptom profiles in specialist mental health care (eg, depression, anxiety, and psychosis). Each symptom profile was linked to potentially relevant EMA items. For example, relevant

constructs for clients in depression treatment could be mood, anhedonia, worry, self-esteem, and resilience.

The decision aid first prompts clients and clinicians to decide on the clinical goal, symptom profile, and preferred diary schedule (see Figure 5 or the PETRA website for a video [49]). In the next 2 steps, based on these decisions, PETRA automatically preselects the relevant EMA items and proposes a diary schedule. This preselection can be adjusted according to personal preferences while remaining within valid boundaries (eg, if the goal is to assess the relationship between activities and symptoms, only a semirandom schedule of at least five assessments per day can be selected [48]). Clients and clinicians can deselect irrelevant items or select additional items. The PETRA decision aid further includes a preview function, which visualizes the diary as the patient views it on their phone, and a *burden indicator*, which is a gradient bar that changes color according to the number of items in the diary to remind clients and clinicians to discuss the burden when constructing diaries (S6 and S7 in Multimedia Appendix 1 [16,17,47,50,51]). Although research currently offers no clear-cut guidelines on the diary length and associated burden [17,50], a study demonstrated that 60 items per diary were perceived as more burdensome than 30 [51], and another study showed that EMA studies, on average, comprised 30 items [16]. It is likely that the number of items and associated burden differs between, as well as within, individuals and across populations and treatment stages. Therefore, the burden indicator was based on our experience in designing EMA studies and will be updated based on the UXs and emerging research.

EMA Item Repository

The EMA item repository was developed as clients and clinicians indicated a preference for a *menu* in which they could select clinically relevant constructs (eg, mood or hearing voices) that automatically tie into preworded items. Although the EMA field is yet to establish the validity of EMA items, best practices have been developed in EMA research on the wording of EMA items [16,47,50]. The PETRA items were developed based on these best practices.

The EMA item repository was developed by the scientific team based on extensive experience with the items used in EMA studies and subsequent evaluations and suggestions by clients and clinicians. First, we determined potentially relevant constructs to be assessed; that is, constructs pertaining to symptoms, strengths, and contextual factors. For each construct, we formulated 4 to 10 diary items.

This resulted in four item sets based on the EMA assessment schedule: (1) at semirandom moments (128 items), (2) at fixed moments (126 items), (3) once daily in the evening (62 items), and (4) once daily in the morning (18 items). EMA assessments can occur either at fixed time points (eg, at noon) or at random points in predefined intervals (eg, somewhere between 10:30 AM and noon), termed fixed or semirandom EMA designs, respectively. Fixed designs are believed to be less burdensome for clients, whereas semirandom designs supposedly provide a more representative overview of daily experiences [2,48]. The item sets were developed as experiences can be expected to operate on different timescales and, therefore, require different wording. For example, mood usually fluctuates throughout the day and requires multiple assessments for valid estimation. However, excessive buying, a symptom of mania, may only vary from day to day and can be sufficiently assessed only in the evening. Most items are answered on 0 to 100 visual analog scales (ranging from “not at all” to “very much”), some via selecting categorical multiple-choice checkboxes (eg, indicating activities, coping strategies, or locations), and some via entering text (eg, descriptions of events). The full lists of EMA items are available in item S9 in Multimedia Appendix 1 [16-18,22,23,26,37,42,47,50-72] (Bos, F, unpublished data, May 2022; Jenner, JGL, unpublished data, October 2006; Bringmann, L, unpublished data, January 2021). In addition to the item sets, clients and clinicians can also formulate a limited number of person-specific items.

Semirandom and fixed diary schedules are mutually exclusive. Once-daily items are often an informative addition to these more frequent assessment schedules. Therefore, in the decision aid, the constructs of the once-per-day schedules are integrated within the construct structure of the semirandom and fixed schedules. This ensures that clients and clinicians can select items based on the construct. For example, the construct *(hypo)mania* provides items assessed in the moment (semirandom or fixed), as well as morning and evening items. Although the EMA items are grouped into constructs in the decision aid, they are reordered in the actual diary to balance positive and negative items and ensure that more fleeting momentary emotions are assessed before contextual experiences [16,50].

EMA Delivery System

Once clients and clinicians have constructed the EMA diary, clients receive SMS text messages on their personal smartphones. The messages are sent according to the chosen diary schedules and contain a link to the EMA diary items, which are presented and filled in via the browser. Clients receive reminders if the diary has not been completed within 20 minutes, and the link is disabled after 30 minutes (semirandom and fixed schedules) or 3 hours (once-daily schedules).

Feedback Module

The gathered EMA data for each client are visualized in the PETRA feedback module, which clients and their clinicians can discuss during a regular treatment session (see Figures 6-8 or the PETRA website for a video [49]). This module is grouped around three themes identified by clients and clinicians as relevant to psychiatric treatment: the fluctuations and changes in symptoms and strengths over time, associations between (social) contexts and symptoms and strengths, and the impact of pleasant and unpleasant events (see item S8 in Multimedia Appendix 1 for more details). The early versions of the feedback module were based on the feedback module developed in the Therap-i trial [52]. The PETRA feedback module is updated in real time and offers dynamic feedback options, and each graph can be adapted and can display any of the assessed EMA items.

Clients and clinicians indicated that the understanding of EMA data is greatly facilitated by qualitative descriptions of the context provided by clients. Therefore, clients are actively

encouraged to provide such descriptions. In the feedback module, these qualitative evaluations can be linked to high and low scores to provide contextual information on moments the client experiences high or low symptoms. Furthermore, all entered text is summarized in word clouds to intuitively summarize the main themes that emerged from the qualitative data and provide context for when these themes emerged. The PETRA feedback is descriptive as statistical analysis of diary data was considered too complex, unclear, and vulnerable to misinterpretation [73,74]. However, the modular structure of the feedback module ensures that novel feedback can be easily added in later stages. For example, if empirical evidence shows that certain patterns in EMA data can be linked to treatment response or outcome, PETRA can offer this information directly in its feedback module [75-77]. Finally, the module has a report function, which enables users to save graphs and comments (eg, therapy notes) for later use.

Figure 6. An interactive graph on the variation in mood and symptoms in the feedback module. All continuous ecological momentary assessment (EMA) diary items can be selected from the menu on the right side of the screen. Clicking on any of the assessment points will activate a sliding pane depicted in Figure 7. For a video, see here [49].

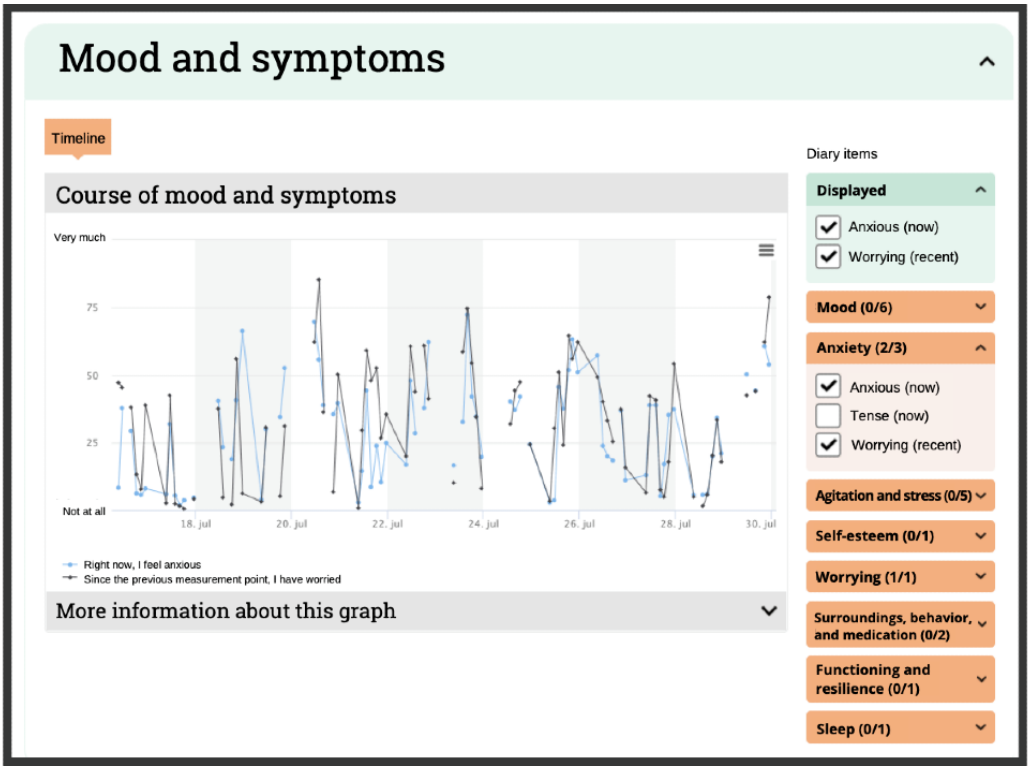


Figure 7. Sliding panel in the feedback module that provides relevant contextual information for a specific moment in the time series and surrounding moments. Moments are selected by clicking on assessment points in the graph depicted in Figure 6. For a video, see here [49].

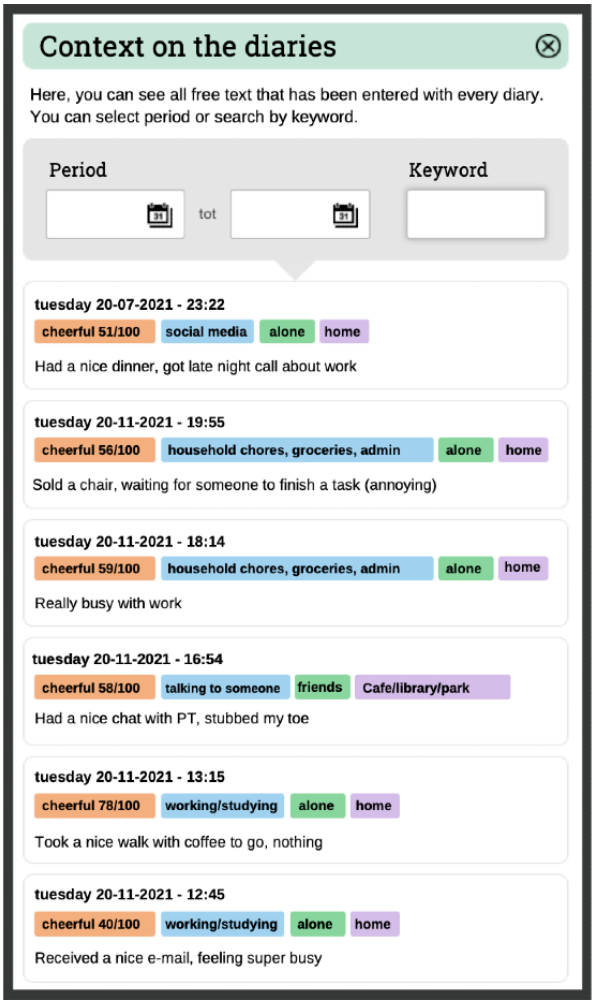
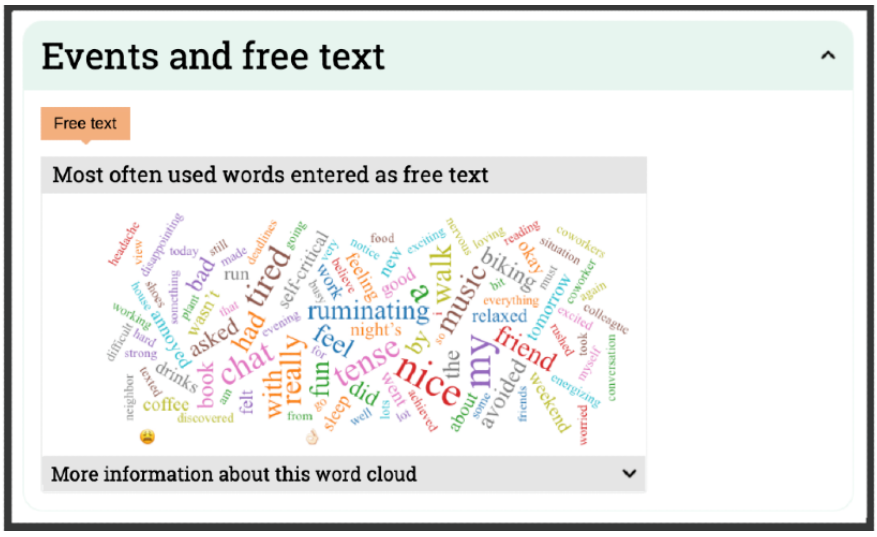


Figure 8. A word cloud of all freely entered text in the feedback module. Clicking on a word provides more context on when this word was used. For a video, see here [49].



Discussion

Principal Findings

In this paper, we outline the development of PETRA, a web-based application for personalized EMA in Dutch psychiatric care. Interviews, focus groups, and usability sessions with clients and clinicians demonstrated that they expected an added value of integrating personalized EMA into treatment. Crucial requirements entailed a personalized and user-friendly diary tool built on scientific foundations, which adheres to privacy regulations. The subsequent thorough cocreation design

process ensured that PETRA was considered intuitive, user-friendly, and useful for clients and clinicians.

By involving clients and clinicians from the start of the development, PETRA can be considered a significant advancement of current EMA (web-based) applications that are mostly targeted at researchers. By systematically integrating the perspectives of clients and clinicians with those of researchers and software developers, the PETRA web application has promise as a tool for assisting both clients and clinicians in the personalized treatment of mental health problems.

Limitations

In this project, we attempted to include a diverse group of clients and clinicians with differing levels of interest in using personalized EMA for treatment. However, this was a relatively small sample, and participants were selected based on their interest in EMA. Therefore, it cannot be ruled out that we overestimated the eventual uptake in clinical practice. As EMA requires significant time and resources from both clients and clinicians, it may not appeal to everyone [78,79], which was also highlighted in the focus groups and interviews. However, research indicates that EMA is feasible for diverse psychiatric populations [47,80,81], including individuals diagnosed with depression [26,27,82], bipolar disorder [83,84], anxiety [85], and psychotic [53,86,87] disorders, from youth [88,89] to older adult populations [90]. Feasibility will likely not depend on diagnosis but rather on the intrinsic motivation of clients, symptom severity, and stage of care [22,23].

Furthermore, there are some limitations that pertain specifically to the PETRA tool. First, PETRA is only accessible to clients via their clinicians via their EHRs. This can be a psychiatrist, psychologist, or psychiatric nurse affiliated with a RoQua mental health care facility. We are currently developing client access to PETRA directly via their personal health record system to enable clients to interpret feedback independently of their clinician. This is in line with clients' desire to have access to their own data [23]. Second, proficiency in the Dutch language and client possession of a smartphone with internet access are required to use PETRA. This implies that PETRA is currently unavailable in other languages.

Comparisons With Prior Studies

Thus far, EMA software development has mainly focused on researchers [91-94], who have different needs from clients and clinicians. However, there are three known EMA platforms for clinical purposes: MindLogger [94], m-Path [95], and PsyMate [96,97]. MindLogger does not solely focus on EMA but also on other self-monitoring techniques and mobile interventions. As m-Path and PsyMate are most similar to PETRA, we will briefly review them here. Both platforms offer a web environment for clinicians and clients and a smartphone app that delivers EMA diaries to clients. PsyMate provides a standardized EMA diary: a set list of EMA items (which may differ slightly according to the target group) and 10 assessments per day on a semirandom schedule. m-Path is an open-source platform that targets both researchers and clinicians. It offers full freedom with creating a diary but no decision aid, requiring knowledge of constructing EMA diaries. Similar to PETRA, PsyMate and m-Path were developed based on best practices for EMA and GDPR compliance. Their smartphone apps enable client access to EMA data, which PETRA is still developing. PETRA differs in that it was codeveloped with clients and clinicians from the beginning, balancing personalization with user-friendly diary construction and feedback. Furthermore, the PETRA decision aid and feedback module were specifically developed for use as shared decision-making tools, integrating EMA with contextual information. Finally, its integration into EHR systems via RoQua [38] ensures that PETRA is sustainably embedded in clinical practice. As such, each platform has a different focus, and its concurrent development can advance our understanding of how to optimally implement EMA in clinical practice.

Qualitative work demonstrates that clients and clinicians consider personalized EMA a useful add-on tool for diagnosis and treatment [22-24,78,98]. Similarly, self-monitoring of symptoms is considered safe and acceptable in diverse psychiatric populations [32,99,100]. Our results suggest the need for a collaborative approach to integrating personalized diaries in care, where clients and clinicians decide together on the focus of the EMA and the interpretation of feedback [22]. This corresponds to research demonstrating the importance of shared decision-making for psychiatric outcomes [10,101-103]. Throughout the development process, we realized the importance of tailoring PETRA's EMA diaries and feedback to diverse clinical settings and client populations. However, much is unknown regarding the optimal type and timing of EMA feedback and how this potentially varies across diagnoses and treatment strategies. This means that PETRA will need to be continuously re-evaluated and adapted based on the diverse needs of clients and clinicians in psychiatric care. This iterative process has also been highlighted in other recent publications on the co-design of self-monitoring technology with users [42,97,104].

Ethical Considerations

Several ethical considerations arise when developing tools for highly personalized self-monitoring. First, the (sometimes qualitative) data are sensitive, and the privacy of the client should be well protected [105]. For PETRA, software and data transmission security are guaranteed by RoQua [38] and are in compliance with the GDPR. Second, clients may expect that the self-monitoring application alerts them or their clinician in the case of elevated EMA scores, which PETRA currently does not provide. Research indicates that this feature is desired by clients [23]; however, scientific evidence is lacking on how to provide EMA-based alerts with adequate sensitivity and specificity. Therefore, it is important to manage the expectations of clients and clinicians through training. Third, self-monitoring can be burdensome: clients can feel disturbed in daily activities or even experience symptom worsening [22,106]. Clients and clinicians will have to decide together whether the benefits of EMA outweigh this burden. The PETRA application helps clients and clinicians with this decision by providing a burden indicator, which helps them make informed decisions on the number of items. Similarly, the burden may be reduced by placing equal emphasis on symptoms and strengths in

personalized diaries, in line with other studies on self-monitoring technology [104,107].

Finally, research suggests that clinician training and support via a helpdesk are important requirements for the implementation of EMA in clinical settings [25,97]. The clinicians in this study also expressed these needs. Training is important to minimize potential health risks and help clinicians construct valid diaries, accurately interpret EMA feedback, and integrate EMA into treatment. Therefore, in the PETRA project, clinicians receive training on the valid use of EMA and the integration of PETRA in treatment, and they have access to a helpdesk. Clinicians are made aware that the feedback should be treated as an additional perspective that is equally (not more) valuable compared with the perspectives of clients and clinicians themselves. Furthermore, short instructive videos are included in PETRA's various steps, explaining (1) how the decision aid works, (2) what clients should know when completing a diary, and (3) how to interpret the feedback [49].

Directions for Future Research

Although self-monitoring is already a prominent feature in several treatment modalities and protocols [14,108], an important next step is to determine the clinical effectiveness of adding personalized EMA to the treatment. This is especially relevant as only standardized diaries have been investigated in randomized controlled trials [26,27,29-31]. Extensive personalization options are clearly desired in clinical practice but may mean that effective components may differ across clients, which may complicate research into the effectiveness of personalized EMA [105]. Furthermore, future trials should not only solely focus on symptom improvement but also on more proximate outcomes relevant to the mechanisms of EMA, such as self-management and the therapeutic alliance [52]. Finally, future research will have to investigate the clinical populations and treatment settings for which personalized EMA can be effective. PETRA enables research on these questions. Within the PETRA project, the effectiveness of personalized EMA is tested both qualitatively and quantitatively. PETRA is currently being tested in pilot sessions with clients and clinicians and is continuously updated. Analytics are collected to assess which features of PETRA are used and how, when, and by whom. Furthermore, clients and clinicians are invited to fill out questionnaires on their positive and negative experiences with PETRA, as well as their effects on self-management [109,110] and the therapeutic alliance [111]. This will allow us to examine how personalized EMA influences psychiatric care and to improve PETRA accordingly. These steps can help to set personalized EMA, such as PETRA, as an eHealth technology apart from the myriad of apps available to clients and clinicians, who will be more inclined to use EMA if it is supported by scientific evidence [112].

The second step concerns further scientific validation of the EMA items [113,114] and the development of more innovative EMA feedback. Such endeavors should explore more descriptive visualizations [115] and further develop statistical methods to analyze EMA data. Previous publications on EMA have often highlighted the potential of person-specific statistical models for informing treatment. These models include, for example, network analysis to demonstrate associations between symptoms [54,116,117], early warning signals to alert clients to impending relapse [118,119], and machine learning to differentiate between diagnoses [120]. However, since then, several publications have shown that the outcomes of these models depend on methodological and statistical choices [73] and that researchers disagree on their implications for treatment [74]. Therefore, PETRA focuses on descriptive feedback, with a large role for qualitative descriptions until more consensus on statistical models has been reached.

Conclusions

PETRA was developed to meet the demands of clients and clinicians for a personalized and user-friendly EMA tool embedded in routine psychiatric care. By collaboratively constructing EMA diaries and interpreting the resulting dynamic feedback, PETRA offers clients and clinicians a new tool to illuminate daily life processes that worsen or alleviate mental health problems. This approach may have beneficial effects on client self-management and the therapeutic alliance and, thus, has the potential to significantly improve clinical care. Our findings demonstrate the importance of a multidisciplinary approach to the development of personalized EMA tools, including clients, clinicians, researchers, and software developers, to ensure that their needs are sufficiently addressed in the design of the tool. PETRA is unique in its codevelopment process, extensive but user-friendly personalization options, integration into EHR systems, transdiagnostic focus, and strong scientific foundation in the design of EMA diaries and feedback. As such, PETRA paves the way for a systematic investigation of the effect of personalized EMA on specialized mental health care.

Acknowledgments

The authors thank all clients and clinicians for their invaluable insights. The authors also thank the RoQua team, in particular, Henk van der Veen and Inge ten Vaarwerk, for integrating the PETRA (Personalized Treatment by Real-time Assessment) web application within RoQua. The authors gratefully acknowledge Meike Bak for assisting with client recruitment and our research assistants Jannie Brouwer and Roan Carlier. The authors thank Judith Rosmalen for her insightful inputs on relevant ecological momentary assessment items and for her help in setting up the project. Finally, the authors thank Anoek Houben for designing the manual and the animation videos for the PETRA web application.

This study was financially supported by Innovation Fund "Stichting De Friesland" (grant DS81 to HR and MW), the charitable foundation "Stichting tot Steun VCVGZ" (grant 239 to HR and Dr Jojanneke Bastiaansen), the European Research Council under

the European Union's Horizon 2020 research and innovation program (ERC-CoG-2015; grant 681466 to MW), and the Rob Giel Research Center. The funding agencies had no role in the interpretation of the results and the drafting of this manuscript.

Authors' Contributions

FMB, HR, and LvK conceptualized the study, with assistance from the scientific advisors MW, ES, client representative GHP, and clinician representative BD.
HR and MW contributed to funding acquisition. FMB contributed to the data acquisition, formal analysis, project coordination, and writing of the original paper. ACE, TV, EV, and FMB contributed to software development. HR contributed to study supervision. FMB, HR, LvK, ACE, TV, EV, ES, BD, and GHP reviewed and edited the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary Materials.
[PDF File (Adobe PDF File), 750 KB - mental_v9i8e36430_app1.pdf]

References

1. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008 Apr 01;4(1):1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
2. Delespaul P. Assessing Schizophrenia in Daily Life : The Experience Sampling Method. Maastricht: Universitaire Pers Maastricht; 1995.
3. aan het Rot M, Hogenelst K, Schoevers RA. Mood disorders in everyday life: a systematic review of experience sampling and ecological momentary assessment studies. Clin Psychol Rev 2012 Aug;32(6):510-523. [doi: 10.1016/j.cpr.2012.05.007] [Medline: 22721999]
4. Walz LC, Nauta MH, Aan Het Rot M. Experience sampling and ecological momentary assessment for studying the daily lives of patients with anxiety disorders: a systematic review. J Anxiety Disord 2014 Dec;28(8):925-937. [doi: 10.1016/j.janxdis.2014.09.022] [Medline: 25445083]
5. Oorschot M, Kwapil T, Delespaul P, Myin-Germeys I. Momentary assessment research in psychosis. Psychol Assess 2009 Dec;21(4):498-505. [doi: 10.1037/a0017077] [Medline: 19947784]
6. Shiffman S. Ecological momentary assessment (EMA) in studies of substance use. Psychol Assess 2009 Dec;21(4):486-497 [FREE Full text] [doi: 10.1037/a0017074] [Medline: 19947783]
7. Wichers M, Hartmann J, Kramer I, Lothmann C, Peeters F, van Bemmel L, et al. Translating assessments of the film of daily life into person-tailored feedback interventions in depression. Acta Psychiatr Scand 2011 May;123(5):402-403. [doi: 10.1111/j.1600-0447.2011.01684.x] [Medline: 21320080]
8. Wichers M, Simons CJ, Kramer IM, Hartmann JA, Lothmann C, Myin-Germeys I, et al. Momentary assessment technology as a tool to help patients with depression help themselves. Acta Psychiatr Scand 2011 Oct;124(4):262-272. [doi: 10.1111/j.1600-0447.2011.01749.x] [Medline: 21838742]
9. van Os J, Delespaul P, Wigman J, Myin-Germeys I, Wichers M. Beyond DSM and ICD: introducing "precision diagnosis" for psychiatry using momentary assessment technology. World Psychiatry 2013 Jun 04;12(2):113-117 [FREE Full text] [doi: 10.1002/wps.20046] [Medline: 23737412]
10. van Os J, Verhagen S, Marsman A, Peeters F, Bak M, Marcelis M, ESM-MERGE Investigators PhD, et al. The experience sampling method as an mHealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice. Depress Anxiety 2017 Jun 23;34(6):481-493. [doi: 10.1002/da.22647] [Medline: 28544391]
11. Kirk JK, Stegner J. Self-monitoring of blood glucose: practical aspects. J Diabetes Sci Technol 2010 Mar 01;4(2):435-439 [FREE Full text] [doi: 10.1177/193229681000400225] [Medline: 20307405]
12. Pickering TG, White WB, American Society of Hypertension Writing Group. When and how to use self (home) and ambulatory blood pressure monitoring. J Am Soc Hypertens 2008 May;2(3):119-124. [doi: 10.1016/j.jash.2008.04.002] [Medline: 20409893]
13. Torous J, Bucci S, Bell IH, Kessing LV, Faurholt-Jepsen M, Whelan P, et al. The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. World Psychiatry 2021 Oct 09;20(3):318-335 [FREE Full text] [doi: 10.1002/wps.20883] [Medline: 34505369]
14. Cohen JS, Edmunds JM, Brodman DM, Benjamin CL, Kendall PC. Using self-monitoring: implementation of collaborative empiricism in cognitive-behavioral therapy. Cognit Behav Pract 2013 Nov;20(4):419-428. [doi: 10.1016/j.cbpra.2012.06.002]
15. Lambert MJ, Whipple JL, Kleinstäuber M. Collecting and delivering progress feedback: a meta-analysis of routine outcome monitoring. Psychotherapy (Chic) 2018 Dec;55(4):520-537. [doi: 10.1037/pst0000167] [Medline: 30335463]

16. Janssens KA, Bos EH, Rosmalen JG, Wichers MC, Riese H. A qualitative approach to guide choices for designing a diary study. BMC Med Res Methodol 2018 Nov 16;18(1):140 [FREE Full text] [doi: 10.1186/s12874-018-0579-6] [Medline: 30445926]
17. Vachon H, Viechtbauer W, Rintala A, Myin-Germeys I. Compliance and retention with the experience sampling method over the continuum of severe mental disorders: meta-analysis and recommendations. J Med Internet Res 2019 Dec 06;21(12):e14475 [FREE Full text] [doi: 10.2196/14475] [Medline: 31808748]
18. Helmich M, Snippe E, Kunkels Y, Riese H, Smit A, Wichers M. Transitions in Depression (TRANS-ID) Recovery: study protocol for a repeated intensive longitudinal n= 1 study design to search for personalized early warning signals of critical transitions towards improvement in depression. PsyArXiv 2020. [doi: 10.31234/osf.io/fertq]
19. Alpay L, van der Boog P, Dumaij A. An empowerment-based approach to developing innovative e-health tools for self-management. Health Informatics J 2011 Dec 21;17(4):247-255 [FREE Full text] [doi: 10.1177/1460458211420089] [Medline: 22193825]
20. Reis H. Why researchers should think "real-world": a conceptual rationale. In: Handbook of Research Methods for Studying Daily Life. New York, NY United States: Guilford Press; 2012.
21. Hatcher RL, Gillaspy JA. Development and validation of a revised short version of the working alliance inventory. Psychotherapy Res 2006 Jan;16(1):12-25. [doi: 10.1080/10503300500352500]
22. Bos FM, Snippe E, Bruggeman R, Doornbos B, Wichers M, van der Krieke L. Recommendations for the use of long-term experience sampling in bipolar disorder care: a qualitative study of patient and clinician experiences. Int J Bipolar Disord 2020 Dec 01;8(1):38 [FREE Full text] [doi: 10.1186/s40345-020-00201-5] [Medline: 33258015]
23. 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]
24. Folkersma W, Veerman V, Ornée DA, Oldehinkel AJ, Alma MA, Bastiaansen JA. Patients' experience of an ecological momentary intervention involving self-monitoring and personalized feedback for depression. Internet Interv 2021 Dec;26:100436 [FREE Full text] [doi: 10.1016/j.invent.2021.100436] [Medline: 34430220]
25. Piot M, Mestdagh M, Riese H, Brouwer J, Weermeier J, Dejonckheere E. Practitioner and researcher perspectives on the utility of ecological momentary assessment in clinical practice: A survey study. OSF. URL: https://osf.io/v9rtd/ [accessed 2022-07-08]
26. Kramer I, Simons CJ, Hartmann JA, Menne-Lothmann C, Viechtbauer W, Peeters F, et al. A therapeutic application of the experience sampling method in the treatment of depression: a randomized controlled trial. World Psychiatry 2014 Mar 04;13(1):68-77 [FREE Full text] [doi: 10.1002/wps.20090] [Medline: 24497255]
27. Simons C, Hartmann J, Kramer I, Menne-Lothmann C, Höhn P, van Bemmel A, et al. Effects of momentary self-monitoring on empowerment in a randomized controlled trial in patients with depression. Eur Psychiatry 2015 Nov 06;30(8):900-906. [doi: 10.1016/j.eurpsy.2015.09.004] [Medline: 26647864]
28. Widdershoven RL, Wichers M, Kuppens P, Hartmann JA, Menne-Lothmann C, Simons CJ, et al. Effect of self-monitoring through experience sampling on emotion differentiation in depression. J Affect Disord 2019 Feb 01;244:71-77. [doi: 10.1016/j.jad.2018.10.092] [Medline: 30321767]
29. Snippe E, Simons CJ, Hartmann JA, Menne-Lothmann C, Kramer I, Booij SH, et al. Change in daily life behaviors and depression: within-person and between-person associations. Health Psychol 2016 May;35(5):433-441. [doi: 10.1037/hea0000312] [Medline: 26690641]
30. Bastiaansen JA, Ornée DA, Meurs M, Oldehinkel AJ. An evaluation of the efficacy of two add-on ecological momentary intervention modules for depression in a pragmatic randomized controlled trial (ZELF-i). Psychol Med 2020 Dec 14:1-10. [doi: 10.1017/s0033291720004845]
31. Van Roekel E, Vrijen C, Heininga VE, Masselink M, Bos EH, Oldehinkel AJ. An exploratory randomized controlled trial of personalized lifestyle advice and tandem skydives as a means to reduce anhedonia. Behav Ther 2017 Jan;48(1):76-96. [doi: 10.1016/j.beth.2016.09.009] [Medline: 28077223]
32. Murnane E, Cosley D, Chang P, Guha S, Frank E, Gay G, et al. Self-monitoring practices, attitudes, and needs of individuals with bipolar disorder: implications for the design of technologies to manage mental health. J Am Med Inform Assoc 2016 May;23(3):477-484. [doi: 10.1093/jamia/ocv165] [Medline: 26911822]
33. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
34. Kip H, Beerlage-de JN, Wentzel J. The contextual inquiry. In: eHealth Research, Theory and Development A Multidisciplinary Approach. London: Routledge; 2018.
35. Holistic development of eHealth technology. In: eHealth Research, Theory and Development A Multidisciplinary Approach. London: Routledge; 2018.
36. Van De Belt TH, Engelen LJ, Berben SA, Schoonhoven L. Definition of health 2.0 and medicine 2.0: a systematic review. J Med Internet Res 2010 Jun 11;12(2):e18 [FREE Full text] [doi: 10.2196/jmir.1350] [Medline: 20542857]
37. Home. iLab. URL: https://ilab-psychiatry.nl/en_US/ [accessed 2021-12-20]

38. roQua Routine Outcome Monitoring. RoQua. URL: https://www.roqua.nl/ [accessed 2021-12-10]
39. The Rob Giel Research center (RGOc) homepage. The Rob Giel Research center (RGOc). URL: https://www.rgoc.nl/english-pages/ [accessed 2021-10-10]
40. Dierckx de Casterlé B, Gastmans C, Bryon E, Denier Y. QUAGOL: a guide for qualitative data analysis. Int J Nurs Stud 2012 Mar;49(3):360-371. [doi: 10.1016/j.ijnurstu.2011.09.012] [Medline: 21996649]
41. Development. PETRA. URL: https://www.petrapsy.nl/en/ontwikkeling/ [accessed 2021-12-09]
42. Rosmalen JG, van Gils A, Acevedo Mesa MA, Schoevers RA, Monden R, Hanssen DJ. Development of grip self-help: an online patient-tailored self-help intervention for functional somatic symptoms in primary care. Internet Interv 2020 Mar;19:100297 [FREE Full text] [doi: 10.1016/j.invent.2019.100297] [Medline: 31879599]
43. Wright AG, Woods WC. Personalized models of psychopathology. Annu Rev Clin Psychol 2020 May 07;16(1):49-74. [doi: 10.1146/annurev-clinpsy-102419-125032] [Medline: 32070120]
44. Nijland N, van Gemert-Pijnen J, Boer H, Steehouder MF, Seydel ER. Evaluation of internet-based technology for supporting self-care: problems encountered by patients and caregivers when using self-care applications. J Med Internet Res 2008 May 15;10(2):e13 [FREE Full text] [doi: 10.2196/jmir.957] [Medline: 18487137]
45. BinDhim NF, Hawkey A, Trevena L. A systematic review of quality assessment methods for smartphone health apps. Telemed J E Health 2015 Mar;21(2):97-104. [doi: 10.1089/tmj.2014.0088] [Medline: 25469795]
46. Marshall JM, Dunstan DA, Bartik W. The digital psychiatrist: in search of evidence-based apps for anxiety and depression. Front Psychiatry 2019 Nov 15;10:831 [FREE Full text] [doi: 10.3389/fpsyt.2019.00831] [Medline: 31803083]
47. Palmier-Claus JE, Myin-Germeys I, Barkus E, Bentley L, Udachina A, Delespaul P, 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]
48. Dejonckheere E, Erbas Y. Designing an experience sampling study. In: The Open Handbook of Experience Sampling Methodology: A Step-by-step Guide to Designing, Conducting, and Analyzing ESM Studies. -: Independently published; 2021.
49. Videos. PETRA. URL: https://www.petrapsy.nl/en/videos/ [accessed 2021-11-03]
50. Eisele E, Kasanova Z, Houben M. Questionnaire designevaluation. In: The Open Handbook of Experience Sampling Methodology: A Step-by-step Guide to Designing, Conducting, and Analyzing ESM Studies. -: Independently published; 2021.
51. 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 2022 Mar;29(2):136-151. [doi: 10.1177/1073191120957102] [Medline: 32909448]
52. Riese H, von Klipstein L, Schoevers RA, van der Veen DC, Servaas MN. Personalized ESM monitoring and feedback to support psychological treatment for depression: a pragmatic randomized controlled trial (Therap-i). BMC Psychiatry 2021 Mar 10;21(1):143 [FREE Full text] [doi: 10.1186/s12888-021-03123-3] [Medline: 33691647]
53. Reininghaus U, Klippel A, Steinhart H, Vaessen T, van Nierop M, Viechtbauer W, et al. Efficacy of Acceptance and Commitment Therapy in Daily Life (ACT-DL) in early psychosis: study protocol for a multi-centre randomized controlled trial. Trials 2019 Dec 26;20(1):769 [FREE Full text] [doi: 10.1186/s13063-019-3912-4] [Medline: 31878966]
54. van der Krieke L, Blaauw FJ, Emerencia AC, Schenk HM, Slaets JP, Bos EH, et al. Temporal dynamics of health and well-being: a crowdsourcing approach to momentary assessments and automated generation of personalized feedback. Psychosom Med 2017;79(2):213-223. [doi: 10.1097/psy.0000000000000378]
55. Krieke LV, Jeronimus BF, Blaauw FJ, Wanders RB, Emerencia AC, Schenk HM, et al. HowNutsAreTheDutch (HoeGekIsNL): a crowdsourcing study of mental symptoms and strengths. Int J Methods Psychiatr Res 2016 Jun;25(2):123-144 [FREE Full text] [doi: 10.1002/mpr.1495] [Medline: 26395198]
56. Schoevers RA, van Borkulo CD, Lamers F, Servaas M, Bastiaansen JA, Beekman AT, et al. Affect fluctuations examined with ecological momentary assessment in patients with current or remitted depression and anxiety disorders. Psychol Med 2020 Apr 01;51(11):1906-1915. [doi: 10.1017/s0033291720000689]
57. 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 01;42(3):712-722 [FREE Full text] [doi: 10.1093/schbul/sbv190] [Medline: 26834027]
58. Knapen S. Rhythm & Blues: Chronobiology in the Pathophysiology and Treatment of Mood Disorders. [Groningen]: Rijksuniversiteit Groningen; 2019.
59. Tyler E, Jones S, Black N, Carter L, Barrowclough C. The relationship between bipolar disorder and cannabis use in daily life: an experience sampling study. PLoS One 2015 Mar 4;10(3):e0118916 [FREE Full text] [doi: 10.1371/journal.pone.0118916] [Medline: 25738578]
60. Myin-Germeys I, Peeters F, Havermans R, Nicolson N, DeVries MW, Delespaul P, et al. Emotional reactivity to daily life stress in psychosis and affective disorder: an experience sampling study. Acta Psychiatr Scand 2003 Mar;107(2):124-131. [doi: 10.1034/j.1600-0447.2003.02025.x] [Medline: 12534438]

61. Barrantes-Vidal N, Chun CA, Myin-Germeys I, Kwapil TR. Psychometric schizotypy predicts psychotic-like, paranoid, and negative symptoms in daily life. J Abnorm Psychol 2013 Nov;122(4):1077-1087. [doi: 10.1037/a0034793] [Medline: 24364610]
62. Schreuder MJ, Groen RN, Wigman JT, Hartman CA, Wichers M. Measuring psychopathology as it unfolds in daily life: addressing key assumptions of intensive longitudinal methods in the TRAILS TRANS-ID study. BMC Psychiatry 2020 Jul 06;20(1):351 [FREE Full text] [doi: 10.1186/s12888-020-02674-1] [Medline: 32631277]
63. Kroeze R, van der Veen DC, Servaas MN, Bastiaansen JA, Oude Voshaar RC, Borsboom D, et al. Personalized feedback on symptom dynamics of psychopathology: a proof-of-principle study. J Pers Oriented Res 2017 Nov 03;3(1):1-10 [FREE Full text] [doi: 10.17505/jpor.2017.01] [Medline: 33569120]
64. Bos F. Ecological Momentary Assessment as A Clinical Tool in Psychiatry Promises, Pitfalls, and Possibilities. Groningen, The Netherlands: University of Groningen; 2021.
65. Zijlema WL, Stolk RP, Löwe B, Rief W, BioSHaRE, White PD, et al. How to assess common somatic symptoms in large-scale studies: a systematic review of questionnaires. J Psychosom Res 2013 Jun;74(6):459-468. [doi: 10.1016/j.jpsychores.2013.03.093] [Medline: 23731742]
66. Wichers M, Peeters F, Rutten BP, Jacobs N, Derom C, Thiery E, et al. A time-lagged momentary assessment study on daily life physical activity and affect. Health Psychol 2012 Mar;31(2):135-144. [doi: 10.1037/a0025688] [Medline: 21988094]
67. Laura Bringmann's intensive longitudinal data lab. LaBLab. URL: https://www.laurabringmannlab.com/ [accessed 2021-12-20]
68. MYIN-GERMEYS I, KRABBENDAM L, DELESPAUL PA, VAN OS J. Do life events have their effect on psychosis by influencing the emotional reactivity to daily life stress? Psychol Med 2003 Feb 14;33(2):327-333. [doi: 10.1017/S0033291702006785]
69. Booij SH, Wichers M, de Jonge P, Sytema S, van Os J, Wunderink L, et al. Study protocol for a prospective cohort study examining the predictive potential of dynamic symptom networks for the onset and progression of psychosis: the Mapping Individual Routes of Risk and Resilience (Mirorr) study. BMJ Open 2018 Jan 21;8(1):e019059 [FREE Full text] [doi: 10.1136/bmjopen-2017-019059] [Medline: 29358438]
70. Leverich G, Post R. The NIMH life chart manual for recurrent affective illness: The LCM. NIMH Monograph. 1993. URL: http:/www.bipolarnews.org/pdfs/Clinician%20Prospective%20Manual.pdf [accessed 2022-06-30]
71. Schiepek G, Stöger-Schmidinge B, Kronberger H, Aichhorn W, Kratzer L, Heinz P, et al. The Therapy Process Questionnaire - Factor analysis and psychometric properties of a multidimensional self-rating scale for high-frequency monitoring of psychotherapeutic processes. Clin Psychol Psychother 2019 Sep 10;26(5):586-602 [FREE Full text] [doi: 10.1002/cpp.2384] [Medline: 31153157]
72. Carney C, Buysse D, Ancoli-Israel S, Edinger J, Krystal A, Lichstein K, et al. The consensus sleep diary: standardizing prospective sleep self-monitoring. Sleep 2012 Mar 01;35(2):287-302 [FREE Full text] [doi: 10.5665/sleep.1642] [Medline: 22294820]
73. Wichers M, Riese H, Hodges TM, Snippe E, Bos FM. A narrative review of network studies in depression: what different methodological approaches tell us about depression. Front Psychiatry 2021 Oct 28;12:719490 [FREE Full text] [doi: 10.3389/fpsyt.2021.719490] [Medline: 34777038]
74. Bastiaansen JA, Kunkels YK, Blaauw FJ, Boker SM, Ceulemans E, Chen M, et al. Time to get personal? The impact of researchers choices on the selection of treatment targets using the experience sampling methodology. J Psychosom Res 2020 Aug 05;137:110211 [FREE Full text] [doi: 10.1016/j.jpsychores.2020.110211] [Medline: 32862062]
75. Lutz W, Rubel JA, Schwartz B, Schilling V, Deisenhofer A. Towards integrating personalized feedback research into clinical practice: development of the Trier Treatment Navigator (TTN). Behav Res Ther 2019 Sep;120:103438. [doi: 10.1016/j.brat.2019.103438] [Medline: 31301550]
76. Lutz W, Schwartz B, Hofmann SG, Fisher AJ, Husen K, Rubel JA. Using network analysis for the prediction of treatment dropout in patients with mood and anxiety disorders: a methodological proof-of-concept study. Sci Rep 2018 May 18;8(1):7819 [FREE Full text] [doi: 10.1038/s41598-018-25953-0] [Medline: 29777110]
77. Rubel J, Fisher A, Husen K, Lutz W. Translating person-specific network models into personalized treatments: development and demonstration of the dynamic assessment treatment algorithm for individual networks (DATA-IN). Psychother Psychosom 2018 Apr 20;87(4):249-251. [doi: 10.1159/000487769] [Medline: 29680835]
78. Frumkin MR, Piccirillo ML, Beck ED, Grossman JT, Rodebaugh TL. Feasibility and utility of idiographic models in the clinic: a pilot study. Psychother Res 2021 Apr 24;31(4):520-534 [FREE Full text] [doi: 10.1080/10503307.2020.1805133] [Medline: 32838671]
79. Ellison WD. An initial study of practicing psychologists' views of the utility of ecological momentary assessment for difficult psychotherapy cases. Adm Policy Ment Health 2021 Jul 12;48(4):597-607. [doi: 10.1007/s10488-020-01093-4] [Medline: 33047277]
80. Myin-Germeys I. Psychiatry. In: Handbook of Research Methods for Studying Daily Life. New York, NY United States: Guilford Press; 2012:636.

81. 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 24;17(2):123-132 [FREE Full text] [doi: 10.1002/wps.20513] [Medline: 29856567]
82. Burns MN, Begale M, Duffecy J, Gergle D, Karr CJ, Giangrande E, et al. Harnessing context sensing to develop a mobile intervention for depression. J Med Internet Res 2011 Aug 12;13(3):e55 [FREE Full text] [doi: 10.2196/jmir.1838] [Medline: 21840837]
83. 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]
84. 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]
85. Fisher AJ, Bosley HG, Fernandez KC, Reeves JW, Soyster PD, Diamond AE, et al. Open trial of a personalized modular treatment for mood and anxiety. Behav Res Ther 2019 May;116:69-79. [doi: 10.1016/j.brat.2019.01.010] [Medline: 30831478]
86. Vaessen T, Steinhart H, Batink T, Klippel A, Van Nierop M, Reininghaus U, et al. ACT in daily life in early psychosis: an ecological momentary intervention approach. Psychosis 2019 Mar 19;11(2):93-104. [doi: 10.1080/17522439.2019.1578401]
87. Bell IH, Rossell SL, Farhall J, Hayward M, Lim MH, Fielding-Smith SF, et al. Pilot randomised controlled trial of a brief coping-focused intervention for hearing voices blended with smartphone-based ecological momentary assessment and intervention (SAVVy): feasibility, acceptability and preliminary clinical outcomes. Schizophr Res 2020 Feb;216:479-487. [doi: 10.1016/j.schres.2019.10.026] [Medline: 31812327]
88. 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 17;29(3):560-577 [FREE Full text] [doi: 10.1111/jora.12471] [Medline: 31573762]
89. Heron K, Everhart R, McHale S, Smyth J. Using mobile-technology-based Ecological Momentary Assessment (EMA) methods with youth: a systematic review and recommendations. J Pediatr Psychol 2017 Nov 01;42(10):1087-1107. [doi: 10.1093/jpepsy/jsx078] [Medline: 28475765]
90. Cain AE, Depp CA, Jeste DV. Ecological momentary assessment in aging research: a critical review. J Psychiatr Res 2009 Jul;43(11):987-996 [FREE Full text] [doi: 10.1016/j.jpsychires.2009.01.014] [Medline: 19272611]
91. Vahabzadeh M, Epstein D, Mezghanni M, Lin J, Preston K. An electronic diary software for ecological momentary assessment (EMA) in clinical trials. In: Proceedings of the 17th IEEE Symposium on Computer-Based Medical Systems. 2004 Presented at: 17th IEEE Symposium on Computer-Based Medical Systems; Jun 25, 2004; Bethesda, MD, USA. [doi: 10.1109/cbms.2004.1311709]
92. Activity monitoring. Movisens. URL: https://www.movisens.com/en/ [accessed 2021-11-11]
93. Ilumivu homepage. Ilumivu. URL: https://ilumivu.com/ [accessed 2021-11-11]
94. Klein A, Clucas J, Krishnakumar A, Ghosh SS, Van Auken W, Thonet B, et al. Remote digital psychiatry for mobile mental health assessment and therapy: mindlogger platform development study. J Med Internet Res 2021 Nov 11;23(11):e22369 [FREE Full text] [doi: 10.2196/22369] [Medline: 34762054]
95. m-Path homepage. m-Path. URL: https://m-path.io/landing/ [accessed 2021-12-10]
96. Verhagen SJ, Berben JA, Leue C, Marsman A, Delespaul PA, van Os J, et al. Demonstrating the reliability of transdiagnostic mHealth Routine Outcome Monitoring in mental health services using experience sampling technology. PLoS One 2017 Oct 12;12(10):e0186294 [FREE Full text] [doi: 10.1371/journal.pone.0186294] [Medline: 29023568]
97. Daniëls NE, Hochstenbach LM, van Bokhoven MA, Beurskens AJ, Delespaul PA. Implementing experience sampling technology for functional analysis in family medicine - a design thinking approach. Front Psychol 2019 Dec 11;10:2782 [FREE Full text] [doi: 10.3389/fpsyg.2019.02782] [Medline: 31920830]
98. Palmier-Claus JE, Rogers A, Ainsworth J, Machin M, Barrowclough C, Laverty L, et al. Integrating mobile-phone based assessment for psychosis into people's everyday lives and clinical care: a qualitative study. BMC Psychiatry 2013 Jan 23;13:34 [FREE Full text] [doi: 10.1186/1471-244X-13-34] [Medline: 23343329]
99. Hetrick SE, Goodall J, Yuen HP, Davey CG, Parker AG, Robinson J, et al. Comprehensive online self-monitoring to support clinicians manage risk of suicide in youth depression. Crisis 2017 May;38(3):147-157. [doi: 10.1027/0227-5910/a000422] [Medline: 27659516]
100. Torous J, Staples P, Shanahan M, Lin C, Peck P, Keshavan M, et al. Utilizing a personal smartphone custom app to assess the patient health questionnaire-9 (PHQ-9) depressive symptoms in patients with major depressive disorder. JMIR Ment Health 2015 Mar 24;2(1):e8 [FREE Full text] [doi: 10.2196/mental.3889] [Medline: 26543914]
101. Drake RE, Cimpean D, Torrey WC. Shared decision making in mental health: prospects for personalized medicine. Dialogues Clin Neuroscience 2022 Apr 01;11(4):455-463. [doi: 10.31887/dcns.2009.11.4/redrake]
102. Hamann J, Leucht S, Kissling W. Shared decision making in psychiatry. Acta Psychiatr Scand 2003 Jun;107(6):403-409. [doi: 10.1034/j.1600-0447.2003.00130.x] [Medline: 12752015]

103. Joosten EA, DeFuentes-Merillas L, de Weert GH, Sensky T, van der Staak CP, de Jong CA. Systematic review of the effects of shared decision-making on patient satisfaction, treatment adherence and health status. Psychother Psychosom 2008;77(4):219-226. [doi: 10.1159/000126073] [Medline: 18418028]
104. Hetrick SE, Robinson J, Burge E, Blandon R, Mobilio B, Rice SM, et al. Youth codesign of a mobile phone app to facilitate self-monitoring and management of mood symptoms in young people with major depression, suicidal ideation, and self-harm. JMIR Ment Health 2018 Jan 23;5(1):e9 [FREE Full text] [doi: 10.2196/mental.9041] [Medline: 29362208]
105. Rubeis G. iHealth: the ethics of artificial intelligence and big data in mental healthcare. Internet Interv 2022 Apr;28:100518 [FREE Full text] [doi: 10.1016/j.invent.2022.100518] [Medline: 35257003]
106. van Genugten CR, Schuurmans J, Lamers F, Riese H, Penninx BW, Schoevers RA, et al. Experienced burden of and adherence to smartphone-based ecological momentary assessment in persons with affective disorders. J Clin Med 2020 Jan 23;9(2):322 [FREE Full text] [doi: 10.3390/jcm9020322] [Medline: 31979340]
107. Drake G, Csipke E, Wykes T. Assessing your mood online: acceptability and use of Moodscope. Psychol Med 2012 Nov 13;43(7):1455-1464. [doi: 10.1017/s0033291712002280]
108. Gibbons C, Porter I, Gonçalves-Bradley DC, Stoilov S, Ricci-Cabello I, Tsangaris E, et al. Routine provision of feedback from patient-reported outcome measurements to healthcare providers and patients in clinical practice. Cochrane Database Syst Rev 2021 Oct 12;10:CD011589. [doi: 10.1002/14651858.CD011589.pub2] [Medline: 34637526]
109. Witteman C, Bolks L, Hutschemaekers G. Development of the illness perception questionnaire mental health. J Ment Health 2011 Apr 20;20(2):115-125. [doi: 10.3109/09638237.2010.507685] [Medline: 20854175]
110. Monger B, Hardie SM, Ion R, Cumming J, Henderson N. The Individual Recovery Outcomes Counter: preliminary validation of a personal recovery measure. Psychiatrist 2018 Jan 02;37(7):221-227. [doi: 10.1192/pb.bp.112.041889]
111. Munder T, Wilmers F, Leonhart R, Linster HW, Barth J. Working Alliance Inventory-Short Revised (WAI-SR): psychometric properties in outpatients and inpatients. Clin Psychol Psychother 2010;17(3):231-239. [doi: 10.1002/cpp.658] [Medline: 20013760]
112. Giota K, Kleftaras G. Mental health apps: innovations, risks and ethical considerations. E-Health telecommun syst Netw 2014;03(03):19-23. [doi: 10.4236/etsn.2014.33003]
113. The Experience Sampling Method (ESM) item repository. OSF HOME. URL: https://osf.io/kg376/ [accessed 2022-06-30]
114. 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 Method Pract Psychol Sci 2021 Mar 02;4(1):251524592092468. [doi: 10.1177/2515245920924686]
115. Bringmann LF, van der Veen DC, Wichers M, Riese H, Stulp G. ESMvis: a tool for visualizing individual Experience Sampling Method (ESM) data. Qual Life Res 2021 Nov 22;30(11):3179-3188 [FREE Full text] [doi: 10.1007/s11136-020-02701-4] [Medline: 33222049]
116. Epskamp S, van Borkulo CD, van der Veen DC, Servaas MN, Isvoranu A, Riese H, et al. Personalized network modeling in psychopathology: the importance of contemporaneous and temporal connections. Clin Psychol Sci 2018 May 19;6(3):416-427 [FREE Full text] [doi: 10.1177/2167702617744325] [Medline: 29805918]
117. von Klipstein L, Riese H, van der Veen DC, Servaas MN, Schoevers RA. Using person-specific networks in psychotherapy: challenges, limitations, and how we could use them anyway. BMC Med 2020 Nov 23;18(1):345 [FREE Full text] [doi: 10.1186/s12916-020-01818-0] [Medline: 33222699]
118. Smit A, Snippe E, Wichers M. Increasing restlessness signals impending increase in depressive symptoms more than 2 months before it happens in individual patients. Psychother Psychosom 2019 Jun 28;88(4):249-251 [FREE Full text] [doi: 10.1159/000500594] [Medline: 31256155]
119. Wichers M, Smit AC, Snippe E. Early warning signals based on momentary affect dynamics can expose nearby transitions in depression: a confirmatory single-subject time-series study. J Pers Oriented Res 2020 Sep 10;6(1):1-15 [FREE Full text] [doi: 10.17505/jpor.2020.22042] [Medline: 33569148]
120. Perez Arribas I, Goodwin GM, Geddes JR, Lyons T, Saunders KE. A signature-based machine learning model for distinguishing bipolar disorder and borderline personality disorder. Transl Psychiatry 2018 Dec 13;8(1):274 [FREE Full text] [doi: 10.1038/s41398-018-0334-0] [Medline: 30546013]

Abbreviations

CeHRes: Center for eHealth Research
EHR: electronic health record
EMA: ecological momentary assessment
GDPR: General Data Protection Regulation
PETRA: Personalized Treatment by Real-time Assessment
RGOc: Rob Giel Research Center
UX: user experience

Edited by J Torous; submitted 14.01.22; peer-reviewed by A Klein, M Richter, C Skirrow, V Wiemker; comments to author 24.02.22; revised version received 11.04.22; accepted 06.05.22; published 09.08.22.

Please cite as:
Bos FM, von Klipstein L, Emerencia AC, Veermans E, Verhage T, Snippe E, Doornbos B, Hadders-Prins G, Wichers M, Riese H
A Web-Based Application for Personalized Ecological Momentary Assessment in Psychiatric Care: User-Centered Development of the PETRA Application
JMIR Ment Health 2022;9(8):e36430
URL: https://mental.jmir.org/2022/8/e36430
doi:10.2196/36430
PMID:35943762

Original Paper

Predicting Multiple Sclerosis Outcomes During the COVID-19 Stay-at-home Period: Observational Study Using Passively Sensed Behaviors and Digital Phenotyping

Prerna Chikersal[1], MS; Shruthi Venkatesh[2], BS; Karman Masown[2], BS; Elizabeth Walker[2], BS; Danyal Quraishi[2], BS; Anind Dey[3*], PhD; Mayank Goel[1*], PhD; Zongqi Xia[2*], MD, PhD

[1]School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, United States

[2]Department of Neurology, University of Pittsburgh, Pittsburgh, PA, United States

[3]Information School, University of Washington, Seattle, Seattle, WA, United States

[*]these authors contributed equally

Corresponding Author:
Zongqi Xia, MD, PhD
Department of Neurology
University of Pittsburgh
3501 Fifth Avenue,
BST3, Suite 7014
Pittsburgh, PA, 15260
United States
Phone: 1 412 383 5377
Email: zxia1@pitt.edu

Abstract

Background: The COVID-19 pandemic has broad negative impact on the physical and mental health of people with chronic neurological disorders such as multiple sclerosis (MS).

Objective: We presented a machine learning approach leveraging passive sensor data from smartphones and fitness trackers of people with MS to predict their health outcomes in a natural experiment during a state-mandated stay-at-home period due to a global pandemic.

Methods: First, we extracted features that capture behavior changes due to the stay-at-home order. Then, we adapted and applied an existing algorithm to these behavior-change features to predict the presence of depression, high global MS symptom burden, severe fatigue, and poor sleep quality during the stay-at-home period.

Results: Using data collected between November 2019 and May 2020, the algorithm detected depression with an accuracy of 82.5% (65% improvement over baseline; F_1-score: 0.84), high global MS symptom burden with an accuracy of 90% (39% improvement over baseline; F_1-score: 0.93), severe fatigue with an accuracy of 75.5% (22% improvement over baseline; F_1-score: 0.80), and poor sleep quality with an accuracy of 84% (28% improvement over baseline; F_1-score: 0.84).

Conclusions: Our approach could help clinicians better triage patients with MS and potentially other chronic neurological disorders for interventions and aid patient self-monitoring in their own environment, particularly during extraordinarily stressful circumstances such as pandemics, which would cause drastic behavior changes.

(JMIR Ment Health 2022;9(8):e38495) doi:10.2196/38495

KEYWORDS

mobile sensing; sensor; sensing; mobile health; mHealth; algorithm; multiple sclerosis; disability; mental health; depression; sleep; fatigue; tiredness; predict; machine learning; feature selection; neurological disorder; COVID-19; isolation; behavior change; health outcome; fitness; movement; physical activity; exercise; tracker; digital phenotyping

Introduction

The COVID-19 pandemic and the ensuing response (eg, lockdown and social distancing) have broad negative impacts on physical and mental health worldwide [1-7]. The effect is more pronounced for people with chronic neurological diseases such as multiple sclerosis (MS) [8-10]. People with MS have a significantly higher burden of mental health comorbidities than the general population. Moreover, people with MS have a 50% lifetime prevalence of depression, 2-3 times higher than the general population [11-13]. Given its association with higher disability and mortality, depression is a major comorbidity that lowers the quality of life [11,14-20]. Further, people with MS have greater COVID-19 risk due to certain immune disease-modifying therapies as well as their physical disability, and many have experienced drastic change in their neurological care due to the pandemic [21]. Concerns for COVID-19, coupled with decreased social support and health care access during the pandemic, have contributed to even higher stress and depression in people with MS [10,22-24].

During the pandemic, digital technologies have become invaluable for supporting social interaction, health care access, and health monitoring. Digital health tools can also measure an individual's mental health profile based on passive (noninvasive) tracking. Given the complexity and heterogeneity of real-world behaviors, models that leverage different aspects of an individual's daily behaviors are necessary to accurately predict mental health status. Relevant to depression in people with MS, clinicians could use this digital passive sensing approach to potentially identify patients who require urgent health interventions.

Past research has leveraged passively generated data from personal digital devices (eg, smartphones and fitness trackers) to capture human behavior and predict health outcomes. This moment-by-moment, in situ quantification of the individual-level human phenotype using data from personal digital devices is known as digital phenotyping [25]. Previous works using passively sensed smartphone and wearable data to predict physical disability and fatigue in people with MS have been exploratory in assessing the feasibility of data collection and the preliminary association between sensed behaviors and outcomes [26-28]. However, the clinical applicability of digital phenotyping to inform clinical outcomes in people with MS in the real world has not yet been established.

Here, we present a machine learning approach leveraging data from the smartphones and fitness trackers of people with MS to predict their health outcomes during a mandatory stay-at-home period of the pandemic. Building on an existing analytical pipeline [29], we quantified behavior changes during the stay-at-home period when compared to the preceding period and used the changes to predict the presence of patient-reported outcomes of depression, neurological disability, fatigue, and poor sleep quality during the stay-at-home period. This study is different from prior studies in that it examines the clinical utility of digital phenotyping with passive sensors for predicting health outcomes during the early wave of the COVID-19 pandemic in a unique natural experiment. The study has relevance for predicting the health outcomes of patients with chronic and complex conditions beyond MS during major stressful scenarios (eg, pandemics and natural disasters) that could considerably alter behaviors.

Methods

Overview

This study was part of a larger study that aimed to examine the clinical utility of passive sensors on smartphones and fitness trackers in predicting clinically relevant outcomes in people with MS. Data collection from participants in this larger study occurred between November 2019 and January 2021. Because data collection for 56 participants spanned the locally mandated stay-at-home period in response to the COVID-19 pandemic, we used this unique natural experiment to test the hypothesis whether machine learning models leveraging passive sensor data can predict the health outcomes of people with a chronic neurological disorder (ie, people with MS) during major stressful scenarios.

To briefly summarize our approach, we used data from 3 sensors in the participants' smartphones (calls, location, and screen activity) and 3 sensors in the participants' fitness trackers (heart rate, sleep, and steps) to predict patient-reported outcomes of depression, global MS symptom burden, fatigue, and sleep quality during the COVID-19 stay-at-home period. We computed behavioral features from these 6 sensors before and during the stay-at-home period and took the difference as a measure of behavior change resulting from the stay-at-home mandate. We then used changes in behavioral features to predict the outcomes.

All methods were performed in accordance with the institutional review board guideline and institutional regulation.

Participants

The study included adults 18 years or older with a neurologist-confirmed MS diagnosis who owned a smartphone (Android or iOS) and enrolled in the Prospective Investigation of Multiple Sclerosis in the Three Rivers Region study, a clinic-based natural history study at the University of Pittsburgh Medical Center [21,30-34].

Ethical Considerations

The institutional review boards of University of Pittsburgh (STUDY19080007) and Carnegie Mellon University (STUDY2019_00000037) approved the study. All participants provided written informed consent.

Study Design

The participants downloaded a mobile app to capture sensor data from their own smartphones and additionally received a Fitbit Inspire HR (Fitbit Inc) to track steps, heart rate, and sleep. Data were continuously collected from smartphone and Fitbit sensors of 56 participants during the study period (16 November 2019 to 15 May 2020, including the local stay-at-home period).

All 56 (100%) participants completed data collection for a predefined period of 12 weeks while 39 (70%) agreed to extend data collection for an additional 12 weeks (for a total of 24

weeks). Moreover, 6 (11%) participants who did not have sufficient data during the period before the stay-at-home mandate were excluded from the machine learning analysis.

Survey Response and Patient-Reported Outcomes

All participants completed a baseline questionnaire, which queried their demographics and baseline health outcomes, on the Saturday following enrollment. During the study, the participants completed additional questionnaires, as described below, at intervals according to each questionnaire. All questionnaires for the overall study were administered via the web using the secure, web-based Research Electronic Data Capture system, including the stay-at-home period [35,36].

Depression

We used the Patient Health Questionnaire-9 (PHQ-9) to measure the severity of depression symptoms once every 2 weeks [37]. PHQ-9 contained 9 questions, with each answer being scored on a scale of 0 to 3. Higher scores indicated more severe depressive symptoms.

Global MS Symptom Burden

We used the Multiple Sclerosis Rating Scale—Revised (MSRS-R) to measure global MS symptom burden and neurological disability once every 4 weeks [38]. MSRS-R assessed 8 neurological domains (walking, upper limb function, vision, speech, swallowing, cognition, sensory, bladder, and bowel function); each domain scored as 0 to 4, with 0 indicating the absence of symptom and 4 indicating higher symptom burden and more severe disability.

Fatigue

We used the 5-item version of the Modified Fatigue Impact Scale (MFIS-5) to measure the impact of fatigue on cognitive, physical, and psychosocial function once every 4 weeks [39]. Each item in MFIS-5 was scored on a 5-point Likert scale from 0 (never) to 4 (almost always). Higher scores indicated more severe fatigue.

Sleep Quality

We used the Pittsburgh Sleep Quality Index (PSQI) to measure sleep disturbances once every 4 weeks [40]. PSQI comprised 19 individual items, with 7 component scores (each on a 0-3 scale) and 1 composite score (0 to 21, where higher scores indicate a poorer sleep quality).

For each outcome, we averaged the measures collected during the stay-at-home-period and then dichotomized the resulting outcomes using thresholds. The binary outcomes would likely have better clinical utility as they are more easily understood by patients (for self-monitoring), volunteers with limited mental health training, or even clinicians. For "Depression," PHQ-9 scores were dichotomized as "≥5: presence of depression" and "<5: absence of depression." For "Global MS symptom burden," MSRS-R scores were dichotomized as "≥6.4: higher burden" and "<6.4: lower burden." For "Fatigue," MSIF-5 scores were dichotomized as "≥8: high fatigue" and "<8: low fatigue." For "Sleep quality," PSQI scores were dichotomized as "≥9: poorer sleep quality" and "<9: better sleep quality." The thresholds for depression and sleep quality were based on previous works [37,41]. Given the lack of consensus from the literature, we calculated the median scores of the global MS symptom burden and fatigue in a larger data set of 104 people with MS, of which the 56 (53.8%) people with MS in this paper represented a subgroup (with data collection encompassing the stay-at-home period) and used the median scores as the thresholds.

Sensor Data Collection

Each participant installed a mobile app based on the AWARE framework [42], which provided backend and network infrastructure that unobtrusively collected from smartphones the location, screen usage (ie, when the screen status changed to on or off and locked or unlocked), and call logs (for incoming, outgoing, and missed calls). Further, participants wore a Fitbit Inspire HR, which captured the number of steps, sleep status (asleep, awake, restless, or unknown), and heart rate. Calls and screen use were event-based sensor streams, whereas location, steps, sleep, and heart rate were time series sensor streams. We sampled location coordinates at 1 sample per 10 minutes, and steps, sleep, and heart rate at 1 sample per minute.

Data from AWARE were deidentified and automatically transferred over WiFi to a study server at regular intervals. Data from the Fitbit were retrieved using the Fitbit application programming interface at the end of the data collection. Participants were asked to keep their devices charged and to always carry their phone and wear Fitbit.

To protect confidentiality, we removed identifiable information (eg, names and contact information) from survey and sensor data prior to analysis. We followed the standard practice for sensor data security.

Mediation Analysis

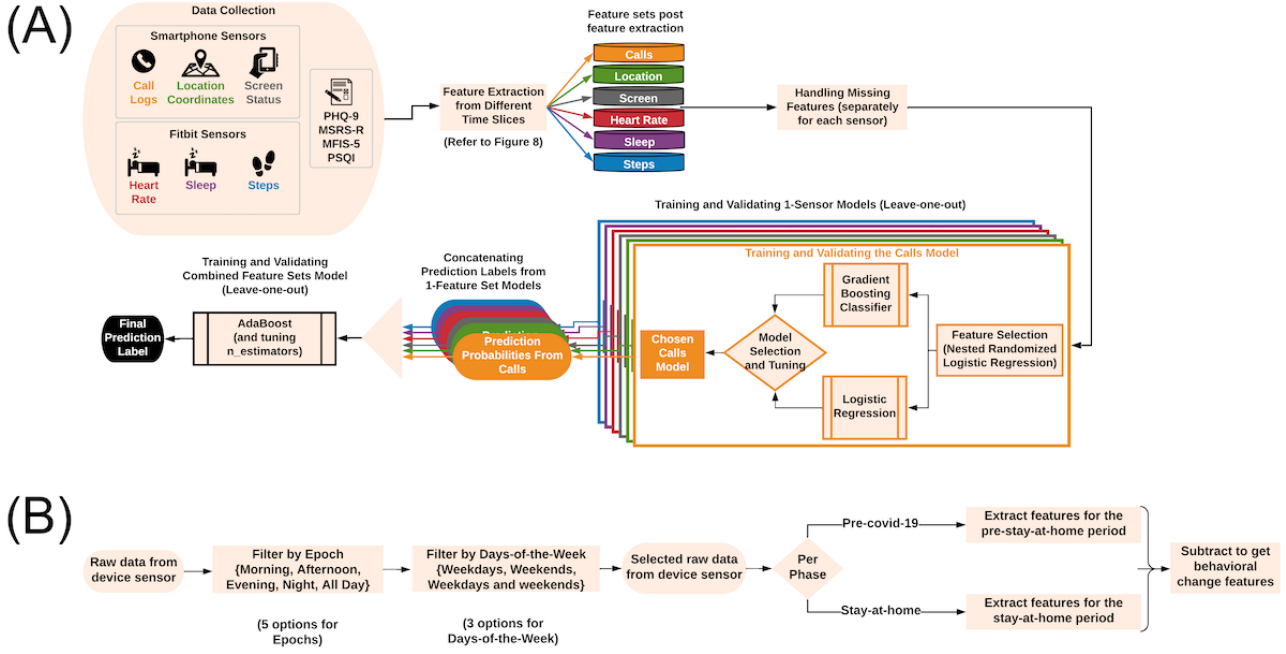
Mediation analysis was performed using the nondichotomized outcomes (ie, the average of the patient-reported outcomes collected during the stay-at-home-period). Process Macro in SPSS (IBM Corp) was used for mediation analysis [43].

Data Processing and Machine Learning

The data processing and analysis pipeline (Figure 1) were built on our prior work [29] and involved several steps:

1. Feature extraction from sensors over time slices to identify behavior changes.
2. Handling missing features.
3. Machine learning to predict patient-reported health outcomes during the stay-at-home period:
    - Using 1-sensor models (ie, models containing features from 1 sensor).
    - Combining 1-sensor models to obtain the best model for each outcome.

Figure 1. Data processing and analysis pipeline. (A) Machine learning pipeline for predicting depression (Patient Health Questionnaire-9 [PHQ-9]), global MS symptom burden (Multiple Sclerosis Rating Scale—Revised [MSRS-R]), fatigue (Modified Fatigue Impact Scale-5 [MFIS-5]), and sleep quality (Pittsburgh Sleep Quality Index [PSQI]) using passive sensors from smartphones and fitness trackers. (B) For each sensor during the pre–stay-at-home period and the stay-at-home period, each feature was extracted from 15 time slices. The pre–stay-at-home features were subtracted from the stay-at-home features to obtain the behavior change features. First, raw data from the device sensor were preprocessed and then filtered by a time-of-the-day epoch and a days-of-the-week option. Features were then extracted from the selected raw data.



Feature Extraction

We computed features from the 6 sensors of calls, heart rate, location, screen, sleep, and steps, given their potential to inform depressive symptoms [29,44-48], as well as fatigue [49], MS symptom burden such as decreased mobility [27], and sleep quality [50,51].

Location features captured mobility patterns. Steps and heart rate captured the extent of physical activities. Calls features captured communication patterns. Screen features might inform the ability for concentration [52,53] and the extent of sedentary behavior [54], despite potential caveats for people with MS and other chronic neurological disorders. Sleep features captured sleeping duration and patterns, which could indicate sleep disturbance (eg, insomnia or hypersomnia) associated with depression [55]. Please see Multimedia Appendix 1 (section A.1 [29,44,46,56-59]) for details of features extracted from each sensor.

Features from the 6 sensors were extracted over a range of temporal slices (Figure 1B) preceding and during the stay-at-home period. For each period, we obtained the daily averages of these features by computing the average of the daily feature values. We computed features of behavior changes by subtracting the daily averages of features during the baseline (pre–stay-at-home) period from the stay-at-home period for the machine learning models.

Temporal Slicing

The temporal slicing approach extracted sensor features from different time segments (Figure 1B). Past work showed that this approach can better define the relationship between a feature and depression. For example, Chow et al [60] found no relationship between depression and the time spent at home during 4-hour time windows, but they found that people with more severe depression tended to spend more time at home between 10 AM and 6 PM. Similarly, Saeb et al [61] found that the same behavioral feature calculated over weekdays and weekends could have a very different association with depression. Here, we obtained all available data (spanning multiple days of the study) from a specific epoch or time segment of the day (all day, night [ie, 12 AM-6 AM], morning [ie, 6 AM-12 PM], afternoon [ie, 12 PM-6 PM], and evening [ie, 6 PM-12 AM]) and for specific days of the week (all days of the week, weekdays only [ie, Monday-Friday], and weekends only [ie, Saturday-Sunday]) to achieve 15 data streams or temporal slices. To extract features from each of the 15 temporal slices, we first computed daily features, averaged daily features from the pre–stay-at-home period, and averaged daily features from the stay-at-home period. We then subtracted the pre–stay-at-home feature matrix from the stay-at-home feature matrix to obtain the behavior change features. We concatenated the resulting 15 temporal slices of behavior change features to derive the final feature matrix.

Feature Matrix

After feature extraction, each of the 6 sensors had a feature matrix, with each sample containing a participant's feature vector comprising behavior change features from 15 different temporal slices.

Handling Missing Data

Missing sensor data can occasionally occur due to several reasons. Our approach for handling missing data is described in Multimedia Appendix 1 (section A.2).

Machine Learning Using Nested Feature Selection

We built machine learning models to predict dichotomized outcomes using the data set, building on a published approach [29], and validated our models using leave-5-participants-out cross-validation to minimize overfitting. The model generation process followed these steps:

1. Stable feature selection using randomized logistic regression, leveraging temporal slices.
2. Training and validating 1-sensor models for each of the 6 feature sets of calls, heart rate, location, screen, sleep, and steps.
3. Obtaining predictions from combinations of sensors by combining detection probabilities from 1-sensor models to identify the best performing model.
4. Classifying different outcomes by running the pipeline for each outcome.

Stable Feature Selection

To enable stable feature selection from a vast number of behavioral features, Chikersal et al [29] proposed an approach called "nested randomized logistic regression," which we deployed in this study. This method decomposed the feature space for each sensor by grouping features from the same time slices and performed randomized logistic regression on each of these groups. The selected features from all groups (ie, all time slices) are then concatenated to give a new and much smaller set of features. Next, we performed randomized logistic regression on this new set of features to extract the final selected features for the sensor. We performed the nested feature selection for each of the six 1-sensor models, thereby nesting the process. This method was performed in a leave-5-participants-out manner such that the model used to detect an outcome for a participant did not include that person during the feature selection process. More details about this method can be found in Multimedia Appendix 1 (section A.3).

Training and Validating 1-Sensor Models

For each sensor, we built a model of the selected features from that sensor to detect an outcome. We used leave-5-participants-out cross-validation to choose the parameters for that model. We trained models using the following 2 machine learning algorithms: logistic regression and gradient boosting classifier [29]. We chose the model with the best F_1-score for a given outcome, which provides the detection probabilities for the outcome. The process is independent of other outcomes.

Obtaining Predictions From Combinations of Sensors

The detection probabilities from all six 1-sensor models were concatenated into a single feature vector and given as input to an ensemble classifier (ie, AdaBoost with gradient boosting classifier as a base estimator), which then outputted the final label for the outcome. For all outcomes, only the detection probabilities of the positive label "1" were concatenated. The positive label was the "presence of depression" for "depression," "high burden" for "global MS symptom burden," "severe fatigue" for "fatigue," and "poor sleep quality" for "sleep quality." The "n_estimators" (the maximum number of estimators at which boosting is terminated) parameter was tuned during leave-5-participants-out cross-validation to achieve the best-performing combined model.

To analyze the usefulness of each sensor, we implemented a feature ablation analysis by generating detection results for all possible combinations of 1-sensor models. For six 1-sensor models, there were 57 combinations of feature sets, as the total combinations = combinations with 2 sensors + ... + combinations with 6 sensors =

Classifying Different Outcomes

This pipeline of training and validating six 1-sensor models and 57 combined models was run independently for each of the 4 outcomes. For each outcome, we reported the performance based on the best combination of sensors. We also reported the performance of baseline models (ie, a simple majority classifier whereby every point is assigned to whichever is in the majority in the training set) as well as models containing all 6 sensors.

Results

Participant Characteristics

The characteristics of the 56 participants were representative of the typical MS study (median age 43.5 years; n=48, 86% women). Table 1 shows the detailed participant characteristics.

Table 1. Study participant characteristics.

Variable	Value
Sex, n (%)	
Female	48 (86)
Male	8 (14)
Race, n (%)	
White	51 (91)
African or African American	5 (9)
Ethnicity, n (%)	
Non-Hispanic or Latino	55 (98)
Hispanic or Latino	1 (2)
Age (years), median (IQR)	43.5 (37-52)
Time elapsed (years) from age of first neurological symptom onset to study participation, median (IQR)	13.0 (6.7-17.4)
PDDS[a] score at start of study, median (IQR)	1 (0-3)
Disease-modifying treatment, n (%)	
Higher efficacy	38 (68)
Standard efficacy	12 (21)
Depression diagnosis, n (%)	
Not diagnosed with clinical depression before study enrollment	39 (70)
Diagnosed with clinical depression before study enrollment	17 (30)
Pharmacotherapy for depression, n (%)	
Not taking medication for depression before study enrollment	39 (70)
Taking medication for depression before study enrollment	17 (30)
Nonpharmacotherapy for depression, n (%)	
Not receiving nonmedication therapy for depression before study enrollment	52 (93)
Receiving nonmedication therapy for depression before study enrollment	4 (7)
Study outcomes: average measures during the stay-at-home period, median (IQR)	
PHQ-9[b] (depression)	3.7 (0.0-7.4)
MSRS-R[c] (global MS[d] symptom burden)	7.5 (3.4-10.3)
MFIS-5[e] (fatigue)	8.0 (4.6-11.0)
PSQI[f] (sleep quality)	11.0 (7.8-14.3)

[a]PDDS: Patient Determined Disease Steps.

[b]PHQ-9: Patient Health Questionnaire-9.

[c]MSRS-R: Multiple Sclerosis Rating Scale—Revised.

[d]MS: multiple sclerosis.

[e]MFIS-5: Modified Fatigue Impact Scale-5.

[f]PSQI: Pittsburgh Sleep Quality Index.

Interrelated Outcomes

The main study outcome is patient-reported depression as well as associated neurological symptom burden, fatigue, and sleep quality. We measured the Pearson correlations among the average values of the 4 outcomes during the stay-at-home period for the participants. Depression severity (PHQ-9) correlated with the global MS symptom burden (MSRS-R), fatigue severity (MFIS-5), and sleep quality (PSQI; Figure 2).

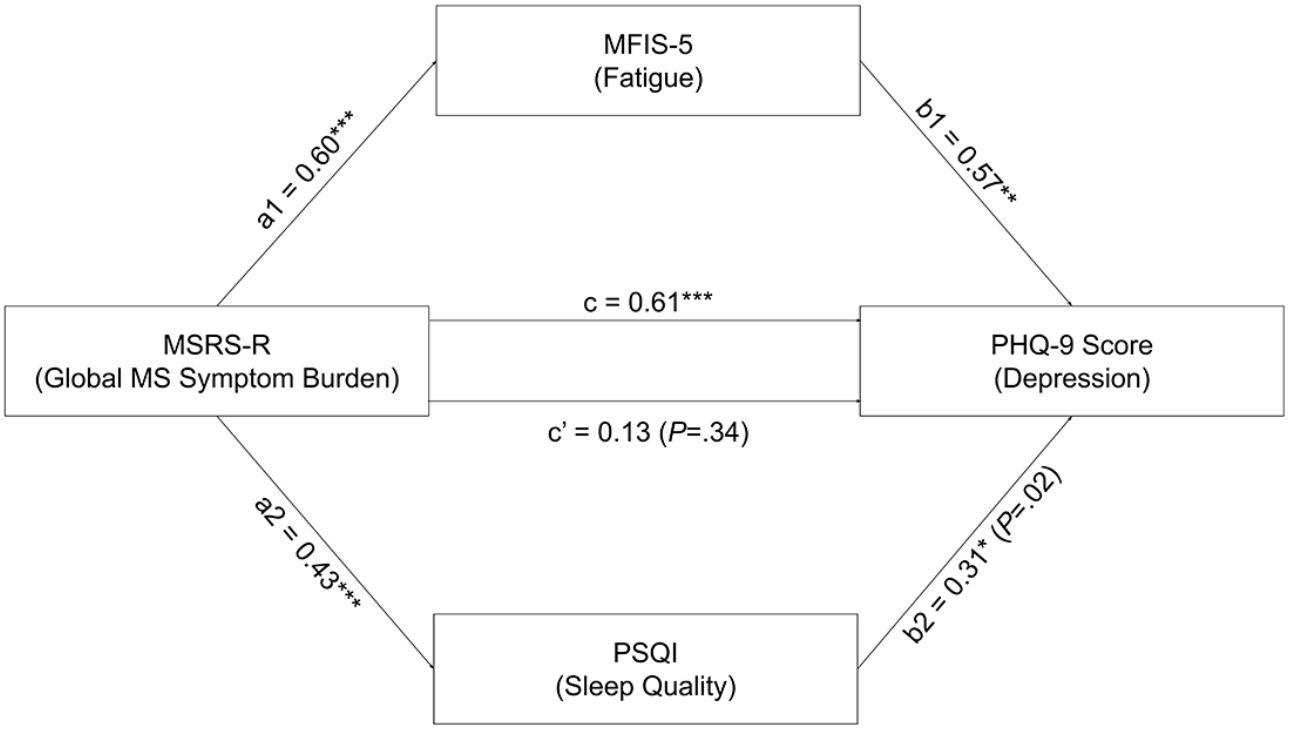
To dissect the complex relationship among these outcomes to inform better patient monitoring and guide potentially more precise interventions, we performed mediation analysis (Figure 3). When MFIS-5 and PSQI were both included as mediators in the model (path c'), the association between MSRS-R and

PHQ-9 was no longer significant (effect size=0.13 and the bias-corrected bootstrap confidence intervals=–0.14 and 0.40). However, the association between MSRS-R and PHQ-9 through MFIS-5 (path a1b1) remained significant (effect size=0.34 and the bias-corrected bootstrap confidence intervals=0.13-0.52). The association between MSRS-R and PHQ-9 through PSQI (path a2b2) also remained significant (effect size=0.13 and the bias-corrected bootstrap confidence intervals=0.02-0.27). Hence, the relationship between the global MS symptom burden and depression might be mediated by both fatigue and sleep quality.

Figure 2. Correlations among the 4 clinically relevant patient-reported outcomes in this study. For all correlations, *P*<.001 (N=56). MFIS-5: Modified Fatigue Impact Scale-5; MSRS-R: Multiple Sclerosis Rating Scale—Revised; PHQ-9: Patient Health Questionnaire-9; PSQI: Pittsburgh Sleep Quality Index.

	1	2	3	4
1. PHQ-9 Score (Depression)	1			
2. MSRS-R (Functional Disability)	0.602	1		
3. MFIS-5 (Fatigue)	0.713	0.73	1	
4. PSQI (Sleep Quality)	0.597	0.486	0.562	1

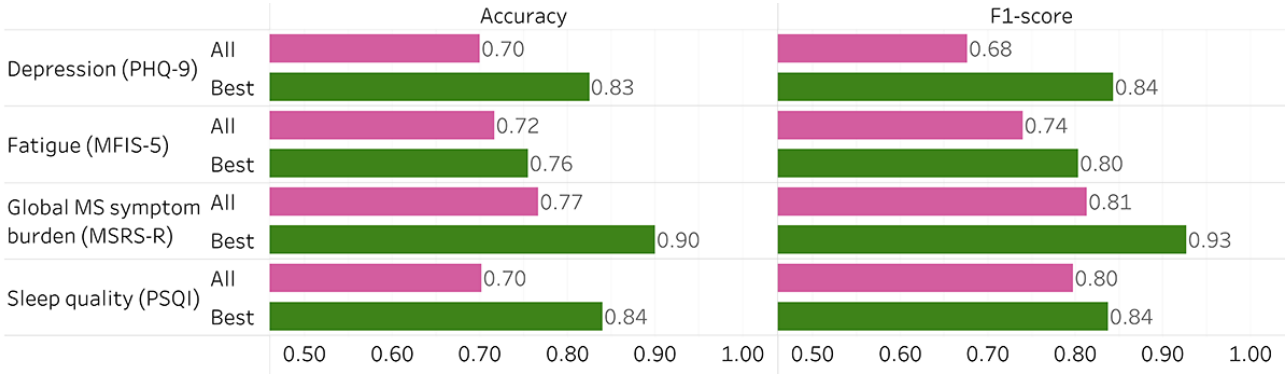
Figure 3. Parallel mediation analysis. Path model showing the effect of Multiple Sclerosis Rating Scale—Revised (MSRS-R; measuring global MS symptom burden) on the Patient Health Questionnaire-9 (PHQ-9) score (measuring depression) as mediated simultaneously by Modified Fatigue Impact Scale-5 (MFIS-5; measuring fatigue) and Pittsburgh Sleep Quality Index (PSQI; measuring sleep quality). Path c represents the effect of MSRS-R on PHQ-9 without mediators in the model. Path c' represents the effect of MSRS-R on PHQ-9 when MFIS-5 and PSQI mediators are included in the model. Paths a1b1 and a2b2 represent the effect of MSRS-R on PHQ-9 through MFIS-5 or PSQI respectively. The figure shows nonstandardized β regression coefficients (*P<.05, **P<.001, ***P<.0001) as reported by PROCESS Macro in SPSS [43]. MS: multiple sclerosis.



Predicting Outcomes During the Stay-at-home Period

Figure 4 shows the performance of the machine learning pipeline for predicting each of the 4 outcomes using the best sensor combinations (ie, the set of sensors that had the best performance for each outcome). Accuracy is the percentage of patients for whom the outcome label was correctly predicted. F_1-score is a metric of model performance that measures the harmonic mean of precision and recall. Precision is the positive predictive value, or the number of true positive labels divided by the number of all positive labels (true positive + false positive). Recall is sensitivity, or the number of true positive labels divided by the number of all patients who should have the positive labels (true positive + false negative). In this study, "positive" label refers to the outcome of interest (eg, presence of depression is the positive label for depression). Figures S1 to S4 in Multimedia Appendix 1 report the performance of individual sensors and when all 6 sensors were included. Tables S1 to S4 in Multimedia Appendix 1 list the features selected by the best models for each outcome, and their corresponding coefficients.

Figure 4. Performance of the machine learning pipeline using all sensors and the best sensor combination for predicting each of the 4 clinically relevant outcomes in people with multiple sclerosis during a state-mandated stay-at-home period. "Accuracy (All Sensors)" and "F1 Score (All Sensors)" are the accuracy (× 0.01) and F1-score obtained by combining all 6 sensors. "Accuracy (Best Sensors)" and "F1 Score (Best Sensors)" are the accuracy (× 0.01) and F1-score obtained by the best combination of sensors. See Multimedia Appendix 1 for additional performance metrics of all models. MFIS-5: Modified Fatigue Impact Scale-5; MSR-R: Multiple Sclerosis Rating Scale—Revised; PHQ-9: Patient Health Questionnaire-9; PSQI: Pittsburgh Sleep Quality Index.



Depression

The baseline model (simple majority classifier) had an accuracy of 50.0% in predicting the presence of depression during the stay-at-home period. The model containing all sensors had an accuracy of 70% (40% improvement over the baseline). The model with the best combination of sensors (calls, heart rate, and location) had an accuracy of 82.5% (65% improvement over the baseline).

Global MS Symptom Burden

The baseline model had an accuracy of 64.7% in predicting high global MS symptom burden (versus “low burden”) during the stay-at-home period. The model containing all sensors had an accuracy of 76.7% (18.5% improvement over the baseline). The model with the best combination of sensors (calls, heart rate, location, and screen) had an accuracy of 90% (39% improvement over the baseline).

Fatigue

The baseline model had an accuracy of 61.8% in predicting severe fatigue (versus “mild fatigue”) during the stay-at-home period. The model containing all sensors had an accuracy of 71.7% (16% improvement over the baseline). The model with the best combination of sensors (calls, heart rate, and location) had an accuracy of 75.5% (22% improvement over the baseline)

Sleep Quality

The baseline model had an accuracy of 65.7% in predicting poor sleep quality (ie, “poor sleep quality” versus “better sleep quality”) during the stay-at-home period. The model containing all sensors had an accuracy of 70.2% (7% improvement over the baseline). The model with the best combination of sensors (location and screen) had an accuracy of 84% (28% improvement over the baseline).

Discussion

Principal Findings

In this unique natural experiment conducted during the early wave of the COVID-19 pandemic, we reported the clinical utility of digital phenotyping for predicting clinically relevant outcomes for people with MS. Using only passively sensed data, our machine-learning models predicted the presence of depression, high global MS symptom burden, severe fatigue, and poor sleep quality during the stay-at-home period with potentially clinically actionable performance.

The best models outperformed not only baseline models (simple majority classifier) but also models containing all sensors. The best sensor combinations for predicting depression and fatigue were the same (ie, calls, heart rate, and location), while these sensors were also included in the best sensor combination for predicting global MS symptom burden (ie, calls, heart rate, location, and screen). Comparably, the best sensor combination for sleep quality (ie, location and screen) had the smallest overlap with the sensor combinations for the other three outcomes. This observation was consistent with the finding that depression, fatigue, and global MS symptom burden were better correlated among themselves than with sleep quality (Figure 2). We also looked at the feature coefficients of the features selected by the best models (Multimedia Appendix 1, section B.2). Examples of the best features of changed behavior selected by the best model for predicting depression (ie, features with the highest absolute coefficients) include increase in number of incoming calls during evenings on weekdays, decrease in average heart rate when the person is at rest or has low activity (outside exercise heart rate zones) during evenings on weekends, and increase in regularity in movement patterns in 24-hour periods with respect to nights on weekends.

Our findings built on a small body of prior work that explored the feasibility of passive sensing in people with MS and preliminary correlations between passively sensed behaviors and MS outcomes. For example, Newland et al [26] explored real-time depth sensors at home to identify gait disturbance and falls in 21 patients with MS. Other studies reported correlations between passively sensed physical activity and disability worsening in people with MS [27,62,63]. Chitnis et al [28] examined the gait, mobility, and sleep of 25 people with MS over 8 weeks using sensors mounted on their wrist, ankle, and sternum, and reported correlations among gait-related features (eg, turn angle and maximum angular velocity), sleep and activity, and disability outcomes.

Previous work on predicting health outcomes for people with MS using passively sensed behaviors is scarce. Tong et al [49] used passively sensed sleep and activity data collected from 198 people with MS over 6 months to predict fatigue severity and overall health scores, achieving good performance in line with acceptable instrument errors. To our knowledge, our study is the first to use passively sensed behavior changes to predict multiple interrelated clinically relevant health outcomes in MS, including depression, disability, fatigue, and sleep quality. While several studies used passively sensed data from the general population to report behavior changes during the COVID-19 pandemic [64-67], our study provides the first real-world evidence of potential clinical utility of passively sensed behavior changes to predict health outcomes during the unique stay-at-home period in a population with a chronic neurological disorder and complex health needs. From a methodological standpoint, the application of behavioral features computed over temporal slices to predict depression and other health outcomes in people with MS is novel. Our approach of using change in features between the period preceding the stay-at-home and stay-at-home periods to predict outcomes during the stay-at-home period is also novel. Finally, we included new heart rate features that can be computed using data from the Fitbit application programming interface.

Our approach has potential clinical utility, particularly during major stressful events (beyond COVID-19) that worsen health outcomes and limit health care access. For instance, predictive models built using our approach could help patients self-monitor their health when access to in-person clinical care becomes suddenly limited and could encourage patients (or their caregivers) to actively seek medical attention sooner when the models predict adverse outcomes. Further, our models could help clinicians better monitor at-risk patients and make triage decisions for patients who require prioritization for interventions (eg, medication and counseling), particularly in the setting of suddenly limited health care access and scarce resources.

Limitations

Our study has 2 limitations. First, the COVID-19 pandemic started during our data collection for an ongoing larger study of people with MS. While it provided a unique opportunity to conduct a natural experiment to assess the utility of digital phenotyping to predict health outcomes in people with MS during the highly unusual stay-at-home period, we had a modest sample size of participants who happened to have sufficient sensor data collected both just before the sudden issue of the stay-at-home order and during the stay-at-home period. We also had limited ability to seek external replication of the drastic behavior changes during the early stage of the pandemic since the stay-at-home order was lifted and has not been reinstated. To reduce the chance of overfitting and improve the validity of the findings, we used leave-5-participants-out cross-validation, such that in each fold, the participants used for training and testing were different. Our approach performed well for not only 1 outcome but all 4 clinically relevant outcomes pertaining to mental health and neurological disability in people with MS. We have reasonable confidence because of the consistently good model performance across all 5 folds and the consistently robust predictions for all 4 outcomes. We are not aware of other published studies with data from before and during the stay-at-home orders, particularly involving patient population with chronic neurological disorders such as MS who are at heightened risk for adverse health outcomes resulting from social isolation, reduced support, and limited health care access. Given the uniqueness of the data set, we believe the findings are clinically relevant despite the relatively modest sample size. Second, the study used patient-reported health outcomes. Given the restriction of in-person clinical visits during the stay-at-home period, rater-performed examination was not feasible. Importantly, these patient-reported outcomes are all validated for people with MS, highly correlated with rater-determined measures, interrelated among themselves, and clinically relevant.

In summary, we reported the potential clinical utility of digital phenotyping in predicting subsequent health outcomes in people with MS during a COVID-19 stay-at-home period. Specifically, we predicted the presence of depression, high global MS symptom burden, severe fatigue, and poor sleep quality in people with MS during the stay-at-home period using passively sensed behavior changes measured by smartphone and wearable fitness tracker. The predictive models achieved potentially clinically actionable performance for all 4 outcomes. This study paved the way for future replication studies during major stressful events and has implications for future patient self-monitoring and clinician screening for urgent interventions in MS and other complex chronic diseases.

Acknowledgments

We would like to thank our undergraduate research assistants: Man Jun (John) Han, Dong Yun Lee, Kasey Park, Phoebe Soong, and Christine Wu, for helping us monitor participant compliance throughout the data collection process. We would also like to thank Yiyi Ren for helping develop the app used for data collection. We would also like to thank the research participants and their treating clinicians. The study is funded by the Department of Defense (CDMRP MS190178).

Authors' Contributions

PC designed and conceptualized the study; analyzed and interpreted data; and drafted and revised the manuscript for intellectual content. SV, KM, EW, and DQ played a major role in the data acquisition. AD and MG designed and conceptualized the study; interpreted the data; and drafted and revised the manuscript for intellectual content. ZX designed and conceptualized the study and had a major role in the data acquisition, data interpretation, drafting, and revision of the manuscript for intellectual content.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary material.
[DOCX File , 18995 KB - mental_v9i8e38495_app1.docx]

References

1. 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]
2. Panchal N, Kamal R, Cox C, Garfield R. The implications of covid-19 for mental health and substance use. Kaiser family foundation. 2020. URL: http://medfam.facmed.unam.mx/wp-content/uploads/2021/05/implicaciones-de-COVID-EN-LA-SALUD-MENTAL.pdf [accessed 2022-07-27]
3. 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]
4. Twenge JM, Joiner TE. U.S. Census Bureau-assessed prevalence of anxiety and depressive symptoms in 2019 and during the 2020 COVID-19 pandemic. Depress Anxiety 2020 Oct;37(10):954-956 [FREE Full text] [doi: 10.1002/da.23077] [Medline: 32667081]
5. Kujawa A, Green H, Compas BE, Dickey L, Pegg S. Exposure to COVID-19 pandemic stress: Associations with depression and anxiety in emerging adults in the United States. Depress Anxiety 2020 Dec 10;37(12):1280-1288. [doi: 10.1002/da.23109] [Medline: 33169481]
6. O'Connor RC, Wetherall K, Cleare S, McClelland H, Melson AJ, Niedzwiedz CL, et al. Mental health and well-being during the COVID-19 pandemic: longitudinal analyses of adults in the UK COVID-19 mental health & wellbeing study. Br J Psychiatry 2021 Jun;218(6):326-333 [FREE Full text] [doi: 10.1192/bjp.2020.212] [Medline: 33081860]
7. Lebel C, MacKinnon A, Bagshawe M, Tomfohr-Madsen L, Giesbrecht G. Elevated depression and anxiety symptoms among pregnant individuals during the COVID-19 pandemic. J Affect Disord 2020 Dec 01;277:5-13 [FREE Full text] [doi: 10.1016/j.jad.2020.07.126] [Medline: 32777604]
8. Motolese F, Rossi M, Albergo G, Stelitano D, Villanova M, Di Lazzaro V, et al. The psychological impact of COVID-19 pandemic on people with multiple sclerosis. Front Neurol 2020 Oct 30;11:580507 [FREE Full text] [doi: 10.3389/fneur.2020.580507] [Medline: 33193033]
9. Zanghì A, D'Amico E, Luca M, Ciaorella M, Basile L, Patti F. Mental health status of relapsing-remitting multiple sclerosis Italian patients returning to work soon after the easing of lockdown during COVID-19 pandemic: A monocentric experience. Mult Scler Relat Disord 2020 Nov;46:102561 [FREE Full text] [doi: 10.1016/j.msard.2020.102561] [Medline: 33045494]
10. Broche-Pérez Y, Jiménez-Morales RM, Monasterio-Ramos LO, Vázquez-Gómez LA, Fernández-Fleites Z. Fear of COVID-19, problems accessing medical appointments, and subjective experience of disease progression, predict anxiety and depression reactions in patients with Multiple Sclerosis. Mult Scler Relat Disord 2021 Aug;53:103070 [FREE Full text] [doi: 10.1016/j.msard.2021.103070] [Medline: 34119745]
11. Patten SB, Marrie RA, Carta MG. Depression in multiple sclerosis. Int Rev Psychiatry 2017 Oct 06;29(5):463-472. [doi: 10.1080/09540261.2017.1322555] [Medline: 28681616]
12. Chan CK, Tian F, Pimentel Maldonado D, Mowry EM, Fitzgerald KC. Depression in multiple sclerosis across the adult lifespan. Mult Scler 2021 Oct 14;27(11):1771-1780 [FREE Full text] [doi: 10.1177/1352458520979304] [Medline: 33307963]
13. Solaro C, Gamberini G, Masuccio FG. Depression in multiple sclerosis: epidemiology, aetiology, diagnosis and treatment. CNS Drugs 2018 Feb 7;32(2):117-133. [doi: 10.1007/s40263-018-0489-5] [Medline: 29417493]
14. Siegert RJ, Abernethy DA. Depression in multiple sclerosis: a review. J Neurol Neurosurg Psychiatry 2005 Apr 01;76(4):469-475 [FREE Full text] [doi: 10.1136/jnnp.2004.054635] [Medline: 15774430]
15. Feinstein A, Magalhaes S, Richard J, Audet B, Moore C. The link between multiple sclerosis and depression. Nat Rev Neurol 2014 Sep 12;10(9):507-517. [doi: 10.1038/nrneurol.2014.139] [Medline: 25112509]
16. Zhang Y, Taylor BV, Simpson S, Blizzard L, Campbell JA, Palmer AJ, et al. Feelings of depression, pain and walking difficulties have the largest impact on the quality of life of people with multiple sclerosis, irrespective of clinical phenotype. Mult Scler 2021 Jul 14;27(8):1262-1275. [doi: 10.1177/1352458520958369] [Medline: 32924841]
17. Diamond BJ, Johnson SK, Kaufman M, Graves L. Relationships between information processing, depression, fatigue and cognition in multiple sclerosis. Arch Clin Neuropsychol 2008 Mar;23(2):189-199. [doi: 10.1016/j.acn.2007.10.002] [Medline: 18053682]
18. Ford H, Trigwell P, Johnson M. The nature of fatigue in multiple sclerosis. Journal of Psychosomatic Research 1998 Jul;45(1):33-38. [doi: 10.1016/s0022-3999(98)00004-x]

19. Bakshi R, Shaikh ZA, Miletich RS, Czarnecki D, Dmochowski J, Henschel K, et al. Fatigue in multiple sclerosis and its relationship to depression and neurologic disability. Mult Scler 2000 Jun 02;6(3):181-185. [doi: 10.1177/135245850000600308] [Medline: 10871830]
20. Strober LB, Arnett PA. An examination of four models predicting fatigue in multiple sclerosis. Arch Clin Neuropsychol 2005 Jul;20(5):631-646. [doi: 10.1016/j.acn.2005.04.002] [Medline: 15894455]
21. Levin SN, Venkatesh S, Nelson KE, Li Y, Aguerre I, Zhu W, Multiple Sclerosis Resilience to COVID-19 (MSReCOV) Collaborative. Manifestations and impact of the COVID-19 pandemic in neuroinflammatory diseases. Ann Clin Transl Neurol 2021 Apr 22;8(4):918-928 [FREE Full text] [doi: 10.1002/acn3.51314] [Medline: 33616290]
22. Vogel AC, Schmidt H, Loud S, McBurney R, Mateen FJ. Impact of the COVID-19 pandemic on the health care of >1,000 People living with multiple sclerosis: A cross-sectional study. Mult Scler Relat Disord 2020 Nov;46:102512 [FREE Full text] [doi: 10.1016/j.msard.2020.102512] [Medline: 32977074]
23. Manacorda T, Bandiera P, Terzuoli F, Ponzio M, Brichetto G, Zaratin P, et al. Impact of the COVID-19 pandemic on persons with multiple sclerosis: Early findings from a survey on disruptions in care and self-reported outcomes. J Health Serv Res Policy 2021 Jul 18;26(3):189-197 [FREE Full text] [doi: 10.1177/1355819620975069] [Medline: 33337256]
24. Levit E, Cohen I, Dahl M, Edwards K, Weinstock-Guttman B, Ishikawa T, Multiple Sclerosis Resilience to COVID-19 (MSReCOV) Collaborative. Worsening physical functioning in patients with neuroinflammatory disease during the COVID-19 pandemic. Mult Scler Relat Disord 2022 Feb;58:103482 [FREE Full text] [doi: 10.1016/j.msard.2021.103482] [Medline: 35016114]
25. Huckvale K, Venkatesh S, Christensen H. Toward clinical digital phenotyping: a timely opportunity to consider purpose, quality, and safety. NPJ Digit Med 2019 Sep 6;2(1):88 [FREE Full text] [doi: 10.1038/s41746-019-0166-1] [Medline: 31508498]
26. Newland P, Wagner JM, Salter A, Thomas FP, Skubic M, Rantz M. Exploring the feasibility and acceptability of sensor monitoring of gait and falls in the homes of persons with multiple sclerosis. Gait Posture 2016 Sep;49:277-282. [doi: 10.1016/j.gaitpost.2016.07.005] [Medline: 27474948]
27. Shammas L, Zentek T, von Haaren B, Schlesinger S, Hey S, Rashid A. Home-based system for physical activity monitoring in patients with multiple sclerosis (Pilot study). Biomed Eng Online 2014 Feb 06;13:10 [FREE Full text] [doi: 10.1186/1475-925X-13-10] [Medline: 24502230]
28. Chitnis T, Glanz BI, Gonzalez C, Healy BC, Saraceno TJ, Sattarnezhad N, et al. Quantifying neurologic disease using biosensor measurements in-clinic and in free-living settings in multiple sclerosis. NPJ Digit Med 2019 Dec 11;2(1):123 [FREE Full text] [doi: 10.1038/s41746-019-0197-7] [Medline: 31840094]
29. Chikersal P, Doryab A, Tumminia M, Villalba DK, Dutcher JM, Liu X, et al. Detecting depression and predicting its onset using longitudinal symptoms captured by passive sensing. ACM Trans. Comput.-Hum. Interact 2021 Feb 28;28(1):1-41. [doi: 10.1145/3422821]
30. Levin SN, Riley CS, Dhand A, White CC, Venkatesh S, Boehm B, et al. Association of social network structure and physical function in patients with multiple sclerosis. Neurology 2020 Aug 07;95(11):e1565-e1574. [doi: 10.1212/wnl.0000000000010460]
31. Mani A, Santini T, Puppala R, Dahl M, Venkatesh S, Walker E, et al. Applying deep learning to accelerated clinical brain magnetic resonance imaging for multiple sclerosis. Front Neurol 2021 Sep 27;12:685276 [FREE Full text] [doi: 10.3389/fneur.2021.685276] [Medline: 34646227]
32. Boorgu DS, Venkatesh S, Lakhani CM, Walker E, Aguerre IM, Riley C, et al. The impact of socioeconomic status on subsequent neurological outcomes in multiple sclerosis. Mult Scler Relat Disord 2022 Jun;65:103994. [doi: 10.1016/j.msard.2022.103994]
33. Kever A, Walker ELS, Riley CS, Heyman RA, Xia Z, Leavitt VM. Association of personality traits with physical function, cognition, and mood in multiple sclerosis.. Mult Scler Relat Disord 2022 Feb;59:103648. [doi: 10.1016/j.msard.2022.103648]
34. Epstein S, Xia Z, Lee AJ, Dahk M, Edwards K, Levit E, Multiple Sclerosis Resilience to COVID-19 (MSReCOV) Collaborative. . Vaccination against SARS-CoV-2 in neuroinflammatory disease: early safety/tolerability data. . Mult Scler Relat Disord 2022 Jan;57:103433. [doi: 10.1016/j.msard.2021.103433]
35. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
36. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
37. 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]
38. Wicks P, Vaughan TE, Massagli MP. The multiple sclerosis rating scale, revised (MSRS-R): development, refinement, and psychometric validation using an online community. Health Qual Life Outcomes 2012 Jun 18;10:70 [FREE Full text] [doi: 10.1186/1477-7525-10-70] [Medline: 22709981]

39. Meca-Lallana V, Brañas-Pampillón M, Higueras Y, Candeliere-Merlicco A, Aladro-Benito Y, Rodríguez-De la Fuente O, et al. Assessing fatigue in multiple sclerosis: Psychometric properties of the five-item Modified Fatigue Impact Scale (MFIS-5). Mult Scler J Exp Transl Clin 2019 Nov 09;5(4):2055217319887987 [FREE Full text] [doi: 10.1177/2055217319887987] [Medline: 31741743]
40. Buysse DJ, Reynolds CIII, Monk TH, Hoch CC, Yeager AL, Kupfer DJ. Quantification of subjective sleep quality in healthy elderly men and women using the Pittsburgh sleep quality index (PSQI). Sleep 1991;14(4):331-338. [doi: 10.1093/sleep/14.4.331]
41. Fictenberg NL, Putnam SH, Mann NR, Zafonte RD, Millard AE. Insomnia screening in postacute traumatic brain injury: utility and validity of the Pittsburgh Sleep Quality Index. Am J Phys Med Rehabil 2001 May;80(5):339-345. [doi: 10.1097/00002060-200105000-00003] [Medline: 11327555]
42. Ferreira D, Kostakos V, Dey AK. AWARE: mobile context instrumentation framework. Front. ICT 2015 Apr 20;2:1-9. [doi: 10.3389/fict.2015.00006]
43. Hayes A. Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. New York, US: Guilford Press; 2017.
44. Saeb S, Zhang M, Kwasny MM, Karr CJ, Kording K, Mohr DC. The relationship between clinical, momentary, and sensor-based assessment of depression. Int Conf Pervasive Comput Technol Healthc 2015 Aug;2015:1-10 [FREE Full text] [doi: 10.4108/icst.pervasivehealth.2015.259034] [Medline: 26640739]
45. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, et al. Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones. 2014 Presented at: ACM international joint conference on pervasive and ubiquitous computing; September 13-17, 2014; Seattle, Washington. [doi: 10.1145/2632048.2632054]
46. Canzian L, Musolesi M. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. 2015 Presented at: ACM international joint conference on pervasive and ubiquitous computing; September 7-11, 2015; Osaka, Japan p. 1293-1304. [doi: 10.1145/2750858.2805845]
47. Xu X, Chikersal P, Dutcher JM, Sefidgar YS, Seo W, Tumminia MJ, et al. Leveraging collaborative-filtering for personalized behavior modeling. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol 2021 Mar 19;5(1):1-27. [doi: 10.1145/3448107]
48. Xu X, Chikersal P, Doryab A, Villalba DK, Dutcher JM, Tumminia MJ, et al. Leveraging routine behavior and contextually-filtered features for depression detection among college students. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol 2019 Sep 09;3(3):1-33. [doi: 10.1145/3351274]
49. Tong C, Craner M, Vegreville M, Lane ND. Tracking fatigue and health state in multiple sclerosis patients using connnected wellness devices. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol 2019 Sep 09;3(3):1-19. [doi: 10.1145/3351264]
50. Min JK, Doryab A, Wiese J, Amini S, Zimmerman J, Hong JI. Toss 'n' turn: smartphone as sleep and sleep quality detector. 2014 Presented at: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; April 26, 2014; Toronto, Canada p. 477-486. [doi: 10.1145/2556288.2557220]
51. Sano A, Phillips AJ, Yu AZ, McHill AW, Taylor S, Jaques N, et al. Recognizing academic performance, sleep quality, stress level, and mental health using personality traits, wearable sensors and mobile phones. 2015 Presented at: IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN); October 19, 2015; Cambridge, MA, USA p. 1-6.
52. Demirci K, Akgönül M, Akpinar A. Relationship of smartphone use severity with sleep quality, depression, and anxiety in university students. J Behav Addict 2015 Jun;4(2):85-92 [FREE Full text] [doi: 10.1556/2006.4.2015.010] [Medline: 26132913]
53. Kwon M, Lee J, Won W, Park J, Min J, Hahn C, et al. Development and validation of a smartphone addiction scale (SAS). PLoS One 2013;8(2):e56936 [FREE Full text] [doi: 10.1371/journal.pone.0056936] [Medline: 23468893]
54. Costigan SA, Barnett L, Plotnikoff RC, Lubans DR. The health indicators associated with screen-based sedentary behavior among adolescent girls: a systematic review. J Adolesc Health 2013 Apr;52(4):382-392. [doi: 10.1016/j.jadohealth.2012.07.018] [Medline: 23299000]
55. Nutt D, Wilson S, Paterson L. Sleep disorders as core symptoms of depression. Dialogues in Clinical Neuroscience 2022 Apr 01;10(3):329-336. [doi: 10.31887/dcns.2008.10.3/dnutt]
56. Press WH, Rybicki GB. Fast algorithm for spectral analysis of unevenly sampled data. The Astrophysical Journal. URL: https://adsabs.harvard.edu/pdf/1989ApJ...338..277P [accessed 2022-07-27]
57. Mantua J, Gravel N, Spencer RMC. Reliability of sleep measures from four personal health monitoring devices compared to research-based actigraphy and polysomnography. Sensors (Basel) 2016 May 05;16(5):646 [FREE Full text] [doi: 10.3390/s16050646] [Medline: 27164110]
58. Cook JD, Prairie ML, Plante DT. Utility of the Fitbit Flex to evaluate sleep in major depressive disorder: A comparison against polysomnography and wrist-worn actigraphy. J Affect Disord 2017 Aug 01;217:299-305 [FREE Full text] [doi: 10.1016/j.jad.2017.04.030] [Medline: 28448949]
59. de Zambotti M, Goldstone A, Claudatos S, Colrain IM, Baker FC. A validation study of Fitbit Charge 2™ compared with polysomnography in adults. Chronobiol Int 2018 Apr;35(4):465-476. [doi: 10.1080/07420528.2017.1413578] [Medline: 29235907]

60. Chow PI, Fua K, Huang Y, Bonelli W, Xiong H, Barnes LE, et al. Using mobile sensing to test clinical models of depression, social anxiety, state affect, and social isolation among college students. J Med Internet Res 2017 Mar 03;19(3):e62 [FREE Full text] [doi: 10.2196/jmir.6820] [Medline: 28258049]
61. Saeb S, Lattie EG, Schueller SM, Kording KP, Mohr DC. The relationship between mobile phone location sensor data and depressive symptom severity. PeerJ 2016;4:e2537 [FREE Full text] [doi: 10.7717/peerj.2537] [Medline: 28344895]
62. Block VJ, Bove R, Zhao C, Garcha P, Graves J, Romeo AR, et al. Association of continuous assessment of step Count by remote monitoring with disability progression among adults with multiple sclerosis. JAMA Netw Open 2019 Mar 01;2(3):e190570 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.0570] [Medline: 30874777]
63. Stuart CM, Varatharaj A, Domjan J, Philip S, Galea I, SIMS study group. Physical activity monitoring to assess disability progression in multiple sclerosis. Mult Scler J Exp Transl Clin 2020 Dec 07;6(4):2055217320975185 [FREE Full text] [doi: 10.1177/2055217320975185] [Medline: 33343919]
64. Sun S, Folarin AA, Ranjan Y, Rashid Z, Conde P, Stewart C, RADAR-CNS Consortium. Using smartphones and wearable devices to monitor behavioral changes during COVID-19. J Med Internet Res 2020 Sep 25;22(9):e19992 [FREE Full text] [doi: 10.2196/19992] [Medline: 32877352]
65. Ong J, Lau T, Massar SAA, Chong ZT, Ng BKL, Koek D, et al. COVID-19-related mobility reduction: heterogenous effects on sleep and physical activity rhythms. Sleep 2021 Feb 12;44(2):zsaa179 [FREE Full text] [doi: 10.1093/sleep/zsaa179] [Medline: 32918076]
66. Pépin JL, Bruno RM, Yang R, Vercamer V, Jouhaud P, Escourrou P, et al. Wearable activity trackers for monitoring adherence to home confinement during the COVID-19 pandemic worldwide: data aggregation and analysis. J Med Internet Res 2020 Jun 19;22(6):e19787 [FREE Full text] [doi: 10.2196/19787] [Medline: 32501803]
67. Huckins JF, daSilva AW, Wang W, Hedlund E, Rogers C, Nepal SK, et al. Mental health and behavior of college students during the early phases of the COVID-19 pandemic: longitudinal smartphone and ecological momentary assessment study. J Med Internet Res 2020 Jun 17;22(6):e20185 [FREE Full text] [doi: 10.2196/20185] [Medline: 32519963]

Abbreviations

MFIS-5: Modified Fatigue Impact Scale-5
MS: multiple sclerosis
MSRS-R: Multiple Sclerosis Rating Scale—Revised
PHQ-9: Patient Health Questionnaire-9
PSQI: Pittsburgh Sleep Quality Index

Edited by J Torous; submitted 06.04.22; peer-reviewed by N Marotta, N Chiaravalloti; comments to author 27.06.22; revised version received 15.07.22; accepted 16.07.22; published 24.08.22.

Please cite as:
Chikersal P, Venkatesh S, Masown K, Walker E, Quraishi D, Dey A, Goel M, Xia Z
Predicting Multiple Sclerosis Outcomes During the COVID-19 Stay-at-home Period: Observational Study Using Passively Sensed Behaviors and Digital Phenotyping
JMIR Ment Health 2022;9(8):e38495
URL: https://mental.jmir.org/2022/8/e38495
doi:10.2196/38495
PMID:35849686

Original Paper

Clinical Targets and Attitudes Toward Implementing Digital Health Tools for Remote Measurement in Treatment for Depression: Focus Groups With Patients and Clinicians

Valeria de Angel[1,2], MSc; Serena Lewis[1,3], BSc; Katie M White[1], BSc; Faith Matcham[1,4], CPsychol, PhD; Matthew Hotopf[1,2], MD, PhD

[1]Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, United Kingdom
[2]NIHR Maudsley Biomedical Research Centre, South London and Maudsley NHS Foundation Trust, London, United Kingdom
[3]Department of Psychology, University of Bath, Bath, United Kingdom
[4]School of Psychology, University of Sussex, Falmer, East Sussex, United Kingdom

Corresponding Author:
Valeria de Angel, MSc
Department of Psychological Medicine
Institute of Psychiatry, Psychology and Neuroscience
King's College London
E3.08, 3rd floor East Wing
16 De Crespigny Park
London, SE5 8AF
United Kingdom
Phone: 44 7415838882
Email: valeria.de_angel@kcl.ac.uk

Abstract

Background: Remote measurement technologies, such as smartphones and wearable devices, can improve treatment outcomes for depression through enhanced illness characterization and monitoring. However, little is known about digital outcomes that are clinically meaningful to patients and clinicians. Moreover, if these technologies are to be successfully implemented within treatment, stakeholders' views on the barriers to and facilitators of their implementation in treatment must be considered.

Objective: This study aims to identify clinically meaningful targets for digital health research in depression and explore attitudes toward their implementation in psychological services.

Methods: A grounded theory approach was used on qualitative data from 3 focus groups of patients with a current diagnosis of depression and clinicians with >6 months of experience with delivering psychotherapy (N=22).

Results: Emerging themes on clinical targets fell into the following two main categories: *promoters* and *markers* of change. The former are behaviors that participants engage in to promote mental health, and the latter signal a change in mood. These themes were further subdivided into external changes (changes in behavior) or internal changes (changes in thoughts or feelings) and mapped with potential digital sensors. The following six implementation acceptability themes emerged: technology-related factors, information and data management, emotional support, cognitive support, increased self-awareness, and clinical utility.

Conclusions: The *promoters* versus *markers* of change differentiation have implications for a causal model of digital phenotyping in depression, which this paper presents. Internal versus external subdivisions are helpful in determining which factors are more susceptible to being measured by using active versus passive methods. The implications for implementation within psychotherapy are discussed with regard to treatment effectiveness, service provision, and patient and clinician experience.

(JMIR Ment Health 2022;9(8):e38934) doi:10.2196/38934

KEYWORDS

depression; digital health tools; implementation; qualitative; wearable devices; smartphone; passive sensing; sensor data; mobile health; mHealth; mood disorders; digital phenotyping; mobile phone

Introduction

The widespread availability of remote measurement technologies (RMTs), such as smartphones and wearables, provides opportunities to assist in the management of patients with long-term health conditions, such as depression. Passive sensing involves the automatic monitoring of behavioral, physiological, and environmental information using multiple digital sensors [1], whereas active sensing requires user input by replying to questionnaires or completing smartphone-based tasks. Used in combination, active and passive monitoring provide ways of capturing continuous, ecologically valid, and high-resolution measures of signs and symptoms related to depression. There are numerous potential uses for such data, including outcome measurements, patient stratification, and clinical decision-making within the treatment.

Cognitive behavioral therapy (CBT) is the first-line treatment for people with mild to moderate depression; however, approximately 50% of patients do not recover from the episode following treatment [2,3]. To date, there are no reliable predictive indicators of treatment outcomes [4]; therefore, a key question is whether RMTs could be used to identify predictors of recovery. For example, RMTs could detect certain behavioral subtypes of depression, which may be more responsive to CBT than antidepressants.

Another application of RMTs is as an outcome measure. Currently, clinical and research outcome measurements rely on infrequent use of symptom scales, which rely on patients remembering and communicating complex mood states, sleep, appetite, and other core symptoms of depression, which is an ability impaired in depression [5]. RMT, by directly measuring key features of depression such as sleep and activity, may provide more valid indicators of core depression phenomenology and therefore provide better measures of treatment outcomes.

A further application of RMTs would be to provide information to clinicians and patients, which could enhance the delivery of care. For example, the impact of interventions such as sleep hygiene or behavioral activation could be observed by patients and clinicians using RMT, which would provide a more direct and continuous indicator of change in targeted behaviors than when relying on diaries or conventional outcome measures.

The measure of behaviors of interest in depression has taken a *bottom-up* or data-driven approach to infer clinical states from digital sensors; digital markers of behavior have adjusted to existing sensors rather than the other way around [6]. These approaches are helpful in identifying potential digital biomarkers of disease; however, they are not always clinically meaningful. To generate clinical targets relevant to a patient's needs, their views on what illness or improvement looks like for them should be included. Clinical outcomes in digital health research are based on mainstream diagnostic scale items; however, even established psychometrics are criticized for overlooking outcome domains that are important to patients [7]. Therefore, patient-centered approaches to digital mental health are needed.

User acceptability is at the core of technology adoption [8], and thus, the patient and clinician acceptability of RMTs is crucial for successful implementation within the treatment. Although RMTs have shown adequate levels of acceptability [9], few studies have explored in-depth views on patient experience [6], and those conducted in a clinical setting are fewer still [10]. This is problematic, given that treatment-seeking populations have the potential added acceptability considerations of increased burden of symptom severity, low cognitive and emotional resources [11], and ethical concerns surrounding personal data sharing [12].

In addition to barriers, identifying and harnessing facilitators of RMT use during treatment may further motivate their use within services. Simblett et al [11] found that patients felt that RMTs provided opportunities to connect with peers and brought about a sense of control and understanding of their condition. However, previous studies have mostly considered participants who were already using health tools as part of a research study [13-15]. Although the experiences of current users are undeniably helpful, understanding pre-use attitudes toward new tools, including barriers to and facilitators of their adoption, is central to their uptake and implementation in health care services.

Therefore, the objectives of this study were 2-fold. Using qualitative methods, we first aimed to understand what outcomes are important for clinicians and patients as they improve with treatment to create clinically meaningful targets for digital health research. Second, we aimed to explore patient and clinician attitudes toward the use of RMTs and identify any perceived barriers to and facilitators of using these methods during psychological treatments for depression.

Methods

Design

This was a qualitative study with a focus group design. A thematic analysis was used to identify overarching themes within the participants' attitudes and experiences. The topic guide was developed based on the research goals, where the 2 main aims were allocated to approximately half of the session each.

Participants

We recruited a total of 22 participants, of whom 16 (73%) were current or recent users of the Improving Access to Psychological Therapies (IAPT) Talking Therapy program, a psychotherapy delivery service provided by the United Kingdom's National Health Service (NHS). A separate group of clinicians within the IAPT program were also recruited (6/22, 27%). They either delivered CBT or were part of the care team for patients undergoing CBT for depression and anxiety.

Clinicians with at least 6 months of experience in their roles were recruited via email through their services. Patients were recruited through advertisements in mental health service waiting rooms and advertisements circulated by the staff. Patients were eligible if they were aged ≥18 years and had received at least three sessions of IAPT-delivered CBT for a depressive disorder in the past year.

Ethics Approval

The study was reviewed by the London Bridge Research Ethics Committee, and approval from the Health Research Authority was obtained (reference 19/LO/0662).

Patient and Public Involvement

This research was reviewed by a team with experience with mental health problems and their caregivers who were specially trained to advise on research proposals and documentation through the Feasibility and Acceptability Support Team for Researchers—a free, confidential service in England provided by the National Institute for Health Research Maudsley Biomedical Research Centre via King's College London, South London, and Maudsley NHS Foundation Trust. The reviewed documentation comprised all participant-facing documents, excluding the interview schedule.

Procedure

A total of 3 separate focus groups led by 2 researchers were conducted in August 2019 at King's College London, 2 for patients and 1 for clinicians. All participants provided written informed consent to participate in this study. Patient sessions ran for 2 hours and clinician groups for 1 hour. The participants were compensated £25 (US $32.50) for participating in the study.

Baseline demographic data, including age, gender, ethnicity, and the type of treatment received or administered, were recorded. During the focus groups, participants were prompted with a series of prespecified questions to explore the outcomes that people value in relation to their mental health (eg, improved sleep, increased socialization, and completion of daily chores). In addition, they were asked about their attitudes toward using smartphones and wearable technology during psychological treatment (see supplementary note 1 for the topic guides in Multimedia Appendix 1). Sessions were audio recorded and transcribed verbatim (excluding filler words, such as "erm" or "um"). The same 2 researchers—a PhD candidate and a postdoctoral health psychologist—facilitated all the sessions.

Data Analysis

Transcriptions were independently coded and analyzed by 2 researchers using the NVivo software (version 12; QSR International). To improve the breadth of perspectives and reduce researcher bias, coding was replicated by a qualitative researcher who was not present in the focus groups.

Following Braun and Clarke [16], an inductive, thematic analysis approach was followed, in which the data drive the generation of themes rather than a previous theoretical basis. The researchers read the transcripts to identify emerging themes. Recurring topics were grouped under the same code until the general patterns of themes were identified. The researchers then met to discuss and consolidate themes, generating a new list of codes, which was used for a further round of coding.

Results

Sample Demographics

There were a total of 22 participants. The main demographic characteristics of the sample are presented in Table 1. All patients had a diagnosis of depression, and 63% (10/16) reported comorbidities. Half of the participants (8/16, 50%) reported having comorbid anxiety disorders, including generalized anxiety disorder and panic disorder. The results were separated into two main sections: (1) relevant clinical outcomes and (2) implementation barriers and facilitators.

Table 1. Demographic characteristics of participants in the clinician focus group and both patient focus groups (N=22).

Characteristics	Clinicians (n=6)	Patients		Total (N=22)
		Group 1 (n=9)	Group 2 (n=7)	
Age (years), mean (SD)	36.7 (9.3)	47.9 (15.7)	47.7 (11.6)	44.6 (13.3)
Gender, n (%)				
Women	5 (83)	6 (67)	7 (100)	18 (82)
Men	1 (17)	3 (33)	—[a]	4 (18)
Ethnicity, n (%)				
Asian or Asian British	1 (16)	—	1 (14)	2 (9)
Black African or Caribbean, or Black British	1 (16)	3 (33)	—	4 (18)
White British	3 (50)	4 (44)	4 (57)	11 (50)
White other	—	1 (11)	1 (14)	2 (9)
Mixed or multiple ethnic groups	1 (16)	1 (11)	1 (14)	3 (14)
Comorbid anxiety diagnosis, n (%)	N/A[b]	4 (44)	4 (57)	8 (50)[c]
Physical health condition, n (%)	N/A	2 (22)	1 (14)	3 (19)[c]

[a]Not available (no participants with these characteristics).

[b]N/A: not applicable (diagnosis information not collected for the clinician group).

[c]Patient data only; total: N=16.

Clinical Outcomes

Overview

The first section of the focus group discussions revolved around identifying which behavioral changes signaled an improvement in, or worsening of, mood to our participants. Two main themes were identified: behaviors that people engage in, which have an impact on mood, named here as *promoters of change*, and behaviors that manifest because of a change in mood, termed as *markers of change*. These themes were further subdivided into either external changes (referring to changes in their behaviors and could, in principle, be measured objectively) or internal changes (referring to changes in thoughts or feelings rather than actions). Figure 1 summarizes the markers and promoters of change discussed in the groups with example quotes.

Figure 1. Internal and external markers and promoters of change and their corresponding remote measurement technology sensors through which they could be measured.

Changes	Internal	External	Potential passive sensors
Promoters of change	• Perfectionism and pressure • Insight into improvement • Engaging with treatment and homework	• Physical activity • Establishing a routine • Sleep hygiene • Setting realistic goals • Psychoeducation • Distractions and hobbies • Screen time and social media • Stigma	→ Accelerometer → GPS, phone use, accelerometer → Accelerometer, microphone, light → Phone use, calendar → App use → App use, microphone, GPS → Phone use → Social media use, call and message logs
Example quotes	*"If I connect with people who are expecting a bit too much from me, that can set me back."* *"If you put too much in your diary then you start to get really kind of stressed, trying to sort of fit it all in."*	"It's been really good for me I mean, you know, forcing myself to run and, just walking or to keep up and just get out and get moving." "Once they realise that they need to put routine back in their lives [...] it makes a huge difference to the individual and they do start feeling better"	

Markers of change	• Cognitive factors
	  ○ Concentration and focus
	  ○ Problem solving
	  ○ Ability to cope
	• Confidence and self-acceptance
	• Plans for the future
		• Socialisation
		• Increased energy and activity
		• Leaving the house
		• Task completion
		• Change in speech
		• Improved self-care
		• Body language
			→ Bluetooth, microphone, call logs
			→ Accelerometer, GPS
			→ GPS
			→ Calendar, app use
			→ Microphone
			→ Camera
			→ Accelerometer, magnetometer, gyroscope

Example quotes	*"People are less overwhelmed, they're able to problem solve, for the most part, and deal with their avoidance, which can increase their problems, especially if they're related to things that have necessary tasks that need to be completed."*	*"At some point due to high stress and exhaustion my body just shut down. And I wasn't able to leave my house at all."* *"[their] responses to questions becoming faster, and their content becoming longer and more detailed".*	

Promoters of Change

Internal

Three main subthemes emerged as internal processes that could help promote or hinder improvements in people's moods. First, participants found that feeling pressure from themselves or striving for perfectionism to be *better* or *do more* would hinder their improvement. Second, they felt that having insights into their illness and awareness of progress was gratifying. Third, experiencing an improvement in mood led some people to gradually disengage with treatment "instead of doing the work again so you do get complacent" (P07) by leaving homework incomplete or stopping the medication, which, in consequence, worsened mental health.

External

A prevalent subtheme for all focus groups as a promoter of mental well-being was staying active, including both vigorous exercise and milder physical activity. This was followed by establishing a routine, which both clinicians and patients found to be central to experiencing improvement. Striving to create better sleep habits to improve mental health and setting realistic goals in the short term gave people a sense of accomplishment. However, unrealistic targets ran the risk of bringing discouragement and worsening the mood if left incomplete.

Psychoeducation, including signing up for and engaging with mental health courses, was perceived as helpful, as was completing the homework assigned during therapy. In addition, participants found value in finding suitable distractions such as comedy shows, music, and pursuing hobbies. Notably, only clinicians discussed social media and screen time as having a potential impact on mental health.

Societal stigma or how other people reacted to participants' mental health difficulties was discussed only by patients as something that affected their mental state but recognized that they had no control over it.

Markers of Change

Internal

As people experienced improvement, they reported a general feeling of being better able to cope, feeling less overwhelmed with daily difficulties, and more motivated to go about their daily lives. A participant summarized it as "I find I'm able to run my day better" (P13). Participants also reported an increase in confidence, self-acceptance, and making more plans for the future, as their symptoms improved.

External

The single most mentioned theme was socialization and how, as mood improved, they "began to reach out to friends again, to reconnect with them socially" (P07). Going out more often and meeting an increased number of people were mentioned.

Lack of energy, both physically and mentally, was discussed as one of the more noticeable things, particularly with reference to being able to leave the house as a combination of both physical and mental motivation. Clinicians pointed out that "people become more active, don't they, as we go through therapy" (C02).

Successful completion of daily tasks and other small daily achievements was highlighted, which clinicians believed represented an improvement in cognitive function. Specifically, people noticed they were able to "get things off my to-do list rather than just postponing them" (P01).

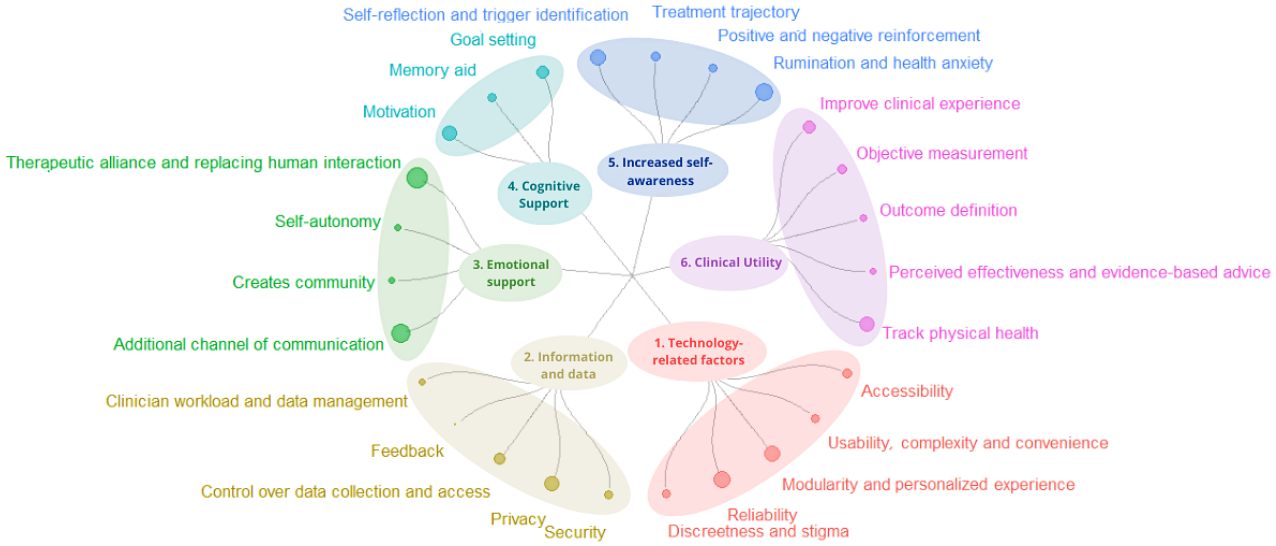
Clinicians reported noticing changes in a patient's speech; however, one of the clinicians pointed out that speech change patterns may be different in anxiety, which, in contrast to those with depression, may be faster to answer when symptoms are more severe. Changes in self-care habits were also reported by both clinicians and patients.

The patients expressed the importance of body language and how others could discern their mood from their physical appearance and facial expressions.

Implementation Barriers and Facilitators

The second part of the study focused on how RMTs, with the potential to measure the aforementioned symptoms, were perceived by patients and clinicians. A total of 6 overarching themes and 26 subthemes were identified (Figure 2), and quotes for each subtheme are presented in Table S1 in Multimedia Appendix 1.

Figure 2. A breakdown of the six themes and subsequent subthemes emerging from the data. End nodes correspond to the number of codes related to each subtheme; the larger the node, the more instances of coding for that subtheme.



Technology-Related Factors

Under the first theme of technology-related factors, we identified the subthemes of accessibility, usability and convenience, personalization of experience and modularity, reliability, and discreetness or stigma. Most participants were concerned about the price of the technology and the surrounding infrastructure,

including access to reliable Wi-Fi. Participants also agreed that technology and software would need to be simple to use and convenient; for example, requiring regular charging or software updates that took time and technological literacy would affect their engagement. Participants valued the option to personalize their experience and adapt the software to their needs by deciding what to measure and manipulating the screen display. Clinicians agreed that modularity within the software would also be useful for them.

Concerns arose regarding the reliability of the technology, especially devices with physiological readings, and how making health decisions based on unreliable sensors could potentially be harmful. Discretion was also important, as wearable devices should not identify people with psychiatric disorders. Participants preferred mobile phone apps to wearable devices; however, commercially available fitness devices or smartwatches would avoid these concerns.

Information and Data

Data and information management emerged among the three focus groups and revolved around five main topics: security, privacy, control over data collection, feedback, and clinician data management and workload. Compared with patients, clinicians verbalized more data considerations.

Within the subtheme of security, participants were mainly concerned about sensitive information being leaked or picked up through digital sensors and going out to third parties and private corporations. Privacy concerns were related to who could gain access to the data and what choice patients would have over their access. In addition, patients wanted control over data collection and the option to engage with this technology or opt out rather than being a prerequisite for treatment. Therefore, many participants preferred the idea of a wearable device over a smartphone app for health or research purposes, as having a separate device gave participants a sense of awareness and control over when data collection took place.

Patients generally wished to review their progress, and thus, receiving feedback from RMTs would be valuable but should be tailored according to personal preferences and in a way that promotes positive reinforcement.

For clinicians, themes emerged related to managing big data, training, and clinical workloads. They were concerned about receiving adequate training, which would affect the effectiveness of the tool and their overall workload, and discussed the necessity of software to help sift through large amounts of information to generate actionable insights. Clinicians raised further concerns about whether RMT data would be used to evaluate risk and whether this would add to their workload.

Emotional Support

Attitudes toward whether RMTs could provide an additional source of emotional support during treatment went in 2 opposite directions. On the positive end of the spectrum, participants believed that digital tools could complement human interaction by creating another channel of communication, which was thought to help with in-treatment discussions. Both clinicians and patients discussed how RMTs could accommodate the provision of lower-intensity extra support after treatment, for example, by having posttreatment check-ins. This technology was also thought to provide an opportunity to create a community by sharing achievements and involving others in their goals, thus providing an additional source of support.

Conversely, 2 further subthemes revolved around the fear of overreliance on technology and a decline in human interaction. Clinicians believed that relying on a wearable device during treatment could affect relapse rates if patients became reliant on these tools and they were then taken away after treatment. Patients echoed this sentiment and added that this overreliance may detract from a feeling of autonomy.

Any replacement of human interactions was almost universally considered detrimental to the therapeutic alliance. Participants believed that no longer feeling accountable to a therapist would detract from their motivation to engage with treatment or homework.

Cognitive Support

The positive motivational impact of automated messages or notifications being delivered to people via smartphones was discussed in all groups. Participants thought positively of RMTs as tools for memory aids in several ways, including helping manage practicalities such as taking medication and keeping track of what improves their mood. Clinicians believed that using RMTs as memory aids could improve adherence to homework or medication. Another way in which it was perceived as a complement to treatment was through planning and goal setting.

Increased Self-awareness

Subthemes within self-awareness revolved around having a reflexive tool to identify triggers and boosters of well-being, assess treatment trajectory, positive or negative reinforcements, and the potential to worsen rumination and health anxiety. Keeping a continuous log of these reflections was thought to help identify changes in symptom trajectories in an ecological context, such that it could provide guidance on what improves people's mood and identify negative triggers.

Clinicians discussed the benefits of potentially identifying treatment trajectories, such that they may adjust strategies depending on changes in behavior. A subtheme of positive reinforcement emerged in the groups, in which people believed that being shown progress could prove motivating. Others advised caution, as being presented with failure to achieve goals could lead to feelings of self-defeat. Being forewarned of a possible dip in mood gave some people a sense of foreboding, which they felt could create a self-fulfilling prophecy.

A prevalent concern was whether monitoring health would lead to increased anxiety as people obsess over data, which could, in turn, worsen existing symptoms such as lack of sleep. Indeed, some people disliked the idea of having a regular request to reflect on their mental state and, therefore, constant reminders of their ill mental health. Clinicians highlighted the risks to those with health anxiety if they were exposed to constant health monitoring and feedback.

Clinical Utility

The theme of clinical utility emerged in relation to how, and to what extent, digital tools could complement treatment. The identified subthemes were related to improving the clinical experience, objective symptom measurement, outcome definitions, treatment effectiveness or concerns over the evidence base, and targeting physical health.

RMTs were thought to improve the clinical experience by replacing questionnaires, offering an electronic alternative to paper, and gamifying data collection. Both patient groups expressed a strong dislike for repeated symptom measurement questionnaires, and automatic data collection was considered a pleasing alternative. For aspects of therapy that cannot be fully automated, such as homework, participants still expressed a stronger disposition to engage with electronic methods.

Clinicians discussed the utility of more objective behavioral or physiological symptom measures that could help contextualize a patient's symptoms. At the same time, they warned that objective benchmarks for outcome definitions would vary drastically across patients, such that symptom improvement may manifest with opposite digital signatures across different conditions.

Both patient groups discussed the issue of trust and whether technology was as effective as human-based treatments in the context of treatment. It was important for all groups that any recommendations based on digital tool readings be evidence based or accredited by the NHS to ensure their reliability.

An additional discussed benefit of using these devices in a clinical population was that they encourage improvement in physical health and sleep, which are items reportedly targeted in treatment.

Discussion

We aimed to understand what improvement-related outcomes are important for clinicians and patients, as well as the attitudes toward the use of RMTs within psychological treatment. The main purpose of this study was to generate clinically meaningful targets for RMT research and help in their implementation in research and clinical practice.

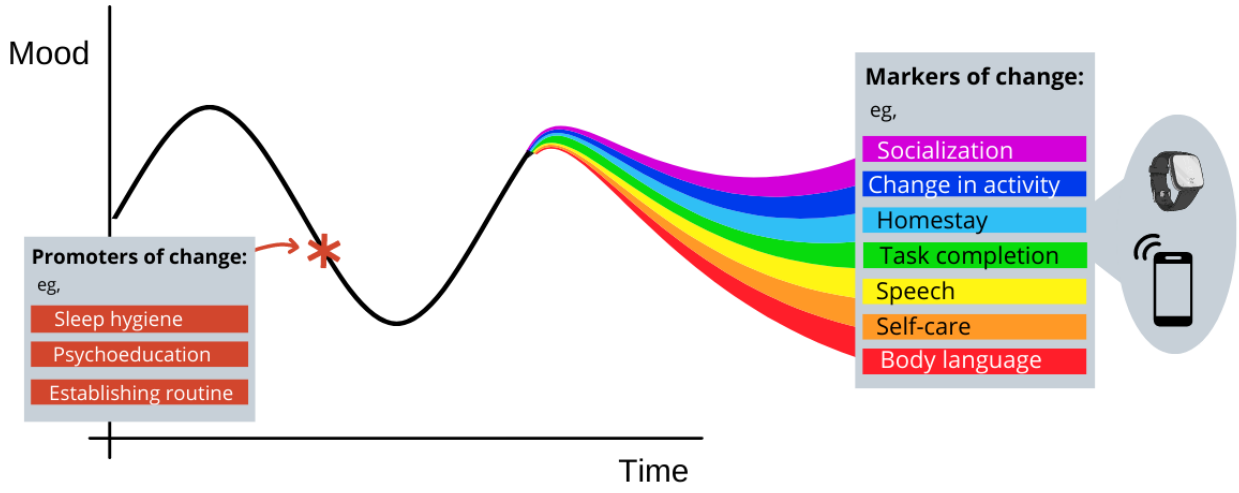
Clinical Outcomes

Participants talked about the behavioral, cognitive, and emotional changes experienced alongside mood fluctuations. Emerging themes were categorized as either markers or promoters of change to differentiate how participants viewed these fluctuations. Markers of change seemed to be referred to as consequences of improvement in mental state, whereas promoters were viewed as thoughts or behaviors that were perceived as affecting changes in mental state.

This differentiation has implications for the causal model of digital phenotyping in depression. Undoubtedly, patient and clinician perspectives do not necessarily reflect the reality of the time order relationship of these symptoms, which is a necessary characteristic of causal inference. However, this model can serve as a useful framework to contextualize the analysis of relationships between digital features. This is especially relevant as, with few exceptions [17], much of the existing work on digital mental health is based on correlations [18], and determining causality is necessary to extract actionable insights, especially as they relate to treatment [19].

In this model, depicted in Figure 3, markers of change could be used as a proxy for changes in mental state—the *effect* in a causal effect model—and, therefore, primary targets for the remote monitoring of symptoms. However, promoters of change may be unrelated to the *current* mental state but be actionable targets for treatment and predictors of mood fluctuations.

Figure 3. Model depiction of promoters and markers of change. As mood fluctuates, so do the markers of change, such as socialization, homestay, or speech, each represented by a colored line. These vary with mood and can be used in combination to assess the current mental state. Promoters of change, such as routine, sleep hygiene, and psychoeducation, can be viewed as clinical targets, which can be actioned at a time of downward mood trend (as depicted by the red asterisk) and promote improved mental health.



The subdivisions of internal versus external factors could be helpful in determining which factors are more susceptible to measurement with active versus passive methods. GPS and accelerometer sensors have been widely used for this purpose, with promising results [20,21]. Other factors, such as body language (gait and posture), maintenance of a routine, self-care,

and task completion, have a very scarce literature base if any (eg, gait [22] and task completion [23]), likely because of the difficulty in operationalizing them via remote sensing.

Despite the advantage of minimal engagement required for passive sensing, data from active apps such as smartphone calendars can provide low-burden information on items such as workload, leisure activities, and completed daily tasks. For example, Wahle et al [23] included calendar events in prediction algorithms to distinguish people with depression from nondepressed people. However, it is likely that a multidomain approach to data collection would be preferable to capture the complex nature of the behaviors of interest.

Questions have arisen regarding the accuracy and face validity of digital sensors that genuinely detect the behavior they claim to measure [24]. However, sensor data have been widely found to have adequate construct validity in that they can predict self-assessed moods [18]. This means that although caution is advised when inferring real behavior from digital features, they may serve as helpful signals for depression.

Implementation Barriers and Facilitators

Overview

The implementation of digital health tools in health care services requires insights into barriers and facilitators from stakeholders. In the second section of our focus group discussions, we identified six main themes: technology-related factors, information and data, emotional support, cognitive support, increased self-awareness, and clinical utility. The implications and impact of these themes on implementation within treatment with regard to treatment effectiveness, service provision, and patient and clinician experiences are discussed in the following sections.

Implications for Treatment Effectiveness

One of the necessary drivers of implementation is that these tools are perceived as improving treatment outcomes. The perception that RMTs could provide better communication and increased self-awareness are 2 routes through which this could be achieved. The third route is through supporting homework [25]. Being able to follow patient trajectories more closely would aid the development of prediction models for recovery and relapse [15] and help clinicians make prompt and better-informed decisions [26]. In addition, integrating a treatment platform where patient-clinician communication is enhanced with physical health tracking, psychoeducation, and memory and motivational aids that are specifically tailored to patient symptoms and treatment schedules could boost effectiveness and move toward a more holistic and personalized approach to treatment.

The patient-clinician therapeutic relationship was one of the most prominent themes, where concerns that it may replace human-delivered treatment can feel dismissive and indeed hinder treatment effectiveness, despite finding value in the extra low-intensity support technology provided. Despite the dearth of research on the effect of digital health on therapeutic relationships [27], a recent narrative review suggests that an alliance within digital mental health can be cultivated, although it may take a different form [28]. Given that a therapeutic alliance is a highly predictive factor for treatment response [28], such concerns should be carefully considered in implementation strategies [29].

Importantly, RMT capabilities in terms of feedback schedules need to be tailored to individual needs so as not to exacerbate symptoms. Generally, feedback was appreciated; however, close monitoring was strongly discouraged for those with health anxiety. Indeed, case studies from the physical health field have found that wearables worsen mental health in such cases [30].

Impact on Service Provision

Service provision could be negatively affected by the broadening of the digital divide—the gap between those who benefit from the digital age and those who do not. This study found that technology-related factors such as accessibility, usability, and complexity could affect the breadth of service provision by increasing the digital divide in three main ways: those with lower technology experience and literacy are less likely to benefit from technologies with more complex designs [31], people with severe depression are more likely to be unemployed and therefore have lower purchasing power for devices or stable internet [32], and reduced cognitive abilities could affect the capacity to effectively interact with the devices.

Despite research linking remote monitoring to increased overall access to health care [33] and an overall increase in the capacity for service provision, inequalities may still appear if those with less access, experience, or capacity are less likely to benefit from this technology.

Patient Experience

The main potential for improvement in patient experience is by replacing recurring symptom questionnaires, which were almost unanimously disliked by patients, who disliked the repetition and negative phrasing of their symptoms. This is supported by previous research showing that patients with depression would value measures of wellness and illness [34]. Digital active data collection methods could allow for different ways of capturing symptoms, such as visual or animated scales. Their use could also be reduced altogether by incorporating passive sensing.

Additional patient concerns revolved around data, specifically around having agency over data collection and access, a finding replicated elsewhere [11]. Therefore, patient-clinician discussions on which behaviors to monitor, and why, could improve attitudes toward RMT use in treatment. Privacy concerns could be reduced for some people if certain features, such as location-based data, were not deemed relevant to a patient's treatment and, therefore, not recorded.

Clinician Experience

A key facilitator for clinicians was the utility of having objective behavioral measures to contextualize symptoms. They believed that this could be used to motivate and engage patients in their care. However, a major factor that could affect clinician buy-in was related to data management and interoperability, a concern also found in primary care settings [35]. Electronic medical record systems are already a source of excess data inputting for clinicians. Adding active and passive data generated from digital

health tools is unlikely to improve care and may well overburden clinicians unless they are provided with adequate tools and training to manage and interpret data in actionable ways. Much of the research on acceptability revolves around patients; however, clinicians are end users who have been shown to express varying attitudes toward the implementation of new technologies, attitudes that can be improved through exposure to technology and training [36].

Strengths and Limitations

This study included clinician views, a key stakeholder with a strong influence on implementation success and whose practice we aim to influence. The inclusion of 3 separate focus groups allowed for different dynamics and a wider range of experiences to be captured in this study.

Some limitations include the fact that despite the importance of capturing attitudes before implementation, the barriers and facilitators reported here may vary after the experience of sustained RMT use. This study relied on a small sample of participants with treatment experience from a single health service, meaning that these results may not be broadly generalizable to other settings.

In addition, the participants' previous experiences with technologies were not included in the analysis.

There may also be confusion surrounding the term *RMTs* and what they can be used for; thus, a dichotomy may arise between the use for self-management, which patients are more likely to envision, or the use for prediction in clinical care. The latter is less understood and thus might have affected the participants' responses.

Future Research

Future longitudinal studies could include active data that capture the emotional and psychological factors presented in this study and could generate passive data features that closely match the behavioral markers reported here. Studies examining the severity of depression and device engagement are already underway [37]. Applying such methods to patients currently in psychotherapy could provide insights into how real engagement and device use vary within mental health services. Finally, the study by Torous et al [38] sheds light on the importance of supporting working alliances in digital health platforms; therefore, similar studies examining remote measurement and digital therapeutics are necessary.

Conclusions

Digital tools bring new methods of data collection and new outcomes; therefore, there is a need for a re-evaluation of the clinical targets that are considered important. Therefore, advice on what type of information should be measured and analyzed, as well as attitudes toward their use from different stakeholders, is central. This study used a qualitative approach to identify clinical targets that are important to patients and clinicians and developed a framework for determining which factors may be more susceptible to be measured using active or passive methods. We also identified 6 main themes surrounding attitudes toward RMTs, which could drive implementation efforts in health care settings.

Acknowledgments

This study represents independent research funded by the National Institute for Health and Care Research Biomedical Research Centre in South London and Maudsley National Health Service Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the National Health Service, the National Institute for Health and Care Research, or the Department of Health and Social Care. The funders of the study had no role in the study design, data collection, data analysis, data interpretation, or writing of the manuscript.

Conflicts of Interest

MH is the principal investigator of the Remote Assessment of Disease and Relapse–Central Nervous System program, a precompetitive public-private partnership funded by the Innovative Medicines Initiative and the European Federation of Pharmaceutical Industries and Associations.

Multimedia Appendix 1
Topic guide and patient and clinician quotes for themes and subthemes.
[DOCX File , 32 KB - mental_v9i8e38934_app1.docx]

References

1. Mohr DC, Shilton K, Hotopf M. Digital phenotyping, behavioral sensing, or personal sensing: names and transparency in the digital age. NPJ Digit Med 2020 Mar 25;3:45 [FREE Full text] [doi: 10.1038/s41746-020-0251-5] [Medline: 32219186]
2. More than half of patients who finished psychological therapy recovered in 2017-18. National Health Service Digital. 2018 Nov 20. URL: https://digital.nhs.uk/news/2018/more-than-half-of-patients-who-finished-psychological-therapy-recovered-in-2017-18 [accessed 2022-08-01]
3. Clark DM. Realizing the mass public benefit of evidence-based psychological therapies: the IAPT program. Annu Rev Clin Psychol 2018 May 07;14:159-183 [FREE Full text] [doi: 10.1146/annurev-clinpsy-050817-084833] [Medline: 29350997]
4. Cohen ZD, DeRubeis RJ. Treatment selection in depression. Annu Rev Clin Psychol 2018 May 07;14:209-236. [doi: 10.1146/annurev-clinpsy-050817-084746] [Medline: 29494258]

5. Wells JE, Horwood LJ. How accurate is recall of key symptoms of depression? A comparison of recall and longitudinal reports. Psychol Med 2004 Aug;34(6):1001-1011. [doi: 10.1017/s0033291703001843] [Medline: 15554571]
6. Taylor KI, Staunton H, Lipsmeier F, Nobbs D, Lindemann M. Outcome measures based on digital health technology sensor data: data- and patient-centric approaches. NPJ Digit Med 2020 Jul 23;3:97 [FREE Full text] [doi: 10.1038/s41746-020-0305-8] [Medline: 32715091]
7. Chevance A, Ravaud P, Tomlinson A, Le Berre C, Teufer B, Touboul S, et al. Identifying outcomes for depression that matter to patients, informal caregivers, and health-care professionals: qualitative content analysis of a large international online survey. Lancet Psychiatry 2020 Aug;7(8):692-702. [doi: 10.1016/S2215-0366(20)30191-7] [Medline: 32711710]
8. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
9. Girolamo G, Barattieri di San Pietro C, Bulgari V, Dagani J, Ferrari C, Hotopf M, et al. The acceptability of real‐time health monitoring among community participants with depression: a systematic review and meta‐analysis of the literature. Depress Anxiety 2020 Apr 27;37(9):885-897. [doi: 10.1002/da.23023]
10. Dogan E, Sander C, Wagner X, Hegerl U, Kohls E. Smartphone-based monitoring of objective and subjective data in affective disorders: where are we and where are we going? Systematic review. J Med Internet Res 2017 Jul 24;19(7):e262 [FREE Full text] [doi: 10.2196/jmir.7006] [Medline: 28739561]
11. Simblett S, Matcham F, Siddi S, Bulgari V, Barattieri di San Pietro C, Hortas López J, RADAR-CNS Consortium. Barriers to and facilitators of engagement with mHealth technology for remote measurement and management of depression: qualitative analysis. JMIR Mhealth Uhealth 2019 Jan 30;7(1):e11325 [FREE Full text] [doi: 10.2196/11325] [Medline: 30698535]
12. Kolovson S, Pratap A, Duffy J, Allred R, Munson SA, Areán PA. Understanding participant needs for engagement and attitudes towards passive sensing in remote digital health studies. Int Conf Pervasive Comput Technol Healthc 2020 May;2020:347-362 [FREE Full text] [doi: 10.1145/3421937.3422025] [Medline: 33717638]
13. Chum J, Kim MS, Zielinski L, Bhatt M, Chung D, Yeung S, et al. Acceptability of the Fitbit in behavioural activation therapy for depression: a qualitative study. Evid Based Ment Health 2017 Nov;20(4):128-133 [FREE Full text] [doi: 10.1136/eb-2017-102763] [Medline: 29056608]
14. Saunders KE, Bilderbeck AC, Panchal P, Atkinson LZ, Geddes JR, Goodwin GM. Experiences of remote mood and activity monitoring in bipolar disorder: a qualitative study. Eur Psychiatry 2017 Mar;41:115-121 [FREE Full text] [doi: 10.1016/j.eurpsy.2016.11.005] [Medline: 28135594]
15. Matcham F, Barattieri di San Pietro C, Bulgari V, de Girolamo G, Dobson R, Eriksson H, RADAR-CNS consortium. Remote assessment of disease and relapse in major depressive disorder (RADAR-MDD): a multi-centre prospective cohort study protocol. BMC Psychiatry 2019 Feb 18;19(1):72 [FREE Full text] [doi: 10.1186/s12888-019-2049-z] [Medline: 30777041]
16. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
17. Meyerhoff J, Liu T, Kording KP, Ungar LH, Kaiser SM, Karr CJ, et al. Evaluation of changes in depression, anxiety, and social anxiety using smartphone sensor features: longitudinal cohort study. J Med Internet Res 2021 Sep 03;23(9):e22844 [FREE Full text] [doi: 10.2196/22844] [Medline: 34477562]
18. De Angel V, Lewis S, White K, Oetzmann C, Leightley D, Oprea E, et al. Digital health tools for the passive monitoring of depression: a systematic review of methods. NPJ Digit Med 2022 Jan 11;5(1):3 [FREE Full text] [doi: 10.1038/s41746-021-00548-8] [Medline: 35017634]
19. Causality in digital medicine. Nat Commun 2021 Sep 15;12(1):5471 [FREE Full text] [doi: 10.1038/s41467-021-25743-9] [Medline: 34526509]
20. Ben-Zeev D, Scherer EA, Wang R, Xie H, Campbell AT. Next-generation psychiatric assessment: using smartphone sensors to monitor behavior and mental health. Psychiatr Rehabil J 2015 Sep;38(3):218-226 [FREE Full text] [doi: 10.1037/prj0000130] [Medline: 25844912]
21. Minaeva O, Booij SH, Lamers F, Antypa N, Schoevers RA, Wichers M, et al. Level and timing of physical activity during normal daily life in depressed and non-depressed individuals. Transl Psychiatry 2020 Jul 30;10(1):259 [FREE Full text] [doi: 10.1038/s41398-020-00952-w] [Medline: 32732880]
22. Wang Y, Wang J, Liu X, Zhu T. Detecting Depression Through Gait Data: Examining the Contribution of Gait Features in Recognizing Depression. Front Psychiatry 2021 Aug;12(6):661213-661211 [FREE Full text] [doi: 10.3389/fpsyt.2021.661213] [Medline: 34025483]
23. Wahle F, Kowatsch T, Fleisch E, Rufer M, Weidt S. Mobile sensing and support for people with depression: a pilot trial in the wild. JMIR Mhealth Uhealth 2016 Sep 21;4(3):e111 [FREE Full text] [doi: 10.2196/mhealth.5960] [Medline: 27655245]
24. Birk RH, Samuel G. Can digital data diagnose mental health problems? A sociological exploration of 'digital phenotyping'. Sociol Health Illn 2020 Nov;42(8):1873-1887. [doi: 10.1111/1467-9566.13175] [Medline: 32914445]
25. Tang W, Kreindler D. Supporting homework compliance in cognitive behavioural therapy: essential features of mobile apps. JMIR Ment Health 2017 Jun 08;4(2):e20 [FREE Full text] [doi: 10.2196/mental.5283] [Medline: 28596145]

26. Saunders R, Buckman JE, Cape J, Fearon P, Leibowitz J, Pilling S. Trajectories of depression and anxiety symptom change during psychological therapy. J Affect Disord 2019 Apr 15;249:327-335 [FREE Full text] [doi: 10.1016/j.jad.2019.02.043] [Medline: 30802698]
27. Henson P, Wisniewski H, Hollis C, Keshavan M, Torous J. Digital mental health apps and the therapeutic alliance: initial review. BJPsych Open 2019 Jan;5(1):e15 [FREE Full text] [doi: 10.1192/bjo.2018.86] [Medline: 30762511]
28. Ardito RB, Rabellino D. Therapeutic alliance and outcome of psychotherapy: historical excursus, measurements, and prospects for research. Front Psychol 2011 Oct 18;2:270 [FREE Full text] [doi: 10.3389/fpsyg.2011.00270] [Medline: 22028698]
29. Henson P, Peck P, Torous J. Considering the therapeutic alliance in digital mental health interventions. Harv Rev Psychiatry 2019;27(4):268-273. [doi: 10.1097/HRP.0000000000000224] [Medline: 30907764]
30. Rosman L, Gehi A, Lampert R. When smartwatches contribute to health anxiety in patients with atrial fibrillation. Cardiovasc Digit Health J 2020;1(1):9-10 [FREE Full text] [doi: 10.1016/j.cvdhj.2020.06.004] [Medline: 34386784]
31. Torous J, Rodriguez J, Powell A. The new digital divide for digital biomarkers. Digit Biomark 2017 Sep;1(1):87-91 [FREE Full text] [doi: 10.1159/000477382] [Medline: 29104959]
32. Lerner D, Adler DA, Chang H, Lapitsky L, Hood MY, Perissinotto C, et al. Unemployment, job retention, and productivity loss among employees with depression. Psychiatr Serv 2004 Dec;55(12):1371-1378 [FREE Full text] [doi: 10.1176/appi.ps.55.12.1371] [Medline: 15572564]
33. Banbury A, Roots A, Nancarrow S. Rapid review of applications of e-health and remote monitoring for rural residents. Aust J Rural Health 2014 Oct;22(5):211-222. [doi: 10.1111/ajr.12127] [Medline: 25303412]
34. Simblett S, Matcham F, Curtis H, Greer B, Polhemus A, Novák J, Remote Assessment of Disease and Relapse – Central Nervous System (RADAR-CNS) Consortium. Patients' measurement priorities for remote measurement technologies to aid chronic health conditions: qualitative analysis. JMIR Mhealth Uhealth 2020 Jun 10;8(6):e15086 [FREE Full text] [doi: 10.2196/15086] [Medline: 32519975]
35. Davis MM, Freeman M, Kaye J, Vuckovic N, Buckley DI. A systematic review of clinician and staff views on the acceptability of incorporating remote monitoring technology into primary care. Telemed J E Health 2014 May;20(5):428-438 [FREE Full text] [doi: 10.1089/tmj.2013.0166] [Medline: 24731239]
36. Bourla A, Ferreri F, Ogorzelec L, Peretti CS, 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]
37. Matcham F, Leightley D, Siddi S, Lamers F, White KM, Annas P, RADAR-CNS consortium. Remote Assessment of Disease and Relapse in Major Depressive Disorder (RADAR-MDD): recruitment, retention, and data availability in a longitudinal remote measurement study. BMC Psychiatry 2022 Feb 21;22(1):136 [FREE Full text] [doi: 10.1186/s12888-022-03753-1] [Medline: 35189842]
38. Torous J, Wisniewski H, Bird B, Carpenter E, David G, Elejalde E, et al. Creating a digital health smartphone app and digital phenotyping platform for mental health and diverse healthcare needs: an interdisciplinary and collaborative approach. J Technol Behav Sci 2019 Apr 27;4(2):73-85. [doi: 10.1007/s41347-019-00095-w]

Abbreviations

CBT: cognitive behavioral therapy
IAPT: Improving Access to Psychological Therapies
NHS: National Health Service
RMT: remote measurement technology

Edited by J Torous; submitted 22.04.22; peer-reviewed by A Chevance, R Armstrong Junior; comments to author 02.06.22; revised version received 13.06.22; accepted 13.06.22; published 15.08.22.

Please cite as:
de Angel V, Lewis S, White KM, Matcham F, Hotopf M
Clinical Targets and Attitudes Toward Implementing Digital Health Tools for Remote Measurement in Treatment for Depression: Focus Groups With Patients and Clinicians
JMIR Ment Health 2022;9(8):e38934
URL: https://mental.jmir.org/2022/8/e38934
doi:10.2196/38934
PMID:35969448

Original Paper

Operationalizing Engagement With an Interpretation Bias Smartphone App Intervention: Case Series

Ramya Ramadurai[1], BA; Erin Beckham[2], BA; R Kathryn McHugh[3], PhD; Thröstur Björgvinsson[4], PhD, ABPP; Courtney Beard[2], PhD

[1]Department of Psychology, American University, Washington, DC, United States

[2]Cognition and Affect Research and Education Lab, McLean Hospital, Belmont, MA, United States

[3]Division of Alcohol, Drugs, and Addiction, McLean Hospital, Harvard Medical School, Belmont, MA, United States

[4]Behavioral Health Partial Hospital Program, McLean Hospital, Harvard Medical School, Belmont, MA, United States

Corresponding Author:
Ramya Ramadurai, BA
Department of Psychology
American University
4400 Massachusetts Avenue NW
Washington, DC, 20016
United States
Phone: 1 202 885 8000
Email: rr4748a@student.american.edu

Abstract

Background: Engagement with mental health smartphone apps is an understudied but critical construct to understand in the pursuit of improved efficacy.

Objective: This study aimed to examine engagement as a multidimensional construct for a novel app called *HabitWorks*. *HabitWorks* delivers a personalized interpretation bias intervention and includes various strategies to enhance engagement such as human support, personalization, and self-monitoring.

Methods: We examined app use in a pilot study (n=31) and identified 5 patterns of behavioral engagement: consistently low, drop-off, adherent, high diary, and superuser.

Results: We present a series of cases (5/31, 16%) from this trial to illustrate the patterns of behavioral engagement and cognitive and affective engagement for each case. With rich participant-level data, we emphasize the diverse engagement patterns and the necessity of studying engagement as a heterogeneous and multifaceted construct.

Conclusions: Our thorough idiographic exploration of engagement with *HabitWorks* provides an example of how to operationalize engagement for other mental health apps.

(JMIR Ment Health 2022;9(8):e33545) doi:10.2196/33545

KEYWORDS

engagement; mental health apps; cognitive bias modification; human support; mobile health; mHealth; mobile phone

Introduction

Background

Over the past 2 decades, the number of available mental health smartphone apps has grown to well over 10,000 [1]. Compared with the number of apps available, research testing the efficacy of apps is extremely limited [2,3]. However, a growing body of research supports the clinical benefits of some mental health apps for a range of emotional disorders (anxiety and depression [4], anxiety [5], depression [6], schizophrenia [7], and alcohol [8]), particularly when app use is supported by some level of human coaching [9,10].

A critical challenge to realizing the potential of mental health apps is attrition; app use has been found to decline significantly after the first 2 weeks [11], and a recent review of health app use from >100,000 users found that the average period of use was just 5.5 days [12]. Mental health app users rarely complete the “full course” of the app intervention [13]. There are many possible explanations for declining app use; for example, users may have “gotten what they need” [4], or the app has “lost its

novelty" [14]. Intervention fatigue (emotional or cognitive weariness attributed to the intervention) [15], lack of accountability, and low alliance with app interventions have been highlighted as the reasons for disengagement [16]. Furthermore, the presence of technical issues [17] or general unhappiness with app features are obvious reasons for discontinued use and emphasize the need to incorporate user input into the app design process [18].

Although there is an implicit assumption of a meaningful relationship between app use and benefit, the relationship between app use and clinical outcomes is complex. Greater app use has not consistently been associated with better clinical outcomes (eg, Lin et al [19] and Bakker et al [20] found association between app use and clinical outcomes, and Graham et al [4] found no association between app use and clinical outcomes). Thus, researchers have called for more attention to engagement [1,21], suggesting that the way in which people use and relate to (ie, connect with and enjoy) the app may have important implications. This user-app relationship occurs both during and outside of actual app use and may be central to app efficacy [22]. Although systematic research on the most effective methods to enhance engagement is lacking, a recent review identified a broad range of factors that may facilitate engagement, including increases in insight, sense of control over one's mental health challenges [23], and human connections incorporated into the intervention [11,24].

Although there are many definitions of engagement, most concur on its multifaceted and dynamic nature [23,25] suggesting that it subsumes the extent of intervention use (amount, frequency, and duration), as well as subjective experience (attention, interest, and affect) with the intervention [26]. Nahum-Shani et al [27] integrated theories of engagement across disciplines (ie, education, industrial or organizational psychology, and computer science) and suggested that engagement may be best thought of as "energy investment involving physical, affective, and cognitive energies directed toward a focal stimulus or task." Recent examinations of engagement have indeed focused on a tripart model: behavioral (physical involvement with the intervention), cognitive (thinking about, attending to, and processing the intervention), and affective (emotional response to the intervention) [23,27-29]. These 3 domains are distinct; an individual can enjoy an intervention (affective) but not complete the suggested amount of use (behavioral), or they can complete intervention sessions (behavioral) and not make connections between the app and their life (cognitive). Nahum-Shani et al [27] asserted that engagement is a *state* that waxes and wanes because of a variety of internal and external factors [30] rather than a relatively stable construct [28,31].

This Study

In this study, we aimed to operationalize the model of engagement by Nahum-Shani et al [27] for a novel mental health app called *HabitWorks*. We developed *HabitWorks* to provide support during the critical transition between acute psychiatric care and outpatient treatment, a time of high risk for symptom deterioration, rehospitalization, and treatment disengagement [32]. *HabitWorks* was initially developed for patients receiving cognitive behavioral therapy (CBT) skills–based partial hospital care and was designed to augment treatment by facilitating the practice of cognitive therapy skills, to promote skill practice in the postacute period, and to ease the transition back into community treatment [33].

HabitWorks delivered a personalized interpretation bias intervention, as well as self-monitoring. This intervention was designed to promote an adaptive interpretive style, as the tendency to interpret ambiguous situations negatively (or not interpret them positively) plays an important role in the maintenance of most emotional disorders [34]. This type of intervention reliably improved interpretation bias and, in some cases, led to improved clinical symptoms [35,36]. The interpretation bias exercise was framed as a way for participants to practice catching themselves when jumping to negative conclusions, ultimately fostering healthier mental habits. The symptom-monitoring component was presented as a way of raising awareness about mood fluctuations.

In a small pilot study, *HabitWorks* was feasible and acceptable for a transdiagnostic sample of patients attending a partial hospital program [33]. Qualitative feedback revealed that participants enjoyed using the app and related the content to their daily lives. Although adherence was excellent during acute care (ie, 78.6% met the 5-session benchmark), similar to many apps [11], use throughout the month after discharge decreased over time (ie, based on the 3-session weekly adherence benchmark, weeks 1-3: approximately 33% adhered; week 4: approximately 0% adhered) [33]. Increasing app use during the month following discharge is likely to be vital to the efficacy of *HabitWorks*, as similar cognitive bias modification interventions have been found to be most effective with practice and repetition [36,37]. Consequently, we made several refinements to the app to enhance engagement during the postdischarge period.

This study aimed to (1) present an operationalization of engagement with *HabitWorks* based on the 3-facet model, (2) identify patterns of behavioral engagement with *HabitWorks* during the month after discharge, and (3) present case examples to illustrate 5 distinct patterns of behavioral engagement. The identification of engagement patterns was based solely on use because of the objectivity of the measurement, precedent regarding the way in which engagement patterns have been categorized in larger studies [11,38] and our project's a priori determination of adherence. Although we primarily focused on behavioral engagement for pattern categorization, we also explored indicators of affective and cognitive engagement. Research indicates that presenting only behavioral outcomes may be simplistic and fail to fully capture engagement as a construct [27]. Examining affective and cognitive engagement is crucial for developing a more thorough and nuanced understanding of the way people interact with apps. An idiographic approach was preferred to achieve a rich understanding of patterns of engagement [39,40] with this new app and as research on other apps has highlighted the heterogeneity in the preference of app features [26]. Exploring individual patterns of engagement with *HabitWorks* may inform further tailoring of the app to enhance its efficacy for high-risk populations, as well as inform the development of similar types of mental health apps.

Methods

Participants and Setting

This study included 31 participants who were randomly assigned to *HabitWorks* in a pilot randomized controlled trial (RCT; Table 1 provides the demographic characteristics of participants). Participants were recruited from a partial hospital program at McLean Hospital in Belmont Massachusetts, which provides intensive, CBT-based, transdiagnostic treatment. Inclusion criteria included at least moderate symptom severity at admission (>9 on the Patient Health Questionnaire-9 [41] or Generalized Anxiety Disorder-7 [42]), at least a minimal level of interpretation bias (<80% accuracy on the Word Sentence Association Paradigm [WSAP]; [43]), having an Apple iPhone (*HabitWorks* was not compatible with Android), and willingness to sign a release form to communicate with outpatient providers in case of any safety concerns.

Table 1. Full sample demographics (N=31).

Characteristics	Values
Age (years), mean (SD)	29.2 (10)
Gender, n (%)	
Nonbinary transmasculine	1 (3)
Woman	19 (61)
Man	11 (36)
Sexual identity, n (%)	
Queer	1 (3)
Bisexual	3 (10)
Gay or lesbian	2 (7)
Heterosexual	25 (81)
Ethno-racial identity, n (%)	
Do not know	1 (3)
Asian and White	3 (10)
Asian	2 (7)
Hispanic or Latinx	2 (7)
Non-Hispanic White	24 (77)

Exclusion criteria included current mania, psychosis, or severe clinical acuity, as judged by clinic staff, which would impair the understanding of consent and research procedures. Forgeard et al [44] provided a thorough overview of the partial hospital program, and Beard et al [33] provided the description of eligibility for the *HabitWorks* study. Eligible participants provided informed consent to participate in the study procedures as an augmentation to their care as usual. The 5 case examples chosen from the larger sample (N=31) have been masked such that they include no identifiable information, and all demographic data (ie, diagnosis and occupation) have been changed.

Ethics Approval

This study was approved by the Mass General Brigham Institutional Review Board (2018P000252).

HabitWorks Intervention

HabitWorks delivered a personalized, transdiagnostic interpretation bias intervention. The app was developed in consultation with content experts and clinic directors for implementation strategy. Given the importance of user involvement in the development process [1], a patient advisory board and open trial participants provided critical feedback throughout the development process, informing modifications to the app and methods to enhance engagement [33]. Table 2 provides a detailed list of *HabitWorks* features and prior evidence supporting their usefulness.

Table 2. Features of *HabitWorks* and strategies used to enhance engagement.

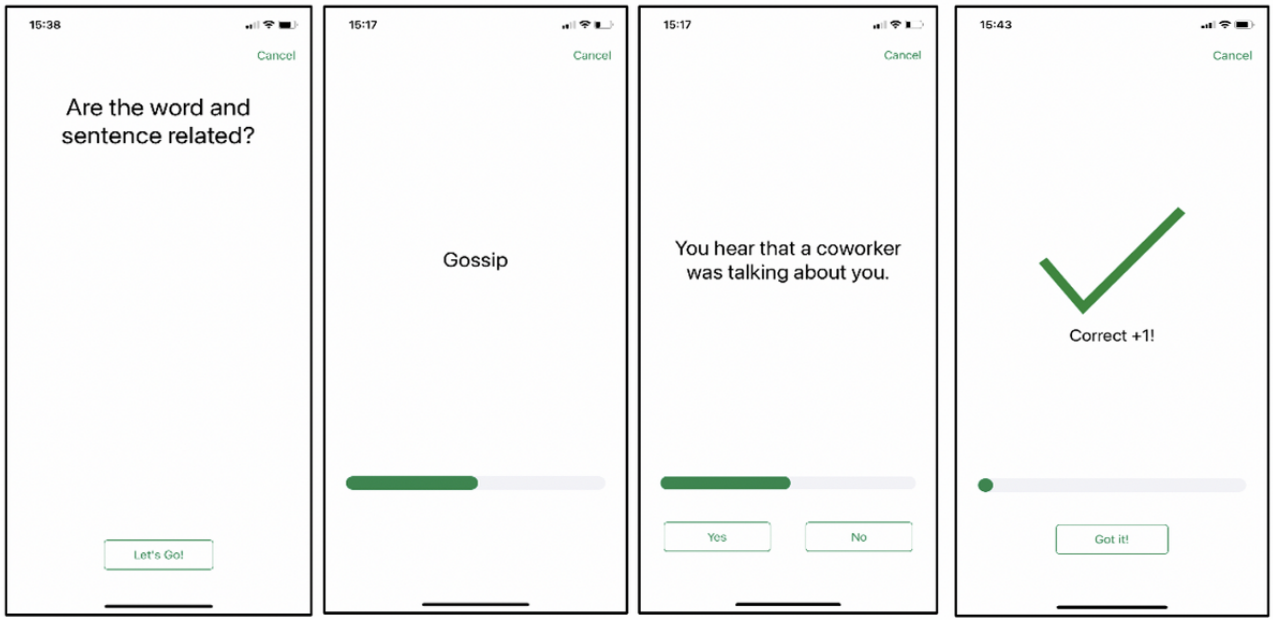
Feature or strategy	Empirical support	What does this look like in *HabitWorks*?
Human support	[11,39,45-50]	• App use was guided during acute care as support staff checked in with participants daily or less frequently if preferred. • Postdischarge support was continued through weekly email check-ins.
Customization and notifications	[51-53]	• Participants were prompted to schedule 3 exercise sessions per week in the month after discharge and were then sent push notifications at the scheduled times. • Exercise scheduling was customizable such that participants could schedule and change exercise session timing, promoting participants' sense of control and feasibility to use in the context of the participant's busy life.
Personalization	[45,54]	• Increased relevancy of HabitWorks by only offering it to those who demonstrated at least a minimal level of interpretation bias. • Participants completed personalization checklists assessing demographic characteristics and worry domains (eg, social situations, panic symptoms, and relationships). The app algorithm then selects relevant word-sentence pairs (see the study by Beard et al [33] for checklists).
Novelty	[55,56]	• HabitWorks presented variations of the interpretation bias exercise in format and length through the "level up" and bonus functions. When participants reached 90% accuracy, they progressed to the next out of 10 levels, which featured increasingly positive interpretations and introduced novel word-sentence pairings [33]. • The app presented 17 randomized encouraging GIFs, such as a celebrity giving a thumbs up, at the end of each exercise session.
Mood and tracking features	[45,50,57-60]	• Participants completed mood surveys prompted by the app weekly and self-initiated surveys as desired. • HabitWorks included progress graphs of mood check-in data, as well as exercise performance. The exercise graphs depicted changes in reaction time and interpretation accuracy over time.
Habitdiary	[61-63]	• The Habitdiary asked participants to reflect on their week and record instances in which they found themselves jumping to negative conclusions or noticed changes in their thinking or behavior. • Participants were prompted to complete entries once weekly during check-ins and could also initiate additional entries as desired.
Feedback	[54,60]	• HabitWorks provided feedback during the exercise to participants immediately following each trial based on the accuracy of their responses (ie, "Correct!" Or "Try Again!"), as well as at the end of each exercise on overall reaction time, accuracy, and percentage improvement (see the study by Beard et al [33] for a description of feedback). • HabitWorks provided PHQ-9[a] and GAD-7[b] scores.
Privacy and data security	[18,64-66]	• Users required a unique passcode to access HabitWorks. • HabitWorks enabled touch ID to access the app and ensured thorough understanding of participant rights, data collected, data storage techniques, and data uses by going over consent documentation and storing this document within the app.

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

[b]GAD-7: Generalized Anxiety Disorder-7.

The interpretation bias exercises were based on the WSAP [43,67]. At the onset of the exercises (Figure 1 provides screenshots), participants were instructed to imagine themselves in each of the upcoming situations. Next, a word was presented that represented a positive (*funny*), neutral (*toast*), or negative (*embarrassing*) interpretation of an ambiguous situation that followed (*during your speech at the wedding, you notice people in the audience laughing*). Participants clicked "yes" or "no" on their phone screen, indicating whether they believed the word and sentence were related. Next, they were presented with corrective feedback (ie, "Correct!") based on the accuracy of their responses. In this task, endorsing neutral or positive interpretations and rejecting negative interpretations were considered as accurate responses.

Figure 1. Supplemental screenshots of a *HabitWorks* exercise trial.



HabitWorks delivered several versions of the WSAP, varying by the length and order of the stimuli. Each exercise after discharge comprised 50 trials. Additional variations of the task included the following: (1) a *reverse exercise* (the sentence was followed by the word), (2) a *bonus session* (only 30 trials), and (3) a *habit test* (personalized assessment version of the task in which there was no corrective feedback).

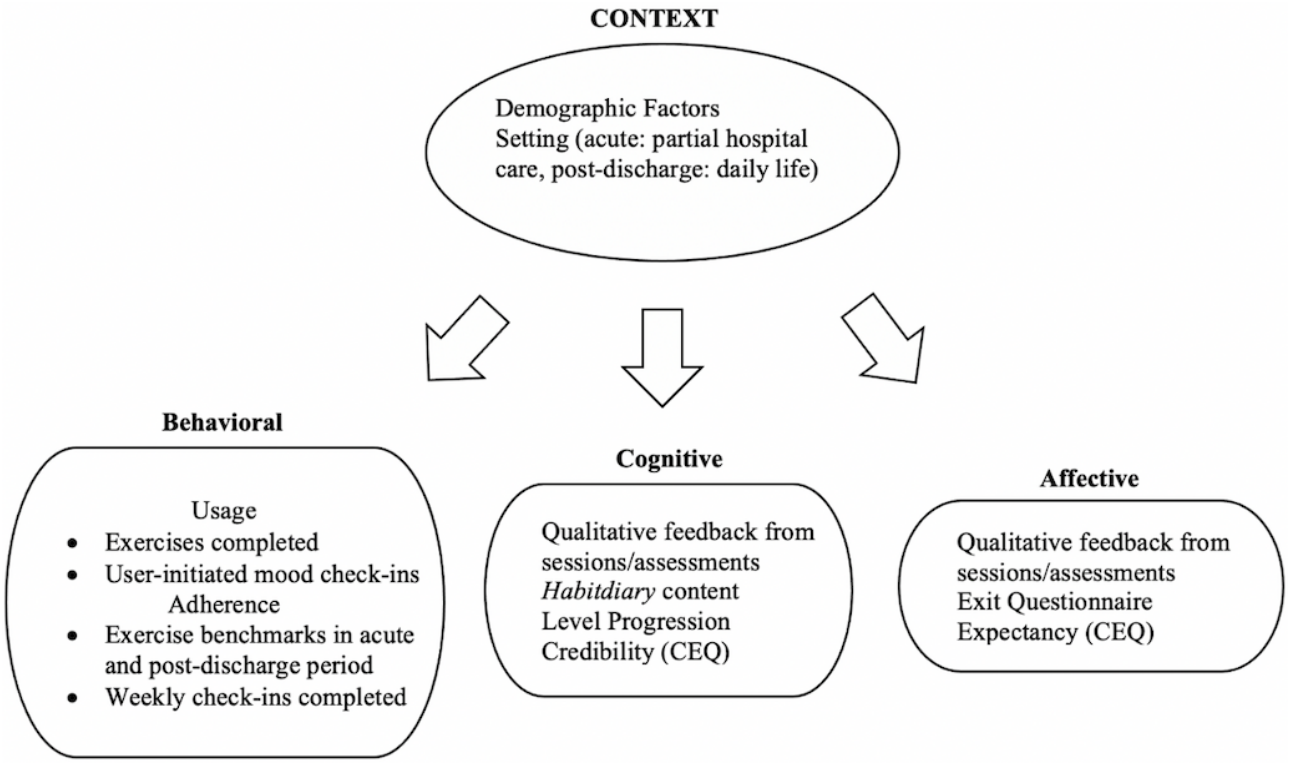
Participants were asked to use the app daily during acute care, with support from bachelor's degree–level research staff as desired. This report focuses on engagement during the month following discharge, during which participants were asked to complete exercises 3 times per week independently, as well as a weekly in-app check-in that included a mood check-in (ie, depression and anxiety scores) and the habit test. During this postdischarge period, participants continued to be supported via weekly email check-ins from the staff. Participants were asked to complete assessments after treatment (1 week after discharge) and after 1 month (1 month after discharge). Participants were compensated US $100 for completing the study assessments but were not compensated for their app use.

Measures

Overview

Measures were administered via the *HabitWorks* app, as well as on the web using REDCap (Research Electronic Data Capture; Vanderbilt University) [68]. Figure 2 [27,69] shows the indicators used for the measurement of each engagement facet. Of note, although some indicators of engagement were planned a priori (eg, number of exercises completed and affective ratings on exit questionnaire), others were selected post hoc based on available data from the RCT (eg, Habitdiary entries).

Figure 2. Operationalization of engagement in HabitWorks based on the visual model used by Nahum-Shani [27] and created by Appleton et al [69]. CEQ: Credibility and Expectancy Questionnaire.



Behavioral Engagement

Use

We calculated the number of exercises completed per week, number of Habitdiary entries completed, and number of self-initiated mood surveys.

Adherence

Adherence to the protocol was defined as the completion of the suggested 12 exercises and 4 weekly check-ins during the 1-month postacute phase of the study.

Cognitive Engagement

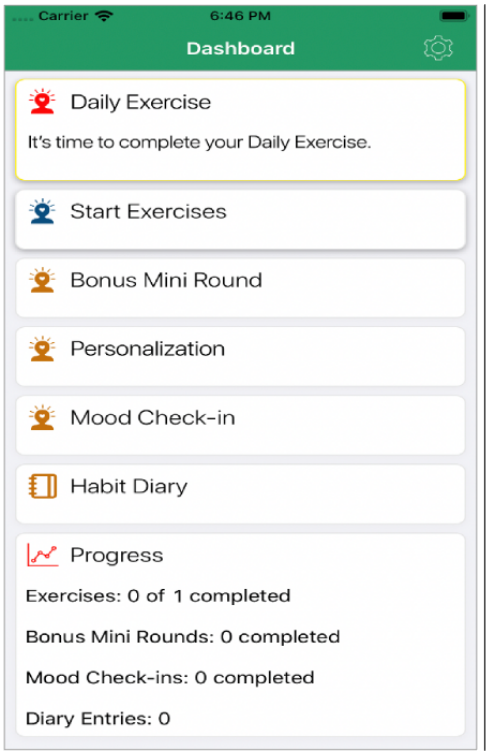
Credibility and Expectancy Questionnaire (Credibility Only)

After the first session of *HabitWorks*, the participants were asked to complete the Credibility and Expectancy Questionnaire (CEQ) [70]. The CEQ is a widely used 6-item self-report measure with items that load on 2 factors: credibility (items 1-3) and expectancy (items 4-6). A rating scale of 1 (*not at all*) to 9 (*completely*) or 0% to 100% is used for each question, depending on the question content. The credibility items from the CEQ assess how logical the participants *believe* the intervention to be. We examined the initial ratings of credibility as a measure of early-stage cognitive engagement with the intervention.

Habitdiary

Participants were asked to complete free-response diary entries weekly during the 1-month postdischarge phase and were able to initiate additional entries as desired from the dashboard of the app (Figure 3). The content of the entries was coded as an indicator of the degree to which participants applied the app content to their lives or used the feature as a free-response diary.

Figure 3. Supplemental screenshot of the *HabitWorks* dashboard.



Level Progression

The participants progressed through a series of 10 levels in the *HabitWorks* app based on exercise performance [33]. As participants progressed through the levels, they were presented with increasingly positive stimuli to endorse compared with more neutral stimuli at the beginning. As such, the achieved levels corresponded with mastery of the task and the content received (ie, more positive stimuli). To progress from one level to the next, the participants had to achieve 90% accuracy in their exercise. Importantly, an accuracy score of 70% on the assessment version of the WSAP (ie, no corrective feedback) reflects a healthy, nonanxious interpretation style [43]. We examined the final achieved level as a marker of cognitive engagement with the app.

Qualitative Feedback

Participants were asked to provide feedback on the *HabitWorks* app verbally during each assessment time point and during weekly check-ins conducted via email. In addition, qualitative interviews were conducted at the 1-month assessment by the senior author (CB). Although qualitative interviews were not initially intended to assess engagement, several prompts (ie, "Do you feel like anything's changed with you since you started the *HabitWorks* app?" and "Are you thinking about yourself or other people differently?") reflect our theoretical understanding of cognitive engagement (Multimedia Appendix 1 provides a full measure). Feedback from assessments and sessions underwent rapid coding qualitative analysis [71] by the first (RR) and second (EB) authors to identify predominant themes related to the ways in which participants connected the app to their other treatment or daily life. These data were subsequently used as indicators of cognitive engagement.

Affective Engagement

Exit Questionnaire

We administered a self-reported measure of satisfaction [35]. This exit questionnaire prompted participants to rate how helpful, relevant, user-friendly, and satisfying they found *HabitWorks* on a scale with options ranging from 1 (*completely disagree*) to 7 (*completely agree*; Multimedia Appendix 1 provides the full measure).

Qualitative Feedback

Several items (ie, "What did you think about the *HabitWorks* app?"; "What did you find beneficial?"; "What was not helpful?") included in the qualitative interview reflect our theoretical understanding of affective engagement (Multimedia Appendix 1 provides the full measure). This qualitative interview along with assessment feedback underwent rapid coding qualitative analysis (described previously). Themes and feedback that were identified as reflective of their experience (eg, enjoyment and irritation) using *HabitWorks* were used as indicators of affective engagement.

CEQ (Expectancy Only)

The expectancy items assessed how participants *feel* regarding the intervention's potential to reduce their symptoms. We explored the ratings of expectancy toward *HabitWorks* as a measure of early-stage affective engagement with the intervention.

Results

Behavioral Engagement Patterns Overview

App use data were passively collected within the app and stored on a secure REDCap server. Upon study completion, data were exported and aggregated by participants for the following

variables: type of use, date, and content related to use (eg, accuracy score for exercises, mood symptom score, and Habitdiary content). Use during the month after discharge was focused on, as many factors (ie, insurance, clinical acuity, and logistics) affected the length of stay in acute care, making comparisons of use during acute care challenging. We calculated the following summary variables for the month after discharge: number of exercises completed per week, number of weekly check-ins completed (of 4), number of Habitdiary entries completed, and number of user-initiated mood surveys completed.

After a thorough visual inspection of the data, the first (RR), second (EB), and last (CB) authors discussed and came to a consensus to identify 5 patterns of engagement in the month after discharge. The 3 authors then independently categorized participants into one of the 5 use patterns: consistently low (0-2 exercises per week; 5/31, 16%), adherent (9-15 exercises during month; 14/31, 45%), drop-off (adherent initially, then dropout; 2/31, 6%), high diary (adherent plus >2 diaries per week; 3/31, 10%), and superuser (>16 exercises during month, 7/31, 23%). We then selected the cases that represented each engagement pattern. Table 3 provides a summary of participant engagement indicator data.

Table 3. Summary of participant engagement.

Facet and indicator	Participant A (consistently low)	Participant B (adherent)	Participant C (drop-off)	Participant D (high diary)	Participant E (superuser)
Behavioral					
Exercises during 1 month after discharge (suggested 12), n (%)	4 (33)	13 (108)	13 (108)	10 (83)	60 (500)
Number of Habitdiaries	4	4	6	11	17
Weekly check-ins (suggested 4), n (%)	3 (75)	3 (75)	2 (50)	4 (100)	4 (100)
Number of user-initiated surveys 1 month after discharge	3	3	7	1	22
Cognitive					
1 (not at all) to 9 (completely)	7	6	5	9	7
Credibility: useful—1 (not at all) to 9 (completely)	6	7	3	5	5
Level completion by 1 month (out of 10 levels), n (%)	4 (40)	8 (80)	10 (100)	1 (10)	10 (100)
Habitdiary content	Relationship functioning, eating behaviors and symptoms, and interpersonal conflict	Dating, current treatment, general mental health status, and awareness of symptom improvement	Symptom improvement and current treatment, social functioning, work, and COVID-19–related worries	Free-response record (ie, monitored with timings): sleep, food, symptoms, and medication	Worries about the future, romantic relationships, family, and health
Affective					
Expected improvement (%)	80	30	10	70	30
Exit questionnaire: 1 (completely disagree) to 7 (completely agree), mean (SD)	6.6 (0.55)	6 (0.71)	N/A[a]	5.6 (0.55)	6.6 (0.55)

[a]N/A: not applicable.

Participant A: "Consistently Low"

Participant A was a college student with a primary diagnosis of bipolar disorder. Participant A maintained a low level of activity in the app throughout the month after discharge and completed the 1-month follow-up assessment.

Behavioral Engagement

During the month after discharge, participant A completed 75% (3/4) of the weekly check-ins, as well as 3 self-initiated mood check-ins. Exercise completion during the month after discharge was low (ie, 4), reflecting low and sporadic use: participant A completed 1 exercise in week 1, a total of 2 exercises in week 2, no exercise in week 3, and 1 exercise in week 4.

Cognitive Engagement

At baseline, participant A's cognitive engagement, assessed by credibility ratings (out of 9="completely"), was good (treatment logicality=7 and usefulness of treatment=6). Participant A completed 4 Habitdiary entries that covered several themes such as relational functioning and interpretations (ie, family, social, and romantic relationships), eating-related symptoms, and interpersonal conflict. Level completion was low; they reached level 4 (out of 10) by the 1-month time point. At the 1-month assessment, participant A indicated that they enjoyed the weekly mood check-ins and that these were "eye opening" with regard to their symptoms.

Affective Engagement

At baseline, affective engagement measured by expectancy was high (80%). At the 1-month follow-up, affective engagement reflected by the exit questionnaire ratings (out of 7="completely agree") was excellent (satisfaction=6, helpfulness=7, and user-friendliness=7). At the 1-month assessment, participant A indicated that they liked the notifications and the ability to schedule and reschedule exercise sessions at their convenience.

Summary

Participant A was considered "Consistently low" as they did not reach an adherent level of use on a weekly basis, or cumulatively, throughout the month following discharge. Despite low use, participant A demonstrated moderate cognitive engagement and strong affective engagement. Therefore, we speculate that other factors may have affected their behavioral engagement. Notably, participant A's month after discharge coincided with the onset of the COVID-19 pandemic. Participant A's lack of activity in week 3 seemed to coincide with an increase in suicidality, for which they received a risk evaluation from the senior author. Their qualitative data revealed other life factors that increased their stress level during their transition out of acute care (ie, moving out of their parents' home during the onset of the COVID-19 pandemic and conflict with family), which may have contributed to their low use.

Participant B: "Adherent"

Participant B had a primary diagnosis of major depression, was living alone, and was preparing to apply to college. Participant B completed all follow-up assessments.

Behavioral Engagement

During the month after discharge, participant B completed 75% (3/4) of the weekly check-ins, as well as 3 self-initiated mood check-ins. Participant B was categorized as adherent as they completed 13 of the 12 suggested exercises.

Cognitive Engagement

At baseline, cognitive engagement, assessed by credibility ratings on a scale out of 9 ("completely"), was good (treatment logicality=6 and usefulness of treatment=7). During the 1-month postdischarge period, participant B completed 4 Habitdiary entries that covered several themes such as dating, current treatment, general mental health status, and awareness of improvement of symptoms. Level completion was good; they completed level 8 by the 1-month time point. At the 1-month assessment, participant B mentioned "[HabitWorks] allowed me to have more control over negative automatic thoughts."

Affective Engagement

At baseline, affective engagement measured by expectancy was low, with the expected symptom improvement rated at 30%. In the daily sessions, participant B consistently reported finding the app easy to use. At the 1-month follow-up, affective engagement reflected by the exit questionnaire ratings (out of 7="completely agree") was excellent (satisfaction=6, helpfulness=6, and user-friendliness=7). In the qualitative interview, participant B said that they found the app easy to use and feasible to fit into the structure of the day.

Summary

Participant B was considered "Adherent" as they met the suggested exercise completion benchmarks. Despite their low expectancy early in treatment, they demonstrated strong behavioral, cognitive, and affective engagement throughout the month. They did not exhibit high initiation to use app features outside of the prompted occasions.

Participant C: "Drop-off"

Participant C was a teacher and had a primary diagnosis of major depression. Participant C adhered to the study protocol through week 3 of the postdischarge month. Drop-off during week 4 coincided with the transition from remote to in-person learning at their school, and participant C subsequently did not complete the 1-month follow-up assessment.

Behavioral Engagement

During the month after discharge, participant C completed 50% (2/4) of the weekly check-ins, as well as 7 self-initiated mood check-ins, all before a drop-off in week 4. Exercise completion during the month after discharge (ie, 13) was adherent but reflected a drop-off in use; participant C completed 5 exercises in weeks 1 and 2, a total of 3 exercises in week 3, and no exercises in week 4.

Cognitive Engagement

At baseline, participant C's cognitive engagement, assessed by credibility ratings (out of 9="completely"), was low to moderate (treatment logicality=5 and usefulness of treatment=3). Participant C commented on having trouble with the WSAP and ambiguous situations related to work. Participant C completed 6 Habitdiary entries that covered several themes such as symptom improvement and current treatment, social functioning, work, and COVID-19–related worries (ie, getting COVID-19 at work and wearing a mask). Level completion was excellent; they reached level 10 by the end of week 3.

Affective Engagement

At baseline, affective engagement measured by expectancy was low, with expected symptom improvement rated at 10%. As participant C did not complete the 1-month follow-up, the exit questionnaire ratings and qualitative interviews could not be used to indicate the level of affective engagement.

Summary

Participant C was considered "Drop-off" as they initially exceeded the suggested number of exercises and then suddenly dropped off in use and did not complete the follow-up assessment. Although participant C was active, they used all app features (ie, diary, mood surveys, and exercises) and showed good cognitive engagement. Participant C's drop-off coincided with the transition from remote to in-person school during the COVID-19 pandemic, and they had previously voiced concerns about this transition because of the fear of contracting COVID-19.

Participant D: "High Diary"

Overview

Participant D had a primary diagnosis of panic disorder. Participant D was excited to participate and "contribute to science" and was attuned to the app, frequently reporting perceived glitches or malfunctions to study staff. Participant D stated that they wanted to be completely adherent and completed all study assessments.

Behavioral Engagement

During the month after discharge, participant D completed 100% (4/4) of the weekly check-ins, as well as 1 self-initiated mood check-in. Exercise completion during the postdischarge month was generally adherent, although slightly less than suggested (ie, 10): a total of 5 exercises in week 1, a total of 2 exercises in weeks 2 and 3, and 1 exercise in week 4.

Cognitive Engagement

At baseline, cognitive engagement, assessed by credibility ratings (out of 9="completely"), was very good (treatment logicality=9 and usefulness of treatment=5). Participant D completed 11 Habitdiary entries and seemed to primarily use this feature as a tool for monitoring sleep, food, symptoms, and medication changes. Level completion was very low, remaining at level 1 by the end of the month after discharge. Despite not improving in exercise accuracy, participant D reported that it was "cool that [the app made me] notice my negative automatic thoughts" and that it was "eye-opening" in that it created greater awareness of interpretive style in daily life.

Affective Engagement

At baseline, affective engagement measured by expectancy was good, with expected symptom improvement rated at 70%. At the 1-month follow-up, affective engagement reflected by the exit questionnaire ratings (out of 7="completely agree") was good (satisfaction=6, helpfulness=5, and user-friendliness=6). In the qualitative interview, participant D reported that they liked the checklists to personalize stimuli and subsequently found all presented stimuli relatable.

Summary

Participant D was considered "High diary" as they clearly developed a preference for the Habitdiary feature. Indeed, although participant D completed 10 exercises during the postdischarge month, they seemed to use *HabitWorks* primarily for its diary function rather than connecting the WSAP exercises to their daily life. Similarly, they did not seem to benefit from the interpretation bias intervention exercises, as indicated by them never progressing beyond level 1 (indicating low interpretation accuracy).

Participant E: "Super User"

Overview

Participant E had a primary diagnosis of major depression. Participant E was extremely interested in participating mentioning past positive experiences with mental health apps and an interest in continuing to use apps to address mental health concerns. Participant E was active throughout the study and completed all the study assessments.

Behavioral Engagement

During the month after discharge, participant E completed 100% (4/4) of the weekly check-ins, as well as 22 self-initiated mood check-ins. Exercise completion during the postdischarge month was extremely high (ie, 60 total, 15 exercises per week).

Cognitive Engagement

At baseline, cognitive engagement, assessed by credibility ratings (out of 9="completely"), was moderate to good (treatment logicality=7 and usefulness of treatment=5). Participant E completed 17 Habitdiary entries, using this feature as intended to track negative automatic thoughts, as well as negative interpretations of events occurring in daily life. Themes present in the diary entries included worries about the future, romantic relationships, family, and health. Level completion was high, reaching level 10 by the end of the month after discharge. During the follow-up assessment, participant E reported that they found the situations personally relevant and noticed that handling some real-life situations was more challenging after they stopped using the app.

Affective Engagement

At baseline, affective engagement measured by expectancy was low, with the expected symptom improvement rated at 30%. At the 1-month follow-up, affective engagement reflected by the exit questionnaire ratings (out of 7="completely agree") was excellent (satisfaction=7, helpfulness=6, and user-friendliness=7). Throughout the study, participant E reported that the exercises were fun and enjoyable. In the 1-month qualitative interview, participant E reported that they enjoyed both the routineness (ie, consistent daily and weekly elements) and the "game component" of the app. They also mentioned sometimes struggling to quantify symptoms over the past 24 hours during weekly check-ins and sometimes found the app stimuli redundant.

Summary

Participant E was considered a "Super user" as they far exceeded benchmarks for exercise completion during the month after discharge. They also completed an extremely high number of Habitdiaries and user-initiated mood surveys during this period.

Discussion

Principal Findings

We examined patterns of behavioral engagement with a new mental health app designed to facilitate a healthier interpretive style as well as cognitive therapy skills practice following discharge from short-term psychiatric care. First, we operationalized engagement using a model that captures its multifaceted and dynamic nature and presented 5 cases reflecting the engagement patterns present in the sample. The data revealed heterogeneity across participants in behavioral use patterns, as well as variability within participants in their behavioral, cognitive, and affective engagement.

Behavioral Engagement

We identified 5 patterns of engagement in our sample: consistently low, adherent, drop-off, high diary, and superuser. Most of the participants (22/31, 71%) were categorized as adherent or superuser. This finding differs from the typical pattern of quick disengagement with mental health apps. Indeed, only 16% (5/31) participants were categorized as consistently low in use. This may be because of the framing of the app as an augmentation and extension of care, motivation and excitement to use the app in our sample, and the engagement enhancement strategies used in *HabitWorks*.

High behavioral engagement may have been because of the use of bachelor's degree–level staff for human support throughout the protocol [50,46]. *HabitWorks* is unique in that it shifts from a guided intervention (during acute care) to a fully automated or user-automated intervention (postdischarge period) [72]. However, even as a user-automated intervention, research staff played an important role, checking in with progress via weekly email, answering any technical or content-related questions regarding the app, and scheduling follow-up assessments. Notably, participant D mentioned the usefulness of staff in handling technical issues that arose, an issue area that often otherwise results in dropout [17].

The evidence supporting the usefulness of human support brings to the forefront the therapeutic alliance within app research, a well-documented, robust predictor of treatment outcome in traditional mental health care [73]. Human support may promote an alliance by creating step-by-step "process accountability" and enhancing agency and investment in treatment [16]. In *HabitWorks,* human support was delivered by research staff who checked in with the participants and monitored their app data (both exercise and mood scores) throughout the study. This type of support in *HabitWorks* cultivated a sense of "teamwork" among the app, staff, and participant, in essence, an alliance. As defined, the therapeutic alliance seems to subsume the aspects of affective (ie, expectancy and liking) and cognitive (ie, trust and credibility) engagement. Overall, our findings suggest that human support may have positively influenced behavioral engagement at various points throughout the study.

Cognitive Engagement

Indicators of cognitive engagement varied across the 5 cases. Although cognitive engagement assessed by initial credibility ratings ranged from average to good, level completion varied dramatically across the cases. Level progression in *HabitWorks* required the achievement of 90% accuracy in the current level. We might expect practice, or exercise completion, to be associated with level achievement. However, participant D "High diary" completed 10 exercises after discharge but still did not progress past level 1. This is surprising, and one might conclude that participant D misunderstood the exercise, was inattentive during the exercise sessions, or was not engaged cognitively with the app.

However, in addition to level completion, cognitive engagement with *HabitWorks* was elicited by the Habitdiary function, which prompted participants to journal briefly about when they noticed themselves jumping to negative conclusions in their daily lives. Participant D ("High diary") used the feature somewhat differently than the other participants (ie, as a free-response diary and self-monitoring record) and completed a high number of diary entries. Their qualitative data indicated that they were aptly applying the principles of the app to their life. Taken together, we may conclude that this participant showed a preference toward the diary feature and was in fact cognitively engaged, despite their lack of level progression. This apparent discrepancy may highlight the importance of measuring each facet of engagement with >1 indicator.

Qualitative data from all 5 cases added further nuance to our understanding of cognitive engagement, indicating that these participants found that the app helped them become aware of and assert control over their negative automatic thoughts, notice their interpretive style in their daily life, and better handle daily life situations. Participants' use of CBT language (ie, negative automatic thoughts) in their feedback may illustrate a useful integration between the app and their CBT-based partial hospital treatment.

Affective Engagement

Affective engagement, measured by expectancy for treatment to improve symptoms, was quite low for participants B, C, and E. However, at the 1-month assessment, all participants rated *HabitWorks* highly across acceptability indicators (ie, user-friendliness, satisfaction, and helpfulness). Qualitative feedback highlighted how participants easily integrated the app into their lives; how the app was relevant to their experiences; and that the app was fun, enjoyable, and game-like. Although it may be intuitive that a focus on subjective user experience is important to successful implementation [72], this focus may also be central to securing clinically meaningful benefits for users [21]. It is also notable that despite the initial low expectancy for some, all users ultimately reported enjoying the app. These findings suggest that *HabitWorks* has room for improvement in generating early "buy-in" in this population and support the conceptualization of affective engagement as a state that fluctuates over time.

Relationships Between the Facets of Engagement

Although early affective engagement (ie, expectancy of app benefits) was low for some participants and high for others, these early ratings did not correspond in the expected direction with behavioral engagement throughout the 1 month. The typical relationship between expectancy and treatment engagement is

such that lower expectancy is associated with lower engagement in treatment [74]. However, participant A had the highest expectancy and exhibited the lowest behavioral engagement, and participant E had low expectancy and exhibited the highest behavioral engagement. Moreover, all cases reported excellent affective engagement on the exit questionnaire. Although we cannot draw any conclusions from a case series, this observation underscores 2 aspects of the model of engagement by Nahum-Shani et al [27]: (1) engagement is dynamic and should be assessed in a corresponding manner and (2) the facets of engagement are related but distinct.

Participant C ("Drop-off") illustrates the connection between cognitive engagement and behavioral engagement and the difficulty of relying on just one or the other to determine meaningful use. Although participant C's use of the app suddenly dropped off after week 3 (ie, behavioral: shorter duration of use), they had already completed the prescribed number of exercises (ie, behavioral: adherent number of exercises) and had achieved the highest level possible in the app (ie, cognitive: interpretation bias accuracy). Their level completion indicates that they reached a "healthy" interpretation level (ie, 90% accuracy) at each level. Considering their behavioral and cognitive engagement together, we can surmise that they effectively used *HabitWorks*, suggesting that a drop-off in use is not necessarily problematic in all instances.

This discussion aligns with previous research suggesting that behavioral engagement alone does not necessitate better outcomes [22]. Indeed, some minimum amount of use may be necessary [75]; however, further use alone may not necessitate larger improvements. Similarly, participant A's use pattern illustrates the proposition that sustained use may not be synonymous with meaningful use, and some participants may benefit from a period of inactivity. Specifically, participant A was inactive during week 3 but became active again later in the treatment month and went on to complete the 1-month assessment. Their period of disengagement may constitute a "recovery period," a period of psychophysiological unwinding thought to be important to meaningful engagement [27], which allowed them to re-engage with the app subsequently. It is possible that this type of sporadic engagement may be a generally healthy or adaptive use style.

Limitations and Future Directions

Our study had some limitations. First, the current case series included some indicators of engagement that were chosen post hoc and were specific to the *HabitWorks* app. Thus, it is difficult to compare engagement patterns across studies. Second, although in the RCT, *HabitWorks* was compared with an active control condition, differences in both features and recommendations for use between conditions prevented comparison of engagement patterns across conditions. Third, in our focus on incorporating strategies to maximize app use, it may be important to consider the potential for app overuse. We did not examine the length of the interaction time, which may be critical to further understanding effective use [76]. Problematic smartphone use that exceeds the necessary clinically meaningful use can become disruptive in the user's life and lead to maintenance or furthering of psychosocial functioning impairment [77]. Fourth, the free-response diaries could be completed as desired by participants, and thus, the total amount of diary use varied across participants, with greater content available for those presumably more engaged in the app. Fifth, the categorization of participants was based solely on behavioral engagement. Future research with larger samples may apply quantitative analyses to identify more nuanced patterns of engagement that comprise all 3 facets, including cognitive and affective. Finally, it is possible that some engagement strategies were more helpful during early app interactions (ie, privacy and security), whereas others encouraged engagement during later interactions with the app (ie, novelty), and some others elicited engagement throughout (ie, human support). Given the conceptualization of engagement as state-like, it is likely that the helpfulness of these strategies was not linear. An important extension of this study will be to understand why users engaged with various app features [78] and which engagement strategies were the most helpful and at which time points.

Conclusions

This case series of *HabitWorks* participants illustrated 5 patterns of engagement seen in our psychiatric sample transitioning out of CBT skills–based care. In the context of an RCT with specific recommendations for use and standardized delivery, 5 distinct patterns of engagement emerged. The study of engagement may be best approached from an individual difference's perspective rather than with aggregated data. To better understand and promote "effective use" or "the extent, frequency, and duration of investment of physical, cognitive, and affective energies...to bring about a prespecified outcome" [27], a focus on multiple facets of engagement and their interactions may be important. This focus may ultimately allow for a better prediction of clinical outcomes.

Acknowledgments

This research was supported by a grant from the National Institute of Mental Health (R34MH113600) awarded to the senior author and was registered on ClinicalTrials.gov (identifier NCT03509181). *HabitWorks* was developed by the senior author. Curiosity Health, specifically JP Pollak and James Kizer, programmed the app and assisted with all the technological issues. The authors are grateful to the entire staff and patients at the Behavioral Health Partial Hospital Program who made this work possible, especially members of the Patient Advisory Board. The authors would additionally like to acknowledge and thank Dr John Torous, Arielle Solomon, Yasaman Salon, Heather Martin, and Alexandra Culhane for their contributions. The authors are grateful to the members of their advisory board: Nader Amir, Bethany Teachman, John Torous, Risa Weisberg, Jaqueline Sperling, Eve Lewandowski, Hilary Connery, Andrew Kuller, Andrew Peckham, Kirsten Christensen, and Ivar Snorrason. The authors would like to thank McLean Hospital's Institute for Technology in Psychiatry for providing helpful guidance at various stages of the

project. The authors would like to thank Lauren Wadsworth from Lauren Page Wadsworth Photography for recording the instructional videos in the *HabitWorks* app. The authors would also like to thank Jason Elias and Adam Reid from CBTeam, LLC, for providing the office space for recording instructional videos. The authors would like to thank the anonymous reviewers for their comments, which lead to substantial improvements in the presentation of the findings.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental measures.
[DOCX File , 15 KB - mental_v9i8e33545_app1.docx]

References

1. Torous JB, Chan SR, Gipson SY, Kim JW, Nguyen TQ, Luo J, et al. A hierarchical framework for evaluation and informed decision making regarding smartphone apps for clinical care. Psychiatr Serv 2018 May 01;69(5):498-500. [doi: 10.1176/appi.ps.201700423] [Medline: 29446337]
2. 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]
3. 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 2;2:118 [FREE Full text] [doi: 10.1038/s41746-019-0188-8] [Medline: 31815193]
4. Graham AK, Greene CJ, Kwasny MJ, Kaiser SM, Lieponis P, Powell T, et al. Coached mobile app platform for the treatment of depression and anxiety among primary care patients: a randomized clinical trial. JAMA Psychiatry 2020 Sep 01;77(9):906-914 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.1011] [Medline: 32432695]
5. 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]
6. 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 [FREE Full text] [doi: 10.1002/wps.20472] [Medline: 28941113]
7. Almeida RF, Sousa TJ, Couto AS, Marques AJ, Queirós CM, Martins CL. Development of weCope, a mobile app for illness self-management in schizophrenia. Arch Clin Psychiatry (São Paulo) 2019 Feb;46(1):1-4. [doi: 10.1590/0101-60830000000182]
8. Fowler LA, Holt SL, Joshi D. Mobile technology-based interventions for adult users of alcohol: a systematic review of the literature. Addict Behav 2016 Nov;62:25-34. [doi: 10.1016/j.addbeh.2016.06.008] [Medline: 27310031]
9. Mohr DC, Azocar F, Bertagnolli A, Choudhury T, Chrisp P, Frank R, Banbury Forum on Digital Mental Health. Banbury forum consensus statement on the path forward for digital mental health treatment. Psychiatr Serv 2021 Jun;72(6):677-683 [FREE Full text] [doi: 10.1176/appi.ps.202000561] [Medline: 33467872]
10. 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 [FREE Full text] [doi: 10.1002/wps.20673] [Medline: 31496095]
11. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
12. Pratap A, Neto EC, Snyder P, Stepnowsky C, Elhadad N, Grant D, et al. Indicators of retention in remote digital health studies: a cross-study evaluation of 100,000 participants. NPJ Digit Med 2020 Feb 17;3:21 [FREE Full text] [doi: 10.1038/s41746-020-0224-8] [Medline: 32128451]
13. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
14. Kwasny MJ, Schueller SM, Lattie E, Gray EL, Mohr DC. Exploring the use of multiple mental health apps within a platform: secondary analysis of the IntelliCare field trial. JMIR Ment Health 2019 Mar 21;6(3):e11572 [FREE Full text] [doi: 10.2196/11572] [Medline: 30896433]
15. Heckman BW, Mathew AR, Carpenter MJ. Treatment burden and treatment fatigue as barriers to health. Curr Opin Psychol 2015 Oct 01;5:31-36 [FREE Full text] [doi: 10.1016/j.copsyc.2015.03.004] [Medline: 26086031]

16. 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]
17. Bohleber L, Crameri A, Eich-Stierli B, Telesko R, von Wyl A. Can we foster a culture of peer support and promote mental health in adolescence using a Web-based app? A control group study. JMIR Ment Health 2016 Sep 23;3(3):e45 [FREE Full text] [doi: 10.2196/mental.5597] [Medline: 27663691]
18. Torous J, Andersson G, Bertagnoli A, Christensen H, Cuijpers P, Firth J, et al. Towards a consensus around standards for smartphone apps and digital mental health. World Psychiatry 2019 Feb;18(1):97-98 [FREE Full text] [doi: 10.1002/wps.20592] [Medline: 30600619]
19. Lin J, Faust B, Ebert DD, Krämer L, Baumeister H. A Web-based acceptance-facilitating intervention for identifying patients' acceptance, uptake, and adherence of internet- and mobile-based pain interventions: randomized controlled trial. J Med Internet Res 2018 Aug 21;20(8):e244 [FREE Full text] [doi: 10.2196/jmir.9925] [Medline: 30131313]
20. Bakker D, Rickard N. Engagement with a cognitive behavioural therapy mobile phone app predicts changes in mental health and wellbeing: MoodMission. Australian Psychologist 2020 Nov 12;54(4):245-260. [doi: 10.1111/ap.12383]
21. Graham AK, Kwasny MJ, Lattie EG, Greene CJ, Gupta NV, Reddy M, et al. Targeting subjective engagement in experimental therapeutics for digital mental health interventions. Internet Interv 2021 May 19;25:100403 [FREE Full text] [doi: 10.1016/j.invent.2021.100403] [Medline: 34401363]
22. Pham Q, Graham G, Carrion C, Morita PP, Seto E, Stinson JN, et al. A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: scoping review. JMIR Mhealth Uhealth 2019 Jan 18;7(1):e11941 [FREE Full text] [doi: 10.2196/11941] [Medline: 30664463]
23. Kelders SM, van Zyl LE, Ludden GD. The concept and components of engagement in different domains applied to eHealth: a systematic scoping review. Front Psychol 2020 May 27;11:926 [FREE Full text] [doi: 10.3389/fpsyg.2020.00926] [Medline: 32536888]
24. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. J Med Internet Res 2021 Mar 24;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
25. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
26. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
27. Nahum-Shani I, Shaw SD, Carpenter SM, Murphy SA, Yoon C. Engagement in digital interventions. Am Psychol (forthcoming) 2022 Mar 17. [doi: 10.1037/amp0000983] [Medline: 35298199]
28. McLean Institute for Technology in Psychiatry. 2020 Jan 23. URL: https://www.youtube.com/watch?v=ziROuXluz2k&list=PLWolxJyENg7ies08SDoj1ksbggo-tIxXa&index=9 [accessed 2020-08-20]
29. King G, Currie M, Petersen P. Child and parent engagement in the mental health intervention process: a motivational framework. Child Adolesc Ment Health 2014 Feb;19(1):2-8. [doi: 10.1111/camh.12015] [Medline: 32878365]
30. Kahn WA. Psychological conditions of personal engagement and disengagement at work. Acad Manag J 1990 Dec;33(4):692-724. [doi: 10.5465/256287]
31. 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]
32. Durbin J, Lin E, Layne C, Teed M. Is readmission a valid indicator of the quality of inpatient psychiatric care? J Behav Health Serv Res 2007 Apr;34(2):137-150. [doi: 10.1007/s11414-007-9055-5] [Medline: 17437186]
33. Beard C, Ramadurai R, McHugh RK, Pollak JP, Björgvinsson T. HabitWorks: development of a CBM-I smartphone app to augment and extend acute treatment. Behav Ther 2021 Mar;52(2):365-378. [doi: 10.1016/j.beth.2020.04.013] [Medline: 33622506]
34. Hirsch CR, Meeten F, Krahé C, Reeder C. Resolving ambiguity in emotional disorders: the nature and role of interpretation biases. Annu Rev Clin Psychol 2016;12:281-305. [doi: 10.1146/annurev-clinpsy-021815-093436] [Medline: 27019398]
35. Beard C, Rifkin LS, Silverman AL, Björgvinsson T. Translating CBM-I into real-world settings: augmenting a CBT-based psychiatric hospital program. Behav Ther 2019 May;50(3):515-530. [doi: 10.1016/j.beth.2018.09.002] [Medline: 31030870]
36. Amir N, Taylor CT. Interpretation training in individuals with generalized social anxiety disorder: a randomized controlled trial. J Consult Clin Psychol 2012 Jun;80(3):497-511 [FREE Full text] [doi: 10.1037/a0026928] [Medline: 22250851]
37. Hirsch CR, Krahé C, Whyte J, Loizou S, Bridge L, Norton S, et al. Interpretation training to target repetitive negative thinking in generalized anxiety disorder and depression. J Consult Clin Psychol 2018 Dec;86(12):1017-1030. [doi: 10.1037/ccp0000310] [Medline: 30507227]

38. Chien I, Enrique A, Palacios J, Regan T, Keegan D, Carter D, et al. A machine learning approach to understanding patterns of engagement with Internet-delivered mental health interventions. JAMA Netw Open 2020 Jul 01;3(7):e2010791 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.10791] [Medline: 32678450]
39. Orr LC, Graham AK, Mohr DC, Greene CJ. Engagement and clinical improvement among older adult primary care patients using a mobile intervention for depression and anxiety: case studies. JMIR Ment Health 2020 Jul 08;7(7):e16341 [FREE Full text] [doi: 10.2196/16341] [Medline: 32673236]
40. Rauseo-Ricupero N, Henson P, Agate-Mays M, Torous J. Case studies from the digital clinic: integrating digital phenotyping and clinical practice into today's world. Int Rev Psychiatry 2021 Jun;33(4):394-403. [doi: 10.1080/09540261.2020.1859465] [Medline: 33792463]
41. 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]
42. Spitzer RL, Kroenke K, Williams JB, 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]
43. Beard C, Amir N. Interpretation in social anxiety: when meaning precedes ambiguity. Cognit Ther Res 2009;33(4):406-415 [FREE Full text] [doi: 10.1007/s10608-009-9235-0] [Medline: 20046862]
44. Forgeard M, Beard C, Kirakosian N, Björgvinsson T. Research in partial hospital settings. In: Codd RT, editor. Practice-Based Research: A Guide for Clinicians. New York, NY, USA: Routledge; 2019:221-239.
45. Schubart JR, Stuckey HL, Ganeshamoorthy A, Sciamanna CN. Chronic health conditions and Internet behavioral interventions: a review of factors to enhance user engagement. Comput Inform Nurs 2011 Feb;29(2):81-92 [FREE Full text] [doi: 10.1097/NCN.0b013e3182065eed] [Medline: 21164337]
46. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. J Affect Disord 2020 Feb 15;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
47. 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]
48. Newman MG, Szkodny LE, Llera SJ, Przeworski A. A review of technology-assisted self-help and minimal contact therapies for anxiety and depression: is human contact necessary for therapeutic efficacy? Clin Psychol Rev 2011 Feb;31(1):89-103. [doi: 10.1016/j.cpr.2010.09.008] [Medline: 21130939]
49. Firth J, Torous J, Carney R, Newby J, Cosco TD, Christensen H, et al. Digital technologies in the treatment of anxiety: recent innovations and future directions. Curr Psychiatry Rep 2018 May 19;20(6):44 [FREE Full text] [doi: 10.1007/s11920-018-0910-2] [Medline: 29779065]
50. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. J Med Internet Res 2020 May 29;22(5):e17572 [FREE Full text] [doi: 10.2196/17572] [Medline: 32348255]
51. Birk MV, Mandryk RL. Improving the efficacy of cognitive training for digital mental health interventions through avatar customization: crowdsourced quasi-experimental study. J Med Internet Res 2019 Jan 08;21(1):e10133 [FREE Full text] [doi: 10.2196/10133] [Medline: 30622095]
52. Sundar SS, Marathe SS. Personalization versus customization: the importance of agency, privacy, and power usage. Hum Commun Res 2010 Jul;36(3):298-322. [doi: 10.1111/j.1468-2958.2010.01377.x]
53. Fry JP, Neff RA. Periodic prompts and reminders in health promotion and health behavior interventions: systematic review. J Med Internet Res 2009 May 14;11(2):e16 [FREE Full text] [doi: 10.2196/jmir.1138] [Medline: 19632970]
54. Strecher VJ, McClure J, Alexander G, Chakraborty B, Nair V, Konkel J, et al. The role of engagement in a tailored Web-based smoking cessation program: randomized controlled trial. J Med Internet Res 2008 Nov 04;10(5):e36 [FREE Full text] [doi: 10.2196/jmir.1002] [Medline: 18984557]
55. Czyz EK, Horwitz AG, Arango A, King CA. Short-term change and prediction of suicidal ideation among adolescents: a daily diary study following psychiatric hospitalization. J Child Psychol Psychiatry 2019 Jul;60(7):732-741 [FREE Full text] [doi: 10.1111/jcpp.12974] [Medline: 30246870]
56. Rabbi M, Philyaw Kotov M, Cunningham R, Bonar EE, Nahum-Shani I, Klasnja P, et al. Toward increasing engagement in substance use data collection: development of the substance abuse research assistant app and protocol for a microrandomized trial using adolescents and emerging adults. JMIR Res Protoc 2018 Jul 18;7(7):e166 [FREE Full text] [doi: 10.2196/resprot.9850] [Medline: 30021714]
57. Dubad M, Winsper C, Meyer C, Livanou M, Marwaha S. A systematic review of the psychometric properties, usability and clinical impacts of mobile mood-monitoring applications in young people. Psychol Med 2018 Jan;48(2):208-228. [doi: 10.1017/S0033291717001659] [Medline: 28641609]
58. 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]
59. Kauer SD, Reid SC, Crooke AH, Khor A, Hearps SJ, Jorm AF, et al. Self-monitoring using mobile phones in the early stages of adolescent depression: randomized controlled trial. J Med Internet Res 2012 Jun 25;14(3):e67 [FREE Full text] [doi: 10.2196/jmir.1858] [Medline: 22732135]

60. Purkayastha S, Addepally SA, Bucher S. Engagement and usability of a cognitive behavioral therapy mobile app compared with Web-based cognitive behavioral therapy among college students: randomized heuristic trial. JMIR Hum Factors 2020 Feb 03;7(1):e14146 [FREE Full text] [doi: 10.2196/14146] [Medline: 32012043]
61. Podina IR, Fodor LA, Cosmoiu A, Boian R. An evidence-based gamified mHealth intervention for overweight young adults with maladaptive eating habits: study protocol for a randomized controlled trial. Trials 2017 Dec 12;18(1):592 [FREE Full text] [doi: 10.1186/s13063-017-2340-6] [Medline: 29233169]
62. Ambeba EJ, Ye L, Sereika SM, Styn MA, Acharya SD, Sevick MA, et al. The use of mHealth to deliver tailored messages reduces reported energy and fat intake. J Cardiovasc Nurs 2015;30(1):35-43 [FREE Full text] [doi: 10.1097/JCN.0000000000000120] [Medline: 24434827]
63. Kristjánsdóttir OB, Fors EA, Eide E, Finset A, Stensrud TL, van Dulmen S, et al. A smartphone-based intervention with diaries and therapist-feedback to reduce catastrophizing and increase functioning in women with chronic widespread pain: randomized controlled trial. J Med Internet Res 2013 Jan 07;15(1):e5 [FREE Full text] [doi: 10.2196/jmir.2249] [Medline: 23291270]
64. Proudfoot J, Parker G, Hadzi Pavlovic D, Manicavasagar V, Adler E, Whitton A. Community attitudes to the appropriation of mobile phones for monitoring and managing depression, anxiety, and stress. J Med Internet Res 2010 Dec 19;12(5):e64 [FREE Full text] [doi: 10.2196/jmir.1475] [Medline: 21169174]
65. Sunyaev A, Dehling T, Taylor PL, Mandl KD. Availability and quality of mobile health app privacy policies. J Am Med Inform Assoc 2015 Apr;22(e1):e28-e33. [doi: 10.1136/amiajnl-2013-002605] [Medline: 25147247]
66. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
67. Beard C, Amir N. A multi-session interpretation modification program: changes in interpretation and social anxiety symptoms. Behav Res Ther 2008 Oct;46(10):1135-1141 [FREE Full text] [doi: 10.1016/j.brat.2008.05.012] [Medline: 18675400]
68. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
69. Appleton JJ, Christenson SL, Kim D, Reschly AL. Measuring cognitive and psychological engagement: validation of the Student Engagement Instrument. J School Psychol 2006 Oct;44(5):427-445. [doi: 10.1016/j.jsp.2006.04.002]
70. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. J Behav Ther Exp Psychiatry 2000 Jun;31(2):73-86. [doi: 10.1016/s0005-7916(00)00012-4] [Medline: 11132119]
71. Nevedal AL, Reardon CM, Opra Widerquist MA, Jackson GL, Cutrona SL, White BS, et al. Rapid versus traditional qualitative analysis using the Consolidated Framework for Implementation Research (CFIR). Implement Sci 2021 Jul 02;16(1):67 [FREE Full text] [doi: 10.1186/s13012-021-01111-5] [Medline: 34215286]
72. Hermes ED, Lyon AR, Schueller SM, Glass JE. Measuring the implementation of behavioral intervention technologies: recharacterization of established outcomes. J Med Internet Res 2019 Jan 25;21(1):e11752 [FREE Full text] [doi: 10.2196/11752] [Medline: 30681966]
73. Castonguay LG, Constantino MJ, Holtforth MG. The working alliance: where are we and where should we go? Psychotherapy (Chic) 2006;43(3):271-279. [doi: 10.1037/0033-3204.43.3.271] [Medline: 22122096]
74. Constantino MJ, Arnkoff DB, Glass CR, Ametrano RM, Smith JZ. Expectations. J Clin Psychol 2011 Feb;67(2):184-192. [doi: 10.1002/jclp.20754] [Medline: 21128304]
75. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
76. Alshurafa N, Jain J, Alharbi R, Iakovlev G, Spring B, Pfammatter A. Is more always better?: discovering incentivized mHealth intervention engagement related to health behavior trends. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Dec;2(4):153 [FREE Full text] [doi: 10.1145/3287031] [Medline: 32318650]
77. Elhai JD, Rozgonjuk D, Yildirim C, Alghraibeh AM, Alafnan AA. Worry and anger are associated with latent classes of problematic smartphone use severity among college students. J Affect Disord 2019 Mar 01;246:209-216. [doi: 10.1016/j.jad.2018.12.047] [Medline: 30583147]
78. Kelders SM, Kip H, Greeff J. Psychometric evaluation of the TWente Engagement with Ehealth Technologies Scale (TWEETS): evaluation study. J Med Internet Res 2020 Oct 09;22(10):e17757 [FREE Full text] [doi: 10.2196/17757] [Medline: 33021487]

Abbreviations

CBT: cognitive behavioral therapy
CEQ: Credibility and Expectancy Questionnaire
RCT: randomized controlled trial

REDCap: Research Electronic Data Capture
WSAP: Word Sentence Association Paradigm

Edited by J Torous; submitted 12.09.21; peer-reviewed by S Schueller, R Marshall; comments to author 28.10.21; revised version received 28.02.22; accepted 20.06.22; published 17.08.22.

Please cite as:
Ramadurai R, Beckham E, McHugh RK, Björgvinsson T, Beard C
Operationalizing Engagement With an Interpretation Bias Smartphone App Intervention: Case Series
JMIR Ment Health 2022;9(8):e33545
URL: https://mental.jmir.org/2022/8/e33545
doi:10.2196/33545
PMID:35976196

Review

Appropriate Use and Operationalization of Adherence to Digital Cognitive Behavioral Therapy for Depression and Anxiety in Youth: Systematic Review

Sophie H Li[1], BPsych (Hons), MClinPsych, PhD; Melinda R Achilles[2], BBPsych; Aliza Werner-Seidler[1], BPsych (Hons), MClinPsych, PhD; Joanne R Beames[2], BPsych (Hons), MClinPsych, PhD; Mirjana Subotic-Kerry[2], BSc (Hons), Grad Dip (Psychology), PhD; Bridianne O'Dea[2], BHSc (Hons), PhD

[1]Black Dog Institute and School of Psychology, The University of New South Wales, Randwick, Australia

[2]Black Dog Institute, The University of New South Wales, Randwick, Australia

Corresponding Author:
Sophie H Li, BPsych (Hons), MClinPsych, PhD
Black Dog Institute and School of Psychology
The University of New South Wales
Hospital Road
Randwick, 2031
Australia
Phone: 61 411116615
Email: s.h.li@blackdog.org.au

Abstract

Background: Digital, self-guided cognitive behavioral therapy (CBT) interventions circumvent many barriers to in-person therapy for young people (aged 12-24 years), although adherence to these interventions is low. The absence or insufficient disclosure of recommendations or instructions for appropriate use may account for this. As such, many young people may not self-administer these interventions appropriately or receive the optimal degree of treatment.

Objective: This systematic review aims to synthesize the literature on digital CBT for depression and anxiety in young people to describe how appropriate use has been defined and communicated to users as instructions for use, to describe how adherence has been measured, and to determine the associations between adherence and treatment outcomes.

Methods: A systematic review was conducted with 2 reviewers (SHL and MRA) extracting data independently. Overall, 4 electronic databases (Embase, MEDLINE, PsycINFO, and Cochrane Library) were searched in April 2021 for studies that met the following inclusion criteria: participants aged between 12 and 24 years, evaluated a digital CBT intervention targeting depression or anxiety, and reported instructions or recommendations for use or measures of adherence. Studies that evaluated non-CBT interventions or cognitive- or behavioral-only interventions were excluded. Methodological quality was assessed using the Cochrane Risk of Bias Tool and the Integrated Quality Criteria for the Review of Multiple Study Designs.

Results: There were 32 manuscripts that met the inclusion criteria, of which 28 (88%) were unique studies (N=16,578 youths). Definitions of appropriate use varied among the different interventions in terms of intended recipients, duration and frequency of use, and the features used to support engagement and adherence to appropriate use definitions. Reporting of appropriate use definitions in studies was inconsistent, with no study systematically describing components of appropriate use or providing information on how recommendations for use were relayed to users. Most often, definitions of appropriate use were derived from the study protocol and descriptions of intervention features. Adherence was mostly operationalized as the degree of intervention completion; however, reporting of adherence data was heterogeneous. There was little evidence of an association between degree of use and outcomes in the 9 studies that examined this.

Conclusions: Definitions of appropriate use are unique to each digital CBT intervention. However, statements of appropriate use are not systematically reported in the literature. Furthermore, the extent to which recommendations for use are communicated to users is not routinely reported. Despite unique definitions of appropriate use, adherence was most often generically operationalized as the degree of intervention completion and was not consistently associated with outcomes. We proposed a framework to promote systematic reporting of definitions of appropriate use for digital interventions to provide guidance to users and to assist the development of appropriate and nuanced measures of adherence.

Trial Registration: PROSPERO CRD42020208668; https://tinyurl.com/4bu2yram

(JMIR Ment Health 2022;9(8):e37640) doi:10.2196/37640

KEYWORDS

adherence; youth; digital; cognitive behavioral therapy; review; mobile phone

Introduction

Background

Cognitive behavioral therapy (CBT) is a structured, skills-based psychotherapy, typically delivered in-person by a trained clinician over a duration of 12 to 18 weeks [1,2]. It is the gold standard psychological treatment for anxiety and depression [3] in both adults and young people, defined here as people aged between 12 and 24 years, consistent with the Australian Institute of Health and Welfare [4-7]. To safeguard the integrity of CBT and ensure that efficacy is maintained [8], CBT manuals provide clinicians with explicit instructions on administration. These manuals detail indications of who will benefit, the treatment targets and goals, the number and duration of sessions required, sequence of content delivery, and between-session practice and real-world enactment of skills (ie, homework [9]). Clinicians also adapt various engagement strategies to promote adherence to CBT, including developing a therapeutic alliance through collaborative goal setting, communicating and enhancing treatment expectancies, and clarifying concepts to ensure that treatment processes and rationale are well understood [1,2,5-7,10,11]. Some CBT manuals also differentiate the critical, core treatment components from optional modules and strategies to measure symptoms and treatment outcomes [1]. To summarize, the components of CBT manuals that guide clinicians' administration as intended are as follows: recipients, target condition, number and duration of sessions, sequence of content, homework activities, engagement-promoting strategies, core therapeutic components, assessment and monitoring, and crisis management.

Despite its efficacy, the uptake of in-person CBT is suboptimal across all ages [12-15]. In young people, this is due in part to the affordability and availability of trained practitioners, perceived stigma, poor mental health literacy, and a preference for self-reliance [16-18]. Among the young people who do seek help, most do not receive CBT [19,20]. Improving access and uptake of CBT is particularly critical for young people, as three-quarter of depression and anxiety cases emerge by late adolescence [21], and these disorders are a leading cause of disability in this age group [22,23]. The transition of CBT from in-person professional administration to digital, self-directed delivery was hypothesized to overcome many of the treatment barriers faced by young people [24]. A recent systematic review and meta-analysis of internet- and computer-delivered CBT for youth confirmed the positive effects of these interventions on symptom reduction [25]. However, adherence to many of these interventions is low despite the alignment of digital CBT with young people's help-seeking preferences [24].

Although personal factors such as motivation may account for low adherence, reliance on the young person to appropriately self-administer digital CBT interventions with none to limited clinician guidance or supervision may also contribute. Thus, similar to any self-administered treatment, it is essential that the young person is provided with clear instructions to ensure appropriate use; that is, to ensure that the intervention is used in a way that generates optimal clinical benefit. Unlike in-person CBT, where instructions for engaging with the treatment are well documented within manuals and directly relayed to youth through their clinician, it is unclear whether instructions or recommendations for the appropriate use of digital CBT are provided to users or consistently documented in the literature. Instructions for use may be provided during intervention onboarding or guided via explicit or implicit software design features. For example, many digital CBT interventions have used the design feature, "tunnelling," to replicate the structure of in-person CBT, whereby therapeutic modules are presented sequentially to guide users through the content in an appropriate sequence [26,27]. Other software design features, such as tailoring content in response to user input, automated feedback, rewards and encouragement, and reminders and notifications, have also been used to replicate clinician guidance and supervision [28]. An understanding of how digital CBT interventions define appropriate use and how instructions for use are relayed to youth requires examination to determine their adequacy in supporting appropriate self-administration.

In addition to ensuring optimal outcomes, clear instructions for the use of digital CBT interventions are required for researchers and clinicians to operationalize users' adherence to these interventions. Adherence is defined as a meaningful measure of the extent to which individuals' intervention use corresponds with creators' recommendations, instructions, and expectations of appropriate use [29,30]. For example, if the recommendation is to complete all modules within a digital intervention, adherence is measured by determining the number of modules completed. Alternatively, if the recommendation is to engage with a particular intervention activity on a specified occasion (eg, complete a mindfulness mediation upon waking), adherence is measured by determining the frequency of this event. In this way, adherence provides a measure of the validity of treatment administration. Thus, clear instructions for appropriate use and corresponding adherence measures are crucial for ensuring that the necessary standards of quality, safety, and efficacy are met. Despite the importance of this, in a review of the literature, Christ et al [25] found that many studies on digital CBT for youth generically operationalized adherence as program completion, heterogeneously reported as the proportion of participants who completed all treatment modules or the average number of modules completed across the sample. Several studies failed to report any adherence data, and definitions of appropriate use were not examined in the study by Christ [25].

Similar variability in the reporting of adherence has been found in reviews of adult digital health interventions. One systematic

review found that almost three-quarter of studies (45/62, 73%) operationalized adherence as the degree of intervention use, reported as the number of completed modules and activities, log-ins, or time spent in the intervention [31]. Furthermore, statements on instructions for use were only reported in one-third (23/62, 37%) of studies, giving little indication of how interventions defined appropriate use [31]. Inadequate reporting of instructions for use and significant heterogeneity in reporting of adherence are further demonstrated in at least 5 systematic and narrative reviews of adherence to digital psychological interventions in adults [32-36]. Collectively, adherence appears to be operationalized in the absence of adequately reported definitions of appropriate use or is generically operationalized as adherence to total intervention exposure, based on the notion of "the more, the better" rather than explicitly measuring the accordance between recommended use and actual use. Studies have provided little justification for this approach [31]. Furthermore, studies of adults and youth have not consistently supported a linear relationship between adherence and outcomes [25,29,37-39].

Despite the prevalence of digital CBT interventions for youth [25], little is known about how young people have been instructed to use these interventions, how adherence to interventions has been operationalized, or whether measures of adherence adequately determine differences between actual and recommended use. It is important to examine this, as young people have different patterns of engagement with technology, preferences, and expectations compared with adults [40-42]. Clearer expectations of use may not only improve the effectiveness of digital CBT but also assist young people in selecting interventions most suited to their circumstances, while also facilitating greater endorsement and dissemination by mental health professionals. Establishing protocols for the disclosure of appropriate use and the operationalization of adherence is imperative for improving the uptake, adherence, and effectiveness of digital CBT among youth.

Objectives

The primary aim of this systematic review was to synthesize the published literature on digital CBT interventions for depression and anxiety in young people to investigate the definitions of appropriate use and the disclosure of instructions for use to users. To this end, descriptions, recommendations or instructions of appropriate use, or intervention features guiding use were extracted from the included studies and mapped onto the 10 components within CBT manuals that typically guide administration as a means of determining definitions of appropriate use. A specific focus on digital CBT differentiated our review from recently published reviews on engagement in digital health interventions that incorporated CBT and non-CBT interventions. This review also aimed to examine how adherence to recommended use has been operationalized and measured and to determine the associations between digital CBT intervention use and outcomes among young people. This information will improve our understanding of how young people have been instructed to self-administer digital CBT interventions and will determine whether the measures of adherence used by researchers have accurately captured the degree to which young people complied with instructions for use.

Methods

Protocol and Registration

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) was used to ensure the quality and consistency of the procedure and reporting [43]. The review protocol was registered with PROSPERO (CRD42020208668). We deviated from the original protocol to exclude cognitive- and behavioral-only interventions and mindfulness-based and acceptance and commitment therapy interventions to confine the focus to digital CBT. In addition, we found that the study manuscripts rarely reported on the instructions of appropriate use. We acknowledge that this information may have been provided to users within the intervention itself, but this is not clear. As such, we derived the information on appropriate use from the study protocols and intervention features.

Ethical Considerations

Conducting a systematic review or peer-reviewed literature is not listed as a research activity that requires human ethics approval by the UNSW Research Ethics Board. As such, ethics approval was not applied for.

Eligibility Criteria

Participants

Participants were young people aged between 12 and 24 years. This age range was selected to cover the full spectrum of youth, as defined by the Australian Institute of Health and Welfare [4]. Where the age range of participants extended 12 to 24 years, studies were excluded if the total sample contained <80% of participants in the target range. In the absence of age range data, a judgment was made based on the recruitment setting (eg, secondary school) and the reported mean age. Diagnostic status was not used as an eligibility criterion for samples.

Interventions

Eligible interventions included those directly targeting anxiety or depression (including transdiagnostic interventions) via a predominantly CBT-based psychological treatment delivered by a computer, smartphone, or internet platform. Interventions were also required to be used on more than one occasion. Purely cognitive or behavioral interventions alone were not included, nor were treatments aimed solely at problem-solving. Interventions that were not predominantly CBT, such as mindfulness-based interventions, acceptance and commitment therapy, and interpersonal psychotherapy, were excluded. Nonpsychological interventions, including exercise or physical activities, music, and art therapy, were excluded. Gratitude-based therapies and journaling were also excluded.

Comparison Groups

No restriction was imposed on the type or use of control or comparison groups.

Outcomes

Primary outcomes were symptoms of depression or anxiety measured using standardized, validated, and reliable instruments or scales, suitable for adolescents. Included studies were required to report at least one measure of adherence and any of the following: descriptions, statements, or instructions for appropriate use or recommendations of use; descriptions of intervention features supporting appropriate use, including adherence-promoting features; or the association between adherence and depression outcomes and adherence and anxiety outcomes.

Studies

Studies were included if they were written in English, published in peer-reviewed journals, and published after January 1, 1991. Studies were not excluded based on study type or quality; both controlled and uncontrolled studies (eg, pre-post studies without a control group) were included. Case studies were excluded.

Search Strategy

Overview

The electronic databases, Embase, MEDLINE, PsycINFO, and Cochrane Library, were searched for articles published from January 1, 1991, to April 7, 2021. The start date was selected to coincide with the year in which the World Wide Web was introduced. The following search terms were used in the title, abstract, and keywords: *(adolescent** OR *youth* OR *child** OR *teen** OR *young adult)* AND *(online* OR *digital* OR *internet* OR *app* OR *mHealth* OR *eHealth* OR *web* OR *web-based* OR *smartphone* OR *smart phone* OR *computer*)* AND *(anx** OR *depress** OR *affect* OR *mood)* AND *(cognitive therapy* OR *cognitive behavioural therapy* OR *cognitive behavior therapy* OR *cognitive behavioral therapy* OR *CBT)*. Additional sources were included through a hand search that comprised examining reference lists of key articles and systematic reviews and authors' knowledge of manuscripts related to digital CBT for youth. Furthermore, the details from included studies were used to conduct a search to identify relevant manuscripts reporting on secondary or adherence outcomes from the included studies' data sets.

Data Extraction and Synthesis

Articles were identified from the search strategy, and after duplicates were removed, titles and then abstracts were reviewed for relevance by the first author (SHL). A second reviewer (AWS) independently screened 10% of the identified studies to ensure the reliability of the eligibility criteria. Full texts of the remaining studies (n=174) were appraised by 2 reviewers (SHL and MRA) to determine eligibility. Reasons for exclusion were discussed, and consensus for eligibility was confirmed for 6 articles. In cases where a consensus between the 2 reviewers could not be reached, a third reviewer (MSK) assessed for eligibility and a consensus was reached. The 2 reviewers (SHL and MRA) independently performed the data extraction using a data extraction template (Multimedia Appendix 1) designed to identify information, including the following: study details, sample details, intervention details, and outcomes of interest. It is worth noting that, for delivery mode, interventions were categorized as *sequential* where modular content was presented sequentially or explicitly delivered in a sequential order with content only becoming available when preceding modules were completed, or *nonsequential* where content was unrestricted allowing complete self-navigation.

Study outcomes of interest included the following: descriptions, statements, or instructions of appropriate use for young people; measures of adherence; and associations between adherence and symptom outcomes. We relied only on the information provided in the study manuscripts. Using the components of manualized CBT, we identified information on appropriate use pertaining to the following: intended recipient of the intervention, intended target condition, intended number of modules to be completed, intended duration and frequency of use, instructions regarding real-world enactment of skills (homework), adherence-promoting features (embedded within the intervention), core components, and symptom assessment and monitoring. We also extracted data on when and how users were recommended to access in-person support, excluding risk management procedures that constituted the study protocol. In lieu of specific descriptions of appropriate use, we extracted any information on intervention design features or other study details that implied appropriate use. This included the following: inclusion and exclusion criteria to indicate intended recipients, duration of intervention access to indicate intended duration of use, and tunneling to indicate intended sequence of content use. The sources of information used to derive aspects of appropriate use were recorded (trial protocol, feature description, or a statement in the study manuscript). Post hoc determination of the number of modules needed to achieve benefits was not considered as constituting any aspect of appropriate use.

Adherence-promoting features included those explicitly used by the authors to improve adherence or those that fit within the following categories: supported use defined as use that involves support, encouragement or guidance from a person, reminders, rewards, gamification, social or peer support, tailoring defined as the capacity to tailor content to meet an individual's requirements, personalized feedback including system-generated personalized feedback, customization of visual features such as avatars and color schemes, and interactive content defined as content requiring active inputs from the user. These categories were derived from past literature that endorsed supported use as an effective adherence promoter [44] and systematic reviews of user preferences [24,45]. Supported use was further categorized into autonomous, supported, or intervention-led blended, using the definitions provided by Fairburn and Patel [46] (Table 1). Interventions delivered in schools were categorized as supported, unless otherwise specified, and studies that did not specify support in the use of the intervention were recorded as autonomous. Interactive content was further coded as follows: activities, quizzes, homework activities, and multimedia content. The same intervention delivered under different conditions (eg, with or without additional in-person therapy sessions) was treated as 2 separate interventions and is presented in different rows in the tables. It is worth noting that, we elected to categorize personal feedback from a clinician or support person as supported use, not tailoring, as supported use is likely to have an element of personalization that does not necessarily involve tailored delivery of the intervention content.

In addition, the included studies were screened for references to registered or published protocols. Where protocols were identified, they underwent the same data extraction and synthesis procedure as the study manuscripts to supplement data on descriptions, statements, or instructions for appropriate use. Finally, the number of aspects of digital self-administration described in each study was tallied to show the degree to which aspects of self-administration were disclosed.

Table 1. Characteristics of included studies (n=32).

Study and year published	Country, setting, and year conducted	Study design	Participants (intervention)			
			Population and setting	Sample size, n	Girl or woman (transgender person), %	Age (years), mean (SD; range)
Berg et al [47], 2020	Sweden, 2018	2×2 factorial design	Clinically relevant anxiety symptoms and comorbid depression	120	81	16.97 (1.20; 15-19)
Bevan Jones et al [48], 2020	United Kingdom	Pre-post feasibility trial	History or risk of depression	35	79	16.3 (2.36; 13-23)
Calear et al [49,50], 2009 and 2013	Australia, 30 schools, 2006	Cluster RCT[a]	Universal sample of secondary school students (30% with prior history of depression)	559	60	14.34 (0.75; 12-17)
Clarke et al [51], 2009	United States, health maintenance organization	RCT	History of depression or risk of depression	83	81	22.6 (2.3; 18-24)
Fleming et al [52], 2012	New Zealand, 2009-2010	RCT	Excluded from mainstream education	20	44	14.9 (0.79; 13-16)
Ip et al [53], 2016	Hong Kong, 3 schools, 2013-2015	RCT	Mild or moderate depressive symptoms	130	68.1	14.63 (0.81; 13-17)
Jaycox et al [54], 2019	United States, 5 schools, year undisclosed	Pre-post open trial	In 4 out of 5 schools, students were identified by school counselors and social workers; in the fifth school, all students in the health class participate.	51	56.9	15.02 (1.86; range NR[b])
Kuosmanen et al [55], 2017	Ireland, Youthreach Centers, 2015-2016	Cluster RCT	From a second-chance education intervention. No symptom eligibility criteria	30	53.4	17.60 (1.24; 14-20)
Lenhard et al [56], 2017	Sweden, 2014-2015	12-week, single-blinded RCT	Obsessive compulsive disorder	33	46	14.6 (1.71; 12-17)
Lillevoll et al [57], 2014	Norway, 4 schools, 2009	4-arm RCT	No symptom eligibility criteria	527 (3 arms received intervention)	50	16.8 (1.0; 15-20)
Lucassen et al [58], 2020	New Zealand, 2014	Open trial—secondary analysis of 5 years of SPARX[c] usage data	No symptom eligibility criteria	9079	65.7 girls, 2.3 transgender person	NR (NR; 12-19
March et al [26], 2018	Australia, 2014-2016	Open trial	Elevated anxiety	4425	66.39	12.95 (2.97; 7-17)
Melnyk et al [59], 2015	United States, large public university, 2012-2013	RCT	No symptom eligibility criteria	82	86.4	18.4 (1.9; range NR)
Merry et al [60], 2012	New Zealand, 12 primary health care youth clinics, general practices, and school-based counseling services	Randomized controlled noninferiority trial	Depressive symptoms	94	65.7	15.55 (1.54; 12-19)
O'Connor et al [61], 2020	Canada, 2014-2016	2-arm, multisite, pilot RCTs	Anxiety concerns	36	90	15.3 (1.2; 13-17)
O'Dea et al [62], 2020	Australia, 2018-2019	RCT	No symptom eligibility criteria	94	86.5	14.82 (0.93; 12-16)
O'Kearney et al [63], 2009	Australia, girls-only schools	Controlled trial	No symptom eligibility criteria	67	100	Year 10

Study and year published	Country, setting, and year conducted	Study design	Participants (intervention)			
			Population and setting	Sample size, n	Girl or woman (transgender person), %	Age (years), mean (SD; range)
Radomski et al [64], 2020	Canada, 2016-2018	RCT	Mild to moderate anxiety	258	71	16.6 (1.7; 13-19)
Smith et al [65], 2015	United Kingdom, schools, 2011-2013	RCT	Significant depression	55	NR	NR (NR; 12−16)
Spence et al [66], 2011	Australia, 2006-2008	3-arm RCT	Anxiety disorder	44	59.13	13.98 (1.63; 12-18)
Stallard et al [67], 2011	United Kingdom, home and school	RCT	Anxiety disorder or mild to moderate depression	10	NR	13.5 (NR; 11-17)
Stasiak et al [68], 2012	New Zealand, 8 urban high schools	RCT	Low mood	17	41	15.2 (1.5; 13-18)
Stjerneklar et al [69], 2019	Denmark, 2015-2017	RCT	Anxiety disorder	33	79	15 (1.3; 13-17)
Tillfors et al [70], 2011	Sweden, 2007-2008	RCT	Social anxiety disorder	10	89	16.5 (1.6; 15-21)
Van Voorhees et al [71], 2009	United States, 2007	RCT	Subclinical depression from primary care sites	43	77	17 (NR; NR)
Van Voorhees et al [72], 2020; Gladstone et al [73,74], 2018 and 2020	United States, 2012-2016	RCT	Subclinical depression from 31 primary care sites	193	68	15.4 (1.5; 13-18)
Whittaker et al [75,76], 2017 and 2012	New Zealand, 15 schools, 2009-2011	RCT	No symptom eligibility criteria	426	68	14.3 (undisclosed; 13-17)
Wuthrich et al [27], 2012	Australia	RCT	Anxiety disorder	24	62.79	15.17 (1.11; 14-17)

[a]RCT: randomized controlled trial.

[b]NR: not reported.

[c]SPARX: Smart, Positive, Active, Realistic, X-factor thoughts.

Methodological Quality Assessment

The methodological quality of the included studies was assessed using the Cochrane Risk of Bias (ROB) Tool for randomized controlled trials (RCTs) [77] and the Integrated Quality Criteria for the Review of Multiple Study Designs [78] for other study designs. Quality ratings were determined independently by 2 reviewers (SHL and JRB) and then compared to resolve differences. Studies were considered methodologically sound if they met the Integrated Quality Criteria for the Review of Multiple Study Designs criteria for inclusion [78] or were rated as low in the ROB overall assessment of bias.

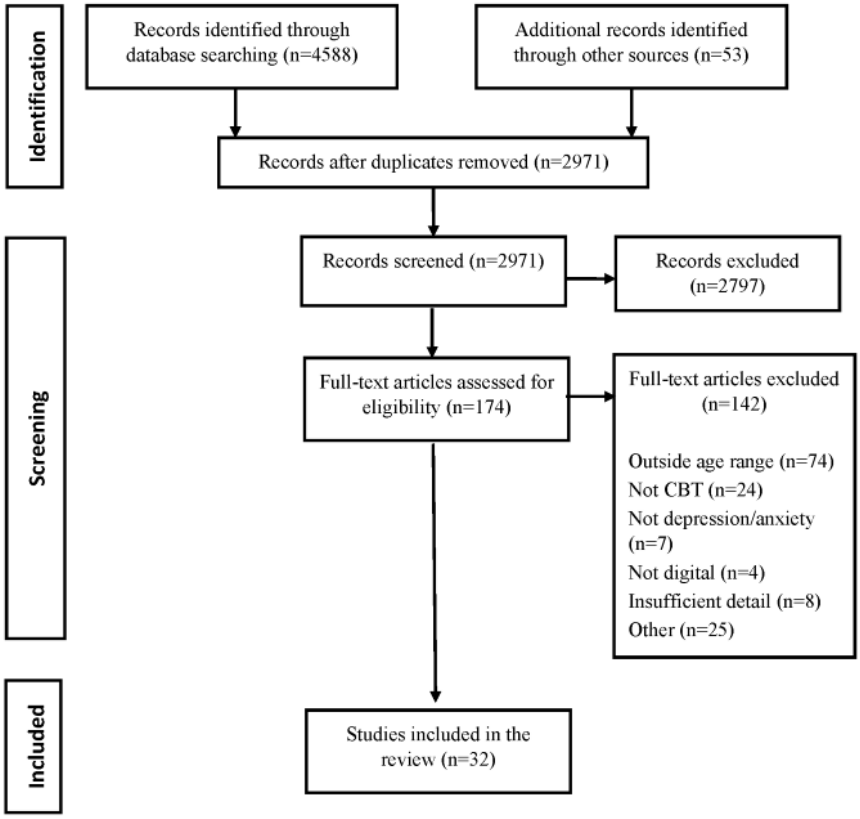
Results

Overview of Included Studies

Overview

The search resulted in 4641 studies, of which 174 full-text articles were identified. A total of 32 studies met all inclusion criteria and were included in the systematic review. The screening process and reasons for exclusion are presented in Figure 1. Table 1 summarizes the study, sample, and intervention characteristics of the 32 included manuscripts. On 3 occasions, >1 manuscript was produced from a single study; as such, these manuscripts were reported together, leaving a total of 28 independent studies.

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



Study Design and Country of Origin (Setting)

Of the 28 studies, most were RCTs (n=22, 78%), and the others were open trials (n=3, 10%), feasibility trials (n=11, 39%), 1 (3%) non-RCT, and 1 (3%) factorial design. The majority were conducted in Australia and New Zealand (11/28, 39%), Europe and the United Kingdom (9/28, 32%), and the United States (5/28, 17%), with others in Canada (2/28, 7%) and Hong Kong (1/28, 3%).

Sample Characteristics

Across the 28 studies, 16,578 participants received the intervention. Sample sizes varied considerably among studies, ranging from 10 to 9079 (mean 592.07, SD 1857.44; median 61). Overall, 42% (12/28) of studies had sample sizes <50, and 57% (16/28) had sample sizes >50. Regarding gender representation, 75% (21/28) of studies had >55% of their sample as girls or women, 10% (3/28) of studies had between 45% and 55% of the sample girls or women, 7% (2/28) of studies had <45% of the sample as girls or women, and 7% (2/28) of studies did not disclose gender representation of their samples. The girl or woman representation within the samples ranged from 41% to 100%. Across all studies, the mean age was 15.74 years (range 12.95-22.60), with a total age range of 7 to 24 years. A total of 10% (3/28) of studies did not report the mean age of the participants, and 14% (4/28) of studies did not report the age range.

Intervention Characteristics

Intervention characteristics are presented in Table 2. A total of 20 interventions from 28 studies were identified and included in the review, with the same intervention used in several studies: SPARX (Smart, Positive, Active, Realistic, X-factor thoughts; University of Auckland; n=4), MoodGYM (Australian National University; n=3), Brave Online (University of Queensland; n=2), CATCH-IT (University of Illinois; Competent Adulthood Transition with Cognitive Behavioral Humanistic and Interpersonal Training; n=2, not including the modified version), and Breathe (University of Alberta; Being Real, Easing Anxiety: Tools Helping Electronically; n=2). A total of 28% (8/28) of interventions targeted depression, 25% (7/28) targeted depression and anxiety, 10% (3/28) targeted unspecified anxiety, and 7% (2/28) targeted a specific anxiety disorder (obsessive compulsive disorder and social anxiety disorder). The digital delivery formats included web-based (n=11), computer or CD-ROM (n=6), multiplatform (n=1), weblinks viewed on a smartphone (n=1), and 1 smartphone app (n=1). Several interventions included elements of other therapeutic approaches in addition to CBT, such as interpersonal therapy (2/28, 7%), family systems therapy (1/28, 3%), and positive psychology (1/28, 3%). The total number of modules ranged from 4 to 15 (mean 8.15, SD 2.74). A total of 10% (3/28) studies did not specify the number of modules contained in the intervention. Most interventions (14/20, 70%) were sequential, 10% (2/20) were nonsequential (WeClick and MoodHwb), and 20% (4/20) were undetermined.

Table 2. Characteristics of interventions (n=20).

Study and year published	Intervention name	Target condition	Delivery mode	Delivery format	Therapeutic model	Number of modules or sessions
Berg et al [47], 2020	Undisclosed	Anxiety and depression	Web	Sequential	CBT[a]	8 modules
Bevan Jones et al [48], 2020	MoodHwb	Depression	Multiplatform (web, app)	Nonsequential	CBT, IPT[b], positive psychology, family systems therapy	Unclear
Calear et al [49,50], 2009 and 2013; Lillevoll et al [57], 2014; O'Kearney et al [63], 2009	MoodGYM	Anxiety and depression	Web	Sequential	CBT	5 modules, 29 exercises
Clarke et al [51], 2009	Undisclosed	Depression	Web	Undisclosed	CBT	Unclear; 4 components
Fleming et al [52], 2012; Kuosmanen et al [55], 2017[c]; Lucassen et al [58], 2020; Merry et al [60], 2012	SPARX[c]	Depression	Computer	Sequential	CBT	7 modules
Ip et al [53], 2016[d]	Grasp the opportunity (translated to Chinese and modified from CATCH-IT)	Depression	Web	Undisclosed	CBT	10 modules
Jaycox et al [54], 2019	The LIFT program	Anxiety and depression	Computer	Sequential	CBT	7 modules
Lenhard et al [56], 2017	BIP OCD[e]	OCD	Web	Undisclosed	CBT	12 modules
March et al [26], 2018; Spence et al [66], 2011	BRAVE Online	Anxiety	Web	Sequential	CBT	10 modules
Melnyk et al [59], 2015	COPE[f]	Anxiety and depression	Web	Sequential	CBT	7 modules
O'Connor et al [61], 2020; Radomski et al [64], 2020	Breathe[g]	Anxiety	Web	Sequential	CBT	8 modules; 6 modules[h]
O'Dea et al [62], 2020	WeClick	Depression, anxiety	Smartphone app	Nonsequential	CBT	4 character stories
Smith et al [65], 2015	Stressbusters	Depression (mild-moderate)	Computer	Sequential	CBT	8 modules
Stallard et al [67], 2011	TFD[i]	Anxiety, depression	CD-ROM	Undisclosed	CBT	6 modules
Stasiak et al [68], 2012	The Journey	Depression	CD-ROM	Sequential	CBT	7 modules
Stjerneklar et al [69], 2019	ChilledOut Online (Danish)	Anxiety, depression	Web	Sequential	CBT	8 modules
Tillfors et al [70], 2011	Undisclosed	Social anxiety disorder	Web	Sequential	CBT	9 modules
Van Voorhees et al [71,72], 2020 and 2009; Gladstone et al [73,74], 2018 and 2020[j]	CATCH-IT	Depression	Web	Sequential	CBT and IPT	14 modules; 14 modules (plus 1 optional anxiety module and 5 parent modules)
Whittaker et al [75,76], 2017 and 2012	MEMO CBT	Depression	Web links viewed on smartphone	Sequential	CBT	2 messages daily, for 9 weeks
Wuthrich et al [27], 2012	Cool Teens	Anxiety	CD-ROM	Sequential	CBT	8 modules

[a]CBT: cognitive behavioral therapy.

[b]IPT: interpersonal psychotherapy.

[c]SPARX-R is a revised version of the original Smart, Positive, Active, Realistic, X-factor (SPARX) intervention. The content of SPARX-R is essentially the same; however, it is framed as a preventive intervention for young people who *feel down, stressed, or angry* rather than focusing exclusively on depression.

[d]The Grasp Opportunity intervention removed all face-to-face components from CATCH-IT, including the motivational interview or brief advice component, and IPT modules were also excluded.

[e]OCD: obsessive compulsive disorder.

[f]COPE: Creating Opportunities for Personal Empowerment.

[g]Breathe: Being real, easing anxiety: tools helping electronically.

[h]Content covered in both interventions seems to be the same; however, Radomski et al 2020 [64] reported 6 modules and O'Connor et al 2020 [61] reported 8 modules.

[i]TFD: think, feel, do.

[j]Motivational interviewing was classified as a form of supported use.

Methodological Quality

All 6 non-RCT studies met the methodological quality criteria for inclusion. Of the 22 RCTs included in the review, 9 (41%) were assessed as having a low ROB, 9 (41%) as having some concerns, and 4 (18%) as having a high ROB. All these studies were included in the analysis.

Digital Self-administration Guidelines as Disclosed in Study Manuscripts

Overview

The aspects of digital self-administration extracted from the included studies are contained in Table 3 and Multimedia Appendix 2 [26,27,47-76]. Only 1 study (the study by O'Dea et al [62]) reported how instructions for appropriate use were relayed to young people. In this study, an animation was used to inform young people that they could use the intervention as they wished.

Table 3. Components of appropriate use described in studies or derived from study protocols or intervention features.

Study and year published	Components of appropriate use (part 1)					
	Recipients	Target condition or therapeutic goals	Amount of intervention completion to benefit	Duration and frequency of use	Sequence of content	Skill enactment
Berg et al [47], 2020	15-19 year olds with clinically significant anxiety, without or without comorbid depression (other comorbidities excluded)[a]	Anxiety and depression[a]	None provided	8-week access[a], 1 module per week	Sequential via numbered modules[b]	None provided
Bevan Jones et al [48], 2020[c]	13-23 year olds with a history or risk of depression[a]	Depression	None provided	Access for a minimum of 2 months[a]. Instructed "they could use the program as they wished"	Nonsequential[d]	None provided
Calear et al [49,50], 2009 and 2013	12-17 year olds[a]	Prevent or decrease depression and anxiety	Completion of all 5 modules	5-week access with 1 module delivered each week[a], each module 20-40 minutes	Sequential via controlled delivery of modules[a]	None provided
Clarke et al [51], 2009	18-24 year olds with a history of depression or risk of depression[a]	Depression[a]	None provided	32-week access[a], unrestricted, self-guided use	None provided	Intervention guides user to create a personalized self-contract to increase the frequency of selected pleasant activities (eg, taking a relaxing bath, going to a restaurant all by yourself)[b]. Prompted to record activities every few days[b]
Fleming et al [52], 2012	13-16 year olds with probable depression; those with severe depression were excluded[a]	Depression[a]	None provided	5-week access[a], 1-2 modules per week, each module approximately 30-minute duration	Sequential[b]	After each level, the digital guide reflected on how the learning could be applied in real life and set homework challenges[b]
Ip et al [53], 2016	13-17 year olds with mild or moderate depressive symptoms[a]	Reduce depressive symptoms (mild to moderate level) or prevent the onset of major depressive episodes.[a] "Improve negative cognition, reduce negative behaviors, strengthen resiliency, and reinforce positive behaviors"	None provided	Access for the study period (12 months)[a], use at anytime, anywhere	None provided	None provided
Jaycox et al [54], 2019	High school students with limited mental health resources[a]	PTSD[e], anxiety, and depression[a]	None provided	1-2 chapters per week	Sequential[b]	Goal setting at the end of each module[b]
Kuosmanen et al [55], 2017	15-20 year olds[a]	Prevent depression; "aimed for young people who feel down, stressed or angry"	None provided	20-30–minute modules, completion of one module each week[a]	Sequential levels[b]	None provided

Study and year published	Components of appropriate use (part 1)					
	Recipients	Target condition or therapeutic goals	Amount of intervention completion to benefit	Duration and frequency of use	Sequence of content	Skill enactment
Lenhard et al [56], 2017	12-17 year olds with a primary OCD[f] diagnosis[a]	OCD[a]	None provided	12 weeks access[a]	Sequential[b]	Young person and parent encouraged to complete ERP[g] exercises together and report back to clinician
Lillevoll et al [57], 2014	Senior high school students[a]	Prevent and reduce depressive symptoms	None provided	45-60 minute modules, 6-7 week access[a]	Sequential via locked content[b]	None provided
Lucassen et al [58], 2020	12-19 year olds[a]	Prevent and treat depressive symptoms	None provided	Open access, modules take 30 minutes	Sequential levels[b]	Set challenges are provided to allow practice and facilitate skill generalization[b]
March et al [26], 2018	7-17 year olds with elevated anxiety[a]	Anxiety[a]	None provided	20-week access, 1 session each fortnight[a]	Sequential[b]	None provided
Melnyk et al [59], 2015	Freshman college students[a]	Depression and anxiety[a]	None provided	10-12–week access,[a] 30-minute modules, 1 module per week	Sequential via locked content[b]	Weekly skill building homework assignments and goal setting logs
Merry et al [60], 2012	12-19 year olds with mild to moderate depressive symptoms[a]	Clinically significant depression[a]	None provided	4-7–week access[a] 30-minute modules	Sequential[b]	"Virtual" guide," sets and monitors real-life challenges, equivalent to homework"[b]
O'Connor et al [61], 2020	13-17 year olds with a self-identified anxiety concern[a]	Anxiety[a]	None provided	8 weeks of website access[a] instructed to use weekly	Sequential via numbered modules[b]	"Try Out feature, which outlined activities for the adolescent to choose to practice the module's key concepts and skills"[b]
O'Dea et al [62], 2020[h]	12-16 year olds[a]	Depression and anxiety	Completion of all 4 character modules	4-week access[a]; self-paced	Nonsequential[b]	None provided
O'Kearney et al [63], 2009	High school girls aged 15-16 years[a]	Aims to reduce depression and vulnerability to depression	None provided	6-week access[a]; self-paced	Sequential[a]	None provided
Radomski et al [64], 2020	13-19 year olds with self-reported anxiety[a]	Aims to address mild to moderate anxiety symptoms	None provided	30 min each, complete one session per week	Numbered modules indicate sequential content[b]	Skill enactment prompted between modules[b]
Smith et al [65], 2015	Designed for adolescents with mild to moderate depression	Depression[a]	None provided	8-week access[a], 30-40 min modules	None provided	"Designs own individualised homework based on specific technique"[b]
Spence et al [66], 2011	12-18 year olds meeting diagnostic criteria for social anxiety disorder, separation anxiety disorder, generalized anxiety disorder, or specific phobia[a]	Reduction in anxiety diagnostic status and severity[a]	Completion of all 10 modules	60 min modules, one module weekly	Sequential[b]	Responses to homework activities are accessed by therapist and feedback is provided via email[a]

Study and year published	Components of appropriate use (part 1)					
	Recipients	Target condition or therapeutic goals	Amount of intervention completion to benefit	Duration and frequency of use	Sequence of content	Skill enactment
Stallard et al [67], 2011	11-16 year olds with depression or anxiety assessed as suitable for CBT[a]	Depression and anxiety[a]	None provided	30-45 minute modules	Sequential via numbered modules[b]	"At the end of each session, participants are given a brief assignment to complete"[b]
Stasiak et al [68], 2012	13-18 year olds self-referred with probable or at risk of depression[a]	Depression[a]	Completion of all 7 modules	25-30 minute modules, complete between 4 and 10 weeks	Sequential via numbered modules[b]	Each module ends with a challenge (homework) for user to complete[b]
Stjerneklar et al [69], 2019	13-17 year olds meeting diagnostic criteria for an anxiety disorder[a]	"Reduce diagnostic severity and anxiety symptoms"	None provided	30-minute modules 14-week access[a]	Sequential via order[b]	"Each module contains homework practice tasks [users]... encouraged to complete"[b]
Tillfors et al [70], 2011	Adolescents (15-21 years) meeting diagnostic criteria for social anxiety disorder[a]	Social anxiety symptoms[a]	Completion of all 9 modules	9-week access,[a] 1 module per week	None provided	None provided
Van Voorhees et al [71], 2009	Adolescents (14-21 years) at risk of depression (persistent subthreshold depression)[a]	"The intervention was intended to reduce multiple thoughts, behaviors, and interpersonal interactions thought to increase vulnerability for depressive disorders... And strengthen behaviors, thoughts and interpersonal relations thought to be protective against depressive disorders"	None provided	None provided	None provided	None provided
Van Voorhees et al [72], 2020; Gladstone et al [73,74], 2018 and 2020	13-18 year olds with elevated depression symptoms or a history of depression or dysthymia[a], at clinically significant risk of depression but not with current depression	Preventing the onset of depressive episode[a]	None provided	15-20 min modules, 12-month access[a]	Sequential[a]	None provided
Whittaker et al [75,76], 2017 and 2012	Nondepressed years, 9-12 years; students (13-17 year old)[a]	"Prevention of the onset of depression"	A minimum of half the intervention completed	2 messages each day for 9 weeks[a]	Sequential[b]	None provided
Wuthrich et al 2012 [27]	14-17 year olds with diagnosed anxiety disorder[a]	Anxiety[a]	Not provided	30-minute modules; 12-week access[a]	Not provided	None provided

[a]Appropriate use derived from the trial protocol.

[b]Appropriate use derived from description of an intervention feature.

[c]All groups received weekly mail feedback via the study platform on exercises from the therapist. In addition, one group was invited to chat with the therapist in a 30-minute session.

[d]Appropriate use reported as a statement in the manuscript.

[e]PTSD: posttraumatic stress disorder.

[f]OCD: obsessive compulsive disorder.

[g]ERP: exposure and response prevention.

[h]All groups received standard digital CBT, designed to be neutral and straightforward. One group additionally received learning support that involved interactive features.

Intended Recipients and Target Condition

A total of 7% (2/28) of study manuscripts contained explicit statements of the intended intervention recipients: adolescents with mild to moderate depression [65] and adolescents with clinically significant risk of depression [72]. For all other manuscripts, intended recipients were derived from the participant inclusion and exclusion criteria. Of these, intended recipients were required to meet the diagnostic criteria for an anxiety disorder in 17% (5/28) of studies, were symptomatic (anxiety, depression, or both) in 32% (9/28) of studies, were at risk of developing depression in 17% (5/28) of studies (ie, symptomatic and/or history of depression), and were asymptomatic in 3% (1/28) of studies. Symptom level was not relevant to intended users in 21% (6/28) of studies (universal sample) and symptom details were not provided in 7% (2/28) of studies. A total of 39% (11/28) of study manuscripts described the intended target condition. For the remaining 43% (17/28) of studies, the intended target condition was derived from the study protocol (primary outcome measure). Depression was the target condition in 50% (14/28) of studies, anxiety in 32% (9/28) and both anxiety and depression in 17% (5/28). Finally, 7% (2/28) of studies provided statements on the intervention's therapeutic goals; both of the goals were to reduce unhelpful cognitions and behaviors, consistent with a CBT approach.

Intended Degree of Completion, Duration and Frequency of Use, and Sequence of Use

A total of 17% (5/28) of studies stated that all intervention modules should be completed, and 3% (1/28) of studies stated that half the modules should be completed. The remaining 82% (23/28) of studies did not provide this information. Although not explicitly stated as intended use by the study authors, 10% (3/28) of studies examined *noncompleters* and *completers*. These concepts appeared to be derived post hoc for analytical purposes. Definitions of the latter included those who completed ≥4 modules out of 8 [64], ≥3 modules out of 5 [63], and more than one-third of all activities [50].

None of the studies explicitly stated the intended duration of intervention use. This information was derived from the duration of intervention access in 82% (23/28) of studies (range 4 weeks to 12 months); however, 17% (5/28) of studies did not provide any information on this (eg, duration between pre- and postintervention assessments). Intended frequency of use was stated in 60% (17/28) of studies and was derived from an intervention feature in 3% (1/28) of studies. Of these 18 studies, 9 (50%) recommended that users complete 1 module per week, 2 (11%) recommended up to 2 modules per week, 1 (5%) recommended 1 module per fortnight, and 1 (5%) recommended that users read 2 messages per day. A total of 21% (6/28) of studies stated that the frequency of intervention use was unrestricted. There were 36% (10/28) of studies that did not provide any information regarding the intended frequency of use. The intended sequence of content was derived from intervention features, including numbered modules, locked content, or module display, in 64% (18/28) of studies, from the study protocols of 11% (3/28) of studies (prescribed delivery of content), and was stated in the manuscript in 3% (1/28) of studies. Of these 22 studies, interventions were intended to be completed sequentially in 20 (90%) studies, with 2 (10%) allowing users unrestricted access to alternate between modules. A total of 27% (6/28) of studies provided no information on the intended sequence of the intervention content.

Intended Skill Enactment

A total of 50% (14/28) of studies described intervention features that prompted homework activities. The remaining 50% (14/28) provided no information regarding recommendations for skill enactment.

Adherence-Promoting Features

Regarding supported use, 17% (5/28) of interventions were autonomous, 42% (12/28) were supported, 21% (6/28) were both supported and intervention-led blended, and 10% (3/28) were intervention-led blended only. A total of 28% (8/28) of interventions used reminders (eg, email or SMS text message), 3% (1/28) implemented rewards (a snack upon completion of a module), and 21% (6/28) contained gamified concepts (eg, leveling up). Regarding interactive content, 82% (23/28) of studies described interventions that used interactive features, including 20 with interactive activities, 8 with quizzes, 7 with homework activities, and 18 with multimedia content. A total of 10% (3/28) of studies used tailoring, 10% (3/28) used personalized feedback, and 14% (4/28) used customization of visual features. None of the interventions were reported to contain peer-support features.

Only 10% (3/28) of studies have examined the influence of an adherence-promoting strategy on adherence to recommended use. Van Voorhees et al [71] examined supported use as an adherence promoter by combining the digital intervention with 3 in-person motivational interviewing sessions. These were found to significantly increase site visits, time in intervention, proportion of exercises completed, and number of characters typed into interactive exercises. Berg et al [47] evaluated 2 strategies: supported use via chat sessions with a therapist and learning support, which included short summaries, pedagogical pictures, videos, and quizzes. These strategies, alone or combined, did not lead to greater adherence when compared with the control condition. Finally, Lillevoll et al [57] evaluated the influence of personalized email reminders on adherence to MoodGYM compared with standard email reminders and no reminders. Neither form of reminder affected adherence to the intervention.

Core Therapeutic Components, Symptom Monitoring, and Accessing Crisis Support

A total of 14% (4/28) of studies reported core therapeutic components, including exposure therapy, exposure and response prevention therapy, behavioral therapy, and cognitive restructuring. The remaining 85% (24/28) of studies did not identify any. A total of 42% (12/28) of studies described intervention features for monitoring mood and assessing safety. Of these, 50% (6/12) of studies described symptom monitoring that occurred *every few days* or weekly. The remainder (6/12, 50%) did not provide any information regarding the frequency of mood monitoring. Furthermore, 28% (8/28) of study manuscripts described intervention features that assessed suicide risk and responded with helping-seeking information. The

remaining 85% (24/28) of studies did not provide any information on when users were recommended to seek crisis support.

Reporting on Aspects of Intended Use

Of the 10 aspects of appropriate use identified to inform instructions for use (presented in Figure 1), no study manuscript reported on all 10 aspects. The average number of aspects reported was 6.29 (SD 1.33) with a range of 3 to 8. Descriptions of intervention features and study protocol details were relied upon heavily to infer aspects of appropriate use (Table 3 and Multimedia Appendix 2).

Measures of Adherence

The measures of adherence used in the included studies are presented in Table 4. To support a clearer interpretation of the results, adherence measures were coded as follows: module completion, intervention visits or log-ins, time spent in intervention, activity or homework completion and practice, and total content completion. Study attrition was not considered to be an appropriate measure of intervention adherence, as it was not possible to differentiate the noncompletion of study assessments from adherence to the intervention. Almost all the studies (24/28, 85%) measured module completion, which was reported as either mean (and SD) module completion or as percentage of the sample completing a defined number of modules (eg, 14/28, 50% of study manuscripts reported the percentage of the sample completing all modules). Moreover, 21% (6/28) of studies reported intervention visits or log-ins, either as the mean or as the percentage of the sample visiting on a defined number of occasions. Furthermore, 25% (7/28) of studies reported time spent in the intervention, either as the mean total intervention time or as the mean time spent per event (eg, visit or module). A total of 32% (9/28) of studies reported activity completion as either the mean, percentage of all activities completed, percentage of the sample that completed a set number of activities, or percentage who engaged in any of the activities. Just under half of all studies (13/28, 46%) reported >1 measure of adherence, indicating multidimensional operationalizations of the construct.

Table 4. Methods for measuring adherence to digital interventions and their association with depression and anxiety outcomes (n=28).

Study and year published	Measure of adherence					Association with depression outcome	Association with anxiety outcomes
	Module completion	Site or pp visits or log-ins	Time spent in intervention (min)	Activity, homework completion or practice	Total intervention completion (%)		
Berg et al [47], 2020	Mean 5.46/8 modules (SD 2.82); all modules=39.2%	NR[a]	NR	NR	39.20	NR	NR
Bevan Jones et al [48], 2020[b]	NR	1-2 per week=21%, 1-2 per month=44%, 1-2 total=26%	Several hours=3% approximately 1 hour=29% approximately 30 minutes=54%, few minutes=20%, no visits=3%	NR	NR	NR	NR
Calear et al [49,50], 2009 and 2013	Mean 3.16/5 modules >3 modules=62%; all modules=32.7%	NR	NR	15% of sample completed at least 20 of 29 exercises	32.70	0[c]	0[c]
Clarke et al [51], 2009	NR	Mean web visits 8.5 (SD 14.2), median web visits 6, range web visits 0-111, page hits (mouse clicks)	Mean 115.1 (SD 176.1), median 52, range 0-1088	NR	NR	Minutes in intervention—Pos[d], page hits—Pos, mean web visits 0	NR
Fleming et al [52], 2012	>4 modules=81%; all modules=69%	NR	NR	NR	69	NR	NR
Ip et al [53], 2016	Median 3 (IQR 5), all modules=10% (n=26/257)	NR	Median=39.3 (IQR=63.4)[e]	NR	10	Neg[f]	Neg
Jaycox et al [54], 2019	Mean 6.37/7 modules	NR	NR	% video watched=63−89%	NR	NR	NR
Kuosmanen et al [55], 2017	Mean 5/7 modules, >4 modules=87%, all modules=30%	NR	NR	55%-65% practiced (variety of skills, eg, thought monitoring, thought challenging, problem-solving)	30	NR	NR
Lenhard et al [56], 2017	Mean 8.52/12 modules [e], 1-5 modules=97%, all modules=27%	NR	NR	NR	27	NR	0
Lillevoll et al [57], 2014	1-2 modules=64%, >3modules=14%, all modules=3%	NR	NR	NR	3	NR	NR
Lucassen et al [58], 2020	1 module=53.7%, 1-3 modules=44.9%, >4 modules=8.8%, 1-6 modules=50.01%, all 7 modules=3.7%	NR	<25 minutes per module	NR	3.70	NR	NR
March et al [26], 2018	Mean 2.21/10 modules (SD 2.44); includes patient that did not start the intervention, no modules=21.65%, 1-2 modules=48.05%, >3 modules=30.31%	NR	NR	17.72/25	NR	NR	NR

Study and year published	Measure of adherence					Association with depression outcome	Association with anxiety outcomes
	Module completion	Site or pp visits or log-ins	Time spent in intervention (min)	Activity, homework completion or practice	Total intervention completion (%)		
Melnyk et al [59], 2015	All modules=99%	NR	NR	NR	NR	NR	NR
Merry et al [60], 2012	>4 modules=86%, all modules=60%	NR	NR	62% completed most or all homework challenges	60	NR	NR
O'Connor et al [61], 2020	No modules=6%, all modules=36%	NR	NR	NR	36% completed all 8 modules (N=13/36)	NR	NR
O'Dea et al [62], 2020	No modules=7.1%, all modules=60%	Mean 4.29, range 1-11	Per story: mean 5.55 min; overall: mean 19 min	NR	60	NR	NR
O'Kearney et al [63], 2009	>3 modules=30% (n=20/67)	NR	NR	NR	NR	0[c]	NR
Radomski et al [64], 2020	Mean 2.2/8 modules (SD 2.3), all modules=19.4%, >4 modules (adherers)=27.9%, <3 modules=72.1%, no modules=35.3%	NR	NR	NR	19.4	NR	0
Smith et al [65], 2015	>4 modules=93%, all modules=86%	NR	NR	NR	86	NR	NR
Spence et al [66], 2011	Mean 7.5/10 modules, all modules=39%	NR	NR	NR	39	NR	NR
Stallard et al [67], 2011	All completed=85%	NR	NR	NR	85	NR	NR
Stasiak et al [68], 2012	All completed=94%	NR	NR	NR	94	NR	NR
Stjerneklar et al [69], 2019	Mean 5.4/8 modules, all modules=30%, >4 modules=69%	Mean 24.4, range 7-51	NR	Mean 74.4	30	NR	NR
Tillfors et al [70], 2011	Mean 2.9/9 modules[e], range 1-6	NR	NR	NR	NR	0	0
Van Voorhees et al [71], 2009	NR	Percentage of sample visiting site at least once=84.1%[g]	Mean 121	Mean percentage of tasks completed=64%[g], mean number of characters typed in tasks=2724	NR	NR	NR
Van Voorhees et al [72], 2020; Gladstone et al [73,74], 2018 and 2020	Mean 3.4/14 modules (SD 4.7)	Days visited the site=3.7 (4.5)	Total time on website (min)=100.2 (143.1)	Characters typed=3071 (4572)	NR	0	NR
Whittaker et al [75,76], 2017 and 2012	NR	NR	NR	Half of the messages viewed by 19% of sample[e,h]	NR	0	NR

Study and year published	Measure of adherence					Association with depression outcome	Association with anxiety outcomes
	Module completion	Site or pp visits or log-ins	Time spent in intervention (min)	Activity, homework completion or practice	Total intervention completion (%)		
Wuthrich et al [27], 2012	All modules—98.4%	NR	NR	NR	98.4	NR	NR

[a]NR: association was not reported.

[b]Time spent in intervention was average per visit [48].

[c]Association between adherence and outcomes determined by completer versus noncompleter analysis.

[d]Pos: positive association.

[e]Adherence measure used to determine the adherence-outcome association.

[f]Neg: negative association.

[g]Data were collated when more than one group received the intervention.

[h]Actual, rather than self-reported outcomes.

Association Between Adherence and Outcomes

The heterogeneity of adherence measures, the different statistical methods used to determine an association between adherence and outcomes, and incomplete reporting of results made it difficult to extract comparable data for a meta-analysis. Associations are, instead, reported descriptively. A total of 19 (67%) studies did not report on the association between adherence and symptom reduction. Of the remaining 8 studies, 6 (21%) examined the relationship between adherence to the intervention and treatment outcomes for depression. A total of 4 studies, 2 (7%) of which compared outcomes for completers and noncompleters (as defined by the authors), found no association between adherence and depression outcomes. Furthermore, 7% (2/28) of studies found an association between depression and adherence, but in opposing directions, such that depression decreased as time in intervention increased as per Ip et al [53], whereas Clarke et al [51] found that depression reduced when the total time spent in intervention decreased. Moreover, 14% (4/28) of studies examined the association between adherence and anxiety outcomes: 10% (3/28) found no association, and 3% (1/28) found a negative association between symptom reduction and total time spent in intervention, whereby more time in the intervention was associated with lower anxiety symptoms [53]. None of the studies that used multiple measures of adherence examined the independent associations of these measures with treatment outcomes.

Discussion

Principal Findings

This systematic review aimed to synthesize how appropriate use has been defined in digital CBT interventions for youth, how instructions for use are relayed to young people and reported in studies, and how adherence to appropriate use has been operationalized and measured. This review found that none of the included studies systematically described definitions of appropriate use, which were instead derived from the study protocol or intervention features. Despite all interventions being based on CBT, the derived definitions of appropriate use varied widely among studies, particularly in terms of intended recipients, frequency and duration of use, and use of adherence-promoting features. The only component of use with some consistency among studies was the sequence of use, with 71% (n=20) of the 28 studies using interventions intended to be used sequentially. Nearly all included studies neglected to disclose how the instructions or recommendations for appropriate intervention use were relayed to young people, with only 3% (1/28) of studies providing such information [62]. The results showed that most studies (24/28, 85%) operationalized adherence as the degree of intervention completion, with significant heterogeneity in how this was measured, regardless of how appropriate use was defined. The most consistently reported measure of adherence was the percentage of the sample that completed the entire intervention (14/28, 50%). There was little evidence of an association between degree of use and improved outcomes.

Appropriate self-administration of digital CBT is critical for achieving the therapeutic goals. Therefore, a lack of reporting on both definitions of appropriate use and how instructions for use are provided to users is somewhat surprising. Indeed, an understanding of how the intervention creators desired the intervention to be used by youth had to be derived from the study protocol and intervention features for almost all components of appropriate use, with the frequency of use the only component explicitly stated in studies. These findings suggest that inadequate consideration has been given to developing nuanced definitions of appropriate use. As early digital CBT interventions were direct translations of in-person CBT, this omission may stem from the assumption that engagement with CBT should be consistent regardless of the delivery mode (ie, full program completion). However, given the current innovative approaches to CBT delivery, particularly the movement away from sequential delivery and expectations of full program completion [38], it is imperative to investigate patterns of digital CBT use that produce the greatest improvements in defining appropriate use. There is also little evidence that young users of digital CBT for depression and anxiety are provided with adequate instructions for use. The impact of this on young people's treatment uptake, engagement, response, and attitudes toward digital mental health care remains

unknown. Poorly relayed instructions for appropriate use to young people, their guardians, and health professionals may contribute to poor intervention adherence in digital CBT. It is possible that more nuanced definitions of appropriate use and clearly delivered instructions on use could improve intervention adherence and ultimately have beneficial effects on the outcomes.

The resulting lack of guidance on how to operationalize and measure adherence is an additional consequence of poorly defined recommendations for appropriate use. Specifically, the degree to which actual use aligns with appropriate use can only be determined if the appropriate use is adequately defined. We found that adherence was most often operationalized as the degree of program completion, demonstrating a lack of correspondence among definitions of appropriate use, which varied widely, and operationalization of adherence. This indicates that definitions of appropriate use are either inadequate or not being used to operationalize adherence. This finding is also consistent with findings in the adult literature, where appropriate use is often undefined but nevertheless operationalized generically as *the more use, the better* [31], with significant heterogeneity in how the amount of use is measured [29,79,80]. Our conceptualization of the appropriate use of digital CBT is made up of 10 components (based on in-person CBT manuals) and is therefore multidimensional. This suggests that the measurement of adherence should be multidimensional and nuanced enough to capture adherence to each aspect of appropriate use. Taken together, adherence definitions and measures that provide a meaningful indication of the extent to which interventions are used appropriately are lacking, and require refinement in future studies.

We found little evidence to suggest an association between amount of use and symptom improvement. Most studies did not examine or report a dose-response effect for the intervention being evaluated. However, of those that did, most found no association between use and either depression or anxiety outcomes, regardless of the type of measure used. This finding demonstrates that the pervasive notion of *the more use, the better* is unsupported for digital CBT. A tenuous association between use and outcomes is consistent with findings reported in a recent meta-analysis of youth digital CBT interventions, which found no association between adherence (measured as the amount of use) and outcomes [25]. It is also somewhat consistent with a review of adult literature, which found that the association between use and depression outcomes depended on the usage measure used [29]. The lack of association between use and outcomes is, however, in contrast to in-person CBT interventions, where there are clear associations between intervention attendance and therapeutic benefits [81,82].

There are several possible explanations for the lack of association between adherence and outcomes. First, it is possible that measuring adherence as the degree of intervention completion or use does not capture the complexity to which a person may or may not adhere to various aspects of appropriate self-administration. Specifically, current one-dimensional measures of adherence are unlikely to quantify adherence to other components of appropriate use that may be more critical to intervention outcomes such as skill enactment or accessing crisis support. However, as digital CBT has been shown to be effective in well-controlled trials, outcomes are unlikely to be completely unrelated to intervention use [80,83]. Another explanation is that users simply stop using the intervention when sufficient benefits have been obtained. This point within the intervention would be expected to vary among individuals, so no association between the amount of intervention use and benefits would be evident. This notion is supported by a study that used graphical modeling to identify a subset of users who showed the greatest improvement in depressive and anxiety symptoms despite spending less time in the intervention than other groups of users [37]. Finally, some critical components of the intervention—the *active ingredients*— may drive the therapeutic response, and it is completion of these components, rather than completion of the whole intervention per se that is associated with outcomes. These explanations suggest that the current narrow definitions of appropriate use as the degree of intervention completion or use are too generic, and a more nuanced approach to the measurement of adherence is required.

Poorly defined recommendations for appropriate use may have several practical implications, including the reduced effectiveness of digital CBT because of inappropriate self-administration. Given the low uptake of evidence-based interventions [19,20], this presents a significant missed opportunity to improve young people's mental health. It may also prevent the self-selection of interventions that are best suited to an individual's circumstances and limit endorsement and dissemination by mental health professionals. Currently, there is little information in the scientific literature to guide the development of definitions of appropriate use or the disclosure of instructions for the use of digital CBT. In response to our findings and to rectify poorly defined appropriate use and generic operationalization of adherence, we developed a framework to support the development and disclosure of definitions of appropriate use and instructions for the self-administration of digital CBT. The results are presented in Table 5. It is based on the components of manualized CBT designed to guide clinicians in delivering in-person CBT and mapped onto definitions of intended use (the purpose of the product), indications for use (who will be using the product and why it will be used), and instructions for use (how to use the product) as defined in the medical arena [84]. The use of this framework to develop nuanced instructions for use could optimize the therapeutic benefits of interventions. Finally, we suggest an active evaluation of appropriate definitions and their refinement based on the outcomes of the evaluation. For example, a method to evaluate the extent to which the degree of completion, duration, and frequency of use incorporated in definitions of appropriate use produces therapeutic benefits is to conduct a survival analysis [85,86]. Comparing instructions for use related to degree of completion, duration, and frequency of use with those demonstrated to produce reliable and clinically significant reductions in symptoms, as determined by the survival analysis, will contribute to the refinement of the appropriate use definition. The proposed development, evaluation, and refinement of definitions of appropriate use and instructions for use is represented in Figure 2.

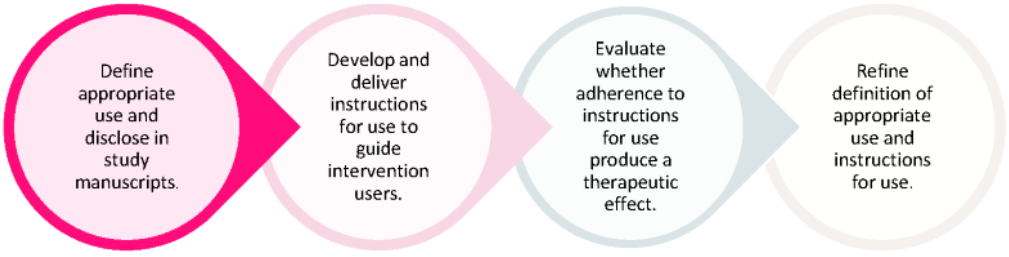
Table 5. Framework to develop definitions of appropriate use and operationalize adherence in digital CBT[a].

Guideline	Description	Measure of adherence	Example—CBT intervention for specific phobia
Indications for use			
Intended recipients	Who should use the intervention? Consider indications and contraindications for who should use the intervention.	Determine who is using the intervention	For 12-16 year olds with a specific phobia, without learning or developmental disorders
Intended target condition	What disorder has the intervention been designed to treat?	Determine symptomatic level of users	Specific phobia, for example, spider phobia
Intended use			
Therapeutic goals	Identify the specific outcomes the treatment aims to achieve.	N/A[b]	Reduce fear of spiders
Core therapeutic components (empirical or theoretical)	Identify intervention components that have been empirically or theoretically demonstrated to be associated with, or to mediate, improved outcomes.	Measure the extent to which core components are completed	Exposure hierarchy
Instructions for use			
Intended intervention completion	Define how much of the intervention should be completed.	What portion of users complete the required amount of the intervention	All modules of the intervention should be completed
Frequency and duration	What frequency and duration of use is required to produce therapeutic outcomes?	What portion of users engage at the required frequency and duration of use	Exposure to spiders for 3-6 months
Intended sequence	What pathway through the content will produce therapeutic outcomes?	Determine navigation through content and the portion of users engaging the required sequence (if relevant)	Content should be completed sequentially
Enactment of skills	What enactment or practice of skills is required beyond actual intervention use to achieve therapeutic outcomes?	Determine the extent to which users practice skills as required	Approach rather than avoid the next time when confronted with feared object or situation
Symptom monitoring	What level of symptom monitoring is required to produce therapeutic outcomes? What behaviors should outcomes of symptom monitoring produce?	Determine the extent to which users monitor symptoms as required (if relevant)	Anxiety symptoms are improving, move on to the next level of the exposure hierarchy (or cease treatment or trigger maintenance planning)
Access to crisis support	When should users be recommended to access crisis support and how will they access crisis support?	Determine the extent to which users access crisis support as suggested	N/A
Supported use	Consider whether self-guidance is unlikely to produce adequate adherence to appropriate self-administration.	Determine the extent to which support is accessed (if relevant)	Access to a clinician to construct exposure hierarchy

[a]CBT: cognitive behavioral therapy.

[b]N/A: not applicable.

Figure 2. Process for defining, evaluating, and refining definitions of appropriate use and instructions for use.



Limitations

In this review, we limited our search to CBT for depression and anxiety, making it difficult to generalize our findings to other interventions or mental health conditions. Future studies could include other interventions and conditions to develop an understanding of how appropriate use and instructions for use have been defined and disclosed more broadly in digital health interventions. Moreover, our data failed to account for the possibility that appropriate use may have been defined or instructions for use provided to participants, but not documented in the study manuscripts or published protocols. However, consistently generic operationalization of adherence across studies suggests that this explanation is unlikely, and adequate and systematic documentation of these constructs is, nevertheless, required.

Conclusions

This review has contributed to our understanding of how appropriate self-administration of digital CBT has been defined, relayed to users, measured, and associated with outcomes across various digital CBT interventions targeting depression and anxiety in young people. Overall, there is a lack of systematic reporting of nuanced definitions of appropriate use, and measures of adherence rarely provide adequate information on the degree of actual use corresponding to recommended appropriate use. There is little evidence that the degree of program completion or use is associated with intervention benefits. Our findings may, in part, explain low engagement in digital CBT; however, more work is required to better understand engagement with digital CBT and find ways to enhance it. A framework to assist in the development of guidelines for the self-administration of digital CBT has been provided.

Acknowledgments

BOD is currently supported by a National Health and Medical Research Council Medical Research Future Fund Investigator Grant (1197249). This project was funded by a noncompetitive philanthropic research grant from the Goodman Foundation. The funders had no role in any aspect of this study.

Authors' Contributions

SHL and AWS designed the study. SHL and MRA conducted literature searches, screened titles, abstracts, and full texts for eligibility for inclusion in the review, and extracted data from the manuscripts. SHL and JRB conducted the quality ratings. MSK provided a third independent review for any articles for which there was disagreement between the first 2 reviewers. SHL wrote the first draft of the manuscript and BOD provided critical revisions. All authors contributed to and have approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data extraction template.
[XLSX File (Microsoft Excel File), 11 KB - mental_v9i8e37640_app1.xlsx]

Multimedia Appendix 2
Components of appropriate use.
[DOCX File , 27 KB - mental_v9i8e37640_app2.docx]

References

1. Barlow D, Farchione T, Fairholme C, Ellard K, Boisseau C, Allen L, et al. Unified Protocol for Transdiagnostic Treatment of Emotional Disorders Therapist Guide. Oxfordshire, England, UK: Oxford University Press; 2010.
2. Zinbarg R, Craske M, Barlow D. Mastery of Your Anxiety and Worry (MAW): Therapist Guide (Treatments That Work) 2nd Edition. Oxfordshire, England, UK: Oxford University Press; 2006.
3. David D, Cristea I. The new great psychotherapy debate: scientific integrated psychotherapy vs. plurality. Why Cognitive-Behavior Therapy is the gold standard in psychotherapy and a platform for scientific integrated psychotherapy. JEBP 2018 Sep 01;18(2):1-17. [doi: 10.24193/jebp.2018.2.11]
4. Health of young people. Australian Institute of Health and Welfare. URL: https://www.aihw.gov.au/reports/children-youth/health-of-young-people [accessed 2022-02-09]
5. Higa-McMillan CK, Francis SE, Rith-Najarian L, Chorpita BF. Evidence base update: 50 years of research on treatment for child and adolescent anxiety. J Clin Child Adolesc Psychol 2016;45(2):91-113. [doi: 10.1080/15374416.2015.1046177] [Medline: 26087438]

6. Watts SE, Turnell A, Kladnitski N, Newby JM, Andrews G. Treatment-as-usual (TAU) is anything but usual: a meta-analysis of CBT versus TAU for anxiety and depression. J Affect Disord 2015 Apr 01;175:152-167. [doi: 10.1016/j.jad.2014.12.025] [Medline: 25618002]
7. Weersing VR, Jeffreys M, Do MT, Schwartz KT, Bolano C. Evidence base update of psychosocial treatments for child and adolescent depression. J Clin Child Adolesc Psychol 2017;46(1):11-43 [FREE Full text] [doi: 10.1080/15374416.2016.1220310] [Medline: 27870579]
8. Perepletchikova F, Kazdin AE. Treatment integrity and therapeutic change: issues and research recommendations. Clin Psychol Sci Pract 2005;12(4):365-383. [doi: 10.1093/clipsy.bpi045]
9. Wilson G. Manual-based treatment and clinical practice. Clin Psychol Sci Pract 1998;5(3):363-375. [doi: 10.1111/j.1468-2850.1998.tb00156.x]
10. Barlow D, Craske M. Mastery of Your Anxiety and Panic. USA: Oxford University Press; 1989.
11. Ehrenreich-May J, Kennedy S, Sherman J, Bilek E, Buzzella B, Bennett S, et al. Unified Protocols for Transdiagnostic Treatment of Emotional Disorders in Children and Adolescents: Therapist Guide (Programs That Work). USA: Oxford University Press; 2018.
12. Fernandez E, Salem D, Swift JK, Ramtahal N. Meta-analysis of dropout from cognitive behavioral therapy: magnitude, timing, and moderators. J Consult Clin Psychol 2015 Dec;83(6):1108-1122. [doi: 10.1037/ccp0000044] [Medline: 26302248]
13. Salmoiraghi A, Sambhi R. Early termination of cognitive-behavioural interventions: literature review. Psychiatrist 2018 Jan 02;34(12):529-532. [doi: 10.1192/pb.bp.110.030775]
14. Swift JK, Greenberg RP. Premature discontinuation in adult psychotherapy: a meta-analysis. J Consult Clin Psychol 2012 Aug;80(4):547-559. [doi: 10.1037/a0028226] [Medline: 22506792]
15. Wergeland GJ, Riise EN, Öst LG. Cognitive behavior therapy for internalizing disorders in children and adolescents in routine clinical care: a systematic review and meta-analysis. Clin Psychol Rev 2021 Feb;83:101918 [FREE Full text] [doi: 10.1016/j.cpr.2020.101918] [Medline: 33186776]
16. Aguirre Velasco A, Cruz IS, Billings J, Jimenez M, Rowe S. What are the barriers, facilitators and interventions targeting help-seeking behaviours for common mental health problems in adolescents? A systematic review. BMC Psychiatry 2020 Jun 11;20(1):293 [FREE Full text] [doi: 10.1186/s12888-020-02659-0] [Medline: 32527236]
17. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
18. Mojtabai R, Olfson M, Sampson NA, Jin R, Druss B, Wang PS, et al. Barriers to mental health treatment: results from the National Comorbidity Survey Replication. Psychol Med 2011 Aug;41(8):1751-1761 [FREE Full text] [doi: 10.1017/S0033291710002291] [Medline: 21134315]
19. Gandhi E, OGradey-Lee M, Jones A, Hudson JL. Receipt of evidence-based care for children and adolescents with anxiety in Australia. Aust N Z J Psychiatry 2022 Jan 19:48674211068780. [doi: 10.1177/00048674211068780] [Medline: 35044881]
20. Sawyer MG, Reece CE, Sawyer AC, Hiscock H, Lawrence D. Adequacy of treatment for child and adolescent mental disorders in Australia: a national study. Aust N Z J Psychiatry 2019 Apr 02;53(4):326-335. [doi: 10.1177/0004867418808895] [Medline: 30387377]
21. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
22. Gore FM, Bloem PJ, Patton GC, Ferguson J, Joseph V, Coffey C, et al. Global burden of disease in young people aged 10–24 years: a systematic analysis. Lancet 2011 Jun 18;377(9783):2093-2102. [doi: 10.1016/S0140-6736(11)60512-6] [Medline: 21652063]
23. The mental health of children and adolescents. Australian Government Department of Health and Aged Care. 2015. URL: https://www.health.gov.au/resources/publications/the-mental-health-of-children-and-adolescents [accessed 2022-02-11]
24. 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 [FREE Full text] [doi: 10.3389/fpsyt.2019.00759] [Medline: 31798468]
25. Christ C, Schouten MJ, Blankers M, van Schaik DJ, Beekman AT, Wisman MA, et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in adolescents and young adults: systematic review and meta-analysis. J Med Internet Res 2020 Sep 25;22(9):e17831 [FREE Full text] [doi: 10.2196/17831] [Medline: 32673212]
26. March S, Spence SH, Donovan CL, Kenardy JA. Large-scale dissemination of internet-based cognitive behavioral therapy for youth anxiety: feasibility and acceptability study. J Med Internet Res 2018 Jul 04;20(7):e234 [FREE Full text] [doi: 10.2196/jmir.9211] [Medline: 29973338]
27. Wuthrich VM, Rapee RM, Cunningham MJ, Lyneham HJ, Hudson JL, Schniering CA. A randomized controlled trial of the Cool Teens CD-ROM computerized program for adolescent anxiety. J Am Acad Child Adolesc Psychiatry 2012 Mar;51(3):261-270. [doi: 10.1016/j.jaac.2011.12.002] [Medline: 22365462]

28. Wozney L, Huguet A, Bennett K, Radomski AD, Hartling L, Dyson M, et al. How do eHealth programs for adolescents with depression work? A realist review of persuasive system design components in internet-based psychological therapies. J Med Internet Res 2017 Aug 09;19(8):e266 [FREE Full text] [doi: 10.2196/jmir.7573] [Medline: 28793983]
29. 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]
30. Adherence to Long-term Therapies Evidence for Action. Geneva: World Health Organization; 2003.
31. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
32. Georgeson A, Highlander A, Loiselle R, Zachary C, Jones DJ. Engagement in technology-enhanced interventions for children and adolescents: current status and recommendations for moving forward. Clin Psychol Rev 2020 Jun;78:101858 [FREE Full text] [doi: 10.1016/j.cpr.2020.101858] [Medline: 32413679]
33. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
34. Ng MM, Firth J, Minen M, Torous J. User engagement in mental health apps: a review of measurement, reporting, and validity. Psychiatr Serv 2019 Jul 01;70(7):538-544 [FREE Full text] [doi: 10.1176/appi.ps.201800519] [Medline: 30914003]
35. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
36. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
37. Chien I, Enrique A, Palacios J, Regan T, Keegan D, Carter D, et al. A machine learning approach to understanding patterns of engagement with internet-delivered mental health interventions. JAMA Netw Open 2020 Jul 01;3(7):e2010791 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.10791] [Medline: 32678450]
38. 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]
39. Raevuori A, Vahlberg T, Korhonen T, Hilgert O, Aittakumpu-Hyden R, Forman-Hoffman V. A therapist-guided smartphone app for major depression in young adults: a randomized clinical trial. J Affect Disord 2021 May 01;286:228-238 [FREE Full text] [doi: 10.1016/j.jad.2021.02.007] [Medline: 33743385]
40. Szymkowiak A, Melović B, Dabić M, Jeganathan K, Kundi GS. Information technology and Gen Z: the role of teachers, the internet, and technology in the education of young people. Technol Soc 2021 May;65:101565. [doi: 10.1016/j.techsoc.2021.101565]
41. Tylee A, Haller DM, Graham T, Churchill R, Sanci LA. Youth-friendly primary-care services: how are we doing and what more needs to be done? Lancet 2007 May 05;369(9572):1565-1573. [doi: 10.1016/S0140-6736(07)60371-7] [Medline: 17482988]
42. Vizcaya-Moreno MF, Pérez-Cañaveras RM. Social media used and teaching methods preferred by generation Z students in the nursing clinical learning environment: a cross-sectional research study. Int J Environ Res Public Health 2020 Nov 09;17(21):8267 [FREE Full text] [doi: 10.3390/ijerph17218267] [Medline: 33182337]
43. Moher D, Liberati A, Tetzlaff J, Altman DG, 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]
44. 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]
45. Jeminiwa RN, Hohmann NS, Fox BI. Developing a theoretical framework for evaluating the quality of mHealth apps for adolescent users: a systematic review. J Pediatr Pharmacol Ther 2019 Jul;24(4):254-269 [FREE Full text] [doi: 10.5863/1551-6776-24.4.254] [Medline: 31337988]
46. Fairburn CG, Patel V. The impact of digital technology on psychological treatments and their dissemination. Behav Res Ther 2017 Jan;88:19-25 [FREE Full text] [doi: 10.1016/j.brat.2016.08.012] [Medline: 28110672]
47. Berg M, Rozental A, de Brun Mangs J, Näsman M, Strömberg K, Viberg L, et al. The role of learning support and chat-sessions in guided internet-based cognitive behavioral therapy for adolescents with anxiety: a factorial design study. Front Psychiatry 2020 Jun 10;11:503 [FREE Full text] [doi: 10.3389/fpsyt.2020.00503] [Medline: 32587533]
48. Bevan Jones R, Thapar A, Rice F, Mars B, Agha SS, Smith D, et al. A digital intervention for adolescent depression (MoodHwb): mixed methods feasibility evaluation. JMIR Ment Health 2020 Jul 17;7(7):e14536 [FREE Full text] [doi: 10.2196/14536] [Medline: 32384053]

49. Calear AL, Christensen H, Mackinnon A, Griffiths KM, O'Kearney R. The YouthMood Project: a cluster randomized controlled trial of an online cognitive behavioral program with adolescents. J Consult Clin Psychol 2009 Dec;77(6):1021-1032. [doi: 10.1037/a0017391] [Medline: 19968379]
50. Calear AL, Christensen H, Mackinnon A, Griffiths KM. Adherence to the MoodGYM program: outcomes and predictors for an adolescent school-based population. J Affect Disord 2013 May;147(1-3):338-344. [doi: 10.1016/j.jad.2012.11.036] [Medline: 23245469]
51. Clarke G, Kelleher C, Hornbrook M, Debar L, Dickerson J, Gullion C. Randomized effectiveness trial of an internet, pure self-help, cognitive behavioral intervention for depressive symptoms in young adults. Cogn Behav Ther 2009;38(4):222-234 [FREE Full text] [doi: 10.1080/16506070802675353] [Medline: 19440896]
52. 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]
53. Ip P, Chim D, Chan KL, Li TM, Ho FK, Van Voorhees BW, et al. Effectiveness of a culturally attuned internet-based depression prevention program for Chinese adolescents: a randomized controlled trial. Depress Anxiety 2016 Dec;33(12):1123-1131. [doi: 10.1002/da.22554] [Medline: 27618799]
54. Jaycox LH, Ayer L, Vona P, Hehman C, Stein BD, Mahmud A, et al. Development and preliminary evaluation of a self-guided, internet-based tool for coping with stress and trauma: Life Improvement for Teens (LIFT). Psychol Serv 2019 Feb;16(1):85-94. [doi: 10.1037/ser0000277] [Medline: 30431308]
55. Kuosmanen T, Fleming TM, Newell J, Barry MM. A pilot evaluation of the SPARX-R gaming intervention for preventing depression and improving wellbeing among adolescents in alternative education. Internet Interv 2017 Jun;8:40-47 [FREE Full text] [doi: 10.1016/j.invent.2017.03.004] [Medline: 30135827]
56. Lenhard F, Andersson E, Mataix-Cols D, Rück C, Vigerland S, Högström J, et al. Therapist-guided, internet-delivered cognitive-behavioral therapy for adolescents with obsessive-compulsive disorder: a randomized controlled trial. J Am Acad Child Adolesc Psychiatry 2017 Jan;56(1):10-9.e2 [FREE Full text] [doi: 10.1016/j.jaac.2016.09.515] [Medline: 27993223]
57. Lillevoll KR, Vangberg HC, Griffiths KM, Waterloo K, Eisemann MR. Uptake and adherence of a self-directed internet-based mental health intervention with tailored e-mail reminders in senior high schools in Norway. BMC Psychiatry 2014 Jan 21;14:14 [FREE Full text] [doi: 10.1186/1471-244X-14-14] [Medline: 24443820]
58. Lucassen MF, Stasiak K, Fleming T, Frampton C, Perry Y, Shepherd M, et al. Computerized cognitive behavioural therapy for gender minority adolescents: analysis of the real-world implementation of SPARX in New Zealand. Aust N Z J Psychiatry 2021 Sep 08;55(9):874-882 [FREE Full text] [doi: 10.1177/0004867420976846] [Medline: 33287554]
59. Melnyk B, Amaya M, Szalacha L, Hoying J, Taylor T, Bowersox K. Feasibility, acceptability, and preliminary effects of the COPE online cognitive-behavioral skill-building program on mental health outcomes and academic performance in freshmen college students: a randomized controlled pilot study. J Child Adolesc Psychiatr Nurs 2015 Aug;28(3):147-154 [FREE Full text] [doi: 10.1111/jcap.12119] [Medline: 26268362]
60. Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MF. The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. BMJ 2012 Apr 18;344(apr18 3):e2598 [FREE Full text] [doi: 10.1136/bmj.e2598] [Medline: 22517917]
61. O'Connor K, Bagnell A, McGrath P, Wozney L, Radomski A, Rosychuk RJ, et al. An internet-based cognitive behavioral program for adolescents with anxiety: pilot randomized controlled trial. JMIR Ment Health 2020 Jul 24;7(7):e13356 [FREE Full text] [doi: 10.2196/13356] [Medline: 32706720]
62. O'Dea B, Han J, Batterham PJ, Achilles MR, Calear AL, Werner-Seidler A, et al. A randomised controlled trial of a relationship-focussed mobile phone application for improving adolescents' mental health. J Child Psychol Psychiatry 2020 Aug 19;61(8):899-913 [FREE Full text] [doi: 10.1111/jcpp.13294] [Medline: 32683737]
63. O'Kearney R, Kang K, Christensen H, Griffiths K. A controlled trial of a school-based internet program for reducing depressive symptoms in adolescent girls. Depress Anxiety 2009;26(1):65-72. [doi: 10.1002/da.20507] [Medline: 18828141]
64. Radomski AD, Bagnell A, Curtis S, Hartling L, Newton AS. Examining the usage, user experience, and perceived impact of an internet-based cognitive behavioral therapy program for adolescents with anxiety: randomized controlled trial. JMIR Ment Health 2020 Feb 07;7(2):e15795 [FREE Full text] [doi: 10.2196/15795] [Medline: 32022692]
65. Smith P, Scott R, Eshkevari E, Jatta F, Leigh E, Harris V, et al. Computerised CBT for depressed adolescents: randomised controlled trial. Behav Res Ther 2015 Oct;73:104-110. [doi: 10.1016/j.brat.2015.07.009] [Medline: 26301756]
66. Spence SH, Donovan CL, March S, Gamble A, Anderson RE, Prosser S, et al. A randomized controlled trial of online versus clinic-based CBT for adolescent anxiety. J Consult Clin Psychol 2011 Oct;79(5):629-642. [doi: 10.1037/a0024512] [Medline: 21744945]
67. Stallard P, Richardson T, Velleman S, Attwood M. Computerized CBT (Think, Feel, Do) for depression and anxiety in children and adolescents: outcomes and feedback from a pilot randomized controlled trial. Behav Cogn Psychother 2011 May;39(3):273-284. [doi: 10.1017/S135246581000086X] [Medline: 21272393]
68. Stasiak K, Hatcher S, Frampton C, Merry SN. A pilot double blind randomized placebo controlled trial of a prototype computer-based cognitive behavioural therapy program for adolescents with symptoms of depression. Behav Cogn Psychother 2012 Dec 20;42(4):385-401. [doi: 10.1017/s1352465812001087]

69. Stjerneklar S, Hougaard E, McLellan LF, Thastum M. A randomized controlled trial examining the efficacy of an internet-based cognitive behavioral therapy program for adolescents with anxiety disorders. PLoS One 2019;14(9):e0222485 [FREE Full text] [doi: 10.1371/journal.pone.0222485] [Medline: 31532802]
70. Tillfors M, Andersson G, Ekselius L, Furmark T, Lewenhaupt S, Karlsson A, et al. A randomized trial of Internet-delivered treatment for social anxiety disorder in high school students. Cogn Behav Ther 2011;40(2):147-157. [doi: 10.1080/16506073.2011.555486] [Medline: 25155815]
71. Van Voorhees BW, Fogel J, Reinecke M, Gladstone T, Stuart S, Gollan J, et al. Randomized clinical trial of an Internet-based depression prevention program for adolescents (Project CATCH-IT) in primary care: 12-week outcomes. J Dev Behav Pediatr 2009 Feb;30(1):23-37 [FREE Full text] [doi: 10.1097/DBP.0b013e3181966c2a] [Medline: 19194326]
72. Van Voorhees B, Gladstone TR, Sobowale K, Brown CH, Aaby DA, Terrizzi DA, et al. 24-month outcomes of primary care web-based depression prevention intervention in adolescents: randomized clinical trial. J Med Internet Res 2020 Oct 28;22(10):e16802 [FREE Full text] [doi: 10.2196/16802] [Medline: 33112254]
73. Gladstone T, Terrizzi D, Stinson A, Nidetz J, Canel J, Ching E, et al. Effect of internet-based cognitive behavioral humanistic and interpersonal training vs. internet-based general health education on adolescent depression in primary care: a randomized clinical trial. JAMA Netw Open 2018 Nov;1(7):e184278 [FREE Full text] [doi: 10.1001/jamanetworkopen.2018.4278] [Medline: 30533601]
74. Gladstone T, Buchholz KR, Fitzgibbon M, Schiffer L, Lee M, Voorhees BW. Randomized clinical trial of an internet-based adolescent depression prevention intervention in primary care: internalizing symptom outcomes. Int J Environ Res Public Health 2020 Oct 22;17(21):7736 [FREE Full text] [doi: 10.3390/ijerph17217736] [Medline: 33105889]
75. Whittaker R, Stasiak K, McDowell H, Doherty I, Shepherd M, Chua S, et al. MEMO: an mHealth intervention to prevent the onset of depression in adolescents: a double-blind, randomised, placebo-controlled trial. J Child Psychol Psychiatry 2017 Sep 02;58(9):1014-1022. [doi: 10.1111/jcpp.12753] [Medline: 28573672]
76. Whittaker R, Merry S, Stasiak K, McDowell H, Doherty I, Shepherd M, et al. MEMO--a mobile phone depression prevention intervention for adolescents: development process and postprogram findings on acceptability from a randomized controlled trial. J Med Internet Res 2012 Jan 24;14(1):e13 [FREE Full text] [doi: 10.2196/jmir.1857] [Medline: 22278284]
77. Higgins JP, 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(oct18 2):d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
78. Zingg W, Castro-Sanchez E, Secci F, Edwards R, Drumright L, Sevdalis N, et al. Innovative tools for quality assessment: integrated quality criteria for review of multiple study designs (ICROMS). Public Health 2016 Apr;133:19-37. [doi: 10.1016/j.puhe.2015.10.012] [Medline: 26704633]
79. Beatty L, Binnion C. A systematic review of predictors of, and reasons for, adherence to online psychological interventions. Int J Behav Med 2016 Dec;23(6):776-794. [doi: 10.1007/s12529-016-9556-9] [Medline: 26957109]
80. Vigerland S, Lenhard F, Bonnert M, Lalouni M, Hedman E, Ahlen J, et al. Internet-delivered cognitive behavior therapy for children and adolescents: a systematic review and meta-analysis. Clin Psychol Rev 2016 Dec;50:1-10 [FREE Full text] [doi: 10.1016/j.cpr.2016.09.005] [Medline: 27668988]
81. Becker KD, Boustani M, Gellatly R, Chorpita BF. Forty years of engagement research in children's mental health services: multidimensional measurement and practice elements. J Clin Child Adolesc Psychol 2018 Jun 02;47(1):1-23. [doi: 10.1080/15374416.2017.1326121] [Medline: 28574780]
82. Becker KD, Lee BR, Daleiden EL, Lindsey M, Brandt NE, Chorpita BF. The common elements of engagement in children's mental health services: which elements for which outcomes? J Clin Child Adolesc Psychol 2015 Jul 23;44(1):30-43. [doi: 10.1080/15374416.2013.814543] [Medline: 23879436]
83. Ebert DD, Zarski A, Christensen H, Stikkelbroek Y, Cuijpers P, Berking M, et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a meta-analysis of randomized controlled outcome trials. PLoS One 2015;10(3):e0119895 [FREE Full text] [doi: 10.1371/journal.pone.0119895] [Medline: 25786025]
84. Technical Guidance Series. World Health Organization. URL: https://extranet.who.int/pqweb/vitro-diagnostics/technical-guidance-series [accessed 2022-02-10]
85. Robinson L, Delgadillo J, Kellett S. The dose-response effect in routinely delivered psychological therapies: a systematic review. Psychother Res 2020 Jan 20;30(1):79-96. [doi: 10.1080/10503307.2019.1566676] [Medline: 30661486]
86. Robinson L, Kellett S, Delgadillo J. Dose-response patterns in low and high intensity cognitive behavioral therapy for common mental health problems. Depress Anxiety 2020 Mar;37(3):285-294 [FREE Full text] [doi: 10.1002/da.22999] [Medline: 32027435]

Abbreviations

CBT: cognitive behavioral therapy
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
ROB: Risk of Bias

SPARX: Smart, Positive, Active, Realistic, X-factor

Edited by J Torous; submitted 28.02.22; peer-reviewed by M Yap, P Shah, R Rondina, S El kefi; comments to author 19.04.22; revised version received 07.06.22; accepted 16.06.22; published 17.08.22.

Please cite as:
Li SH, Achilles MR, Werner-Seidler A, Beames JR, Subotic-Kerry M, O'Dea B
Appropriate Use and Operationalization of Adherence to Digital Cognitive Behavioral Therapy for Depression and Anxiety in Youth: Systematic Review
JMIR Ment Health 2022;9(8):e37640
URL: https://mental.jmir.org/2022/8/e37640
doi:10.2196/37640
PMID:35976180

Review

Self-help Digital Interventions Targeted at Improving Psychological Well-being in Young People With Perceived or Clinically Diagnosed Reduced Well-being: Systematic Review

Camilla M Babbage[1*], PhD; Georgina M Jackson[2*], PhD; E Bethan Davies[1], PhD; Elena Nixon[2*], PhD

[1]National Institute for Health and Care Research (NIHR) MindTech Medtech Co-operative, Mental Health & Clinical Neurosciences, Institute of Mental Health, University of Nottingham, Nottingham, United Kingdom

[2]Institute of Mental Health, Mental Health & Clinical Neurosciences, University of Nottingham, Nottingham, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Camilla M Babbage, PhD
National Institute for Health and Care Research (NIHR) MindTech Medtech Co-operative
Mental Health & Clinical Neurosciences
Institute of Mental Health, University of Nottingham
University Of Nottingham Innovation Park
Nottingham, NG7 2TU
United Kingdom
Phone: 44 1158232438
Email: Camilla.Babbage@nottingham.ac.uk

Abstract

Background: Levels of well-being are declining, whereas rates of mental health problems remain high in young people. The World Health Organization defines mental health as not merely the absence of mental disorder but also includes social and psychological well-being as integral to positive mental health, highlighting that mental health is applicable to young people with mental health conditions *and* those without a diagnosis of a mental health condition. Reduced mental well-being have been identified in studies of young people with clinical populations, as well as in populations consisting of nonclinical young people. Self-help digital interventions can be delivered at mass at a low cost and without the need for trained input, thereby facilitating access to support for well-being. Self-help interventions are effective in young people with mental health conditions, but systematic reviews of such studies have been limited to randomized controlled trials, have not included reduced well-being as an inclusion criterion, and do not consider engagement factors such as retention.

Objective: The objective of this study was to systematically review all controlled studies of digitally delivered, self-administered interventions for young people aged 9 to 25 years, with perceived or clinically diagnosed reduced psychological well-being. Participant retention and effectiveness of the interventions were also explored.

Methods: A systematic search of the PsycInfo, EMBASE, Cochrane, Scopus, and MEDLINE databases from inception to 2021, reference searches of relevant papers, and gray literature was carried out for digitally controlled studies conducted with young people with perceived or clinically diagnosed reduced well-being, aimed at improving psychological well-being. Data were extracted to identify the effectiveness and retention rates of the interventions and the quality of the studies.

Results: Overall, 1.04% (12/1153) of studies met the inclusion criteria: 83% (10/12) of studies were randomized controlled trials and 17% (2/12) were controlled pre-post studies. Most (6/12, 50%) studies aimed to improve symptoms of depression; 3 interventions aimed at both anxiety and depressive symptoms and 2 studies aimed at improving social functioning difficulties. Owing to the high risk of bias across interventions and lack of similar outcome measures, a meta-analysis was not conducted. Retention rates across studies were regarded as good, with moderate to high retention. Overall, the findings indicated that predominantly self-administered self-help interventions improved well-being in the areas targeted by the intervention and identified additional areas of well-being that were positively affected by interventions. Few interventions supported psychological well-being that was different from those used by young people with a clinical diagnosis of mental illness or young people from neurodiverse backgrounds.

Conclusions: The findings, along with the advantages of self-help interventions, highlight the need for upscaling self-help interventions to better support vulnerable populations of young people who experience poor psychological well-being.

Trial Registration: PROSPERO CRD42019129321; https://tinyurl.com/4fb2t4fz

(JMIR Ment Health 2022;9(8):e25716) doi:10.2196/25716

KEYWORDS

digital health interventions; psychological well-being; mental well-being; mental disorders; mental health; children and young people; self-help; systematic review

Introduction

Background

Well-being is a multifaceted construct consisting of a person's physical, psychological, cognitive, social, and economic well-being [1,2]. Psychological and social well-being (termed hereon as psychological well-being) are identified as being integral to a person's health by the World Health Organization, which highlights that mental health does not merely require the absence of a mental health condition but also extends to a need to experience positive well-being, which is inclusive of those not living with a mental health condition. Poor well-being has been identified in both populations of young people living with mental health conditions and those who are not [3-6]. Indicators of psychological well-being and mental health have been found to be similar in youth populations, for example, relationships with teachers, supportive families, and engaging with school are related to good psychological well-being and mental health, and bullying, single parent families, peer problems, and parent-child arguments are indicators of poor psychological well-being and mental illness [1,2,7]. Those living with a mental health condition are more likely to experience reduced well-being compared with those not living with a mental disorder, but there is alarming evidence to suggest that nonclinical youth populations also show trends toward reduced well-being [4,8]. In light of the rising decline in psychological well-being and increasing levels of mental illness in young people [9,10], the efficacy and suitability of well-being interventions that are already available for those with mental health conditions should also be considered in young people who experience low psychological well-being without mental health conditions [11].

The calls for increased resources for young people with reduced well-being may be at odds with the current state of the health care system. Young people aged 10 to 24 years have the lowest health care coverage compared with other age groups [12], and service providers tend to face significant barriers to delivering support, including issues stemming from a lack of investment, resources, and training, resulting in decreased service-user access and increased waiting times for child mental health services [13,14]. The use of self-help interventions is recommended by health care professionals for young people with depression according to the national guidelines in the United Kingdom [15], and increased recognition of self-help is leading to its greater implementation within mental health care [16]. This follows the stepped-care approach which proposes that an individual should initially be treated with the least intensive and least expensive intervention; for example, self-help, which can increase at a rate proportional to the individual's need [17]. Considering the dearth of resources available for psychiatry, resources available for well-being would likely be even fewer [18]. Thus, self-help interventions would be appropriate for young people experiencing reduced well-being. In addition, the introduction of digital health interventions (DHIs) could overcome the current challenges experienced by health care systems worldwide by improving access to care and promoting healthier populations [14], as digitally delivered self-help enables large-scale delivery at a low cost [19]. Furthermore, DHIs help overcome other barriers experienced by young service users in accessing care services, including stigma and embarrassment [20]. Digital self-help for adolescents who prefer to be self-reliant may give them feelings of autonomy [21,22], and concerns about stigma or embarrassment could be reduced, as DHIs can be used privately and anonymously [20,23,24]. The provision of DHI support is consistent with international guidelines that recognize the value of DHI in advancing universal health care coverage [14] and supports the call for early well-being interventions for young people with mild or low-level needs who would benefit from support [25].

The recent surge in self-help DHI has resulted in an increasing number of reviews of studies assessing their effectiveness [26]. Most of these reviews have examined guided self-help interventions, which entail a degree of web-based or offline guidance from a therapist rather than purely self-help interventions. Self-help interventions, by strict definition, are delivered without the need for any professional support through text, audio, video, group, or individual exercises [27]. Newman et al [28] categorized the degree of therapist involvement into 3 levels of support. The lowest level included completely self-guided interventions, with the second level additionally including the therapist providing rationale and instructions for using the intervention. The third level of involvement entailed the greatest amount of support, with the therapist actively involved in providing therapeutic support during the intervention. Considering the level and type of support provided by professionals is important when trying to balance the accessibility, adherence, and efficacy of therapy [29]. Digital self-help interventions vary in their content, usually adapting psychological therapy into a digital format, such as computerized cognitive behavioral therapy (CBT) [30], and also include other types of interventions, such as bibliotherapy [31], serious games [32], peer-to-peer support [33], self-monitoring, and medication adherence [34], all aimed at improving well-being in young people. Current systematic reviews have been limited in terms of their scope and coverage of the level of self-help support involved, with reviews tending to focus on therapist-guided self-help interventions in young people and in young people with clinically diagnosed depression and anxiety [35-39].

The only published systematic reviews to explore the efficacy of therapist support in self-help intervention studies were by Bennett et al [16] and Grist et al [36]. The study by Bennett et al [16] reviewed both paper-based (eg, bibliotherapy) and digital self-help interventions, whereas the study by Grist et al [36] identified digitally delivered interventions only, for young people with diagnosed mental health conditions. Both reviews found improved outcomes in randomized controlled trials (RCTs) with increased levels of therapist support; however, we do not know if this would also apply to young people who may experience reduced psychological well-being without a clinical diagnosis of mental illness. With concerns about the impact of the COVID-19 pandemic on young people, the use of digital technology to support young people's well-being has been promoted [40,41]. Therefore, there could be an increase in the number of interventions available for young people experiencing reduced well-being, causing these reviews to become outdated. Considering the narrower focus on digital interventions fitting the 2 lower levels of support for self-guided interventions, as proposed by Newman et al [28], increasing the search criteria to potentially influential controlled before and after studies and to all common mental disorders in young people, could also be useful in identifying further relevant trials.

Objectives

The objective of this systematic review was to explore the efficacy of predominantly self-administered digital interventions for young people experiencing perceived reduced well-being, whether or not they are diagnosed with a common mental disorder. Through the focus on young people with perceived reduced well-being, not excluding controlled pre-post studies and the inclusion of only digital self-administered interventions, this review will expand the scope of both studies by Bennett et al [16] and Grist et al [36]. The level of retention of self-help interventions will also be reported, which has not been explored in previous studies. If effective, such interventions would not rely on professional support or clinical diagnoses, expanding their scope and enhancing access to well-being support for vulnerable young people with reduced well-being.

Methods

Overview

Studies that assessed the effectiveness and acceptability of self-help digital interventions targeting young people with reduced well-being, with or without a diagnosis of a mental disorder, were included in the search.

A protocol was created and registered on PROSPERO, the International Register of Systematic Reviews (ID: CRD42019129321). Initially, the search focused on the neurodevelopmental conditions of tic disorders and associated conditions. However, because of the lack of self-help digital intervention studies for this condition and the relevance of the literature on all common mental health conditions in youth, the protocol was revised to apply the search criteria to all common youth mental disorders, as well as to perceived poor psychological well-being in the absence of a clinical diagnosis.

Types of Studies

The search was limited to publications in English because translation was not possible. The search was intended for quantitative studies; other study types, including qualitative studies, reviews, commentaries, theses, and protocols, were excluded. The study designs were required to be controlled pre-post intervention designs to be included, but randomization or control groups were not mandatory. Studies using secondary data, in which the primary paper had already been included, were removed to prevent bias.

Types of Participants

Age

Studies were included if the sample was aged between 9 and 25 years and had a mean population age of ≥18.51 years. This ensured that all interventions were performed by young people [16].

Conditions

To be included, the sample of young people had to be identified by the authors of the study or identified as having perceived reduced well-being through self-selection. Authors of the study may have used a threshold for a well-being measure, symptom severity or rating scales for mental conditions, or diagnosis of a mental disorder indicative of reduced well-being. Studies in which well-being was not measured (ie, prevention studies) were excluded. Data from subgroups that met the aforementioned criteria were included. Patients with physical health conditions (eg, epilepsy, pain management, and asthma) were excluded.

Types of Interventions

The interventions had to meet the criterion of being a digital self-help intervention for well-being. This included both web-based and offline digital interventions.

Studies adopting a combined design of digital and nondigital interventions were excluded. No exclusion criteria were applied for the type of digital delivery, length, or number of intervention sessions.

To meet the definition of self-help, interventions should require minimal support from a therapist [27]. Using the categories proposed in the study by Newman et al [28], interventions should fit the 2 lower levels of support and be completely or predominantly self-administered. Therefore, interventions with greater than predominantly self-administered support (ie, therapeutic support from a trained professional) were excluded. Interventions with support from gatekeepers or research staff for technical assistance or overseeing the practical provision of the intervention were included, as this did not provide therapeutic support. Studies including automated emails or feedback were included; however, responses containing therapeutic feedback or guidance were excluded. Many studies on young people have included safety checks. As safety checks do not offer therapeutic support and are necessary for ethical conduct, studies were not excluded based on their use.

The aim of the intervention required improving the psychological well-being of young people, identified by an

outcome measure that pertained to a respective measure of well-being.

Types of Outcome Measures

The primary outcome of interest was psychological well-being. Studies that did not measure changes in well-being (ie, pre-post intervention) were excluded, as this would not allow the assessment of the effectiveness of the intervention.

The outcome measures used in each study were documented to explore how psychological well-being was measured. In defining *psychological well-being* this study used the 2 domains of *psychological* and *social* well-being, as they are areas of functioning greatly affected when young people have a mental health condition which presents with psychological symptoms. The study by Pollard and Lee [42] documented indicators of psychosocial domains of well-being, including aggression, anxiety, emotional problems, loneliness, psychological distress, coping, fulfillment, happiness, purpose in life, and self-esteem. Indicators of social well-being included antisocial behavior, negative life events, peer problems, troubled home relationships, prosocial behaviors, quality of life, relationships with peers, social skills, and support. For a full list of indicators, refer to the study by Pollard and Lee [42]. The secondary outcomes of the effectiveness and trial retention were also recorded.

Search Methods

Overview

The search strategy covered the following constructs: (1) young people, (2) mental health conditions, (3) modes of digital intervention and delivery, and (4) self-help (Multimedia Appendix 1). These constructs were searched using the "AND" Boolean term, meaning papers had to meet each of the above constructs to be returned by the database. Each concept included wide-reaching terms, aiming to capture as much literature as possible. Furthermore, terms were adapted depending on the relevant subject headings or Medical Subject Headings terms to include plurals, singular words, and different spellings of the words.

Electronic Searches

Five bibliographic databases (PsycInfo, Embase, Cochrane, Scopus, and MEDLINE) were electronically searched for relevant articles from the conception of the database until September 24, 2021, when the search was conducted. An example of the search strategy is presented in Multimedia Appendix 2.

Other Searches

Reference lists of relevant systematic reviews (Multimedia Appendix 3 [16,21,31,32,34,35,37,39,43-45]) were hand searched by CMB and imported into EndNote (Clarivate) with other references. The OpenGray database was searched to identify any relevant gray literature.

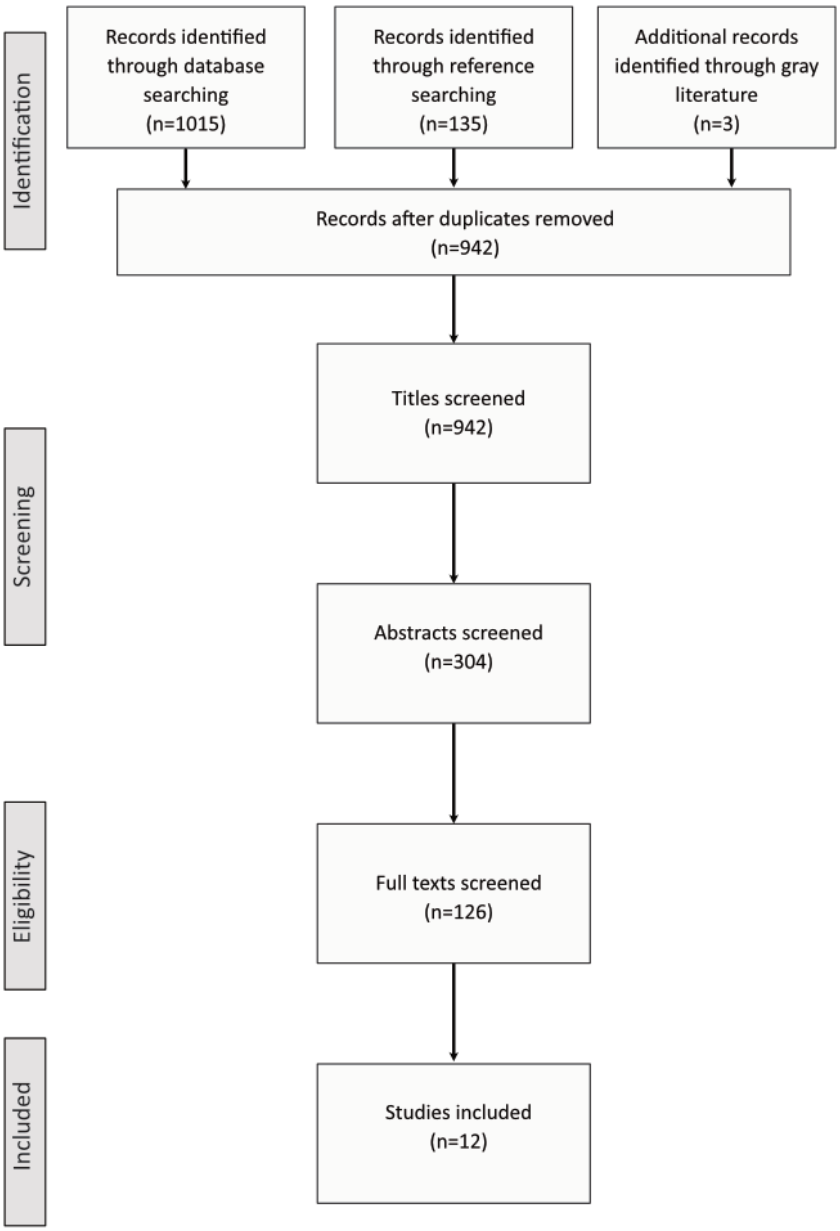
Data Synthesis

Selection of Studies

All references were imported into EndNote X9, and duplicate references were removed. CMB and EBD screened the references based on the title, excluding references that did not fit the inclusion criteria and documenting reasoning in the flowchart shown in Figure 1. This process was repeated based on abstracts and full papers. During full-paper screening, all remaining articles were reviewed by both CMB and EBD. The authors were contacted when more information was required to indicate inclusion or exclusion. Disagreements were resolved through discussions between CMB and EBD.

Full-text articles were screened according to eligibility criteria. Overall, 20.6% (26/126) of the relevant full-text articles were independently screened by the authors GMJ and EN to check for agreement. This systematic review used the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement to ensure transparent and comprehensive reporting [46].

Figure 1. Flowchart of studies in line with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines.



Data Extraction and Management

Data extraction of papers meeting the inclusion criteria was extracted into RevMan (version 5.3) by CMB. Data extracted included sample characteristics (age range, total number, gender, and inclusion criteria), trial retention (by noting the number of participants retained based on those completing outcome measures after the intervention and at follow-up and weighting this by sample size) and intervention retention (ie, the percentage of participants who completed the intervention), study design, intervention and comparison intervention details (aim, length, number of sessions, components of intervention, professional contact, and delivery method), outcomes measured, and when these were measured; key findings (including means, SDs, CIs, and *P* values) were taken from each study and placed into a summary of findings table.

Assessment of Risk of Bias

The risk of bias for each study was assessed by CMB using the Cochrane Collaboration tool [47], which covers 6 domains of bias including selection bias, performance bias, detection bias, attrition bias, reporting bias, and other biases. Disagreements were discussed with EN and GMJ and resolved by consensus. Experimenter bias was assessed as a source of bias. This tool was developed for RCT studies, but for this review, the tool was also used in controlled pre-post studies, with adaptations to the tool suggested in previous protocols [48]. CMB made judgments on the risk of bias, including a supporting quote from the text when possible, and if unknown, it should be stated. The attribution of low risk means that the bias was unlikely to have caused an effect on the findings, and high risk suggests that bias may have had a significant effect on the findings. No publications were excluded based on quality because a meta-analysis was not performed on the data. Studies with high

bias were given less weighting when applying the findings to narrative synthesis.

Results

Study Selection

In total, 1153 studies were returned from the search, of which 63 (5.46%) were duplicates and were removed. Subsequently, of 1153 studies, 942 (81.7%) were screened by title: 618 (53.6%) were excluded based on title, 178 (15.44%) were excluded based on abstract, and 114 (9.89%) were excluded based on screening of full texts, leaving 12 (1.04%) papers included in the review. The reason for exclusion at each stage was documented, and the flow diagram is shown in Figure 1.

Study Characteristics

Table 1 summarizes the study designs for all studies. In all, 10 studies were RCTs and 2 were controlled pre-post intervention studies, all published between 2006 and 2020. The sample size of young people with reduced well-being at baseline ranged from 23 to 240 (mean 107). Studies were categorized according to the primary aim of the interventions, which included improving depression, depression and anxiety, and social functioning. Half of the studies (6/12, 50%) aimed to improve symptoms of depression in young people experiencing depression and used 3 different interventions: *SPARX* [49-51], *The Journey* [52], and *MoodGYM* [53,54]. Three interventions, *Mobiletype*, *Stressbusters*, and *Shamiri-Digital* were aimed at both anxiety and depression; *Shamiri-Digital* and *Mobiletype* were explored in one study [55,56]. *Mobiletype* aimed to improve levels of anxiety, depression, and stress in young people showing elevated depression [56], and Shamiri-Digital was a universal intervention aimed at improving anxiety and depression [55]. *Stressbusters* aimed to improve anxiety and depression in young people experiencing low mood and depression, as assessed in 2 studies [57,58]. The final 2 studies explored social functioning difficulties with an intervention on blogging about social difficulties for a group of young people scoring below average on peer relationships [59] and a cognitive bias modification training (CBMT) intervention for participants scoring above the cutoff for social phobia [60]. The studies were conducted in Israel (n=1), New Zealand (n=3), Australia (n=2), the Netherlands (n=2), Norway (n=1), Kenya (n=1), and the United Kingdom (n=2).

Table 1. Summary of included studies, including sample age, study design, outcomes and key findings, and average age calculated as mean weighted by sample size.

Aim and authors; intervention name	Sample		Intervention components; control components	Delivery method; design	Key findings
	Participants (males); age range (mean)	Inclusion criteria			
Depression					
Merry et al [50]; SPARX[a]	N=187 (64); 12-19 (15.6)	>30 on CDRS-R[b]	CBT[c], psychoeducation, relaxation skills, problem-solving, activity scheduling, challenging and replacing negative thinking, and social skills (n=104); treatment as usual (n=83)	PC; RCT[d]	• SPARX group showed improvements on the CDRS-R (Cohen d=−0.293; *P*=.08) and significant improvements on MFQ[e] [61] (*P*=.03), hopelessness (K-10; *P*=.04), and anxiety (SCAS[f] 0.075); maintained at follow-up • SPARX would be recommended and felt it had appeal
Fleming et al [49]; SPARX	N=32 (18); 13-16 (14.9)	>30 on CDRS-R	CBT, psychoeducation, relaxation skills, problem-solving, activity scheduling, challenging and replacing negative thinking, and social skills (n=20); WL[g] (n=12)	PC; RCT	• SPARX showed significant improvements for depression (CDRS-R: F=18.11; *P*<.001; RADS[h]: F=4.13; *P*=.05) • SPARX may be effective in treating students in special education services with minimal symptoms of depression and anxiety for depression • No differences found at follow-up
Poppelaars et al [51]; SPARX	N=208 (0); 11-16 (13.4)	>59 on RCADS[i]	CBT, psychoeducation, relaxation skills, problem-solving, activity scheduling, challenging and replacing negative thinking, and social skills; CBT program (n=50); SPARX+CBT program (n=56); MC[j] (n=51)	PC; RCT	• Significant reductions in RADS across groups with a medium effect to the 1-year follow-up • No significant differences between groups for SPARX (Cohen d=−0.283; *P*=.11), CBT, or control
Stasiak et al [52]; The Journey	N=34 (20); 13-18 (15.2)	>30 on CDRS-R or >76 on RADS-2	CBT, linking thoughts and actions to feelings, behavioral activation, pleasant activity scheduling, problem-solving and conflict resolution, cognitive restructuring, challenging unhelpful thoughts, thought stopping, relaxation techniques and, relapse prevention (n=17); AC[k] (n=17)	PC; RCT	• Significant improvement for The Journey depression ratings (CDRS-R: Cohen d=−0.532; *P*=.001) and problem-solving (Adolescent Coping Scale-Short [62]) • Nonsignificant reductions found in the RADS for depression and quality of life (Pediatric Quality of Life Enjoyment and Satisfaction Questionnaire [63]) • Suggests short-term efficacy and good adherence • Secondary measures rarely met significance
O'Kearney et al [54]; MoodGYM	N=23/78[l] (23); 15-16 (—[m])	>16 on the CES-D[n]	CBT, Information, relaxation, problem-solving, dysfunctional thoughts, negative thinking, assesses self-esteem, cognitive restructuring, assertiveness, and coping with relationships (n=23); AC (n=24)	Web; pre-post	• Reduction of depression vulnerability after treatment for participants at risk compared with control group (CES-D: Cohen d=0.042; *P*=.87 and The Revised Children's Attributional Style Questionnaire) • Reductions maintained at follow-up compared with before the intervention • Small sample size, so caution is noted

Aim and authors; intervention name	Sample		Intervention components; control components	Delivery method; design	Key findings
	Participants (males); age range (mean)	Inclusion criteria			
Lillevoll et al [53]; MoodGYM	N=198/1337 (—); 15-20 (16.8)[l]	>16 on the CES-D	CBT, Information, relaxation, problem-solving, dysfunctional thoughts, negative thinking, assesses self-esteem, cognitive restructuring, assertiveness, and coping with relationships; MoodGYM without reminders (n=176); MoodGYM with reminders (n=176); MoodGYM with tailored reminders (n=175); WL (n=180)	Web; RCT	• Participants with elevated depression (CES-D) did not show increased self-esteem (Norwegian version of the General Self-Efficacy Scale [64]) or reduced risk of depression (CES-D: *P*=.36) • High attrition and adherence problems
Anxiety and depression					
Reid et al [56]; Mobile-type	N=118 (32); 14-24 (18.1)	Mild or severe mental health indicated by a general practitioner or >16 on The Kazdin hopelessness scale for children [65].	Prompted to complete an entry on current activity; a beep emitted from the phone at random intervals with a reminder beep 5 minutes later; stress and mood 4 times a day; alcohol, drug use, sleep, diet, and exercise once a day (n=69); AC (n=49)	Mobile; RCT	• The intervention group showed a main effect of time on the Emotional Self-Awareness Scale compared with attentional control. A significant effect of time was found at 6-week follow-up. • The sample as a whole showed a decrease in mood-related scores for the Depression Anxiety Stress Scale [66], which may have been because of 91% of the sample receiving a pathways to care before the intervention in their pretest medical review.
Wright et al [57]; Stressbusters	N=91 (31); 12-18 (15.4)	>20 on the MFQ	Goal-setting, getting activated, emotional recognition, noticing thoughts, thought challenging, problem-solving, improving social skills, and relapse prevention (n=46); AC (n=45)	PC; RCT	• Nonsignificant MFQ reduction scores for Stressbusters (Cohen d=−0.172; nonsignificant), which plateaued after the first 4 sessions • Nonsignificant increase in MFQ seen in the control group • No significant differences at the 4-month follow-up (Beck Depression Inventory, SCAS, or MFQ)
Smith et al [58]; Stressbusters	N=112 (—); 12-16 (—)	>20 on the MFQ-C	Goal-setting, getting activated, emotional recognition, noticing thoughts, thought challenging, problem-solving, improving social skills, and relapse prevention (n=55) WL (n=57)	PC; RCT	• Stressbusters showed significant decrease in MFQ (Cohen d=0.097; *P*=.001) and Screening for Child Anxiety Related Emotional Disorders [67] for self-rated and parent-rated scores, compared with the control • Attainment was significantly improved for Stressbusters
Osborn et al [55]; Shamiri-Digital	N=56/103; N=49/103; 13-18 (—)	>10 on the Patient Health Questionnaire-8 [68]; >10 on Generalized Anxiety Disorder-7	Growth mindset, gratitude, and value or virtue affirmation (n=28; n=24); AC (n=28; n=25)	PC; RCT	• A significant time×condition effect was found, suggesting greater reductions in those with depressive symptoms from baseline to follow-up (2 weeks) than those in the control group. A significant time effect was found for those with elevated anxiety; showed declines, regardless of group • For both the elevated depression and anxiety group, changes surpassed the reliable-change index suggesting they met the standard for clinically reliable change
Social functioning					

Aim and authors; intervention name	Sample		Intervention components; control components	Delivery method; design	Key findings
	Participants (males); age range (mean)	Inclusion criteria			
Boniel-Nissim and Barak [59]; Blogging	N=161 (37); 14-17 (15.5)	Scored lower on SD of index of peer relationship; interested in starting a blog	Blogging about social difficulties in open response blog; blogging about social difficulties (closed responses; n=27); blogging about social difficulties (open responses; n=26); blogging about general subjects (open responses; n=28); blogging about general subjects (closed responses; n=27); writing a private diary about social difficulties (n=26); MC (n=27)	Web; pre-post	• Blogging about social-emotional difficulties improved writer's social-emotional condition (ratings by independent judges on the Judgment of social-emotional condition) • Improvements in the Rosenberg Self-Esteem Scale [69] (Cohen d=0.211; *P*=.001), Interpersonal Activities Checklist, and Index of Peer Relationship [70] compared with other blogging groups • Social-emotional difficulty blogs open to public responses had improved judge-rated outcomes • Findings remained stable at the 2-month follow-up
Sportel et al [60]; CBMT[o]	N=240 (66); 13-15 (14.1)	>10 RCADS social phobia; (girls)>9 (boys)	Attentional bias modification tasks, strengthening the association between social-evaluative situations and positive outcomes, enhancing implicit self-esteem (n=86); AC (n=84); MC (n=70)	Web; RCT	• CBMT showed greatest improvements (Single Target Implicit Association Test, Adolescent Interpretation and Belief Questionnaire) • CBT and CBMT showed significant improvements after the test for social RCADS (Cohen d=0.051; *P*=.001) and test anxiety (Spielberger Test Anxiety Inventory [71]), significantly stronger for the CBT group • Follow-up scores suggest effects remained at 12 months, with Cognitive Bias Modification showing lower negative associations than other groups

[a]SPARX: Smart, Positive, Active, Realistic, X-factor thoughts [50].

[b]CDRS-R: The Children's Depression Rating Scale-revised [72].

[c]CBT: cognitive behavioral therapy.

[d]RCT: randomized controlled trial.

[e]MFQ: Moods and Feelings Questionnaire [61].

[f]SCAS: Spence Children's Anxiety Scale [73].

[g]WL: wait-list.

[h]RADS: Reynolds Adolescent Depression Scale [74].

[i]RCADS: Revised Child Anxiety and Depression Scale [75].

[j]MC: monitoring control.

[k]AC: attentional control.

[l]Subgroup analysis: numbers are presented so that one can see how many of the same were included in the subgroup analysis.

[m]Missing data.

[n]CES-D: Centre for Epidemiological Studies Depression Scale [76].

[o]CBMT: cognitive bias modification training [60].

Descriptions of Interventions

In total, this review found papers on 8 different interventions (*SPARX*, *The Journey*, *MoodGYM*, *Mobiletype*, *Shamiri-Digital*, *Stressbusters*, blogging, and CBMT) delivered by mobile phones, the internet, and computers; see Table 1 for an overview of the interventions and their components. Of the 3 interventions that were freely available, 2 (67%) could be used worldwide (blogging and *Shamiri-Digital*), and 1 (33%) was restricted to those living in the country where it was developed (*SPARX*). A more detailed description of the content of the interventions is provided in Multimedia Appendix 4 [49,50,52-60]. One of the interventions required an annual fee (not required by participants in this study; *MoodGYM*), and 4 of the interventions were available only through research institutes (*Mobiletype*, *Stressbusters*, CBMT, and *The Journey*). Professional contact during these self-help interventions was mostly classroom supervision, whereas young people completed the task and site visits to check safety concerns. The length of interventions varied from 1 to 28 sessions, with the duration of sessions

ranging between 3 and 60 minutes. All studies had a follow-up, with the shortest being 4 weeks and the longest at 12 months. All the studies had a control intervention, as shown in Table 1.

Outcome Measures

Outcome measures, both primary and associated, across interventions were highly heterogenous, with a variety of scales used to assess clinical symptoms or psychological and social well-being. The most commonly used outcome measures across the included studies reflected 2 mental health conditions: anxiety and depression. The most commonly used scale was the Revised Child Anxiety and Depression Scale, which was included across 4 of the 6 studies aimed at improving depression [49-52]. A total of 3 studies used the Mood and Feelings Questionnaire [50,57,58], the Children's Depression Rating Scale-revised [49,50,54], and Spence Children's Anxiety Scale [49,50,57], and 2 studies used the Center for Epidemiological Studies Depression Scale [53,54] across interventions aimed at depression and anxiety and depression. Furthermore, outside of measures of anxiety and depression, the Pediatric Quality of Life Enjoyment and Satisfaction Questionnaire, a measure of quality of life, was used by 3 interventions for improving depression [49,50,52]. The Rosenberg Self-Esteem Scale, a measure of self-esteem, was used in 3 separate studies for interventions aimed at improving anxiety and depression and social functioning [53,54,59], and the Adolescent Coping Scale-Short to rate coping was used in interventions for both depression and anxiety interventions [52,56]. In total, there were 36 different types of outcome measures, 28 of which were used only by individual studies.

Owing to a few studies using the same or similar outcome measures (ie, across all interventions, 3 studies had no depression outcomes and those that did measure depression used a variety of scales), 2 studies reporting on the same data set, skewed data, and moderate to high levels of potential bias across the studies, it was decided that conducting a meta-analysis would not reflect the data appropriately, and thus, a meta-analysis was not undertaken.

Characteristics of Participants

In total there were 1460 participants, ranging from 11 to 24 years, with a mean average of 15.4 years (for a given average age, weighted by sample size). On starting the intervention, a clinical threshold on a questionnaire for anxiety or mood disorder was used by nearly all the studies other than the study by Boniel-Nissim and Barak [59], who used an index of peer relationships, that is, a well-being measure rather than a clinical measure. Subgroup analyses from 3 studies with school-based populations showed elevated scores for depression symptoms [53,54] and one with elevated anxiety and depression [55].

Risk of Bias Within and Between Studies

Of the 7 standard quality criteria, 7 studies met 4 or more of the 7 criteria, deeming them to have moderate to low risk of bias [50,52,53,55,56,59,60]. No studies met 6 or more standard quality criteria. Of the remaining studies meeting less than half of the criteria, 4 studies met 3 of the criteria [49,51,57,58], and the study by O'Kearney et al [54] only met 1 of the criteria. Most studies (10/12, 83%) applied adequate random sequence allocation, which is expected to create comparable groups through randomization that could be replicated. Fewer studies (6/12, 50%) applied group allocation without any prior knowledge of the participants, which increased the risk of groups being selected for or against using an intervention. There was a similarly high risk of performance bias, where participants were not blind to the intervention (7/12, 58%). Less than half of the studies used blinded outcome assessors (7/12, 58%), leading to potential detection bias. A total of 3 studies published a protocol that was available on the web before publishing their papers; therefore, there was insufficient information to detect reporting bias. Attrition bias was of higher quality across studies, with most studies reporting the flow of participants throughout the process (10/12, 83%). In *other biases* the highest risk was found, as almost all the studies displayed experimenter bias, whereby the intervention developer was the author of the study (11/12, 92%).

Trial retention varied from a very high score of 100% to a very low score of 10.4% (Table 2). Similar variations were seen in intervention retention scores, from 100% of the sample reported to have completed the full intervention to 0.04% completing all modules of the intervention. Outlier analysis indicated that the very low scores obtained in the study by Lillevoll et al [53] were outliers in an outlier analysis based on postintervention trial retention (95% CI 53-115). Not including the study by Lillevoll et al [53], across studies of a total of 1256 participants, 1071 were retained giving an average trial retention of 85.2%. At follow-up, for the studies who provided the information, the range of retention varied from 57.9% to 100%. Of the 3 school-based populations using a universal intervention, with pooled subgroup analysis, 2 (67%) showed trial and intervention retention rates at the lower end of the scale [53,54], whereas 1 (33%) showed the highest trial retention, retaining all young people who had elevated depression throughout the follow-up [55]. The CBMT also showed low levels of intervention retention [60]. One study did not provide sufficient details to calculate intervention retention [59]. Looking at trial retention compared with intervention retention within trials, the rates had an average difference of 12.4 %.

Table 2. A table of the numbers and percentages of participants enrolled in the studies at baseline, poststudy, and follow-up to display rates of retention, in descending order, with an amount of professional contact. Percentages for participants retained were calculated by taking the total number of participants at baseline and dividing it by the participants at poststudy or follow-up. These scores were used to calculate the overall mean and median percentages.

Author	Professional contact	Participants at baseline, n	Trial retention, n (%)		Intervention retention (%)
			After the intervention	Follow-up	
Osborn et al [55][a]	Shamiri-digital was completed in classrooms over 1 session, where teachers supervised.	56	56 (100)	56 (100)	100
Fleming et al [49]	During class time under minimal supervision from school staff. Weekly check-ins from researcher for safety checks and practical support.	32	31 (96.8)	25 (78.1)	69
Poppelaars et al [51]	Completed at home.	208	201 (96.6)	159 (76.5)	93 (self-report)
Merry et al [50]	Safety checks at all time points plus an additional check at 1 month.	187	170 (90.9)	168 (89.8)	60
Smith et al [58]	Study information given by a clinical psychologist before enrollment. Stressbusters completed at school in an assigned room.	112	100 (89.2)	99 (88.4)	86
Stasiak et al [52]	School counselors instructed to give practical support. Therapeutic support provided only if requested from young people. Completed in the counselor's office.	34	30 (88.2)	25 (73.5)	82
Sportel et al [60]	Weekly emails with links to complete cognitive bias modification training and a reminder if not completed.	240	200 (83)	139 (57.9)	42
Boniel-Nissim and Barak [59]	Email checks conducted for diary entries.	161	124 (78)	—[b]	—
Reid et al [56]	High-risk alert activated by psychologist if a young person indicated suicide or self-harm; young person's community team informed.	112	87 (77.6)	86 (76.7)	52.9
O'Kearney et al [54][a]	Completed at school during tutor period under teacher supervision. Researchers attended the first session to check successful log-in.	23	17 (73.9)	18 (72)	40 (half or more modules)
Wright et al [57]	Practical support to accessing Stressbusters from researchers.	91	55 (60.4)	No follow-up data	70
Lillevoll et al [53][a]	Automated email reminders of level of MoodGYM use (tailored email group not included).	198	19 (10.4)	No follow-up data	0.04

[a]Studies using subgroup analysis data only and not the original sample number.

[b]Data not available.

Self-help Interventions for Depression or Depressive Symptoms

Of the 3 interventions for depression, *SPARX* and *The Journey* were delivered by an offline computer program, and *MoodGYM* was accessed through the internet. All the interventions consisted of key components of CBT, which included cognitive restructuring, problem-solving, relaxation training, social and communication training, and practice. Other components involved in the interventions included psychoeducation (*SPARX* and *MoodGYM*) or explicit homework tasks to reflect on learned skills and apply them to real life (*SPARX*).

Self-help interventions aimed at depressive symptoms for young people showing that symptoms of depression were effective when using fantasy-world games: *The Journey* and *SPARX* [49-52]. These programs were found to reduce depression after the intervention, and improvements were maintained at the 3-month [50] and 1-year follow-up [51]. Furthermore, of these 4 studies, at least 88% of their participants completed postintervention outcome measures and 73% completed follow-up.

Other self-help intervention studies aimed at reducing depressive symptoms have used *MoodGYM*, a traditional linear-style approach of computerized CBT [53,54]. O'Kearney et al [54] found risk and vulnerability reduction for depression after the intervention and at follow-up. These findings were not supported by the study by Lillevoll et al [53], whose sample had high attrition rates of nearly 70% and reported no significant changes, likely because of loss of power. Although the study by O'Kearney et al [54] did show trends toward improvement in depressive symptoms, both studies had small final sample sizes and showed a high risk of bias.

Self-help Interventions for Depression and Anxiety

Three interventions were found for anxiety and depression: *Mobiletype* [56], *Shamiri-Digital* [55], and *Stressbusters* [57,58]. *Mobiletype* is delivered through a mobile phone using an app that self-monitors to indicate to general practitioners and patients whether there is a need for greater intervention. The results suggest that the use of self-monitoring for young people with depression and anxiety symptoms improves emotional self-awareness. These results are promising given that the control intervention was an attentional control similar to the intervention, and the study scored highly for a low risk of bias. Furthermore, retention rates in the trial were good considering that this intervention was relatively demanding, requiring participants to make entries 5 times a day over 2 to 4 weeks, with results being maintained at the 6-week follow-up.

Shamiri-Digital is delivered through a web browser via a PC and includes a single component which aims to be brief and requires minimal training to deliver. The intervention involves components related to well-being rather than clinical components, such as growth mindset, gratitude, and value affirmations. The results are promising as the attentional control mirrored the structure and was similar to the Shamiri intervention but focused on note-taking and effective study. In addition, retention in the trial was very successful, as all participants completed the trial and a 2-week follow-up. This study showed a fairly low risk of bias, with a high risk only seen in selective reporting and the use of intervention developers evaluating the program.

Stressbusters is a CBT-based offline computer intervention that uses a traditional linear-style approach to teach how to recognize and challenge emotions and thoughts. A total of 2 studies showed a reduction in scores for anxiety and depression after the intervention and at the 4- and 6-month follow-up; however, in the study by Wright et al [57], this was not significant compared with the control group, who also showed improvement. Furthermore, there was nearly a 30% difference between the rates of attrition for both studies with the study by Wright et al [57] having 60.4% remaining in their sample at follow-up, compared with 89.2% for the study by Smith et al [58]. Finally, both studies scored poorly with a high or unknown risk of bias across the 4 criteria for each study, in particular, relating to allocation concealment, blinding of participants and researchers, and a lack of independent researchers conducting the research.

Social Functioning

Boniel-Nissim and Barak [59] found improved outcomes when blogging about social difficulties and greater improvements when the blog was open to peers to respond to measures of social-emotional distress and social behavior. These findings were retained at the follow-up and showed a low risk of bias. The study by Sportel et al [60] used CBMT to modify the negative interpretations of people with social anxiety. This intervention reduced the negative associations for up to 12 months after the intervention. Although the CBT control group showed greater improvements than the CBMT group, the CBMT group showed greater improvements compared with the no intervention control group. A low risk of bias was also associated with this intervention.

Well-being Outcomes Not Related to the Initial Aim of the Intervention

Across studies, the reported effects were not always limited to those associated with the main aim of the intervention. Of the 2 studies exploring *SPARX* for depression, one study found improvements in scales for hopelessness and anxiety, alongside improvements in depression scales [50]. *The Journey* succeeded in improving depression scores while also showing improvements in problem-solving skills for coping [52]. As anticipated, one of the *MoodGYM* studies showed patterns toward improvements in depression and vulnerability to depression but also found improved self-esteem and attitudes toward depression, although none of these reached significance [54], and this was not matched by the other *MoodGYM* study [53]. In *Mobiletype* users, although a significant increase in depression and anxiety was not found, a general trend toward improvement was noted, and a significant increase in emotional self-awareness compared with control groups was shown [53]. In the study by Smith et al [58] on *Stressbusters*, a significant increase in school attendance was also seen alongside improvements in anxiety and depression scores. Blogging about social difficulties improved outcomes in social dimensions such as social-emotional distress and engagement in social behaviors and self-esteem [59]. The CBMT group showed improvements in implicit associations and social interpretations over the other groups. Both the CBMT and CBT groups showed reductions in social phobia and test anxiety [60].

Discussion

Summary of Evidence

This review examined the effectiveness of self-help interventions aimed at improving psychological well-being in young people with low psychological well-being, identified through self-selection of perceived poor well-being, clinical diagnosis of anxiety or mood disorders, or meeting a threshold suggestive of mental health symptoms. A total of 12 studies met the search criteria, 10 (83%) of which were RCTs and 2 (17%) were controlled pre-post studies. Most interventions [49-52,56-58] were aimed at young people who showed elevated symptoms indicative of depressive or mood disorders on a symptom rating scale, including interventions aimed at both anxiety and depression. One study focused on those displaying elevated symptoms of anxiety using anxiety symptom rating scales [60], and the other included young people who scored below average within the sample on a social functioning scale [59]. This was the only study to use a measure of well-being which was not also a clinical measure, that is, depression or anxiety. A total of 3 studies were universal interventions for school-based populations, with subgroup analysis of young people showing elevated depression [53,54] and anxiety and depression [55]. No studies included young people who had self-perceived reduced well-being or a diagnosis of a mental health condition as inclusion criteria. Overall, the narrative evaluation indicated that predominately or entirely self-administered self-help

interventions improved psychological well-being in the areas intended for intervention.

The results also highlighted that such interventions can lead to improvements in other areas associated with psychological well-being, such as self-esteem, emotional self-awareness, and problem-solving skills, which had not been the primary outcome of the intervention. The studies indicated reasonable levels of trial retention, with over three-fourths of participants being retained after the intervention for 9 of the 12 studies and at follow-up for half of the studies. Furthermore, the intervention retention levels were promising, as they suggested that young people completed over half of the intervention. Although the relationship between trial retention (ie, completing all outcome measures of a trial) and intervention retention (ie, completing all intervention modules) is not the same, the results seem to support the notion that intervention and trial retention follow a similar relationship in terms of indicating the level of young people's engagement [77].

Most of the included studies provided sufficient details to calculate trial (12/12, 100%) and intervention retention rates (11/12, 92%). Often, these data were not readily available, and sometimes this was not available at all for follow-up data (11/12, 92%), even though participant flowcharts are recommended for transparent reporting of RCTs [78]. For example, a participant may be regarded as a "completer," because they completed all the questionnaires in the study, but this does not ascertain whether they completed all the modules or active time spent in the intervention. Trial settings are believed to consist of *push factors* that influence attrition rates, with more *open* or pragmatic trials tending to have greater dropout [77]. A sample in the study by Lillevoll et al [53] showed notably reduced adherence compared with other studies, possibly as a result of running a trial in a more naturalistic setting. Therefore, it is highly recommended that future interventions, studies, and systematic reviews provide details on both intervention and trial retention. Nonetheless, the findings presented here suggest that young people engage well in self-help intervention trials.

The greatest promise for self-help digital interventions seemed to be CBT-based fantasy-style games (eg, *SPARX* and *The Journey)* for reducing symptoms of depression, which demonstrated good trial retention and follow-up rates, whereas traditional linear-style interventions showed greater variation in improvement (eg, *MoodGYM)*. Interventions used by young people with anxiety and depressive symptomatology showed general improvements in well-being, with mood monitoring (*Mobiletype*) leading to the maintenance of increased emotional self-awareness at follow-up. This was similar for linear-style interventions with one of the *Stressbusters* studies and the *Shamiri-Digital* trial, reaching significance for reductions in depressive symptoms. Finally, interventions aimed at improving social functioning were effective after the intervention and remained significantly so at follow-up for the blogging group. Some of the seemingly less complex interventions (*Shamiri-Digital* and blogging) are freely available and open source, which could present interventions that are truly accessible to young people, raising an important point about the need for available and effective interventions. Previous reviews have suggested that the use of guidance within self-help interventions improves the magnitude of the effect [16,79], but these findings suggest that stand-alone interventions can support young people with elevated symptoms or social functioning difficulties without the need for therapist assistance. Self-administered therapies have major advantages to guided self-help because of their scalability, although not necessitating further demands from already overburdened services, including cost, location, and trained professionals [80,81].

A meta-analysis of the studies included in this report was not deemed appropriate because of a lack of high-quality studies, as nearly half of the papers met less than half of the quality criteria required to protect against biases. Notably, 92% (11/12) of studies demonstrated experimenter bias, and only 8% (1/12) of studies were identified as having a low risk of bias for selective reporting. Conducting a meta-analysis with biased data can lead to misleading conclusions [82]. For example, inappropriate concealment of participant allocation or selection bias, as found in more than half of the studies included in this review, can lead to 30% overestimation of the treatment effect [83]. Although there is a need for more high-quality assessments of self-help interventions, the time frame around running an empirical study against the delivery time of commercial interventions makes this difficult. As of March 2021, there are 53,979 and 53,054 medical health apps available in the Apple App Store and Google Play, respectively, which have shown increases of 4.86% and 6.51%, respectively, since the last quarter of 2020 [84,85]. In light of the long time it takes to complete and publish RCTs, it is unrealistic for researchers to complete controlled assessments of interventions with the unprecedented rate of growth of apps and digital technology development [34,86]. Consequently, the use of RCTs to assess research has been questioned, with iterative approaches to development proposing that RCTs should be used only to assess the overall functionality of an intervention or its theoretical basis rather than for minor intervention modifications [86,87]. For these reasons, this review recommends conducting higher quality controlled before and after studies that do not overlook the bias criteria. This suggestion also meets the National Institute for Health and Care Excellence best practice standards for self-management of DHIs, requiring high-quality studies with a comparison group [88]. One crucial method to improve quality calls for researchers outside the developing group of interventions to carry out controlled trials to minimize experimenter bias. Furthermore, the inclusion of non-RCT studies in future systematic reviews will enable the evaluation of the most recent interventions to provide a comprehensive overview of whether self-help interventions can effectively support young people with mental disorders.

Despite a broad literature search, most trials were interested in young people with clinical symptoms rather than reduced psychological well-being, which was also mirrored in the outcomes of the trials. This highlights the lack of recognition of the importance of well-being in achieving positive mental health outcomes. In contrast, self-help interventions were fairly narrowly aimed at young people with mood, anxiety, or social difficulties, and none were found for other underexplored disorders such as neurodevelopmental, obsessive-compulsive, sleep, eating, and anger conditions. Those living with such

conditions would likely benefit from self-help interventions to support well-being difficulties related to the clinical symptomology of their diagnosis or because of living with their condition [4]. As mentioned earlier, this review calls for higher quality-controlled studies on self-help interventions with populations of young people with such disorders and includes measures of psychological well-being in the sample criteria and outcome measures. Nevertheless, considering the findings of Bennett et al [16], parent-involved self-help interventions may be more useful than predominantly self-administered interventions; the inclusion of parental involvement alongside self-help interventions may overcome some of the aforementioned barriers to interventions while retaining young person engagement, compliance, and increasing the effectiveness of the intervention. This is in line with evidence that parent-involved multimedia and bibliotherapy interventions are as beneficial for behavioral disorders as therapist-led interventions in the long term [89] and that parent involvement within CBT interventions is effective for adolescent anxiety disorders [90]. This area of research warrants further investigation as the digital delivery of parent training could facilitate access to these interventions.

Limitations

This review had several limitations. A small number of studies matched the inclusion criteria, and only 1 intervention was tested with those experiencing reduced psychological well-being outside of clinically diagnosed illnesses (ie, anxiety and depression). This could be because of the specific focus on psychological and social well-being, not including outcome measures related to the cognitive, economic, and physical domains of well-being [42]. Although the authors recognize that these domains relate to and are important aspects of a young person's well-being, psychological and social functioning were felt to be the most impaired when living with perceived reduced well-being or a diagnosed mental health condition, and identifying interventions aimed at improving these aspects was a priority. By including other aspects of well-being, it is possible that more interventions would have been found. As this review only considered common childhood disorders, conditions such as psychosis were not included, although it is recognized that children with uncommon conditions would also experience reduced well-being. Therefore, these results cannot be generalized to less common disorders that were not included in this review. The included studies were mainly developed by and run in societies with a Western cultural influence, meaning that few of the findings can be generalized to other cultures. Finally, data were missing from certain studies which affected study inclusion, conclusions that could be drawn for retention, and the ability to conduct a meta-analysis. Although the authors were contacted to request information related to the eligibility criteria, a lack of response meant that they could not be included. Alongside narrowing the search to English studies on the web and including quantitative studies only, this may have reduced the pool of the data set.

Conclusions

The findings of this review support the utility of digital self-help interventions for young people with elevated symptoms of depression and anxiety. Only 2 interventions were identified for young people experiencing social functioning difficulties; therefore, it is not possible to generalize the findings to understand whether self-help interventions improve well-being in those experiencing reduced well-being. Interventions were found to be effective and had reasonable levels of retention, suggesting that they were acceptable for youth populations. The greatest promise seems to lie in fantasy-style interventions for young people who experience symptoms of depression. Nonetheless, traditional linear-style interventions were still beneficial to users with depression and anxiety symptoms and social dysfunction. The collective advantages of self-administered self-help interventions included low costs, ease of access, and reduced need for trained professionals, but simple interventions were deemed to be an especially feasible option in helping overcome barriers to accessing mental health care for youth populations, as some of these were also readily available to access. Greater efforts are warranted to improve the quality of studies, and greater consensus is required on the use of outcome measures in relation to retention and adherence, as well as different aspects of well-being, to help determine the impact of these interventions on broader well-being and in *real-world* settings. Specifically, further research in this area should focus on improving the quality of studies within the research of predominantly or entirely self-administered digital interventions for young people who show elevated symptomatology of mental health problems and reduced psychological well-being, and have mental, behavioral, and neurodevelopmental disorders. With the improved quality of studies, a meta-analysis could be performed which would provide more precise indicators of the effectiveness of self-help interventions. Research should also explore the use of alternatives to RCTs in the assessment of digital interventions to help close the gap between the progression of technology and the dissemination of empirical studies.

Acknowledgments

This work was supported by the Economic and Social Research Council (grant ES/J500100/1) and Tourettes Action, which funded CMB's Economic and Social Research Council Doctoral Training Centre Collaborative Studentship.

CMB and EBD also acknowledge the financial support of National Institute of Health and Care Research MindTech MedTech Co-operative and National Institute of Health and Care Research Nottingham Biomedical Research Centre.

Conflicts of Interest

GMJ is a nonexecutive director of Neurotherapeutics Ltd, which as part of its portfolio, develops apps. Currently, it has no apps on the market and GMJ was not involved in the development of any of the apps reviewed in the paper, nor was any member of the company.

Multimedia Appendix 1
Population, intervention, comparators, and outcomes table.
[DOCX File , 19 KB - mental_v9i8e25716_app1.docx]

Multimedia Appendix 2
Example search strategy for MEDLINE database.
[DOCX File , 15 KB - mental_v9i8e25716_app2.docx]

Multimedia Appendix 3
Reference list of hand searched systematic reviews.
[DOCX File , 15 KB - mental_v9i8e25716_app3.docx]

Multimedia Appendix 4
Detailed description of intervention content.
[DOCX File , 19 KB - mental_v9i8e25716_app4.docx]

References

1. Patalay P, Fitzsimons E. Correlates of mental illness and wellbeing in children: are they the same? Results from the UK millennium cohort study. J Am Acad Child Adolesc Psychiatry 2016 Sep;55(9):771-783 [FREE Full text] [doi: 10.1016/j.jaac.2016.05.019] [Medline: 27566118]
2. Moore GF, Cox R, Evans RE, Hallingberg B, Hawkins J, Littlecott HJ, et al. School, peer and family relationships and adolescent substance use, subjective wellbeing and mental health symptoms in Wales: a cross sectional study. Child Indic Res 2018;11(6):1951-1965 [FREE Full text] [doi: 10.1007/s12187-017-9524-1] [Medline: 30524519]
3. Mittal D, Fortney JC, Pyne JM, Edlund MJ, Wetherell JL. Impact of comorbid anxiety disorders on health-related quality of life among patients with major depressive disorder. Psychiatr Serv 2006 Dec;57(12):1731-1737. [doi: 10.1176/ps.2006.57.12.1731] [Medline: 17158487]
4. Dey M, Landolt MA, Mohler-Kuo M. Health-related quality of life among children with mental disorders: a systematic review. Qual Life Res 2012 Dec;21(10):1797-1814. [doi: 10.1007/s11136-012-0109-7] [Medline: 22298200]
5. Mental Health of Children and Young People in England: Official Statistics 2017. National Health Service Digital. 2018 Nov 22. URL: https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-of-children-and-young-people-in-england/2017/2017#key-facts [accessed 2022-01-14]
6. The Good Childhood Report 2021 – The Children's Society. The Children's Society. 2021. URL: https://www.childrenssociety.org.uk/good-childhood [accessed 2022-08-11]
7. Universal approaches to improving children and young people's mental health and wellbeing: report of the findings of a Special Interest Group. Public Health England. 2019 Oct. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/842176/SIG_report.pdf [accessed 2022-08-22]
8. Patalay P, Fitzsimons E. Mental ill-health among children of the new century: trends across childhood with a focus on age 14. Centre for Longitudinal Studies. 2017 Sep. URL: https://www.ucl.ac.uk/ioe/sites/ioe/files/patalay_fitzsimons_mental_ill-health_among_children_of_the_new_century_-_september_2017.pdf [accessed 2022-08-11]
9. Chapman KS. Adolescents: the missing population in universal health coverage. Plan International. 2019 May 31. URL: https://plan-uk.org/blogs/adolescents-the-missing-population-in-universal-health-coverage [accessed 2022-08-12]
10. Adolescent mental health: Time for action - PMNCH Knowledge Summary #36. Partnership for Maternal Newborn & Child Health. Geneva, Switzerland: World Health Organization; 2020 Jul 14. URL: https://pmnch.who.int/resources/publications/m/item/pmnch-knowledge-summary-36-adolescent-mental-health-time-for-action [accessed 2020-09-02]
11. Black L, Panayiotou M, Humphrey N. The dimensionality and latent structure of mental health difficulties and wellbeing in early adolescence. PLoS One 2019 Feb 26;14(2):e0213018 [FREE Full text] [doi: 10.1371/journal.pone.0213018] [Medline: 30807608]
12. Patton G, Sawyer S, Santelli J, Ross DA, Afifi R, Allen NB, et al. Our future: a Lancet commission on adolescent health and wellbeing. Lancet 2016 Jun 11;387(10036):2423-2478 [FREE Full text] [doi: 10.1016/S0140-6736(16)00579-1] [Medline: 27174304]

13. Breaking down barriers: the challenge of improving mental health outcomes. British Medical Association. 2017. URL: https://unitementalhealth.files.wordpress.com/2018/02/breaking-down-barriers-mental-health-briefing-apr2017.pdf [accessed 2022-08-16]
14. Recommendations on digital interventions for health system strengthening: WHO guideline. World Health Organization. 2019 Jun 6. URL: https://www.who.int/publications/i/item/9789241550505 [accessed 2022-08-12]
15. Common mental health problems: identification and pathways to care — Clinical guideline [CG123]. National Institute for Health and Care Excellence. 2011 May 25. URL: https://www.nice.org.uk/guidance/cg123 [accessed 2022-08-12]
16. Bennett SD, Cuijpers P, Ebert DD, McKenzie Smith M, Coughtrey AE, Heyman I, et al. Practitioner Review: unguided and guided self-help interventions for common mental health disorders in children and adolescents: a systematic review and meta-analysis. J Child Psychol Psychiatry 2019 Aug;60(8):828-847. [doi: 10.1111/jcpp.13010] [Medline: 30775782]
17. Rapee RM, Lyneham HJ, Wuthrich V, Chatterton ML, Hudson JL, Kangas M, et al. Comparison of stepped care delivery against a single, empirically validated cognitive-behavioral therapy program for youth with anxiety: a randomized clinical trial. J Am Acad Child Adolesc Psychiatry 2017 Oct;56(10):841-848. [doi: 10.1016/j.jaac.2017.08.001] [Medline: 28942806]
18. Cloninger C. The science of well being: an integrated approach to mental health and its disorders. Psychiatr Danub 2006 Dec;18(3-4):218-224. [Medline: 17183779]
19. Andersson G, Cuijpers P. Pros and cons of online cognitive-behavioural therapy. Br J Psychiatry 2008 Oct;193(4):270-271. [doi: 10.1192/bjp.bp.108.054080] [Medline: 18827286]
20. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: a systematic review. BMC Psychiatry 2010 Dec 30;10:113 [FREE Full text] [doi: 10.1186/1471-244X-10-113] [Medline: 21192795]
21. Pennant ME, Loucas CE, Whittington C, Creswell C, Fonagy P, Fuggle P, Expert Advisory Group. Computerised therapies for anxiety and depression in children and young people: a systematic review and meta-analysis. Behav Res Ther 2015 Apr;67:1-18. [doi: 10.1016/j.brat.2015.01.009] [Medline: 25727678]
22. Shechtman Z, Vogel DL, Strass HA, Heath PJ. Stigma in help-seeking: the case of adolescents. Br J Guid Couns 2018;46(1):104-119 [FREE Full text] [doi: 10.1080/03069885.2016.1255717]
23. Kauer SD, Mangan C, Sanci L. Do online mental health services improve help-seeking for young people? A systematic review. J Med Internet Res 2014 Mar 04;16(3):e66 [FREE Full text] [doi: 10.2196/jmir.3103] [Medline: 24594922]
24. 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 [FREE Full text] [doi: 10.3389/fpsyt.2019.00759] [Medline: 31798468]
25. Transforming children and young people's mental health provision: a green paper. Department of Health and Social care. 2018. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/728894/impact-assessment-for-tranforming-cy-mental-health-provision-green-paper.pdf [accessed 2022-08-12]
26. Rickwood D, Bradford S. The role of self-help in the treatment of mild anxiety disorders in young people: an evidence-based review. Psychol Res Behav Manag 2012;5:25-36 [FREE Full text] [doi: 10.2147/PRBM.S23357] [Medline: 22427736]
27. Bower P, Richards D, Lovell K. The clinical and cost-effectiveness of self-help treatments for anxiety and depressive disorders in primary care: a systematic review. Br J Gen Pract 2001 Oct;51(471):838-845 [FREE Full text] [Medline: 11677710]
28. Newman MG, Szkodny LE, Llera SJ, Przeworski A. A review of technology-assisted self-help and minimal contact therapies for drug and alcohol abuse and smoking addiction: is human contact necessary for therapeutic efficacy? Clin Psychol Rev 2011 Feb;31(1):178-186. [doi: 10.1016/j.cpr.2010.10.002] [Medline: 21095051]
29. Talbot F. Client contact in self-help therapy for anxiety and depression: necessary but can take a variety of forms beside therapist contact. Behav Change 2012 Jul 06;29(2):63-76 [FREE Full text] [doi: 10.1017/bec.2012.4]
30. Richardson T, Stallard P, Velleman S. Computerised cognitive behavioural therapy for the prevention and treatment of depression and anxiety in children and adolescents: a systematic review. Clin Child Fam Psychol Rev 2010 Sep;13(3):275-290. [doi: 10.1007/s10567-010-0069-9] [Medline: 20532980]
31. Ahmead M, Bower P. The effectiveness of self help technologies for emotional problems in adolescents: a systematic review. Child Adolesc Psychiatry Ment Health 2008 Jul 23;2(1):20 [FREE Full text] [doi: 10.1186/1753-2000-2-20] [Medline: 18651962]
32. Fleming TM, Cheek C, Merry SN, Thabrew H, Bridgman H, Stasiak K, et al. Serious games for the treatment or prevention of depression: a systematic review. Rev Psicopatología y Psicol Clínica 2015 Jan 15;19(3):227-242 [FREE Full text] [doi: 10.5944/rppc.vol.19.num.3.2014.13904]
33. Ali K, Farrer L, Gulliver A, Griffiths KM. Online peer-to-peer support for young people with mental health problems: a systematic review. JMIR Ment Health 2015 May 19;2(2):e19 [FREE Full text] [doi: 10.2196/mental.4418] [Medline: 26543923]
34. Grist R, Porter J, Stallard P. Mental health mobile apps for preadolescents and adolescents: a systematic review. J Med Internet Res 2017 May 25;19(5):e176 [FREE Full text] [doi: 10.2196/jmir.7332] [Medline: 28546138]

35. Ebert DD, Zarski AC, Christensen H, Stikkelbroek Y, Cuijpers P, Berking M, et al. Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a meta-analysis of randomized controlled outcome trials. PLoS One 2015 Mar 18;10(3):e0119895 [FREE Full text] [doi: 10.1371/journal.pone.0119895] [Medline: 25786025]
36. Grist R, Croker A, Denne M, Stallard P. Technology delivered interventions for depression and anxiety in children and adolescents: a systematic review and meta-analysis. Clin Child Fam Psychol Rev 2019 Jun;22(2):147-171 [FREE Full text] [doi: 10.1007/s10567-018-0271-8] [Medline: 30229343]
37. 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;58(4):474-503. [doi: 10.1111/jcpp.12663] [Medline: 27943285]
38. Rice SM, Goodall J, Hetrick SE, Parker AG, Gilbertson T, Amminger GP, et al. Online and social networking interventions for the treatment of depression in young people: a systematic review. J Med Internet Res 2014 Sep 16;16(9):e206 [FREE Full text] [doi: 10.2196/jmir.3304] [Medline: 25226790]
39. Rooksby M, Elouafkaoui P, Humphris G, Clarkson J, Freeman R. Internet-assisted delivery of cognitive behavioural therapy (CBT) for childhood anxiety: systematic review and meta-analysis. J Anxiety Disord 2015 Jan;29:83-92. [doi: 10.1016/j.janxdis.2014.11.006] [Medline: 25527900]
40. Budd J, Miller BS, Manning EM, Lampos V, Zhuang M, Edelstein M, et al. Digital technologies in the public-health response to COVID-19. Nat Med 2020 Aug;26(8):1183-1192. [doi: 10.1038/s41591-020-1011-4] [Medline: 32770165]
41. Goldschmidt K. The COVID-19 pandemic: technology use to support the wellbeing of children. J Pediatr Nurs 2020;53:88-90 [FREE Full text] [doi: 10.1016/j.pedn.2020.04.013] [Medline: 32317129]
42. Pollard EL, Lee PD. Child well-being: a systematic review of the literature. Soc Indic Res 2003;61(1):59-78. [doi: 10.1023/A:1021284215801]
43. 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]
44. Podina IR, Mogoase C, David D, Szentagotai A, Dobrean A. A meta-analysis on the efficacy of technology mediated CBT for anxious children and adolescents. J Rat-Emo Cognitive-Behav Ther 2016;34(1):31-50 [FREE Full text] [doi: 10.1007/s10942-015-0228-5]
45. Schlegl S, Bürger C, Schmidt L, Herbst N, Voderholzer U. The potential of technology-based psychological interventions for anorexia and bulimia nervosa: a systematic review and recommendations for future research. J Med Internet Res 2015 Mar 31;17(3):e85 [FREE Full text] [doi: 10.2196/jmir.3554] [Medline: 25840591]
46. Moher D, Liberati A, Tetzlaff J, Altman DG, 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]
47. Higgins JP, 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]
48. Ryan R, Hill S, Prictor M, McKenzie J. Cochrane Consumers and Communication Review Group. Study Quality Guide. Cochrane. 2013 May. URL: https://cccrg.cochrane.org/sites/cccrg.cochrane.org/files/public/uploads/StudyQualityGuide_May%202013.pdf [accessed 2022-08-12]
49. 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]
50. Merry SN, Stasiak K, Shepherd M, Frampton C, Fleming T, Lucassen MF. The effectiveness of SPARX, a computerised self help intervention for adolescents seeking help for depression: randomised controlled non-inferiority trial. BMJ 2012 Apr 18;344:e2598 [FREE Full text] [doi: 10.1136/bmj.e2598] [Medline: 22517917]
51. Poppelaars M, Tak YR, Lichtwarck-Aschoff A, Engels RC, Lobel A, Merry SN, et al. A randomized controlled trial comparing two cognitive-behavioral programs for adolescent girls with subclinical depression: a school-based program (Op Volle Kracht) and a computerized program (SPARX). Behav Res Ther 2016 May;80:33-42. [doi: 10.1016/j.brat.2016.03.005] [Medline: 27019280]
52. Stasiak K, Hatcher S, Frampton C, Merry SN. A pilot double blind randomized placebo controlled trial of a prototype computer-based cognitive behavioural therapy program for adolescents with symptoms of depression. Behav Cogn Psychother 2014 Jul;42(4):385-401. [doi: 10.1017/S1352465812001087] [Medline: 23253641]
53. Lillevoll KR, Vangberg HC, Griffiths KM, Waterloo K, Eisemann MR. Uptake and adherence of a self-directed Internet-based mental health intervention with tailored e-mail reminders in senior high schools in Norway. BMC Psychiatry 2014 Jan 21;14:14 [FREE Full text] [doi: 10.1186/1471-244X-14-14] [Medline: 24443820]
54. O'Kearney R, Gibson M, Christensen H, Griffiths KM. Effects of a cognitive-behavioural internet program on depression, vulnerability to depression and stigma in adolescent males: a school-based controlled trial. Cogn Behav Ther 2006;35(1):43-54. [doi: 10.1080/16506070500303456] [Medline: 16500776]

55. Osborn TL, Rodriguez M, Wasil AR, Venturo-Conerly KE, Gan J, Alemu RG, et al. Single-session digital intervention for adolescent depression, anxiety, and well-being: outcomes of a randomized controlled trial with Kenyan adolescents. J Consult Clin Psychol 2020 Jul;88(7):657-668. [doi: 10.1037/ccp0000505] [Medline: 32391709]
56. Reid SC, Kauer SD, Hearps SJ, Crooke AH, Khor AS, Sanci LA, et al. A mobile phone application for the assessment and management of youth mental health problems in primary care: a randomised controlled trial. BMC Fam Pract 2011 Nov 29;12:131 [FREE Full text] [doi: 10.1186/1471-2296-12-131] [Medline: 22123031]
57. Wright B, Tindall L, Littlewood E, Allgar V, Abeles P, Trépel D, et al. Computerised cognitive-behavioural therapy for depression in adolescents: feasibility results and 4-month outcomes of a UK randomised controlled trial. BMJ Open 2017 Jan 27;7(1):e012834 [FREE Full text] [doi: 10.1136/bmjopen-2016-012834] [Medline: 28132000]
58. Smith P, Scott R, Eshkevari E, Jatta F, Leigh E, Harris V, et al. Computerised CBT for depressed adolescents: randomised controlled trial. Behav Res Ther 2015 Oct;73:104-110. [doi: 10.1016/j.brat.2015.07.009] [Medline: 26301756]
59. Boniel-Nissim M, Barak A. The therapeutic value of adolescents' blogging about social-emotional difficulties. Psychol Serv 2013 Aug;10(3):333-341. [doi: 10.1037/a0026664] [Medline: 23937092]
60. Sportel BE, de Hullu E, de Jong PJ, Nauta MH. Cognitive bias modification versus CBT in reducing adolescent social anxiety: a randomized controlled trial. PLoS One 2013 May 14;8(5):e64355 [FREE Full text] [doi: 10.1371/journal.pone.0064355] [Medline: 23691203]
61. Costello EJ, Angold A. Scales to assess child and adolescent depression: checklists, screens, and nets. J Am Acad Child Adolesc Psychiatry 1988 Nov;27(6):726-737. [doi: 10.1097/00004583-198811000-00011] [Medline: 3058677]
62. Frydenberg E, Lewis R. The Adolescent Coping Scale: Second Edition (ACS-2). Australian Council for Adult Educational Research. 1993. URL: https://www.acer.org/files/adolescent-coping-scale-2-long-form-kylie-sample.pdf [accessed 2022-08-12]
63. Endicott J, Nee J, Yang R, Wohlberg C. Pediatric Quality of Life Enjoyment and Satisfaction Questionnaire (PQ-LES-Q): reliability and validity. J Am Acad Child Adolesc Psychiatry 2006 Apr;45(4):401-407. [doi: 10.1097/01.chi.0000198590.38325.81] [Medline: 16601644]
64. Røysamb E, Schwarzer R, Jerusalem M. Norwegian Version of the General Perceived Self-Efficacy Scale. Freien Universität Berlin. 1998. URL: https://userpage.fu-berlin.de/~health/norway.htm [accessed 2022-08-12]
65. Andrews G, Slade T. Interpreting scores on the Kessler Psychological Distress Scale (K10). Aust N Z J Public Health 2001 Dec;25(6):494-497. [doi: 10.1111/j.1467-842x.2001.tb00310.x] [Medline: 11824981]
66. 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]
67. Birmaher B, Khetarpal S, Brent D, Cully M, Balach L, Kaufman J, et al. The Screen for Child Anxiety Related Emotional Disorders (SCARED): scale construction and psychometric characteristics. J Am Acad Child Adolesc Psychiatry 1997 Apr;36(4):545-553. [doi: 10.1097/00004583-199704000-00018] [Medline: 9100430]
68. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatr Ann 2002 Sep;32(9):509-515. [doi: 10.3928/0048-5713-20020901-06]
69. Rosenberg M. Society and the Adolescent Self-Image. Princeton, NJ, USA: Princeton University Press; 2015.
70. Hudson WW. Clinical Measurement Package: A Field Manual. Homewood, IL, USA: Irwin Professional Pub; 1982.
71. Van der Ploeg HM. Handleiding bij de Examen/Toets attitude vragenlijst ETAV: Een nederlandstalige bewerking van de spielberger test anxiety inventory [A dutch version of the spielberger test anxiety inventory TAI]. Lisse, The Netherlands: Swets en Zeitlinger; 1988.
72. Poznanski EO, Mokros HB. Children's Depression Rating Scale, Revised (CDRS-R) Manual. Torrance, CA, USA: Western Psychological Services; 1996.
73. Spence SH, Barrett PM, Turner CM. Psychometric properties of the Spence Children's Anxiety Scale with young adolescents. J Anxiety Disord 2003;17(6):605-625. [doi: 10.1016/s0887-6185(02)00236-0] [Medline: 14624814]
74. Reynolds WM. Reynolds Adolescent Depression Scale, 2nd Ed (RADS-2): Professional Manual. Lutz: Psychological Assessment Resources; 2002.
75. Chorpita BF, Yim L, Moffitt C, Umemoto LA, Francis SE. Assessment of symptoms of DSM-IV anxiety and depression in children: a revised child anxiety and depression scale. Behav Res Ther 2000 Aug;38(8):835-855. [doi: 10.1016/s0005-7967(99)00130-8] [Medline: 10937431]
76. Radloff LS. The use of the Center for Epidemiologic Studies Depression Scale in adolescents and young adults. J Youth Adolesc 1991 Apr;20(2):149-166. [doi: 10.1007/BF01537606] [Medline: 24265004]
77. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
78. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ 2010 Mar 23;340:c869 [FREE Full text] [doi: 10.1136/bmj.c869] [Medline: 20332511]

79. Ye X, Bapuji SB, Winters SE, Struthers A, Raynard M, Metge C, et al. Effectiveness of internet-based interventions for children, youth, and young adults with anxiety and/or depression: a systematic review and meta-analysis. BMC Health Serv Res 2014 Jul 18;14:313 [FREE Full text] [doi: 10.1186/1472-6963-14-313] [Medline: 25037951]
80. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
81. Fleming TM, de Beurs D, Khazaal Y, Gaggioli A, Riva G, Botella C, et al. Maximizing the impact of e-therapy and serious gaming: time for a paradigm shift. Front Psychiatry 2016 Apr 18;7:65 [FREE Full text] [doi: 10.3389/fpsyt.2016.00065] [Medline: 27148094]
82. Deeks JJ, Higgins JP, Altman DG. Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins JP, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions. Version 6.1. London, UK: Cochrane Collaboration; 2020.
83. Egger M, Ebrahim S, Smith GD. Where now for meta-analysis? Int J Epidemiol 2002 Feb;31(1):1-5. [doi: 10.1093/ije/31.1.1] [Medline: 11914281]
84. Number of mHealth apps available in the Apple App Store from 1st quarter 2015 to 1st quarter 2021. Statista. 2021. URL: https://www.statista.com/statistics/779910/health-apps-available-ios-worldwide/ [accessed 2022-08-12]
85. Number of mHealth apps available in the Google Play Store from 1st quarter 2015 to 1st quarter 2021. Statista. 2021. URL: https://www.statista.com/statistics/779919/health-apps-available-google-play-worldwide/ [accessed 2022-08-12]
86. Blandford A, Gibbs J, Newhouse N, Perski O, Singh A, Murray E. Seven lessons for interdisciplinary research on interactive digital health interventions. Digit Health 2018 May 3;4:2055207618770325 [FREE Full text] [doi: 10.1177/2055207618770325] [Medline: 29942629]
87. 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]
88. Evidence standards framework (ESF) for digital health technologies. National Institute for Health and Care Excellence. 2019. URL: https://www.nice.org.uk/about/what-we-do/our-programmes/evidence-standards-framework-for-digital-health-technologies [accessed 2022-08-12]
89. O'Brien M, Daley D. Self-help parenting interventions for childhood behaviour disorders: a review of the evidence. Child Care Health Dev 2011 Sep;37(5):623-637. [doi: 10.1111/j.1365-2214.2011.01231.x] [Medline: 21585419]
90. Cardy JL, Waite P, Cocks F, Creswell C. A systematic review of parental involvement in cognitive behavioural therapy for adolescent anxiety disorders. Clin Child Fam Psychol Rev 2020 Dec;23(4):483-509 [FREE Full text] [doi: 10.1007/s10567-020-00324-2] [Medline: 32862330]

Abbreviations

CBMT: cognitive bias modification training
CBT: cognitive behavioral therapy
DHI: digital health intervention
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial

Edited by J Torous; submitted 13.11.20; peer-reviewed by J Tighe, H McCall, Y Amanvermez, K Kruzan, A Knapp; comments to author 26.01.21; revised version received 29.06.21; accepted 29.04.22; published 26.08.22.

Please cite as:
Babbage CM, Jackson GM, Davies EB, Nixon E
Self-help Digital Interventions Targeted at Improving Psychological Well-being in Young People With Perceived or Clinically Diagnosed Reduced Well-being: Systematic Review
JMIR Ment Health 2022;9(8):e25716
URL: https://mental.jmir.org/2022/8/e25716
doi:10.2196/25716
PMID:36018675

Original Paper

Multi-operator Self-exclusion as a Harm Reduction Measure in Problem Gambling: Retrospective Clinical Study on Gambling Relapse Despite Self-exclusion

Anders Håkansson[1,2], Prof Dr; Gunny Åkesson[1,2], MSc

[1]Clinical Research Unit, Competence Center Addiction, Malmö Addiction Center, Region Skåne, Malmö, Sweden

[2]Division of Psychiatry, Department of Clinical Sciences Lund, Lund University, Lund, Sweden

Corresponding Author:
Anders Håkansson, Prof Dr
Clinical Research Unit
Competence Center Addiction, Malmö Addiction Center
Region Skåne
Södra Tullgatan 4, plan 6
Malmö, 20502
Sweden
Phone: 46 40333883
Email: anders_c.hakansson@med.lu.se

Abstract

Background: Voluntary self-exclusion from gambling is a common harm reduction option for individuals with gambling problems. Multi-operator, nationwide self-exclusion services are rare, and a system introduced in the highly web-based gambling market of Sweden is a rare and recent example. However, where web-based casino gambling and web-based betting are the predominate gambling types in those seeking treatment, the risk of breaching one's own self-exclusion through overseas web-based operators may also be high.

Objective: This study aims to assess the prevalence of a nationwide *Spelpaus* ("gambling break") self-exclusion and the prevalence of gambling despite self-exclusion in patients seeking treatment for gambling disorder in 2021.

Methods: Health care documentation of recent treatment seekers (January 1 through September 1, 2021, N=85) in a Swedish treatment facility was reviewed for data regarding problematic gambling types reported, history of self-exclusion, and history of breaching of that self-exclusion.

Results: Common problem gambling types were web-based casino gambling (49/74, 66%) and sports betting (19/74, 26%). The majority who participated in this study (62/85, 73%) were men. All women reported web-based casino gambling. Self-exclusion through *Spelpaus* was common (60/74, 81%). Among self-excluders, gambling despite self-exclusion was common (41/60, 68%), most commonly on unlicensed gambling websites.

Conclusions: The nationwide, multi-operator self-exclusion service of Sweden appears to reach many patients with a gambling disorder. However, the remaining gambling options in an web-based gambling setting present a major challenge despite self-exclusion. The recent data calls for further treatment efforts and potential improvements in services aiming to help voluntary self-excluders abstain from gambling.

(JMIR Ment Health 2022;9(8):e37837) doi:10.2196/37837

KEYWORDS

web-based gambling; gambling disorder; behavioral addiction; self-exclusion; addiction; gambling; prevalence; harm reduction; abstain; problem gambling

Introduction

Gambling disorder is an addictive condition known to have major deteriorating effects on individuals' mental health and social and financial situation [1]. One of the interventions available in the management of problem gambling is the individual's own possibility to self-exclude from gambling, that is, to limit one's own access to a web-based gambling site or a physical gambling venue, to prevent one's own relapse into

gambling. This is a method available in many land-based or web-based gambling settings [2,3] and may serve as a viable option for gamblers who perceive severe problems and who may or may not seek formal treatment and for whom treatment seeking may often be too limited [4].

In Sweden, a previously unreported type of self-exclusion service, *Spelpaus*, meaning "gambling break" in direct translation, was introduced in the gambling legislation from January 2019, allowing gamblers to self-exclude from all licensed gambling operators operating either via the internet or physically in Sweden, including a large number of operators in web-based and land-based casinos, sports and horse racing betting, web-based card games, web-based lotteries, and web-based bingos. The self-exclusion period can be chosen to be either 1 month, 3 months, 6 months, or a year, and although each period is discontinued automatically, one can enter the system at any time to prolong it. *Spelpaus* is accessed at a specific website [5] belonging to the Swedish Gambling Authority, which is a national government authority, and the provision of this service is therefore independent of the individual gambling operators. Hitherto, it has been reported that a majority of problem gamblers likely have not self-excluded, although a substantial number of them do, and a substantial number of individuals without gambling problems also may choose to self-exclude in this system [6]. *Spelpaus* is promoted by gambling operators through radio, television, newspapers, and advertisements on the internet, as well as by preventive and therapeutic institutions. As the service is frequently mentioned in advertisement breaks in popular media, it may have relatively good potential to be heard of by both gamblers and nongamblers in the population.

Any system based on voluntary self-exclusion from gambling is associated with a certain risk of breaching the exclusion, as part of an addictive gambling pattern with lowered self-control. However, the risk of breaching the self-exclusion is very little described for web-based gambling services. Also, such data are virtually unavailable for the present type of nationwide self-exclusion services that involve a large number of gambling types and operators. In a population survey in Sweden, among online gamblers responding to a web-based survey, 38% of those who had ever self-excluded via this Swedish national self-exclusion service had gambled at some point during the exclusion period [7]. Likewise, recent data have shown that in a policy-based measure excluding gamblers from each gambling operator after reaching a maximum loss limit at that operator, it was common for gamblers with moderate risk or problem gamblers to gamble on other operators; this indicates a risk that individuals with intense gambling patterns may migrate to other operators in case of responsible gambling measures putting a limit to their gambling [8]. Thus, given this type of possible migration between operators, it can be suspected that even patients with gambling disorder who choose voluntary self-exclusion may be at risk of switching to other operators thereafter.

Sweden has a highly web-based gambling market, with web-based casino gambling and sports betting representing the vast majority of patients seeking treatment [9] and the vast majority of commercial advertisements seen on television [10]. One major task of policy making in the area has been to exclude unlicensed overseas gambling operators from the market to keep Swedish gambling within the legal operators, which are likely to follow government regulations including self-exclusion [11]. After the introduction of the *Spelpaus* self-exclusion service, a policy debate has addressed the risk of breaching this self-exclusion [7], typically because of the risk of migrating to overseas gambling services, and thereby, limiting the efficiency of self-exclusion. Based on this debate, it is important for policy making in the field to assess whether it is common for patients with gambling disorder to self-exclude from gambling, and whether it is common for them to gamble despite this self-exclusion. For this reason, this subanalysis was carried out as a substudy from a larger retrospective chart review study in a gambling disorder treatment unit. This substudy aimed to evaluate the prevalence of *Spelpaus* self-exclusion and the prevalence of *Spelpaus* breaching in treatment-seeking patients with gambling disorder.

Methods

Study Participants

This study is a substudy of a larger systematic, retrospective hospital chart study. The treatment episodes included are from the regional gambling disorder unit of Region Skåne, where both authors are employed, within the public health care services. This facility has a responsibility for the assessment and treatment of gambling disorder in the whole county of Skåne in southern Sweden, a county with around 1.4 million inhabitants. Other treatment providers in the county include the social services of local municipalities, which may offer different types of interventions for treatment-seeking clients with problem gambling, as well as the voluntary, nongovernmental peer support provided by local patient organizations. The Region Skåne health care unit is specialized in gambling and assesses treatment-seeking adults aged 18 years or older with gambling problems, typically on a diagnostic level corresponding to the World Health Organization diagnostic code of gambling disorder (F63.0 [12]). This treatment facility has been previously described in publications discussing the gambling types and comorbidity in patients with gambling disorder [9] and lack of evident changes in treatment seeking at the facility after the onset of the COVID-19 pandemic [13]. Predominate gambling types at the facility are web-based casino gambling, which is also the gambling type most commonly seen in televised gambling advertising in Sweden [10], and sports betting. The majority of patients (80%) are men, and 58% are diagnosed with another psychiatric comorbidity in parallel with the gambling disorder [9]. The facility opened in December 2015 and has therefore been operating both before and after the formal introduction of gambling into the treatment responsibilities of Swedish health care and social service institutions in January 2018.

Study Procedures

This is a substudy originating from a larger one that aimed to study characteristics and changes in patients seeking treatment at the facility before and after the introduction of gambling into the Swedish addiction treatment legislation, before and after

the outbreak of COVID-19, and before and after the introduction of specific gambling legislations by the Swedish government due to COVID-19. The COVID-19–related aspect of this study was the suspicion during the COVID-19 pandemic that societal changes during the pandemic may be associated with increased gambling patterns, at least in individuals with gambling problems [14]. Given the specific picture of the data collected from the most recent year (2021) and given the political and media attention to the role of the *Spelpaus* self-exclusion service in current trends in web-based gambling, a specific subanalysis was carried out, involving only the treatment episodes from January 1 through September 1, 2021.

This subanalysis includes all patients who applied for treatment and received an appointment at the unit from January 1 through September 1, 2021, that is, during a period when both authors were employed by and actively working at the facility as the medically responsible physician (AH) and as one of the therapists (GÅ). The second author of the paper (GÅ) reviewed each record of health care documentation from the facility. Data extracted for this descriptive analysis include gender, type of problematic gambling reported, whether the patient had reported national *Spelpaus* self-exclusion upon treatment seeking, and whether the patient reported gambling (and the type of modality) despite this self-exclusion.

Ethics Approval

The overall study was approved by the Swedish Ethics Review Authority (file number 2021-03636) on August 11, 2021. Also, the specific subanalysis included in this paper along with its specific aim was approved in an amendment to the same authority (file number 2022-00911-02) on February 22, 2022. In line with the ethics approval, no information or consent from patients was required in this retrospective chart review study.

Statistical Methods

Results are reported as descriptive frequencies, including absolute numbers and percentages, and gambling despite using *Spelpaus* was compared between women and men, using the Fisher exact test with significance defined as $P<.05$.

Results

A total of 85 individuals were included in the study. Among them, 62 (73%) were men and 23 (27%) were women. A total of 7 patients (2 men and 5 women) did not start the formal treatment. Full data were not found in the chart review in 4 cases (4 men). Further data were available for 74 individuals, that is, 56 (76%) men and 18 (24%) women.

Problematic gambling types reported were web-based casino gambling (49/74, 66%), sports betting (19/74, 26%), and other land-based gambling (1/74, 1%); the gambling type was missing in 5 cases (7%). Web-based casino gambling was reported by 100% of women (18/18) and 55% of men (31/56), whereas sports betting was reported by 34% (19/56) of men and none of the women.

In total, 81% (60/74) had self-excluded through *Spelpaus*, 4% (3/74) had not, and data were unavailable from the chart text in 15% (11/74) of the cases. All 3 individuals who did not report self-exclusion were men. Out of those having self-excluded, 68% (41/60) reported having gambled despite using *Spelpaus*, that is, 63% of men (27/43) and 82% of women (14/17), without any significant gender difference ($P=.22$). Among clients reporting gambling despite using *Spelpaus*, this was reported to occur on unlicensed gambling websites in 68% of the cases (28/41), and gambling with somebody else's identity or unlicensed gambling in land-based venues in 22% of the cases (9/41); the information was missing for 10% (4/41) of the cases.

Discussion

This substudy aimed to provide an up-to-date observation regarding the dissemination and feasibility of a harm reduction service (*Spelpaus*) in use since 2019 in the Swedish gambling market. This unique, nationwide, multi-operator self-exclusion service, operating in a highly web-based gambling market, has theoretically faced with a substantial challenge from the hypothetical competition with overseas web-based operators that are not included in this self-exclusion service. This analysis demonstrates that among patients seeking help for gambling disorder in recent months, adherence to the self-exclusion service is very common; but among those who self-excluded, as many as 68% had gambled despite self-exclusion. Thus, given the fact that these patients are actively seeking treatment for a severe gambling problem, it must be assumed that their gambling problem was associated with the gambling activity that breached the self-exclusion.

Therefore, this subanalysis, from a larger clinical study, can be seen as an alert for improved measures against gambling in people with gambling problems who seek measures to control their gambling patterns. It can be concluded that in the present setting, where web-based gambling is highly predominant among patients with gambling disorder [9], this self-exclusion service is insufficient for these patients to cope with their addictive behavior. Structured treatment of the condition is therefore required over and above this self-exclusion service, but also, policy makers may need to consider technical or legal possibilities to expand the self-exclusion possibility beyond the present setting to gambling markets that are web-based and cannot be expected to respect geographical borders.

A nationwide, multi-operator self-exclusion system appears to attract people with severe gambling problems, as a clear majority of treatment seekers in this study reported having enrolled in this system. This can be seen as a major merit of this self-exclusion system, such that it reaches the individuals for whom it is intended. It goes beyond this study, which was carried out in a clinical setting, to analyze whether the service also successfully reaches individuals with subdiagnostic but significant gambling problems, as they also must be considered as a target group of this harm reduction service.

Our study clearly indicates a major challenge of this type of self-exclusion service in a setting where web-based gambling is common, as individuals with gambling disorders may be attracted to overseas gambling operators, which are unlicensed in Sweden, and therefore, beyond the jurisdiction of its gambling regulations. Policy makers should address further technical and judiciary possibilities of promoting self-exclusion from other

gambling operators and not only those operating within the country's geographical setting. Additionally, treatment providers and preventive institutions should address the fact that self-exclusion from gambling is an insufficient measure for individuals with gambling disorders, who also may need structured treatment for their addictive behavior.

In addition to the findings regarding the self-exclusion service, it must be noted that a clear majority of the patients seeking treatment during these 8 months had a problematic gambling pattern associated with web-based casinos. In particular, this was the only problematic gambling type reported by the women in the study, pointing to the potential risk of more female gamblers being recruited to problem gambling due to this modern gambling type. It confirms the previously demonstrated addiction potential of web-based casino gambling, as this is the most common gambling type in patients seeking treatment in this study and previous clinical studies [9]; however, this type of gambling is in fact used by a very small minority of gamblers in the general population [15].

This study, a timely subanalysis from a larger project, aimed to inform stakeholders in the area about an ongoing situation in a gambling market where several different measures of responsible gambling and preventive strategies have been tested. The study results also suggest the need for further, more in-depth studies about how gambling self-exclusion is perceived by users and whether the choice to self-exclude, in gamblers who breach their self-exclusion, may still be beneficial or have negative effects. Such further research may partly require qualitative studies to describe in more detail the considerations made by gamblers in relation to self-exclusion. Also, the use of objective gambling data or financial loan data in those who self-exclude, in relation to their exclusion date and in relation to their breaching, may highlight more closely the potential effects of the self-exclusion system. However, in contrast to its timely findings, this study clearly has some limitations. The study was conducted in a single treatment facility and only for a period of 8 months, which means that the study sample is limited in size. This clearly limits the potential for statistical calculations and the generalizability of findings to other settings. In addition, the number of study variables derived from a retrospective analysis of medical records, intended not primarily for research but for routine treatment purposes, is limited.

In conclusion, this brief report confirms the fact that patients with gambling disorder are successfully reached by the relatively unique Swedish nationwide self-exclusion service, but it also clearly demonstrates the current challenge of self-exclusion in a modern, highly digitalized gambling market. Therefore, although the study supports the self-exclusion service used here, it also concludes that formal treatment, further preventive work, and more advanced services for self-exclusion from unlicensed gambling operators are important. In addition, the overwhelming majority of web-based casino gamblers in a single treatment facility also confirms the previously reported addiction potential of this gambling type.

Conflicts of Interest

AH is employed as a professor at Lund University in Sweden, and his position and research are financially supported by the entirely state-owned gambling operator AB Svenska Spel; he has also obtained research funding from Svenska Spel's research council. None of these organizations had any role in this research. The other author reports no conflicts of interest.

References

1. Potenza MN, Balodis IM, Derevensky J, Grant JE, Petry NM, Verdejo-Garcia A, et al. Gambling disorder. Nat Rev Dis Primers 2019 Jul 25;5(1):51. [doi: 10.1038/s41572-019-0099-7] [Medline: 31346179]
2. Hayer T, Meyer G. Internet self-exclusion: characteristics of self-excluded gamblers and preliminary evidence for its effectiveness. Int J Ment Health Addiction 2010 Aug 4;9(3):296-307. [doi: 10.1007/s11469-010-9288-z]
3. Matheson FI, Hamilton-Wright S, Kryszajtys DT, Wiese JL, Cadel L, Ziegler C, et al. The use of self-management strategies for problem gambling: a scoping review. BMC Public Health 2019 Apr 29;19(1):445 [FREE Full text] [doi: 10.1186/s12889-019-6755-8] [Medline: 31035978]
4. Gainsbury S, Hing N, Suhonen N. Professional help-seeking for gambling problems: awareness, barriers and motivators for treatment. J Gambl Stud 2014 Jun 15;30(2):503-519. [doi: 10.1007/s10899-013-9373-x] [Medline: 23494244]
5. Håkansson A, Widinghoff C, Berge J. Self-exclusion from gambling-a measure of covid-19 impact on gambling in a highly online-based gambling market? Int J Environ Res Public Health 2021 Jul 09;18(14):7367 [FREE Full text] [doi: 10.3390/ijerph18147367] [Medline: 34299817]
6. Håkansson A, Henzel V. Who chooses to enroll in a new national gambling self-exclusion system? A general population survey in Sweden. Harm Reduct J 2020 Oct 21;17(1):82 [FREE Full text] [doi: 10.1186/s12954-020-00423-x] [Medline: 33087113]
7. Håkansson A, Widinghoff C. Gambling despite nationwide self-exclusion-a survey in online gamblers in Sweden. Front Psychiatry 2020 Dec 2;11:599967 [FREE Full text] [doi: 10.3389/fpsyt.2020.599967] [Medline: 33343428]
8. Håkansson A, Sundvall A, Lyckberg A. Effects of a national preventive intervention against potential COVID-19-related gambling problems in online gamblers: self-report survey study. JMIR Form Res 2022 Mar 09;6(3):e33066 [FREE Full text] [doi: 10.2196/33066] [Medline: 34678751]

9. Håkansson A, Mårdhed E, Zaar M. Who seeks treatment when medicine opens the door to pathological gambling patients-psychiatric comorbidity and heavy predominance of online gambling. Front Psychiatry 2017 Nov 29;8:255 [FREE Full text] [doi: 10.3389/fpsyt.2017.00255] [Medline: 29238309]
10. Håkansson A, Widinghoff C. Television gambling advertisements: Extent and content of gambling advertisements with a focus on potential high-risk commercial messages. Addict Behav Rep 2019 Jun;9:100182 [FREE Full text] [doi: 10.1016/j.abrep.2019.100182] [Medline: 31193826]
11. About the government's work for a re-regulated gambling market. Government of Sweden. 2021. URL: https://www.government.se/articles/2021/01/about-the-governments-work-for-a-re-regulated-gambling-market/ [accessed 2022-06-17]
12. ICD-10 : international statistical classification of diseases and related health problems : tenth revision, 2nd ed. World Health Organization. 2004. URL: https://apps.who.int/iris/bitstream/handle/10665/42980/9241546530_eng.pdf?sequence=1&isAllowed=y [accessed 2022-06-17]
13. Håkansson A, Åkesson G, Grudet C, Broman N. No apparent increase in treatment uptake for gambling disorder during ten months of the COVID-19 pandemic-analysis of a regional specialized treatment unit in Sweden. Int J Environ Res Public Health 2021 Feb 17;18(4):1918 [FREE Full text] [doi: 10.3390/ijerph18041918] [Medline: 33671210]
14. Håkansson A, Widinghoff C. Changes of gambling patterns during covid-19 in Sweden, and potential for preventive policy changes. A second look nine months into the pandemic. Int J Environ Res Public Health 2021 Feb 27;18(5):2342 [FREE Full text] [doi: 10.3390/ijerph18052342] [Medline: 33673575]
15. Resultat från Swelogs 2018. Folkhalsomyndigheten. URL: https://www.folkhalsomyndigheten.se/globalassets/livsvillkor-levnadsvanor/andts/spel/swelogs/resultat-swelogs-2018-2019.pdf [accessed 2022-06-17]

Edited by J Torous; submitted 08.03.22; peer-reviewed by K Uludag, I Savolainen; comments to author 26.04.22; revised version received 01.05.22; accepted 05.05.22; published 19.08.22.

Please cite as:
Håkansson A, Åkesson G
Multi-operator Self-exclusion as a Harm Reduction Measure in Problem Gambling: Retrospective Clinical Study on Gambling Relapse Despite Self-exclusion
JMIR Ment Health 2022;9(8):e37837
URL: https://mental.jmir.org/2022/8/e37837
doi:10.2196/37837
PMID:35984678

Original Paper

An e–Mental Health Resource for COVID-19–Associated Stress Reduction: Mixed Methods Study of Reach, Usability, and User Perceptions

Nadia Minian[1,2,3,4], PhD; Allison Gayapersad[1], PhD; Anika Saiva[1], MPH; Rosa Dragonetti[1,2], MSc; Sean A Kidd[5,6], PhD; Gillian Strudwick[5,7], RN, PhD; Peter Selby[1,2,6,8,9], MBBS, MHSc

[1]Nicotine Dependence Service, Centre for Addiction and Mental Health, Toronto, ON, Canada
[2]Department of Family and Community Medicine, University of Toronto, Toronto, ON, Canada
[3]Institute of Medical Science, University of Toronto, Toronto, ON, Canada
[4]Department of Pharmacology and Toxicology, University of Toronto, Toronto, ON, Canada
[5]Centre for Addiction and Mental Health, Toronto, ON, Canada
[6]Department of Psychiatry, University of Toronto, Toronto, ON, Canada
[7]Institute of Health Policy, Management and Evaluation, University of Toronto, Toronto, ON, Canada
[8]Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, ON, Canada
[9]Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada

Corresponding Author:
Peter Selby, MBBS, MHSc
Nicotine Dependence Service
Centre for Addiction and Mental Health
1025 Queen Street West
Toronto, ON, M6H 1H4
Canada
Phone: 1 416 535 8501
Email: peter_selby@camh.net

Abstract

Background: COVID-19 and its public health response are having a profound effect on people's mental health. To provide support during these times, Canada's largest mental health and addiction teaching hospital (Centre for Addiction and Mental Health [CAMH]) launched the Mental Health and COVID-19 Pandemic website on March 18, 2020. This website was designed to be a nonstigmatizing psychoeducational resource for people experiencing mild to moderate distress due to COVID-19 and the public health response to the pandemic.

Objective: The aim of this study was to examine the reach, usability, and user perceptions of the CAMH Mental Health and COVID-19 Pandemic website.

Methods: This study used a mixed methods sequential explanatory design approach, which consisted of the following 2 distinct phases: (1) quantitative data collection and analysis and (2) qualitative semistructured interviews. In phase 1, we analyzed Google Analytics data to understand how many people visited the website and which were the most visited pages. We conducted a survey to identify users' sociodemographic backgrounds, and assess the usability of the website using the System Usability Scale and users' subjective stress levels using the Perceived Stress Scale (PSS-10). For phase 2, we conducted semistructured interviews to explore user experiences; user motivation, engagement, satisfaction, and perception of the stress reduction strategies; reflections of the website's functionality, ease of use, navigation, and design; and recommendations for improvement.

Results: Google Analytics results showed 146,978 unique users from June 2020 to March 2021. Most users were from Canada (130,066, 88.5%). Between February 20, 2021, and June 4, 2021, 152 users completed the survey. Most users identified as white, female, and having at least a college degree. Based on the PSS-10 scores, most participants were experiencing moderate to high stress when they visited the website. Users rated the usability of the website as acceptable. Ten users completed in-depth interviews between May 2021 and June 2021. Positive feedback related to the content was that the website was a trustworthy source of mental health information with helpful evidence-based stress reduction strategies. Areas for improvement included the text heavy

design of the website, wider dissemination/marketing, and greater accessibility of the website to meet the needs of diverse populations.

Conclusions: Adding stress reduction resources to a website from a well-respected institution may be a practical method to increase awareness and access to evidence-based stress reduction resources during times of crisis, where there is severe disruption to usual health care contacts. Efforts to ensure that these resources are more widely accessed, especially by diverse populations, are needed.

(JMIR Ment Health 2022;9(8):e39885) doi:10.2196/39885

KEYWORDS

COVID-19; website; stress; mental health; eHealth; internet-based intervention; mixed methods evaluation; usability; digital health; health informatics

Introduction

The COVID-19 pandemic and its public health response are having a profound effect on all aspects of society, including mental health [1]. Introductions of public health measures, such as quarantine, travel restrictions, and physical closure of places of learning and businesses, were part of the response to reduce the spread of COVID-19 [2-4]. Quarantine, social isolation, and uncertainty surrounding the pandemic are known to negatively impact mental health [5]. For example, one study of individuals quarantined during the 2003 SARS pandemic in Toronto, Ontario, found that 31% of the sample developed posttraumatic stress disorder symptoms, and 29% demonstrated depressive symptom severity similar to that of a patient with clinically diagnosed depression [6].

In response to the disruption of in-person care during the current COVID-19 pandemic, health systems rapidly adopted virtual care options [7]. Where possible, care was delivered virtually by telephone or through video-conferencing platforms and web applications, which was especially true and possible with mental health conditions [7]. Not everyone needed or wanted in-person clinical services, but rather wanted accessible trustworthy psychoeducation to manage the uncertainty-related stress and anxiety associated with the disease itself and the impacts on their lives [8,9]. This led to a concern that in the absence of credible Canadian sources, most would turn to uncurated information through social media [10-12]. In Canada, the capacity of mental health and addiction services in the province of Ontario was also of concern [13]. Even before the pandemic, Ontario's infrastructure for mental health and addiction services was overburdened and underfunded [14]. Thus, expecting this system to respond to people with mild to moderate distress was not deemed to be appropriate.

Given the growing evidence suggesting that positive outcomes can be achieved through the use of digital mental health technologies [15-20], on March 18, 2020 (the same month the World Health Organization declared COVID-19 a pandemic), the Centre for Addiction and Mental Health (CAMH), Toronto, Canada [21] launched a COVID-19 resource page on the CAMH website [22]. The website was designed to be a digital, plain language, nonstigmatizing psychoeducational resource for Canadians experiencing mild to moderate distress due to COVID-19. Stress was assumed to be universal due to the uncertainty associated with an evolving pandemic. The response would affect those with and without pre-existing mental illness to varying degrees, with the majority of the population experiencing mild to moderate distress, but in the absence of resources, this could overwhelm the health care system that was focusing on containing COVID-19. The website provides practical problem-solving tips based on cognitive behavioral therapy in both English and French. Specifically, it contains (1) a moderated discussion board, (2) written content on managing mental health during COVID-19, (3) self-assessments for anxiety using the Generalized Anxiety Disorder-7 (GAD-7) scale and perceived stress and resilience using the Perceived Stress Scale (PSS), and (4) a series of answers to frequently asked questions (FAQs) by the lay public on specific topics including intimate partner violence and grief. Those with diagnosable anxiety and depression might have more severe stress responses, which in turn hampers engagement with digital mental health interventions [23]. The lead developers (PS and RD) felt that these individuals should be directed to seek clinical services either in-person or virtually because they might need more intensive services.

The rapid development of digital strategies to minimize the negative impact of COVID-19 on public mental health has outpaced the evaluation of digital health interventions [24-26], creating a gap in our understanding of the utility and quality of the interventions [27]. In this study, we report on the reach, usability, and user perceptions of the CAMH COVID-19 website.

Methods

Study Design

This study used a mixed methods sequential explanatory design approach, which consisted of the following 2 distinct phases: (1) quantitative data collection and analysis and (2) qualitative semistructured interviews. To report this study, we adhered to the guidelines for Good Reporting of A Mixed Methods Study (GRAMMS) [28].

Recruitment and Data Collection

Phase 1: Reach (Google Analytics) and Usability (Self-administered Survey)

In phase 1, Google Analytics data were used to determine the number of unique users, the countries and cities associated with the users, what pages they visited and in what order, and the channels (eg, Twitter and Facebook) that drove the most traffic.

We collected Google Analytics data from June 2020 to March 2021.

In order to obtain additional descriptive information on who uses CAMH's COVID-19 website tools, we placed an advertisement for this study on the website and sent a tweet letting people know. Tweets were also sent from the organizational accounts, @CAMHnews and @PSQuitSmoking, asking users/readers if they were interested in participating in the study. After consenting, participants completed a survey via Research Electronic Data Capture (REDCap), which included questions about sociodemographics; the PSS-10, a 10-item tool scored on a 5-point Likert scale ranging from 0 (never) to 4 (very often), which assesses subjective stress levels [29]; and the System Usability Scale (SUS) for assessing perceived usability of applications and services [30,31]. The SUS is a 10-item questionnaire containing response options ranging from strongly agree to strongly disagree. The decision to use the SUS tool for measuring usability was based on its validity to effectively distinguish between usable and nonusable systems [30]. The SUS can be easily administered to participants and has been shown to yield reliable results, even with small sample sizes [30]. Once the participants completed the survey, they were redirected to a new page via REDCap to provide their contact details to enter into a draw for a gift card. This was optional, and participants were able to choose not to enter the draw. Participants were also asked about their interest in participating in an interview and about consent for future communication (both optional), and contact information (email and phone number) was collected from those who were interested. Participants were informed that information collected from the survey might be used by CAMH for future research, in which case, the data were deidentified and only aggregated data were reported. We collected survey data from February 20, 2021, to June 4, 2021.

Phase 2: User Perceptions (Semistructured Interviews)

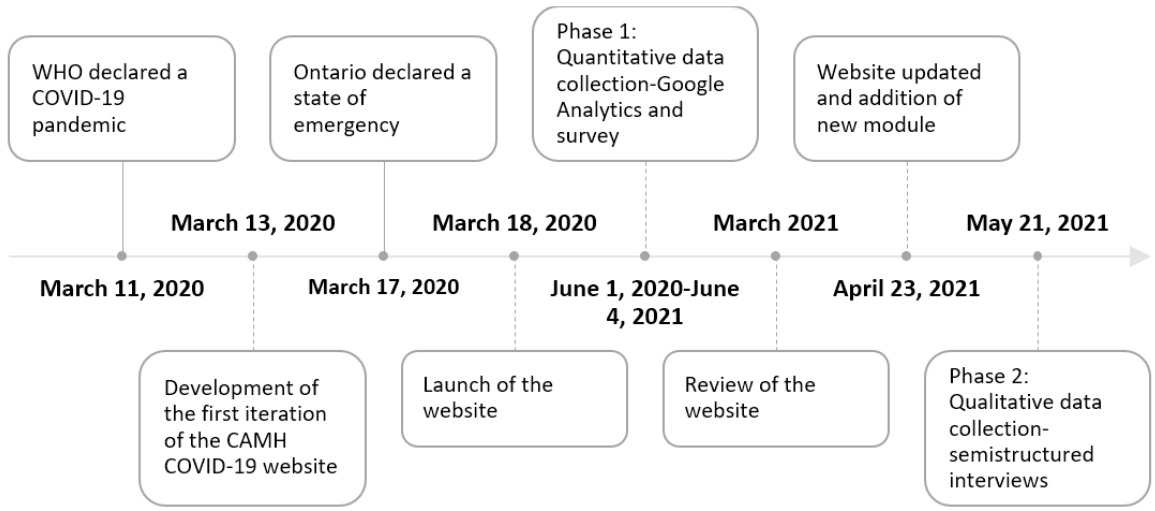
To explore and describe participant perceptions of the CAMH COVID-19 website, semistructured interviews were conducted from May 25, 2021, to June 29, 2021. Of the 152 participants who completed the CAMH COVID-19 website evaluation survey, 53 indicated that they were interested in being interviewed. All 53 participants were eligible to be interviewed as they visited the CAMH COVID-19 website and were over the age of 18 years. They were contacted by email/telephone, and a consent discussion was scheduled with individuals who agreed to proceed with an interview. Once participants consented and signed the informed consent using REDCap, they were booked for a 30-minute semistructured interview over the phone or a video-conferencing system. All interviews were audio recorded and transcribed verbatim. Ten interviews were conducted, and the data were assessed as adequate to permit deep case-oriented analysis that resulted in an in-depth understanding of users' experiences with the website [32]. Several attempts were made to reach and recruit a heterogeneous sample of users for interviews. However, those who consented to participate were primarily college educated, white, and female. All participants who consented to an interview were from Ontario; thus, our sample was not representative of the overall population of Canada.

After updates were made, qualitative semistructured interviews were then used to further elaborate on the usability of the website. Usability was defined as the degree to which a program can be used easily, efficiently, and with satisfaction [33]. Specifically, the interview topic guide explored participants' experiences with the website, including their motivation, engagement, satisfaction, or dissatisfaction, and their perceived experiences on whether the website helped to reduce their stress. It also explored participants' reflections on the CAMH COVID-19 website's functionality and ease of use/navigation, the overall look/design, and how the intervention could be improved.

Timeline of Project Activities

Figure 1 provides a timeline of project activities. The website was launched on March 18, 2020. Content was added over the year including making the website available offline through a downloadable app available on the App Store for Apple devices and Google Play Store for Android devices. Topics were added, based on the evolution of the pandemic and perceived needs to be addressed, by the planning committee (eg, parenting, grief, trauma and loss, parent support, and back to school). On April 23, 2021, based on preliminary feedback from this study and a review of the website by a trained behavioral psychologist and professional designer, the website was updated to improve navigation, eliminate duplication, and include additional content (eg, support for caregivers), based on what was going on at the time.

Figure 1. Project timeline. CAMH: Centre for Addiction and Mental Health; WHO: World Health Organization.



Data Analysis

Phase 1: Google Analytics and Self-administered Survey

Google Analytics data collected from June 2020 to March 2021 were exported to Microsoft Excel, which we used to conduct descriptive statistics. We calculated means, standard deviations, and medians where applicable.

Quantitative data from the survey were analyzed using SPSS (version 25; IBM Corp). The scoring guidelines provided by the US Department of Health and Human Services were used to inform the analysis of the SUS questionnaire in the survey [30]. A SUS score between 0 and 100 was generated for each participant by converting the score for each question into a new number, summing the converted scores, and multiplying the score by 2.5. Descriptive statistics were used to report sociodemographic characteristics, usability, and engagement. The PSS-10 was analyzed by first reversing scores for 4 of the positive items on the scale and then summing across all 10 items to receive a score between 0 and 40 for each participant.

Phase 2: Semistructured Interviews

Qualitative data analysis for the study involved an iterative team-based process. Transcribed interviews were entered into NVivo 12 (QSR International) for qualitative data management and analysis. Two research staff read interview transcripts multiple times to achieve immersion prior to code development. The author AG crafted the initial codebook, and then, AG and AS jointly coded the first 3 transcripts to refine the codebook and defined the codes through consensus. Two other transcripts were coded independently by AG and AS, and the calculated intercoder reliability was assessed using the kappa coefficient in NVivo 12. There was 93% agreement on 96% of codes [34]. The remaining transcripts were independently coded either by AG or AS. Data were analyzed using thematic analysis [35]. A deductive approach was used to identify the coding scheme for the transcripts that allowed for the development of codes corresponding to the components of usability and stress management.

We merged both the quantitative and qualitative analysis results to provide a more comprehensive overview of participants' perceptions of the CAMH COVID-19 website.

Ethics

This study, including the recruitment and consent process, received ethics approval from the CAMH Quality Project Ethics Review Board (QPER #: 2020_34) [36].

Results

Phase 1: Google Analytics and Self-administered Survey

Google Analytics

From June 2020 to March 2021, there were 146,978 unique users (unduplicated/counted only once) of the website. Of the 146,978 users from all over the world, most were from Canada (130,066, 88.5%), and of these, 30,964 (23.8%) were from Toronto and 6538 (5.0%) were from Ottawa. The homepage was the most visited page of the website, which had 91,866 unique page views during the 10 months, followed by the page on "Coping with Stress and Anxiety," which had 29,889 unique page views during the 10 months [22,37]. Organic search traffic (ie, entry through a search engine) accounted for the highest proportion of user visits to the website. From June 2020 to March 2021, 108,332 users (73.7% of all unique visitors) visited the site through organic search and 36,630 users (24.9% of all unique visitors) visited the site by typing the URL directly into their address bar.

Characteristics of Survey and Interview Participants

Survey responses are shown in Table 1. We have divided participant responses into 3 columns. The first column includes participants who responded between February 20, 2021, and April 23, 2021 (before the website was updated); the second includes participants who responded between April 24, 2021, and June 4, 2021 (after the website was updated); and the third includes participants we interviewed.

Table 1. Demographic data, Perceived Stress Scale, and System Usability Scale (SUS) scores.

Variable	Survey participants (February 20-April 23, 2021) (N=101[a])	Survey participants (April 24-June 4, 2021) (N=51[b])	Interview participants (May-June 2021) (N=10)
Race/ethnicity, n (%)			
White	76 (75.3)	36 (70.6)	5 (50.0)
Latin American	5 (5.0)	1 (2.0)	1 (10.0)
Mixed (eg, African and White)	2 (2.0)	2 (3.9)	1 (10.0)
East/South/South East Asian	6 (5.9)	6 (11.8)	2 (20.0)
First Nations/Indigenous	0 (0.0)	1 (2.0)	0 (0.0)
Black	3 (3.0)	1 (2.0)	0 (0.0)
Indian–Caribbean	1 (1.0)	0 (0.0)	1 (10.0)
Did not answer	8 (7.9)	4 (7.8)	0 (0.0)
Age (years), mean (SD)	44 (18)	45 (18)	45 (15)
Gender, n (%)			
Female	86 (85.1)	39 (76.5)	10 (100.0)
Male	10 (9.9)	7 (13.7)	0 (0.0)
Nonbinary	0 (0.0)	4 (7.8)	0 (0.0)
Unknown/did not know/did not answer	5 (5.0)	1 (2.0)	0 (0.0)
Education, n (%)			
Graduate degree	16 (15.8)	17 (33.3)	7 (70.0)
University degree	33 (32.7)	9 (17.6)	1 (10.0)
College diploma	24 (23.8)	12 (23.5)	1 (10.0)
Trade/vocational school	6 (5.9)	5 (9.8)	1 (10.0)
High school	18 (17.8)	7 (13.7)	0 (0.0)
Elementary school	3 (3.0)	0 (0.0)	0 (0.0)
Did not answer	1 (1.0)	1 (2.0)	0 (0.0)
Income (CAD $)[c], n (%)			
$0-$29,999	19 (18.8)	7 (13.7)	2 (20.0)
$30,000-$59,999	19 (18.8)	12 (23.5)	0 (0.0)
$60,000-$89,999	24 (13.9)	5 (9.8)	2 (20.0)
$90,000-$119,999	12 (11.9)	7 (13.7)	3 (30.0)
$120,000-$149,999	5 (5.0)	2 (3.9)	0 (0.0)
$150,000 or more	15 (14.9)	9 (17.6)	2 (20.0)
Do not know/prefer not to disclose	17 (16.8)	9 (17.6)	1 (10.0)
Marital status, n (%)			
Married/in a relationship	47 (46.5)	27 (54.0)	7 (70.0)
Divorced/separated	14 (13.9)	2 (4.0)	1 (10.0)
Single/never married	37 (36.6)	17 (34.0)	1 (10.0)
Widowed	2 (2.0)	2 (4.0)	1 (10.0)
Do not know/did not answer	1 (1.0)	2 (4.0)	0 (0.0)
Level of perceived stress (Perceived Stress Scale score range), n (%)			
Low stress (0-13)	8 (9.0)	6 (12.5)	0 (0.0)
Moderate stress (14-26)	44 (49.4)	25 (52.1)	6 (60.0)
High perceived stress (27-40)	37 (41.6)	17 (35.4)	4 (40.0)

Variable	Survey participants (February 20-April 23, 2021) (N=101[a])	Survey participants (April 24-June 4, 2021) (N=51[b])	Interview participants (May-June 2021) (N=10)
System Usability Scale, n (%)			
Above average (greater than 68)	42 (48.3)	28 (58.3)	7 (70.0)
Below average (lower than 68)	45 (51.7)	20 (41.7)	3 (30.0)

[a]For the variables age, level of perceived stress, and System Usability Scale finding, the N values were 99, 89, and 87, respectively.

[b]For the variables age, marital status, level of perceived stress, and System Usability Scale finding, the N values were 47, 50, 48, and 48, respectively.

[c]A currency exchange rate of CAD $1=US $0.77 is applicable.

Survey Results

In total, 152 participants completed the survey. Most participants identified as white and female, and had at least a college degree. Based on the PSS-10 scores, most participants were experiencing moderate to high stress when they visited the website.

Survey data from February 20, 2021, to April 23, 2021 (before the website was updated) indicated that participants reported a mean SUS score of 66.44 (SD 16.72; IQR 10-97.5; N=87). According to the SUS, this score was below average for a website. Specific SUS items that received a low score included the website being well-integrated (41/87, 47% agreed or strongly agreed) and the likelihood of using the website frequently (34/87, 39% agreed or strongly agreed).

After changes to the website were implemented, survey data from April 24, 2021, to June 4, 2021, indicated that the participants reported a mean SUS score of 70 (SD 17.02; IQR 22.5-100; N=48). According to the SUS, this score was above average for a website. Improved scores were reported for the website being well-integrated (31/48, 65% agreed or strongly agreed) and the likelihood of using the website frequently (26/48, 55% agreed or strongly agreed).

Phase 2: Interview Results

Background on Participants Interviewed

Interview participants (n=10) were experiencing moderate to high levels of stress at the time they visited the website. Four participants had a history of mental illness or had a family member with a mental illness who received services from CAMH. Of 2 participants, whose family members received services from CAMH, one was an essential worker and the other a retired health care worker looking for mental health resources. The 2 participants with a diagnosed mental illness had recently contracted and recovered from COVID-19. Both stated that they were frequent users of the general CAMH website and visited the CAMH COVID-19 page for COVID-19–related information. Five other participants were researchers or professionals in a health-related field working with individuals who required mental health services. One participant was a stay-at-home parent with young children.

Functionality, Ease of Use, and Design

Most participants who were interviewed reported having visited the CAMH COVID-19 website once or twice, and having gone there to look for resources on how to deal with stress (either for themselves or for others). All participants (n=10) shared that their first impression of the website was positive, and that there were many helpful resources. Generally, participants found the website visually appealing, and liked the diversity in the images.

> *I liked the overall design, sort of the esthetic of the website. So, I found all of the photos that are in there to be quite diverse, in terms of the way that people look and the sort of different activities that they're doing. They all seem very relatable….* [29-year-old participant]

Several participants (n=5) reported that the information was helpful, but it was not well organized. Two participants specifically mentioned feeling overwhelmed by the amount of information on the website.

> *Well, my immediate impressions were that it was really overwhelming. … There's so many options…. Great…I mean it's great to have so many options, but because of the pandemic, I think people are already overwhelmed… There's a lot of good stuff there.* [72-year-old participant]

Another common theme, mentioned by 4 participants, was that the information might not be accessible to many users who might benefit from the information on the website.

> *My parents are refugees, like a lot of these words would be difficult for them to understand…like stigma and prejudice…I don't know that those words mean anything to them. Maybe making things a bit more layman's terms. Like if you want to just, kind of just speak to the masses, versus people that are in a clinical setting. Or have a background in mental health and social work.* [35-year-old participant]

In addition, 3 participants mentioned that there was a need to disseminate the website to those who might need it the most.

> *My partner, for example, he's dealt with mental illness and homelessness for over ten years, and he didn't know that this website existed even.* [43-year-old participant]

The Website and Stress

When asked specifically about the website as a stress reduction strategy for people in their province, all participants (n=10) responded that the website had the potential to reduce stress for individuals seeking help with stress-related issues. All participants were from Ontario. Participants noted that CAMH is a trusted authority where people can access evidence-based resources without judgement.

> *Well, I think it's great that, I think it's great that you, you know, that you're a trusted authority. And I think it's great…I think it's really wonderful that all…that this website has been put up. Because, you know, I mean people need, need to get stuff from a trusted place, you know. And a place where they won't…and where they won't be judged, which is another huge thing is judgment, right.* [72-year-old participant]

> *I think, I really liked that you can take the tests. You know the…you know like I'm familiar with the GAD7 just because I'm a Psychologist....Like, it's just nice to see that, you know there's evidence-based research supported measures...* [37-year-old participant]

Consistent with the Google Analytics data, participants mentioned that the section they visited the most was "Stress and Anxiety." Health professionals (n=3) indicated that the self-assessment resources found on this page would be helpful to their clients.

> *And, I think, when I went to the website, there were like different advice kind of little columns that you can read about how to like stay active and how to decrease your stress and your burnout and whatnot. And then there's like different little tests you can take. And so, I thought that was…it was just really nicely laid out. Yeah. For…I mean, not for myself, like I already, you know, I'm kind of like at the point where I'm kind of feeling anxious, as given my profession, but for clients and whatnot…I think it's helpful for them...* [37-year-old participant]

A participant, who was single, mentioned that she found the stress resources needed and was able to employ the recommended stress and coping strategies outlined on the website. This was in direct contrast to what parents said, as noted below.

> *Well, it was very good. Because it tells you, you know you need to be…keep yourself busy. You need to enjoy things like, you know hobbies…. Yes. Definitely. And I think for people who, you know like myself, who lives alone. So, I read that…I read it basically to give me information as a single person what is going on.* [62-year-old participant]

The need for information tailored to parents was mentioned by 2 participants as follows:

> *I think if you had a section for parents and that's something I think that would be helpful for sure. When you have kids it's like, you, you are responsible for other people than yourself. And it's just so much harder, because there is no break …so it was something that was useful, but for me in my situation, it wasn't quite enough, just because, you know you can't really control your children. So, like there's only so much that, so much that's limited to you having control over. Right. …so the coping strategies, I think, are great, if you are a single person with no other people dependent on you.* [32-year-old participant]

Participants with higher levels of stress or those who described their situation as more complex noted that the stress management strategies on the website were good for "normal stress," but that these strategies were not able to address the specific mental health needs during the COVID-19 pandemic.

> *I think that the stress from COVID-19, is just very…it's very unique, in the fact that like obviously none of us have really been through a pandemic like this. But I think normal coping mechanisms may not work as well, just because the nature of COVID. …you can't get out of the pandemic. There's no escaping it. So, I feel like the stress management strategies were very good stress management strategies for if you had normal stress.* [32-year-old participant]

A suggestion by 3 participants was to have a feature that allows the user to connect to a live counsellor for more personalized support, in order to get immediate help. While access to a live counsellor was not available, the website included an online peer-to-peer "Discussion Forum" where moderators were available during business hours to provide general information and encourage the use of available resources. However, only 1 of the 10 interview participants was aware of the discussion forum feature on the website.

> *…the only thing that you could improve, it would if you could connect somebody with a live person. I think it should lead to someone, if somebody really needs to talk to someone, I think that they should be able to…. And I…believe me, I've tried a lot… I think that they should be able to connect to a live person. Because people who are desperate, need humanity.* [72-year-old participant]

> *I think one of the big things that I was kind of like getting to, was just that there wasn't…there wasn't an option to get help right away.* [35-year-old participant]

Discussion

Principal Findings

The goals of the study were to collect information on the reach, usability, and user perceptions of the CAMH COVID-19 website, a web page designed to be a digital, plain-language, nonstigmatizing psychoeducational resource for people experiencing mild to moderate distress due to COVID-19.

Compared with similar websites [38], the CAMH COVID-19 website [22] is a highly accessed website, with 146,978 unique individuals visiting the website in a 10-month period. The majority of the 146,978 users were from Toronto and Ottawa, which is consistent with the findings of a similar study evaluating a Canadian mental health website portal [39] and may be reflective of the increase in the percentage of Ontario residents who indicated a poor or fair mental health status during the second wave of the pandemic [40]. The most visited section was the home page [22], followed by the Coping with Stress and Anxiety [37] section, which shares different coping ideas and strategies, and includes various self-assessment tests that

help users understand/determine the stress levels and anxiety they are experiencing, such as the PSS and GAD-7. Other studies have similar findings that the home page was viewed most frequently [41,42], with the second being a "self-check page, on which users could complete a depression screening tool" [41].

Users rated the usability of the website higher after changes were made on April 23, 2021. During the first period of the survey (February 20, 2021, to April 23, 2021), the average SUS score was 66.44 (SD 16.72). The SUS increased by 3.6 points after changes were made to the website (SUS score 70, SD 17.02). Both these scores correlated with an acceptable range [31,43,44] and were similar to results obtained in other evaluations of mental health websites [45].

Interview participants indicated that in the first iteration of the website, they felt overwhelmed with the volume of information and text-heavy web pages. Other studies have confirmed that text-heavy interventions can make content difficult to digest, thereby acting as a barrier to use [46], and website design related to page complexity, navigational simplicity, and comprehensibility should be considered to "reduce the cognitive effort required to effectively use eHealth applications" [47].

Demographic data indicated that the majority of website users were white women with higher education aged between 20 and 59 years. This is consistent with the findings of other studies that females and adults aged 18-64 years, with high levels of socioeconomic status, are more engaged, than males and other age groups, in using health content–related websites [39,48-54]. The website's reach may be a reflection of recent evidence showing that women perceived a decline in their mental health in 2020 during the second wave of the pandemic and experienced worse mental health outcomes in comparison with men [1,25,40,55]. Moreover, the gender differences in the website's reach may be attributed to men's reluctance to access mental health care because the intervention is not sufficiently sensitive to masculine identity–related factors [56]. Given the disproportionate negative impact of COVID-19 on diverse groups [40], there is a need to reach a wider audience. Targeted strategies need to be informed by gender sensitized services [57] and are needed to provide mental health services to communities who have been disproportionately impacted by the COVID-19 pandemic, such as older adults, indigenous people, and other racial and ethnic minority groups [58]. However, research shows that these communities (older adults, indigenous peoples, and other racial and ethnic minority groups) are the least likely to seek web-based health information [25,52,54]. Proposed strategies to increase user engagement and technology acceptance of e-mental health among diverse populations include the implementation of a human-centered design approach that integrates designers and users in the development process from beginning to end [59].

Interview findings showed that most participants felt positively about the website, as it was a trustworthy source of mental health information and had helpful stress reduction features. Studies have confirmed that individuals seeking mental health resources online are concerned about the trustworthiness and credibility of e-mental health, as well as the effectiveness of self-help strategies [60-62], and as shown by another study, a lack of trust in a single online source can contribute to health information seekers' stress [63].

Several interview participants wanted to connect to a live counsellor for more personalized support and immediate help. While moderators of the website's "Discussion Forum" [64] offered informational support, though not live counselling and support limited to office hours, most interview participants were unaware of this feature. This may be related to the design of the website and participants' feedback that they found the website overwhelming and text heavy. Making the discussion forum more visible might be important, especially given that research has shown that well-designed self-guided interventions can be successful without added human support and their associated costs [65-68].

The website was designed to address the needs of individuals with low to moderate levels of stress. However, the results indicated that most of the individuals visiting the website had moderate to high levels of stress. Further research is needed to understand the characteristics of individuals with low levels of stress and their low engagement with the website. Technology-enabled care coordination, which involves the use of online technology to facilitate delivery of multidisciplinary team-based care, for patients with low to moderate levels of distress has significant potential for improving population mental health [69]. Improvements to the website are needed to address the needs of individuals with high levels of stress, since studies have shown that individuals with more severe mental health symptoms are motivated to use digital mental health interventions, but symptom severity can also form barriers to engagement [23]. A guided approach, which includes the provision of content that is relevant and customizable according to users' personal preferences and offers technical assistance or training, may be suited to address the mental health of individuals with higher levels of stress [23].

Our findings demonstrate that the CAMH COVID-19 website was useful for disseminating information about evidence-based stress reduction strategies. The website is a highly accessed website with above average usability. Users found the website to be a helpful and trustworthy source of mental health information. The recommendations suggest that the design and reach of the website need to align more closely with user needs and with greater and more equitable access to mental health care. These evaluations should be conducted with an equity-oriented approach. For example, examination of who was exposed to the website (ie, reach) allowed us to understand the need to focus on equitable reach throughout any future updates. Feedback from this evaluation will be useful for further improvements to the website, which will be beneficial to both the general public and health care professionals.

Strengths and Limitations

An important strength of this project was the mixed methods design, as it allowed us to understand the complex problem of stress during the first waves of COVID-19 in Ontario and understand what users need as part of a website intervention. Another strength of this project is that based on direct feedback from users, we were able to make significant improvements to

the usability of the website. Our study has several limitations. First, we had a convenient sample and a low number of respondents compared with the number of users in both our survey and interview. Thus, they may not have been representative samples, and it is possible that those who perceived to have gained benefits from the website were more likely to participate in the surveys and interviews. In addition, the small sample of people who agreed to be interviewed (all women) also limited the ability to stratify findings by demographics or other variables of interest. Second, participants were interviewed a number of weeks after having used the website and completing the survey. As such, for some, recalling the program details was difficult. Third, there was a lack of longitudinal follow-up data on the effects of the website. Our study was restricted to the reach, usability, and user perception of the website and did not evaluate clinical efficacy. Lastly, anxiety and depression were not addressed, as the developers felt that individuals with these diagnosable illnesses needed either in-person or virtual clinical services.

Conclusion

Adding stress reduction resources to a website from a well-respected institution may be a practical method to increase awareness and access to evidence-based stress-reduction resources during times of crisis, where there is severe disruption to usual health care contacts. Efforts are needed to ensure that these resources are more widely accessed, especially by diverse populations. Further research is required on the health impact and role of e-mental health initiatives within the mental health services system. Given the promise and significant investment in digital mental health technologies, there is a growing need to evaluate them for low-intensity psychoeducational interventions.

Acknowledgments

We thank all the people who participated in this study. We would also like to thank Freni Jivani for her support with organizing the in-application analytics data. We thank Chantal Fougere, Mathangee Lingam, Stephanie Duench, Iqra Ashfaq, Lenka Vojtila, Huma Taj, and Victoria Kim from the Nicotine Dependence Service department for their contributions to the website's discussion board. We thank Shirley Lin and Alan Tang (infrastructure support), Rino La Grassa and Gurpal Bubbra (app), and Anthony Gallucci (web page) from IMG for technical set up; Samantha Delenardo, Bruno Marsiaj, Paris Semanksy, and Miguel Amante from Public Affairs; and Cameron Norman (https://www.cense.ca/contact/principal/) for review and update of the website.

Authors' Contributions

PS conceptualized the resource for people with mild to moderate distress during a pandemic; PS and RD designed the website; PS and NM conceptualized, designed, and supervised the evaluation with input from RD, GS, and SAK. AG and AS coordinated the study recruitment and conducted the interviews. AG and AS coded and analyzed the interviews under the supervision of NM. NM, AG, and AS drafted the manuscript. All authors participated in the critical revision of the manuscript, and read and approved the final manuscript.

Conflicts of Interest

PS reports receiving grants, salary, and/or research support from the Centre for Addiction and Mental Health, Health Canada, Ontario Ministry of Health, Canadian Institutes of Health Research (CIHR), Canadian Centre on Substance Use and Addiction, Public Health Agency of Canada (PHAC), Medical Psychiatry Alliance, Canadian Cancer Society Research Institute (CCSRI), Cancer Care Ontario, and National Institutes of Health (NIH). PS also reports receiving funding and/or honoraria from the following commercial organizations: Pfizer Inc./Canada and Patient-Centered Outcomes Research Institute. Through an open-tender process Johnson & Johnson, Novartis, and Pfizer Inc are vendors of record for providing smoking cessation pharmacotherapy, free or discounted, for research studies in which PS is the principal investigator or co-investigator.

References

1. Xiong J, Lipsitz O, Nasri F, Lui LMW, Gill H, Phan L, et al. Impact of COVID-19 pandemic on mental health in the general population: A systematic review. J Affect Disord 2020 Dec 01;277:55-64 [FREE Full text] [doi: 10.1016/j.jad.2020.08.001] [Medline: 32799105]
2. Lowe C, Keown-Gerrard J, Ng CF, Gilbert TH, Ross KM. COVID-19 pandemic mental health trajectories: Patterns from a sample of Canadians primarily recruited from Alberta and Ontario. Canadian Journal of Behavioural Science-Revue Canadienne Des Sciences Du Comportement 2022 Jan 27:1665682. [doi: 10.1037/cbs0000313]
3. Cindrich SL, Lansing JE, Brower CS, McDowell CP, Herring MP, Meyer JD. Associations Between Change in Outside Time Pre- and Post-COVID-19 Public Health Restrictions and Mental Health: Brief Research Report. Front Public Health 2021;9:619129 [FREE Full text] [doi: 10.3389/fpubh.2021.619129] [Medline: 33585393]
4. Chum A, Nielsen A, Bellows Z, Farrell E, Durette P, Banda JM, et al. Changes in Public Response Associated With Various COVID-19 Restrictions in Ontario, Canada: Observational Infoveillance Study Using Social Media Time Series Data. J Med Internet Res 2021 Aug 25;23(8):e28716 [FREE Full text] [doi: 10.2196/28716] [Medline: 34227996]
5. 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;57(1):101-110 [FREE Full text] [doi: 10.1007/s10597-020-00728-y] [Medline: 33108569]
6. Hawryluck L, Gold WL, Robinson S, Pogorski S, Galea S, Styra R. SARS control and psychological effects of quarantine, Toronto, Canada. Emerg Infect Dis 2004 Jul;10(7):1206-1212 [FREE Full text] [doi: 10.3201/eid1007.030703] [Medline: 15324539]
7. Mental Health and COVID-19: Early evidence of the pandemic's impact: Scientific brief, 2 March 2022. World Health Organization. 2022. URL: https://www.who.int/publications/i/item/WHO-2019-nCoV-Sci_Brief-Mental_health-2022.1 [accessed 2022-08-20]
8. Sorkin DH, Janio EA, Eikey EV, Schneider M, Davis K, Schueller SM, et al. Rise in Use of Digital Mental Health Tools and Technologies in the United States During the COVID-19 Pandemic: Survey Study. J Med Internet Res 2021 Apr 16;23(4):e26994 [FREE Full text] [doi: 10.2196/26994] [Medline: 33822737]
9. Mahoney AEJ, Elders A, Li I, David C, Haskelberg H, Guiney H, et al. A tale of two countries: Increased uptake of digital mental health services during the COVID-19 pandemic in Australia and New Zealand. Internet Interv 2021 Sep;25:100439 [FREE Full text] [doi: 10.1016/j.invent.2021.100439] [Medline: 34401395]
10. Tsao S, Chen H, Tisseverasinghe T, Yang Y, Li L, Butt ZA. What social media told us in the time of COVID-19: a scoping review. The Lancet Digital Health 2021 Mar;3(3):e175-e194. [doi: 10.1016/s2589-7500(20)30315-0]
11. Htay MNN, Parial LL, Tolabing MC, Dadaczynski K, Okan O, Leung AYM, et al. Digital health literacy, online information-seeking behaviour, and satisfaction of Covid-19 information among the university students of East and South-East Asia. PLoS One 2022;17(4):e0266276 [FREE Full text] [doi: 10.1371/journal.pone.0266276] [Medline: 35417478]
12. Su Z, McDonnell D, Wen J, Kozak M, Abbas J, Šegalo S, et al. Mental health consequences of COVID-19 media coverage: the need for effective crisis communication practices. Global Health 2021 Jan 05;17(1):4 [FREE Full text] [doi: 10.1186/s12992-020-00654-4] [Medline: 33402169]
13. Sockalingam S, Arena A, Serhal E, Mohri L, Alloo J, Crawford A. Building Provincial Mental Health Capacity in Primary Care: An Evaluation of a Project ECHO Mental Health Program. Acad Psychiatry 2018 Aug;42(4):451-457. [doi: 10.1007/s40596-017-0735-z] [Medline: 28593537]
14. Addressing the Access Gap: Leveraging the Potential of e-Mental Health in Canada. Mental Health Commission of Canada. 2017. URL: https://mentalhealthcommission.ca/resource/addressing-the-access-gap-leveraging-the-potential-of-e-mental-health-in-canada/ [accessed 2022-08-20]
15. Buchanan H, Coulson NS. Accessing dental anxiety online support groups: an exploratory qualitative study of motives and experiences. Patient Educ Couns 2007 Jun;66(3):263-269. [doi: 10.1016/j.pec.2006.12.011] [Medline: 17320336]
16. Meier A, Lyons EJ, Frydman G, Forlenza M, Rimer BK. How cancer survivors provide support on cancer-related Internet mailing lists. J Med Internet Res 2007 May 14;9(2):e12 [FREE Full text] [doi: 10.2196/jmir.9.2.e12] [Medline: 17513283]
17. Eysenbach G, Powell J, Englesakis M, Rizo C, Stern A. Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. BMJ 2004 May 15;328(7449):1166 [FREE Full text] [doi: 10.1136/bmj.328.7449.1166] [Medline: 15142921]
18. Klemm P, Bunnell D, Cullen M, Soneji R, Gibbons P, Holecek A. Online cancer support groups: a review of the research literature. Comput Inform Nurs 2003;21(3):136-142. [doi: 10.1097/00024665-200305000-00010] [Medline: 12792194]
19. Finn J. An exploration of helping processes in an online self-help group focusing on issues of disability. Health Soc Work 1999 Aug;24(3):220-231. [doi: 10.1093/hsw/24.3.220] [Medline: 10505283]
20. Bäuerle A, Jahre L, Teufel M, Jansen C, Musche V, Schweda A, et al. Evaluation of the E-Mental Health Mindfulness-Based and Skills-Based "CoPE It" Intervention to Reduce Psychological Distress in Times of COVID-19: Results of a Bicentre Longitudinal Study. Front Psychiatry 2021;12:768132 [FREE Full text] [doi: 10.3389/fpsyt.2021.768132] [Medline: 34803775]
21. Centre for Addiction and Mental Health. URL: https://www.camh.ca/ [accessed 2022-08-20]
22. Mental Health and the COVID-19 Pandemic. Centre for Addiction and Mental Health. URL: https://www.camh.ca/en/health-info/mental-health-and-covid-19 [accessed 2022-08-20]
23. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and Facilitators of User Engagement With Digital Mental Health Interventions: Systematic Review. J Med Internet Res 2021 Mar 24;23(3):e24387 [FREE Full text] [doi: 10.2196/24387] [Medline: 33759801]
24. Chang BP, Kessler RC, Pincus HA, Nock MK. Digital approaches for mental health in the age of covid-19. BMJ 2020 Jun 29;369:m2541. [doi: 10.1136/bmj.m2541] [Medline: 32601049]
25. Strudwick G, Sockalingam S, Kassam I, Sequeira L, Bonato S, Youssef A, et al. Digital Interventions to Support Population Mental Health in Canada During the COVID-19 Pandemic: Rapid Review. JMIR Ment Health 2021 Mar 02;8(3):e26550 [FREE Full text] [doi: 10.2196/26550] [Medline: 33650985]
26. Balcombe L, De Leo D. Digital Mental Health Challenges and the Horizon Ahead for Solutions. JMIR Ment Health 2021 Mar 29;8(3):e26811 [FREE Full text] [doi: 10.2196/26811] [Medline: 33779570]
27. Balcombe L, De Leo D. An Integrated Blueprint for Digital Mental Health Services Amidst COVID-19. JMIR Ment Health 2020 Jul 22;7(7):e21718 [FREE Full text] [doi: 10.2196/21718] [Medline: 32668402]

28. O'Cathain A, Murphy E, Nicholl J. The quality of mixed methods studies in health services research. J Health Serv Res Policy 2008 Apr;13(2):92-98. [doi: 10.1258/jhsrp.2007.007074] [Medline: 18416914]
29. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [Medline: 6668417]
30. System Usability Scale (SUS). Usability.Gov. URL: https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html [accessed 2022-08-20]
31. Chin WSY, Kurowski A, Gore R, Chen G, Punnett L, SHIFT Research Team. Use of a Mobile App for the Process Evaluation of an Intervention in Health Care: Development and Usability Study. JMIR Form Res 2021 Oct 28;5(10):e20739 [FREE Full text] [doi: 10.2196/20739] [Medline: 34709186]
32. Sandelowski M. Sample size in qualitative research. Res Nurs Health 1995 Apr;18(2):179-183. [doi: 10.1002/nur.4770180211] [Medline: 7899572]
33. ISO 9241-420:2011 Ergonomics of human-system interaction — Part 420: Selection of physical input devices. ISO. URL: https://www.iso.org/standard/52938.html [accessed 2022-08-20]
34. O'Connor C, Joffe H. Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines. International Journal of Qualitative Methods 2020 Jan 22;19:160940691989922. [doi: 10.1177/1609406919899220]
35. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
36. Research Ethics. Centre for Addiction and Mental Health. URL: https://www.camh.ca/en/science-and-research/research-ethics [accessed 2022-08-20]
37. Coping with stress and anxiety. Centre for Addiction and Mental Health. URL: https://www.camh.ca/en/health-info/mental-health-and-covid-19/coping-with-stress-and-anxiety [accessed 2022-08-20]
38. Dirmaier J, Liebherz S, Sänger S, Härter M, Tlach L. Psychenet.de: Development and process evaluation of an e-mental health portal. Inform Health Soc Care 2016;41(3):267-285. [doi: 10.3109/17538157.2015.1008486] [Medline: 25710352]
39. Jeong D, Cheng M, St-Jean M, Jalali A. Evaluation of eMentalHealth.ca, a Canadian Mental Health Website Portal: Mixed Methods Assessment. JMIR Ment Health 2019 Sep 06;6(9):e13639 [FREE Full text] [doi: 10.2196/13639] [Medline: 31493328]
40. Self-perceived mental health and mental health care needs during the COVID-19 pandemic. Statistics Canada. URL: https://www150.statcan.gc.ca/n1/pub/45-28-0001/2021001/article/00031-eng.htm [accessed 2022-08-20]
41. Ogrodniczuk JS, Beharry J, Oliffe JL. An Evaluation of 5-Year Web Analytics for HeadsUpGuys: A Men's Depression E-Mental Health Resource. Am J Mens Health 2021 Dec 03;15(6):15579883211063322 [FREE Full text] [doi: 10.1177/15579883211063322] [Medline: 34861812]
42. King K, Schlichthorst M, Turnure J, Phelps A, Spittal MJ, Pirkis J. Evaluating the effectiveness of a website about masculinity and suicide to prompt help-seeking. Health Promot J Austr 2019 Sep;30(3):381-389. [doi: 10.1002/hpja.237] [Medline: 30803102]
43. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. Journal of Usability Studies 2009;4(3):114-123 [FREE Full text]
44. Mol M, van Schaik A, Dozeman E, Ruwaard J, Vis C, Ebert DD, Ince, et al. Dimensionality of the system usability scale among professionals using internet-based interventions for depression: a confirmatory factor analysis. BMC Psychiatry 2020 May 12;20(1):218 [FREE Full text] [doi: 10.1186/s12888-020-02627-8] [Medline: 32398111]
45. Radovic A, Gmelin T, Hua J, Long C, Stein BD, Miller E. Supporting Our Valued Adolescents (SOVA), a Social Media Website for Adolescents with Depression and/or Anxiety: Technological Feasibility, Usability, and Acceptability Study. JMIR Ment Health 2018 Feb 26;5(1):e17 [FREE Full text] [doi: 10.2196/mental.9441] [Medline: 29483067]
46. Eccles H, Nannarone M, Lashewicz B, Attridge M, Marchand A, Aiken A, et al. Barriers to the Use of Web-Based Mental Health Programs for Preventing Depression: Qualitative Study. JMIR Form Res 2021 Jul 15;5(7):e16949 [FREE Full text] [doi: 10.2196/16949] [Medline: 34264195]
47. Rotondi AJ, Spring MR, Hanusa BH, Eack SM, Haas GL. Designing eHealth Applications to Reduce Cognitive Effort for Persons With Severe Mental Illness: Page Complexity, Navigation Simplicity, and Comprehensibility. JMIR Hum Factors 2017 Jan 05;4(1):e1 [FREE Full text] [doi: 10.2196/humanfactors.6221] [Medline: 28057610]
48. Baumann E, Czerwinski F, Reifegerste D. Gender-Specific Determinants and Patterns of Online Health Information Seeking: Results From a Representative German Health Survey. J Med Internet Res 2017 Apr 04;19(4):e92 [FREE Full text] [doi: 10.2196/jmir.6668] [Medline: 28377367]
49. Bidmon S, Terlutter R. Gender Differences in Searching for Health Information on the Internet and the Virtual Patient-Physician Relationship in Germany: Exploratory Results on How Men and Women Differ and Why. J Med Internet Res 2015 Jun 22;17(6):e156 [FREE Full text] [doi: 10.2196/jmir.4127] [Medline: 26099325]
50. Gallagher S, Doherty DT. Searching for health information online: characteristics of online health seekers. J Evid Based Med 2009 May;2(2):99-106. [doi: 10.1111/j.1756-5391.2009.01021.x] [Medline: 21348996]
51. Escoffery C. Gender Similarities and Differences for e-Health Behaviors Among U.S. Adults. Telemed J E Health 2018 May;24(5):335-343. [doi: 10.1089/tmj.2017.0136] [Medline: 28813630]

52. Yoon H, Jang Y, Vaughan PW, Garcia M. Older Adults' Internet Use for Health Information: Digital Divide by Race/Ethnicity and Socioeconomic Status. J Appl Gerontol 2020 Jan;39(1):105-110. [doi: 10.1177/0733464818770772] [Medline: 29661052]
53. Gandrakota N, Ali MK, Shah MK. Trends in Health Information Technology Use Among the US Population With and Without Cardiovascular Risk Factors, 2012-2018: Evidence From the National Health Interview Survey. JMIR Public Health Surveill 2021 Sep 30;7(9):e29990 [FREE Full text] [doi: 10.2196/29990] [Medline: 34591026]
54. Mitchell UA, Chebli PG, Ruggiero L, Muramatsu N. The Digital Divide in Health-Related Technology Use: The Significance of Race/Ethnicity. Gerontologist 2019 Jan 09;59(1):6-14. [doi: 10.1093/geront/gny138] [Medline: 30452660]
55. Anxiety, feelings of depression and loneliness among Canadians spikes to highest levels since spring 2020. Centre for Addiction and Mental Health. 2022. URL: https://www.camh.ca/en/camh-news-and-stories/anxiety-depression-loneliness-among-canadians-spikes-to-highest-levels [accessed 2022-08-20]
56. Rice SM, Oliffe JL, Kealy D, Seidler ZE, Ogrodniczuk JS. Men's Help-Seeking for Depression: Attitudinal and Structural Barriers in Symptomatic Men. J Prim Care Community Health 2020;11:2150132720921686 [FREE Full text] [doi: 10.1177/2150132720921686] [Medline: 32410489]
57. Oliffe JL, Rice S, Kelly MT, Ogrodniczuk JS, Broom A, Robertson S, et al. A mixed-methods study of the health-related masculine values among young Canadian men. Psychology of Men & Masculinities 2019 Jul;20(3):310-323. [doi: 10.1037/men0000157]
58. Jenkins EK, McAuliffe C, Hirani S, Richardson C, Thomson KC, McGuinness L, et al. A portrait of the early and differential mental health impacts of the COVID-19 pandemic in Canada: Findings from the first wave of a nationally representative cross-sectional survey. Prev Med 2021 Apr;145:106333 [FREE Full text] [doi: 10.1016/j.ypmed.2020.106333] [Medline: 33509605]
59. Vial S, Boudhraâ S, Dumont M. Human-Centered Design Approaches in Digital Mental Health Interventions: Exploratory Mapping Review. JMIR Ment Health 2022 Jun 07;9(6):e35591 [FREE Full text] [doi: 10.2196/35591] [Medline: 35671081]
60. Christensen H. E-Mental Health in Australia: Implications of the Internet and Related Technologies for Policy. Canberra, Australia: Commonwealth Department of Health and Ageing; 2002.
61. Moock J. Support from the Internet for Individuals with Mental Disorders: Advantages and Disadvantages of e-Mental Health Service Delivery. Front Public Health 2014;2:65 [FREE Full text] [doi: 10.3389/fpubh.2014.00065] [Medline: 24967221]
62. 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]
63. Choudhury A, Asan O, Alelyani T. Exploring the Role of the Internet, Care Quality and Communication in Shaping Mental Health: Analysis of the Health Information National Trends Survey. IEEE J Biomed Health Inform 2022 Jan;26(1):468-477. [doi: 10.1109/JBHI.2021.3087083] [Medline: 34097623]
64. Coronavirus 2019 (COVID-19) Discussion Forum. Centre for Addiction and Mental Health. URL: https://covid19.camhx.ca/mod/forum/view.php?id=1 [accessed 2022-08-20]
65. 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]
66. Shim M, Mahaffey B, Bleidistel M, Gonzalez A. A scoping review of human-support factors in the context of Internet-based psychological interventions (IPIs) for depression and anxiety disorders. Clin Psychol Rev 2017 Nov;57:129-140. [doi: 10.1016/j.cpr.2017.09.003] [Medline: 28934623]
67. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — A systematic review. Internet Interventions 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
68. Heber E, Ebert DD, Lehr D, Cuijpers P, Berking M, Nobis S, et al. The Benefit of Web- and Computer-Based Interventions for Stress: A Systematic Review and Meta-Analysis. J Med Internet Res 2017 Feb 17;19(2):e32 [FREE Full text] [doi: 10.2196/jmir.5774] [Medline: 28213341]
69. Iorfino F, Occhipinti J, Skinner A, Davenport T, Rowe S, Prodan A, et al. The Impact of Technology-Enabled Care Coordination in a Complex Mental Health System: A Local System Dynamics Model. J Med Internet Res 2021 Jun 30;23(6):e25331 [FREE Full text] [doi: 10.2196/25331] [Medline: 34077384]

Abbreviations

CAMH: Centre for Addiction and Mental Health
GAD-7: Generalized Anxiety Disorder-7
PSS-10: Perceived Stress Scale-10
REDCap: Research Electronic Data Capture
SUS: System Usability Scale

Edited by J Torous; submitted 27.05.22; peer-reviewed by R Poss-Doering, L Balcombe; comments to author 01.07.22; revised version received 05.08.22; accepted 07.08.22; published 26.08.22.

Please cite as:
Minian N, Gayapersad A, Saiva A, Dragonetti R, Kidd SA, Strudwick G, Selby P
An e–Mental Health Resource for COVID-19–Associated Stress Reduction: Mixed Methods Study of Reach, Usability, and User Perceptions
JMIR Ment Health 2022;9(8):e39885
URL: https://mental.jmir.org/2022/8/e39885
doi:10.2196/39885
PMID:35960596

Review

Impact of the COVID-19 Pandemic on the Global Delivery of Mental Health Services and Telemental Health: Systematic Review

Caroline Zangani[1], MD; Edoardo G Ostinelli[1], MD; Katharine A Smith[1], MD, PhD; James S W Hong[1], MD; Orla Macdonald[2], MSc; Gurpreet Reen[3], PhD; Katherine Reid[2], MD; Charles Vincent[3], PhD; Rebecca Syed Sheriff[1], MD, PhD; Paul J Harrison[1], MD; Keith Hawton[1], DSc, FMedSci; Alexandra Pitman[4], MD, PhD; Rob Bale[2], MD; Seena Fazel[1], MD, PhD; John R Geddes[1], MD; Andrea Cipriani[1,2], MD, PhD

[1]Department of Psychiatry, University of Oxford, Oxford, United Kingdom

[2]Oxford Health NHS Foundation Trust, Warneford Hospital, Oxford, United Kingdom

[3]Department of Experimental Psychology, University of Oxford, Oxford, United Kingdom

[4]Division of Psychiatry, University College London, London, United Kingdom

Corresponding Author:
Andrea Cipriani, MD, PhD
Oxford Health NHS Foundation Trust
Warneford Hospital
Oxford, OX3 7JX
United Kingdom
Phone: 44 1865 618228
Email: andrea.cipriani@psych.ox.ac.uk

Abstract

Background: The COVID-19 pandemic required mental health services around the world to adapt quickly to the new restrictions and regulations put in place to reduce the risk of transmission. As face-to-face contact became difficult, virtual methods were implemented to continue to safely provide mental health care. However, it is unclear to what extent service provision transitioned to telemental health worldwide.

Objective: We aimed to systematically review the global research literature on how mental health service provision adapted during the first year of the pandemic.

Methods: We searched systematically for quantitative papers focusing on the impact of the COVID-19 pandemic on mental health services published until April 13, 2021, in the PubMed, Embase, medRxiv, and bioXriv electronic bibliographic databases, using the COVID-19 Open Access Project online platform. The screening process and data extraction were independently completed by at least two authors, and any disagreement was resolved by discussion with a senior member of the team. The findings were summarized narratively in the context of each country's COVID-19 Stringency Index, which reflects the stringency of a government's response to COVID-19 restrictions at a specific time.

Results: Of the identified 24,339 records, 101 papers were included after the screening process. Reports on general services (n=72) showed that several countries' face-to-face services reduced their activities at the start of the pandemic, with reductions in the total number of delivered visits and with some services forced to close. In contrast, telemental health use rapidly increased in many countries across the world at the beginning of the pandemic (n=55), with almost complete virtualization of general and specialistic care services by the end of the first year. Considering the reported COVID-19 Stringency Index values, the increased use of virtual means seems to correspond to periods when the Stringency Index values were at their highest in several countries. However, due to specific care requirements, telemental health could not be used in certain subgroups of patients, such as those on clozapine or depot treatments and those who continued to need face-to-face visits.

Conclusions: During the pandemic, mental health services had to adapt quickly in the short term, implementing or increasing the use of telemental health services across the globe. Limited access to digital means, poor digital skills, and patients' preferences and individual needs may have contributed to differences in implementing and accessing telemental health services during the pandemic. In the long term, a blended approach, combining in-person and virtual modalities, that takes into consideration the needs, preferences, and digital skills of patients may better support the future development of mental health services. It will be required to improve confidence with digital device use, training, and experience in all modalities for both clinicians and service users.

(JMIR Ment Health 2022;9(8):e38600) doi:10.2196/38600

KEYWORDS

COVID-19; coronavirus; mental health services; telemental health; telepsychiatry; face-to-face

Introduction

At the start of the COVID-19 pandemic, mental health services around the world, along with health services in general, needed to adapt quickly to reduce the risk of infection and transmission while continuing to support those with mental health problems [1]. However, as identified early in the COVID-19 pandemic by the World Health Organization (WHO), the pandemic went beyond the physical threat to also affect mental health [2].

The link between COVID-19 and mental illness is increasingly supported by robust evidence. The association is bidirectional; mental illness increases the risk of subsequent COVID-19 infection, and the risk of a new mental health diagnosis increases up to 180 days following COVID-19 infection [3,4]. COVID-19 and related coronavirus infections, such as SARS and Middle East respiratory syndrome (MERS), are associated with increased subsequent risk of depression, anxiety, and other neuropsychiatric consequences, such as dysexecutive syndrome [5-7]. This is in addition to any indirect effects on mental health caused by restrictions and lockdowns imposed by the COVID-19 pandemic [8]. Unemployment, financial insecurity, and other socioeconomic effects of the COVID-19 pandemic also adversely impact mental health. Mental health consequences can occur both at the time of a crisis and afterwards [9]. For instance, it has been suggested that suicide rates in some countries may have a delayed increase as a consequence of the social, health, and economic disruption associated with the pandemic [9]. As the pandemic evolved, access to vaccines and their uptake have altered the pattern of infection [10], and therefore, restrictive measures adopted by public health institutions have changed [11]. Taken together, all these factors suggest that there will be a rise in mental health treatment needs, and this is likely to be long-lasting. Consequently, mental health services have also started to plan or implement measures to prepare for an anticipated increase in demand related both directly to COVID-19 infection and indirectly to the sequelae of its associated restrictions.

During the pandemic, the WHO proposed that the goals of mental health care services need to focus not only on responding to the acute mental health emergency, but also on recovery thereafter and preparations for future emergencies [2]. How and to what extent this has happened is yet to be seen.

In contrast to the volume of data on the mental health impact of COVID-19, there have been few systematic considerations of the response of mental health services. Some reports reflect consultation with experts and service users to consider what changes in services might be needed to meet the mental health consequences of COVID-19 [12]. Organizations, such as the WHO, have suggested an integrated approach including mental health and psychosocial support in the COVID-19 response [2], while the United Nations has highlighted the need for changes and investment immediately to reduce mental health effects later [13].

Preliminary data from a WHO survey suggested a considerable detrimental impact on mental health services [14], with telemental health (TMH) considered critical to maintaining delivery of mental health services in response to pandemic-related social distancing measures and confinements [15]. Prompted by this, several countries modified national telemedicine regulations to promote its spread [16]. However, it is unclear whether the apparently rapid transition to TMH observed in many high-income countries [1] has occurred globally, which services have been affected, and whether this change has been sustained through the different phases of the pandemic.

To understand the degree to which service provision has changed during the first year of the pandemic and the extend of the transition to TMH in different countries, we systematically reviewed the available literature on mental health services during the pandemic. We also assessed these changes across the different phases of the COVID-19 pandemic and in the context of the local restrictions imposed.

Methods

Overview

We performed a systematic review of studies describing the impact of the COVID-19 pandemic on mental health services up to 1 year after the pandemic declaration on March 11, 2020 [17]. This is part of a wider assessment of the impact of the COVID-19 pandemic across the mental health field (previously published protocol for the full project is available on PROSPERO, CRD42020178819) [18].

Search Strategy and Screening Process

The search strategy for the full project (see above) included terms relating to mental health and the COVID-19 pandemic, SARS, and MERS, with specific and generic specifiers (eg, Medical Subject Headings [MeSH] terms; for full details, see Multimedia Appendix 1). We searched PubMed, Embase, medRxiv, and bioXriv electronic bibliographic databases using the COAP (COVID-19 Open Access Project) online platform [19] for relevant reports from inception until April 13, 2021. Records on SARS and MERS were independently searched on PubMed and Embase from inception to April 13, 2021.

At least two members of the review team (CZ, EGO, GR, JSWH, KR, KAS, OM, or RS) independently screened the title and abstract of the retrieved records. Full texts of the potentially eligible records were then independently assessed against the eligibility criteria. Any disagreement about eligibility was discussed with a third member of the research team (AC, AP, CV, or KH). The included articles and relevant systematic reviews were also screened for references to identify additional

records. References from all sources were cross-checked to ensure that all referenced documents had been searched.

Eligibility Criteria

To examine the consequences of the COVID-19 pandemic on mental health services, we included only reports providing quantitative data on changes in service organization and delivery (eg, how services were delivered: remote *versus* face to face; the number of patients accessing the service) and the creation of new services (eg, *ad hoc* created digital technologies and TMH) to overcome the challenges related to the COVID-19 pandemic. Documents solely reporting diagnostic aspects (eg, differential rates of two or more subgroups of mental health diagnoses accessing the service and frequency of mental health symptoms during the pandemic) were excluded, as these data describing part of a service do not provide information about the overall level of activity of a service. We included all types of primary research reports (randomized studies, observational studies, case reports, etc) focusing on mental health services and coronaviruses. No time or language restriction was applied.

Data Extraction

Data extraction was performed independently by at least two reviewers (CZ, EGO, GR, JSWH, KR, KAS, OM, or RS). Relevant data on country of origin, data collection period, service delivery method (ie, face-to-face or TMH), and service provision were extracted from the included papers. We categorized the identified records according to the following categories: randomized controlled trials, cohort studies, before-and-after studies, case-control studies, case reports/case series, and survey/audit. For each paper, we also extracted the key quantitative findings and collected them in a descriptive summary.

COVID-19 Stringency Index

Given that several waves of infections and consequent restrictions related to the COVID-19 pandemic affected different countries at different times, we considered the timing of the study period of the individual studies to be of limited use in contextualizing findings locally. Instead, we reported findings in the context of the country-specific COVID-19 Stringency Index, a composite measure developed within the Our World in Data project by the Global Change Data Lab (a nonprofit organization) and the University of Oxford [20] to reflect the stringency of a government's response to COVID-19 restrictions at any specific time. The score is based on 9 response indicators, including school closures, workplace closures, and travel bans, rescaled to a value from 0 to 100 (where 100 is the strictest measure of response). This is therefore a dynamic measure, capturing changes in a country's policies at any point throughout the pandemic. Where policies vary within a particular country, the index reflects the response level of the strictest subregion [11]. For each individual study, we report the minimum and maximum Stringency Index score for the relevant time period. The country-specific graphical representation of the COVID-19 Stringency Index starting from January 2020 can be accessed online [20].

Results

Overview

We identified a total of 24,339 records, and the number reduced by 3699 records after the removal of duplicates. After the abstract and full-text screening, an additional 20,539 records were excluded (Multimedia Appendix 2). A total of 101 papers were therefore included in the review. Of these, 60 were before-and-after studies, 23 were case reports and case series, and 18 were cross-sectional surveys. A summary of the 101 included articles is reported in Multimedia Appendix 3. Overall, the studies were concentrated in high-income countries, with the majority in Europe (39.6%) and North America (29.7%) (Multimedia Appendix 4).

Face-to-Face Mental Health Services

Of the 101 articles, 72 reported data on face-to-face mental health services. A total of 52 of these were before-and-after studies, 7 were case reports and case series, and 14 were cross-sectional surveys. Regionally, 31 studies reported data from Europe, 20 from North America, 1 from South America, 11 from Asia, and 9 from Australia (Multimedia Appendix 5). No studies reported quantitative data from low-income countries, and none were from the African subcontinent. As shown in Multimedia Appendix 3, several studies reported a reduction in face-to-face visits [21,22] and emergency department presentations [23], with a reduction in the activity of specialized settings, such as electroconvulsive therapy clinics [24-27]. Data on hospitalizations were less clear, with some studies reporting evidence of a reduced number of admissions (eg, [28,29]), others reporting evidence of a rapid increase in admissions soon after the pandemic onset (eg, [23,30]), and still others reporting no differences compared to the same period in the previous year (eg, [31,32]).

TMH Services

Table 1 [33-84] summarizes the characteristics of the subgroup of 55 articles reporting data on TMH services, which includes 23 before-after studies, 19 case reports or case series, and 13 cross-sectional surveys. Most of the collected data within these primary sources refer to the timeframe between January and October 2020, with most focusing on the period around March 2020.

The world distribution of the studies on TMH services (Figure 1) showed a higher representation in North America (n=18) and Europe (n=16, including 19 countries), while fewer articles reported data from Central or South America, Asia, and Australia. No studies reported quantitative data from low-income countries, and none were from the African subcontinent.

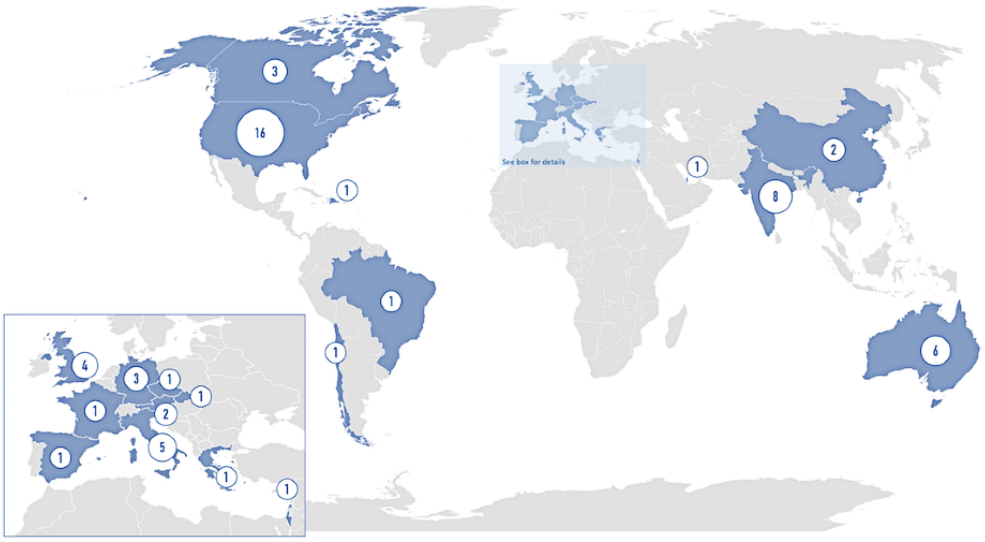
Table 1. Characteristics of the included studies on telemental health (N=55).

Study ID (authors, year)	Study design	Study setting	Country	Study period		Stringency Index, value or range
				Start date	End date	
Arendt et al, 2020 [33]	Before-after study	Crisis hotline services for the general population	Germany and Austria	January 1, 2020	April 20, 2020	0.00-81.48
Beran et al, 2020 [34]	Before-after study	A large, academic, consultation-liaison service	United States	January 2, 2020	June 6, 2020	5.56-72.69
Berthaut and Chamignon, 2021 [35]	Descriptive study	A health care and education center for adolescents	France	March 17, 2020	May 18, 2020	76.85-87.96
Cantini et al, 2020 [36]	Descriptive study	A psycho-oncology unit	Italy	March 3, 2020	June 3, 2020	67.59-93.52
Carpiniello et al, 2020 [37]	Survey	71 departments of mental health and 107 psychiatric wards in general hospitals	Italy	April 1, 2020	April 11, 2020	85.19-87.96
Connoly et al, 2020 [38]	Before-after study	The Department of Veterans Affairs	United States	March 11, 2020	April 22, 2020	21.76-72.69
Datta et al, 2020 [39]	Before-after study	An eating disorder multidisciplinary unit	United States	March 1, 2020	May 31, 2020	8.33-72.69
Garcia-Huidobro et al, 2020 [40]	Before-after study	A large private academic health network	Chile	March 1, 2020	April 30, 2020	0.00-73.15
Ghiretti et al, 2020 [41]	Descriptive study	A free telephone-based psychological support service	Italy	March 18, 2020	June 3, 2020	67.59-93.52
Ghosh et al, 2021 [42]	Before-after study	An outpatient service for substance use disorder	India	May 18, 2020	August 31, 2020	81.94-87.50
Graell et al, 2020 [43]	Before-after study	A hospital outpatient service	Spain	March 16, 2020	May 10, 2020	68.98-85.19
Grover et al, 2020 [26]	Survey	109 private and public mental health services	India	May 1, 2020	May 18, 2020	81.94-96.30
Grover et al, 2020 [27]	Survey	396 private mental health services	India	May 1, 2020	May 15, 2020	81.94-96.30
Guinart et al, 2021 [44]	Survey	818 mental health care professionals	United States	April 1, 2020	May 31, 2020	72.69
Hall and Sukhera, 2020 [45]	Descriptive study	A virtual emergency psychiatric consultation service	Canada	Unknown	Unknown	Unknown
Hames et al, 2020 [46]	Survey	93 clinic directors or designated representatives of the Association of Psychology Training Clinics	United States	March 11, 2020	March 31, 2020	21.76-72.69
Harrison et al, 2020 [47]	Descriptive study	A telephone-based program to address substance abuse within emergency departments	Canada and United States	Unknown	Unknown	Unknown
Hazarika et al, 2021 [48]	Survey	A psychological helpline for the general population	India	April 7, 2020	April 24, 2020	96.30-100.00
Hoffnung et al, 2021 [49]	Before-after study	A community behavioral health center	United States	January 1, 2020	June 30, 2020	0.00-72.69
Humer et al, 2020 [50]	Before-after study	338 health care professionals	Germany, Czech Republic, and Slovakia	March 24, 2020	May 20, 2020	54.63-87.04
Johnson et al, 2020 [51]	Survey	2180 mental health care workers	United Kingdom	January 1, 2020	May 1, 2020	0.00-72.69
Karim et al, 2020 [52]	Before-after study	Mental health outpatient services in Qatar	Qatar	March 1, 2020	June 30, 2020	13.89-86.11

Study ID (authors, year)	Study design	Study setting	Country	Study period		Stringency Index, value or range
				Start date	End date	
Khanra et al, 2021 [53]	Descriptive study	A large psychiatric hospital	India	April 1, 2020	September 30, 2020	81.94-100.00
Lian et al, 2020 [54]	Before-after study	A hospital-based psychological counseling program	China	January 23, 2020	June 30, 2020	44.91-81.94
Looi et al, 2020 [55]	Before-after study	Private psychiatric consultation services in Australia	Australia	April 1, 2020	June 30, 2020	62.04-73.15
Looi et al, 2020 [56]	Before-after study	Psychiatric consultation services in Australian rural settings (Australian Capital Territory, Northern Territory, South Australia, and Tasmania)	Australia	April 1, 2020	May 31, 2020	64.35-73.15
Looi et al, 2020 [57]	Before-after study	Private psychiatric consultation services in New South Wales, Queensland, Victoria, and Western Australia regions	Australia	April 1, 2020	May 31, 2020	64.35-73.15
Looi et al, 2021 [58]	Before-after study	Private psychiatric consultation services in Australia	Australia	July 1, 2020	September 30, 2020	68.06-75.46
Lunsky et al, 2021 [59]	Survey	942 direct support professionals working with people with intellectual and development disabilities	Canada	July 2, 2020	August 10, 2020	67.13-68.98
Mehtani et al, 2021 [60]	Descriptive study	A telephone-based program for people with substance use disorders staying at San Francisco's COVID-19 isolation and quarantine site	United States	April 10, 2020	May 25, 2020	72.69
Mishkind et al, 2020 [61]	Descriptive study	An outpatient clinic	United States	March 2, 2020	April 10, 2020	11.11-72.69
Moreland et al, 2021 [62]	Before-after study	An outpatient service for mental health and substance use treatment to pregnant and postpartum women	United States	March 1, 2020	May 31, 2020	8.33-72.69
Myers Virtue et al, 2021 [63]	Descriptive study	A psychosocial oncology service	United States	March 1, 2020	September 30, 2020	8.33-72.69
Naik et al, 2021 [64]	Descriptive study	An outpatient service	India	March 25, 2020	May 31, 2020	81.94-100.00
Peppou et al, 2020 [65]	Descriptive study	A helpline for the general population	Greece	March 22, 2020	April 13, 2020	74.04-84.26
Peralta et al, 2020 [66]	Descriptive study	A counseling service for the general population	Dominican Republic	March 25, 2020	May 17, 2020	92.59-100.00
Perricone et al, 2021 [67]	Descriptive study	A national psychological counseling service	Italy	April 1, 2020	June 30, 2020	67.59-93.52
Pierce et al, 2020 [68]	Survey	2619 psychologists	United States	May 11, 2020	May 25, 2020	72.69
Prior, 2020 [69]	Descriptive study	A psycho-oncology unit	Italy	Unknown	Unknown	Unknown
Probst et al, 2020 [70]	Survey	1547 psychotherapist	Austria	March 24, 2020	April 1, 2020	81.48
Rainwater et al, 2020 [71]	Before-after study	An outpatient service for consultation and counseling of cancer patients	United States	April 1, 2020	September 30, 2020	62.50-72.69
Ravindran et al, 2020 [72]	Descriptive study	A national helpline for the general population	India	Unknown	Unknown	Unknown
Rosen et al, 2020 [73]	Before-after study	The Veterans Health Administration – the largest integrated service in the United States	United States	January 1, 2020	June 30, 2020	0.00-72.69
Salum et al, 2020 [74]	Descriptive study	A community mental health service	Brazil	March 23, 2020	March 23, 2020	71.76

Study ID (authors, year)	Study design	Study setting	Country	Study period		Stringency Index, value or range
				Start date	End date	
Sampaio et al, 2021 [75]	Survey	768 mental health professionals	United States	April 24, 2020	May 18, 2020	72.69
Sharma et al, 2020 [76]	Descriptive study	A large child psychiatry department	United States	February 28, 2020	April 3, 2020	5.56-72.69
Singh Bhandari, 2020 [77]	Descriptive study	An outpatient clinic	India	March 25, 2020	May 26, 2020	81.94-100.00
Staples et al, 2020 [78]	Before-after study	A national digital mental health service for people experiencing anxiety and depression	Australia	March 19, 2020	June 10, 2020	44.44-73.15
Steeg et al, 2021 [79]	Before-after study	General practice services in England	United Kingdom	March 10, 2020	June 10, 2020	11.11-79.63
Stewart and Broadbent, 2020 [22]	Before-after study	A large mental health service	United Kingdom	February 16, 2020	April 16, 2020	11.11-79.63
Stewart et al, 2020 [80]	Before-after study	A large mental health service	United Kingdom	February 1, 2020	May 15, 2020	11.11-79.63
Vonderlin et al, 2021 [81]	Descriptive study	A hotline for psychological first aid for the general population	Germany	April 22, 2020	July 24, 2020	55.09-76.85
Yaffa et al, 2021 [82]	Before-after study	An eating disorder treatment center for adolescents	Israel	January 1, 2020	October 31, 2020	0.00-94.4
Zhong et al, 2020 [83]	Survey	108 Chinese residents	China	January 27, 2020	February 2, 2020	69.91-77.31
Zulfic et al, 2020 [84]	Survey	A community outpatient service	Australia	Unknown	Unknown	Unknown

Figure 1. The distribution of articles reporting data on telemental health. Values in the circles indicate the number of articles.



The key findings of included papers involving the impact of COVID-19 on the use of TMH are shown in Textbox 1. Overall, studies set in 19 countries reported the implementation of TMH services locally at the beginning of the pandemic. They include inpatient and outpatient services (eg, [71]), specialized services (eg, [38,82]), adolescent services (eg, [82]), and services in both private and public sectors (eg, [27]). For those services in which TMH was already available, an increase in its use during the pandemic was reported in several countries (eg, [21,50,70]). The specific COVID-19 Stringency Index chart for each country throughout the pandemic can be accessed online [20].

Several studies reported the creation of helplines both locally and within national programs in several countries to provide psychiatric and psychological help to the general population (eg, [48,65,67,72]).

Textbox 1. Telemental health services around the world. The specific COVID-19 Stringency Index chart for each country can be accessed online [20].

Asia

China

Before-after study

• A newly created hospital-based psychological counseling program in China served 474 users (7 users per day) during the lockdown, and the number reduced to 162 (2 users per day) after April 2020 [54].

Survey/audit

• 42.6% of 108 survey respondents who received a psychiatric consultation during lockdown attended an internet-based service (n=46), 37.0% attended a face-to-face visit (n=40), and 20.4% attended a telephone-based service (n=22) [83].

India

Before-after study

• In India, a telemedicine-assisted stepped-care outpatient service for substance use disorder was developed and started on the May 18, 2020. By August 31, 2020, 160 new and 219 follow-up patients were registered in the teleaddiction service. Among them, 128 (80%) and 198 (90.2%) patients received teleconsultations, respectively [42].

• Compared to the same period of 2019, the number of patients seen in a psychiatric outpatient service for patients with substance abuse dropped (May: 170 vs 22; June: 351 vs 43; July: 467 vs 44; August: 436 vs 51) [42].

Descriptive study

• A local newly created psychological helpline received 239 calls over the 18 days of lockdown (April 7-24, 2020) [48].

• The Indian Government helpline received 20,475 calls during the first month of activity [72].

• An outpatient service in India created a new telephone outpatient service and delivered 60.1% of the planned follow-up visits using telehealth (1049 of 1748) in the first 2 months of lockdown [64].

• During the first months of the lockdown, a team of psychiatrists of an Indian hospital made 78 teleconsultations with new and followed-up patients. In 6 cases, the patients were referred to their local hospital for further face-to-face assessment [77].

• In a large psychiatric hospital in India, 168 consultations were made using telepsychiatry between April and September 2020. The number of teleconsultations during lockdown was positively correlated with travel cost savings (ρ=0.47, P<.01) and time savings (ρ=0.49, P<.01) [53].

Survey/audit

• A survey comparing the services in both public and private institutes during the lockdown period in India (n=109) showed that telecommunication service use increased from 19.3% to 45.9% comparing the period before and during the pandemic [26].

• A survey evaluating the services in the private sector provided during the lockdown period in India (n=396) showed that telecommunication service use increased from 26.3% to 52.0% comparing the period before and during the pandemic [27].

Israel

Before-after study

• In an eating disorder treatment center for adolescents, telemedicine meetings comprised 37% of all sessions during January to October 2020 (2193 sessions), while they were not used during the respective period between 2015 and 2019 [82].

Qatar

Before-after study

• Analyzing data on the number of visits in the mental health outpatient setting in Qatar in the 4 months following the introduction of telepsychiatry (March-June 2020), the total number of individuals accessing mental health services (both face-to-face and telepsychiatry) increased by approximately 36.5% compared to the same period in 2019 (22,086 and 16,175 cases, respectively) [52].

Australia

Australia

Before-after study

• The use of the MindSpot Clinic, a national digital mental health service (DMHS) providing services to people experiencing anxiety and depression, increased by 16.7% (mean number of assessments per week=455) during the pandemic (March 19 to June 10, 2020) compared to the period between September 1 and September 28, 2019 (390 assessments per week) [78].

• Comparing data on private psychiatrists' visits held in April-June 2020 and the same period in 2019 in Australia, the number of psychiatry consultations (telehealth and face-to-face) rose during the pandemic by 14%, with telehealth representing nearly half of this total. Face-to-face consultations in 2020 were only 56% of the comparative number of 2019 consultations. Most telehealth involved short telephone consultations of ≤15-30 min. Video consultations comprised 38% of total telehealth provision [55].

• Comparing data on visits held in Australian rural settings (Australian Capital Territory [ACT], Northern Territory [NT], South Australia [SA], and Tasmania [TAS]), the overall rate of consultations (face-to-face and telehealth) increased during March and April 2020, compared to the monthly face-to-face consultation average of July 2018 to June 2019, except TAS (ACT: 1724 in April, increased 114%; 2061 in May, increased 136%; NT: 296 in April, increased 108%; 337 in May, increased 123%; SA: 12,864 in April, increased 116%; 12,876 in May, increased 116%; TAS: 1886 in April, reduced 85%; 2189 in May, reduced 98%). For total video and telephone telehealth consultations combined, video consultations were lower in April 2020 and higher in May 2020 [56].

• The total combined use of telehealth and face-to-face private psychiatric consultation services in New South Wales, Queensland, Victoria, and Western Australia regions (Australia) in April and May 2020 increased by 10%-20% of the average monthly face-to-face consultations in the 2018/19 financial year [57].

• Comparing data on private psychiatrists' visits held in July-September 2020 and the same period of 2019 in Australia, the number of psychiatry consultations (telehealth and face-to-face) rose during the pandemic by 14%, with telehealth representing 43% of this total. Face-to-face consultations in 2020 were only 64% of the comparative number of 2019 consultations. Most telehealth involved short telephone consultations of ≤15-30 min. Video consultations comprised 42% of total telehealth provision [58].

Survey/audit

• An audit of 314 community patients to examine the potential implications of telephone support found that 21 (7%) did not have access to a phone, and a further 58 (18%) were reported by the original authors as "unreliable in responding to contact over the phone based on past clinician experience." Moreover, during the first wave, regular face-to-face reviews were necessary for a group of patients, including 91 patients (29%) treated with depot medications, and 71 (23%) taking clozapine [84].

Europe

Austria

Survey/audit

• During the lockdown, the number of patients treated daily via telephone increased from 0.42 ± 3.01 to 4.53 ± 5.77 (+979%, P<.001), and via internet from 0.18 ± 1.35 to 2.99 ± 4.44 (+1561%, P<.001) [70].

Austria and Germany

Before-after study

• The number of crisis hotline calls increased during lockdown in both Austria and Germany [31].

Czech Republic, Germany, and Slovakia

Survey/audit

• A survey on 338 health care professionals showed that the number of patients treated via telephone increased from 0.92 ± 3.16 to 3.28 ± 5.22 per week (+257%, t −8.717, P<.001), and the number of patients treated via the internet increased from 0.59 ± 2.54 to 5.83 ± 6.82 per week (+888%, t −15.346, P<.001) during the 2-month period of confinement [50].

France

Descriptive study

• During lockdown, a health care and education center for adolescents that provides long-term psychiatric care, created a virtual ward using Discord, a platform used to communicate using chat, calls, and video calls, to ensure continuity of care. Of the 38 patients who used this service, only 6 withdrew early [33].

Germany

Descriptive study

• A total of 1292 telephone consultations were made by a telephone hotline for psychological first aid for COVID-19–related burden in 4 months covering a period during and after the German lockdown. In 42% of all consultations, a short therapeutic intervention was performed; in 26%, psychotherapeutic treatment was recommended; and in 11%, referral to other specialized telephone services was made [81].

Greece

Descriptive study

• In a report from a mental health helpline service in Greece during lockdown, most calls pertained to the quarantine (n=482, 83.7%) and more specifically to feelings of "restraint" (56%) and "loneliness" (53%) [65].

Italy

Descriptive study

• In Emilia Romagna region (Italy), a free telephone-based psychological support service was established during lockdown, receiving 312 calls in the first 11 weeks of activity [41].

• A national telephone psychological counseling service established during the COVID-19 pandemic had 193 users between April and June 2020 [67].

• A psycho-oncology unit reported having treated 28 cancer patients, 9 caregivers, and 7 family members during April-September 2020, switching from face-to-face visits to virtual consultations [36].

• In Treviso (Italy), the oncology unit switched their psychological monitoring to a virtual modality during the pandemic. In the first 9 weeks, they conducted 123 online visits (12% were video calls using Google Hangouts, and 88% were telephone calls) [69].

Survey/audit

• Data from 71 departments of mental health and 107 psychiatric wards in general hospitals showed that during the lockdown (April 2020), scheduled psychiatric consultations, both at home and on-site, went ahead for selected cases, being replaced in approximately 75% of cases by scheduled remote contact, mainly telephone calls (100%), video calls (67%), or emails (19%), with 41% of units adopting all these means of contact [37].

Spain

Before-after study

• During the early COVID-19 pandemic (March-May 2020) in a hospital in Madrid, out of 1818 outpatient consultations carried out, 1329 (73.10%) were delivered by telephone or videoconferencing and 489 (26.9%) were face-to-face, corresponding to 365 patients who were receiving treatment at the time in the outpatient clinic or day hospital [43].

United Kingdom

Before-after study

• The likelihood of receiving a remote general practitioner/practice nurse consultation within 3 months of a self-harm episode was higher in the COVID-19 pandemic (67.7%; March-April 2020) than in the prepandemic period (32.3%; ratio 2.10, CI 2.05-2.15; same period of 2010-2019). The overall likelihood of having a general practitioner/practice nurse consultation was slightly lower (80.3% vs 83.2%; ratio 0.97, CI 0.96-0.98) [69].

• Comparing the 31-day periods before and after March 16, 2020 (lockdown announcement), virtual contacts increase by 117% compared to a 3%-22% reduction observed comparing the same periods between 2015 and 2019 [25].

• Comparing the period before March 16 to that between March 16 and May 15, 2020, virtual contacts with community mental health teams increased from 154 ± 17 to 380 ± 97 (+147%). In the same period, mean virtual contacts increased by 102.7% (26.7 ± 7.3 vs 54.2 ± 14.8) and mean total contacts reduced by 24.9% (161.9 ± 30.7 vs 121.5 ± 26.1). Daily caseloads reduced by 2.1% (8729 ± 24 vs 8539 ± 124) and by 26.4% (221.8 ± 8.5 vs 163.3 ± 20.0) [80].

Survey/audit

• During the first COVID-19 wave, 61.1% of 2180 mental health care workers in the United Kingdom rated the adoption of new digital ways of working as “very or extremely important” in the management of the impact of COVID-19 at work [51].

North America

Canada

Descriptive study

• In Ontario, the implementation of a virtual emergency psychiatric consultation service (ie, Emergency Diversion Clinic) during the pandemic allowed provision of consultations for 60% of all youth presenting to the emergency department for a mental health issue. After the assessment, 56% of patients were linked to community mental health support [45].

Survey/audit

• A Canadian survey of direct support professionals (n=942) working during the pandemic with people with intellectual and development disabilities, showed that the majority of them completed telephone-based visits (n=549, 58%), and only 22% (n=204) made at least one videocall. A total of 225 (24%) responders attended at least one face-to-face visit [59].

Canada and United States

Survey/audit

• In March 2020, an online survey involving 93 clinic directors or designated representatives of the Association of Psychology Training Clinics mainly in the United States (n=89) and Canada showed that 23.7% (n=22) were forced to close face-to-face clinics and discontinue services, at least temporarily. Of those that remained open for services, 61 training clinics (65.6% of the total sample, 86% of those that remained open) reported that their sites remained open primarily using telepsychology [46].

United States

Before-after study

• Psychiatric consultations and patient volume in North Carolina decreased with the onset of the pandemic (March 2020) by 66.9% and 25.2%, respectively. After the introduction of video consultations, psychiatric consultations were 39.1% lower compared to prepandemic data [34].

• Daily telemental health visits to the Department of Veterans Affairs rose from 1739 to 11,406 (+556%) in the weeks following the pandemic declaration (March 2020). Daily in-person encounters fell from 57,296 to 10,931 (−81%) [38].

• In an American eating disorder multidisciplinary unit, the yearly average inpatient census by month changed from 31.92 (SD 5.33; before the COVID-19 pandemic) to 19.33 (SD 4.5; March 2020 to May 2020) (59% of the usual census), owing to the implementation of telehealth strategies in most aspects of the services (eg, new admission evaluation, psychotherapy, and group therapy) [39].

• After the onset of COVID-19 (April-May 2020), the average monthly utilization of telehealth services within the Medical University of South Carolina’s Women’s Reproductive Behavioral Health Program (which provides outpatient mental health and substance use treatment to pregnant and postpartum women within obstetric practices) increased by 90% compared to the data prior to the pandemic [62].

• Comparing delivery of telemental health services in New York (USA) before, during, and after the lockdown, there was a tendency for adult patients (n=1115, 21,131 sessions) to prefer telehealth compared to children (n=1374, 22,163 sessions) (P<.001). In this service, telehealth was implemented in March 2020 (onset of lockdown) but patients returned to prefer face-to-face visits when in-person services resumed in May and June 2020 (χ^2=21.745, P<.01) [49].

• Data were compared for the number of mental health consultation visits made by cancer patients face-to-face in April-September 2019 and both face-to-face and virtually in April-September 2020. In 2020, the number of inpatient consultations (330 vs 623), referrals (127 vs 175), and outpatient visits (448 vs 550) decreased. Of the outpatient visits, 359 were telephone contacts, 69 were video visits, and only 20 were in person [71].

• Data from the Veterans Health Administration (VHA), the largest integrated health care system in the United States, showed that before the pandemic (October 2019 to February 2020), VHA had 1.5 to 1.8 million visits per month, of which 85% were in-person visits, 11% were by telephone, and 5% were by video. The use of telepsychiatry increased with the pandemic, and in June 2020, there were 1.5 million visits, of which 19% were in person, 59% were by telephone, and 20% were by video [73].

Descriptive study

• During March 2020, a large child psychiatry department registered an overall reduced number of study visits (645 vs 358) over a 1-month period, with a shift toward phone and home-based telemental health (n=171 and n=160, respectively), compared to in-person visits (n=27) [76].

• In the 2 weeks before the implementation of the "Addiction Telehealth Program" (ATP) in April 2020, which is a telephone-based program to reduce treatment access barriers for people with substance use disorders staying at San Francisco's COVID-19 isolation and quarantine site, the program received 10 calls from other health care providers, whereas in the 6 weeks after the implementation (April 10-May 25, 2020), there were 59 consultations for the isolation and quarantine site [60].

• Data from the Johnson Depression Center and the Steven A. Cohen Military Family Clinic at the University of Colorado Anschutz Medical Campus showed that the change to telemental health helped reduce no-show rates from 11.4% (n=57) in the 2 weeks before the pandemic (March 2-13, 2020) to 7.8% (n=38) after implementation of telehealth (March 30-April 10, 2020). In the following months (April-September 2020), the no-show rate was stable, with a rate between 5.5% and 8.5%, and a 26.2% increase in overall completed visits [61].

• During the first wave (March-September 2020), a psychosocial oncology service in the United States continued to provide visits through telepsychology for 93% of established patients (n=85), while 4 patients left the service and 2 patients preferred to wait for the reintroduction of face-to-face visits. The service also received 263 new referrals and delivered a first visit for 50.6% (n=133), with 82% (n=109) as digital visits and 18% (n=24) as face-to-face visits [63].

• In New York, a telephone-based model was implemented during the pandemic to continue to address substance use within emergency departments (SBIRT program). In 13 weeks, there were 228 incoming calls, 190 outgoing calls, and 4 voicemails processed. Moreover, 108 (26%) calls were with patients, 13 (3%) with family/friends, 224 (53%) with staff members, and 79 (19%) with treatment providers [47].

Survey/audit

• Of the 818 mental health care professionals in the United States who reported using telepsychiatry during the pandemic (April-May 2020), 500 (61%) used both video and telephone, 273 (33%) used only telephone, and 45 (6%) used only video. Among the advantages cited, flexible scheduling or rescheduling, timely appointment starts, and lack or reduction of no-shows were reported. Among the raised concerns, there were technical difficulties and difficult access to video platforms, forcing the provider and patient to conduct visits telephonically despite both parties preferring 2-way video [44].

• A survey of psychologists showed that the use of telepsychology increased with the pandemic, compared to data prior to the pandemic (before January 2020), from 7.07% (SD 14.86) to 85.53% (SD 29.24) of the total activity [68].

• In a survey about the use of telepsychology, data showed that most therapists who already used telepsychology before COVID-19 reported an overall increase in requests for therapy services from current patients (36.5%), whereas 41% of those who had started to use telepsychology during the pandemic reported an overall decrease. Both groups reported a decrease in the number of requests from new clients (45% and 53%, respectively) [75].

Central and South America

Brazil

Descriptive study

• In a community mental health service in Brazil, telephone contact was possible for 61% of 154 treated patients, with 29% being advised face-to-face visits through regular service attendance, and around 7% of service users were unable to be contacted, despite several attempts [74].

Chile

Before-after study

• During the early pandemic, a large private academic health network implemented a new telemental health service. In Santiago, the number of visits delivered during this period via telepsychiatry was comparable to the 24.7% of visits delivered in 2019 by the face-to-face service [40].

Dominican Republic

Descriptive study

• In Santo Domingo, a team of volunteers consisting of 598 psychologists and 70 psychiatrists provided telephone counseling during the pandemic. In the period from March 25 to May 17, 2020, they conducted 6800 phone interventions [66].

Discussion

Principal Findings

Our systematic review found that face-to-face delivery of mental health services was reduced at the start of the pandemic period, with a reduction in the total number of delivered visits and few mental health services forced to close. To meet mental health needs, the use of TMH rapidly increased across high-income countries and low- and middle-income countries at the beginning of the pandemic, with almost complete virtualization of many services by the end of the first year. To the best of our knowledge, this is the first review that, using a preplanned systematic approach, has collected quantitative reports on the extent to which the COVID-19 pandemic impacted mental health services internationally. Moreover, in presenting findings within the context of the relevant COVID-19 Stringency Index, service changes can be understood in the context of how strict national restrictions were throughout the pandemic.

At the beginning of the pandemic, TMH was increasingly adopted by mental health services to overcome the challenges related to COVID-19 restrictions, with a steep rise during the confinement period and, in general, during periods when the COVID-19 Stringency Index was higher. In some settings, this implementation was followed by a greater number of total health care consultations than in previous years, when the only option was face-to-face visits (eg, [34,52,55-58]). Moreover, it also involved specific and specialist (tertiary) mental health service provisions, such as emergency department or services dedicated to postpartum women and patients with cancer. This adaptation was essential to continue following up patients who may have been considered at high risk or as a vulnerable category during the pandemic due to their specific conditions.

The transition to the use of TMH was reported across high- and middle-income countries, but there were no studies from low-income countries. TMH provision has great potential in low-and middle-income countries due to large geographical distances and limited availability of mental health services to cover the population needs [85], combined with reduced health system capacity during the COVID-19 pandemic [86]. However, very few studies quantitively assessed the impact of the COVID-19 pandemic in middle-income countries (Iran, Turkey, Dominican Republic, India, China, and Brazil) and none assessed the impact specifically in low-income or African subcontinent countries. This is despite reports of the rollout of TMH in African nations to provide quicker and more extensive access to mental health services [87]. Additionally, models of service transformation in high-income countries may be difficult to replicate or be less relevant in these settings [88]. Clearly, quantitative data are needed to understand the impact of the COVID-19 pandemic on mental health services and longer-term changes, both positive and negative, in low- and middle-income countries [89].

Although some services provided evidence of a gradual shift back to face-to-face contact with a reduction in the COVID-19 Stringency Index (eg, after the initial lockdown period), virtual delivery methods continued to be used more than in the prepandemic period and, in some cases, remained the method for the majority of patients. This may not have been universal across all services and may have varied with patient preferences and specific characteristics. For example, a US study found that telehealth was less preferred for children after the confinement period compared to adults due to difficulties delivering therapy to them remotely [49]. Moreover, findings indicated that TMH was favored for psychiatric consultations and support services, but less for psychotherapy in both adult and child services [49]. This was despite the findings of a recent systematic review that videoconferencing was an accessible and feasible modality for therapy delivery to adult patients, with comparable levels of therapeutic alliance between virtual and in‐person therapy [90]. This apparent discrepancy might be due to the challenging settings that clinicians and patients faced using the newly created TMH services during the pandemic, but it also suggests that the problems might be remedied by providing training (for both patients and clinicians) in TMH to improve delivery of therapies [15]. Clinicians' attitudes toward TMH may also have an impact on its implementation in daily practice. As shown in a recent systematic review on barriers to and facilitators of TMH during the pandemic, clinicians reported a reduction in their ability to develop and maintain a therapeutic relationship during televisits [91]. This aligns with prepandemic data that showed that clinicians' perspectives on TMH can be a major barrier to the uptake of the virtual modality [92]. Other factors influencing clinicians' adoption of TMH include perceptions of how effectively this can be delivered, fears that aspects of in-person care may be missed in virtual encounters (eg, observation of physical signs such as tremor, fidgeting, and anxiety), and satisfaction with plans for handling clinical emergencies at a distance [92,93]. Moreover, a survey conducted in Hong Kong during the pandemic highlighted the need for mental health staff to receive dedicated training on clinical, technological, and program-specific aspects when providing TMH [94]. There is well-established guidance on virtual consultations to support health care professionals in the pandemic [95-97]. This includes physical examination, risk assessment, and management of emergency situations such as agitation/aggression, suicidality, and domestic violence [98]. During the COVID-19 pandemic, charities and other organizations reported an increase in cases of domestic violence and provided guidance on the safe assessment of domestic abuse during virtual contacts [99]. Although these recommendations may have been helpful to guide staff through the first period of the pandemic, training for clinicians will be crucial to take advantage of TMH in standard clinical care beyond the pandemic [15]. For example, implementing telehealth as part of the formal training for psychiatric residents has been proposed as an effective means of developing these skills [100].

While TMH was implemented for most patients, there were a small number of specific groups where it was not possible to be used alone, and instead, a hybrid model was used, combining face-to-face and digital strategies. This included services providing treatment with clozapine or depot medication, and electroconvulsive therapy. During the pandemic, such services applied several strategies to reduce the risk of exposure for these patients (eg, increasing the timing between blood monitoring for clozapine, new infection control measures, and social distancing) [101], but face-to-face contact could not be avoided

completely. This highlighted that while teleassessment may have similar advantages and disadvantages independently to patient diagnoses, implementation of telemedicine in treatment management will face different barriers in different diagnostic groups. For instance, telemedicine has been typically used for the treatment of some diagnoses, such as anxiety and eating disorders [102,103], while programs addressing other diagnostic groups are now emerging [104].

Due to financial barriers, digital literacy, and features of chronic mental illness, a reliance on TMH services is not feasible for a minority of community patients [84]. Studies describing these clinical groups report a higher prevalence of certain diagnoses (eg, schizophrenia or schizoaffective disorder) [84], while others indicated several barriers, such as lack of digital literacy or confidence to make full use of a telephone, and lack of an appropriate device or broadband required to connect with clinicians [105]. In resource-limited settings during the COVID-19 pandemic, mental health staff working for Médecins Sans Frontières used audio-only platforms for 80% of TMH care interventions [106]. Despite this, over half of the patients were unreachable using these interventions, mainly due to poor network coverage, lack of communication devices, or lack of a private space at home [106]. For these patients, specific provision of a device or broadband might be necessary. Investment in providing these resources will have to be balanced at a country level against the costs associated with the anticipated increase in mental health problems. Moreover, appropriate training in TMH as well as the use of digital tools specifically for patients may increase access to and acceptability of virtual means. For instance, the DOORS (Digital Opportunities for Outcomes in Recovery Services) program provides an example of training to increase digital literacy and confidence in patients with first-episode psychosis and serious mental illness [104]. Such models will need to be developed locally to match the needs of local populations and clinical groups.

Limitations

Our study has some limitations. A systematic review relies on the quality of the primary papers, and in this case, although we found 55 articles on TMH, 19 were case reports or case series presenting descriptive data on local services, with limited generalizability. To address this, we grouped our findings according to the country of origin, considering differences in the local service structure, local needs, and stringency of confinement measures during the pandemic. In addition, we found limited reports on middle-income countries, and no reports from the African subcontinent or from low-income countries. Absent data do not represent evidence for the lack of effectiveness or feasibility, as services (such as those in African nations) may have implemented TMH without publishing quantitative data. The lack of representation of low- and middle-income countries in published health research is well documented [107], but without data, it is difficult to assess the impact of the pandemic on mental health services in these areas. Since the literature on COVID-19 is growing at a rapid pace, it is likely that relevant articles have been published after our final search and are therefore not included in our review. Finally, the aim of our project was to collect information about the changes and difficulties that mental health services faced at the start of the pandemic. Therefore, we limited the data collection to the first year of the pandemic. Future studies should evaluate whether these implementations will be confirmed or further refined.

Conclusions

During sudden health emergencies or disasters, such as the COVID-19 pandemic, mental health services cannot rely on face-to-face activities to provide care to all patients. The implementation of TMH has been demonstrated as feasible and widespread in different countries, allowing mental health services to continue the provision of essential care for patients, even during times of extreme confinement. Although many high-income countries and some middle-income countries were able to transition mental health services to digital delivery, this brought some challenges. While TMH was predominantly used in the acute pandemic, some findings showed a reduction in its use with time. Further training and support will be needed to maintain the benefits of TMH and to address barriers, such as digital exclusion. The pandemic can be seen as an opportunity to explore the benefits of TMH in enabling broader access to high-quality psychiatric treatment in the longer term [108]. Telehealth delivery methods are not only useful tools for the acute pandemic, but also have a role in addressing the anticipated increase in mental health needs after the immediate disruption and in preparation for future emergencies [15]. However, specific subgroups of patients will likely need or prefer face-to-face care, while others may not have access to virtual means. A blended approach, combining in-person and virtual modalities, that takes into consideration the needs, preferences, and digital skills of patients may suit the future development of mental health services, but this requires further investigation that considers acceptability to patients, carers, and clinicians. Delivering effective and equitable blended approaches will require confidence in using digital technologies, training, and experience in all modalities [108,109] to ensure that the benefits of TMH and associated eHealth technologies spark a transformation that is sustained beyond the immediate crisis.

Acknowledgments

AC and EGO are supported by the National Institute for Health Research (NIHR) Research Professorship to AC (grant RP-2017-08-ST2-006). EGO is also supported by the National Institute for Health and Care Research (NIHR) Applied Research Collaboration Oxford and Thames Valley (ARC OxTV) at Oxford Health NHS Foundation Trust. AC, CZ, EGO, and KAS are supported by the NIHR Oxford Cognitive Health Clinical Research Facility and by the NIHR Oxford Health Biomedical Research Centre (grant BRC-1215-20005). The views expressed are those of the authors and not necessarily those of the UK National Health Service, the NIHR, or the Department of Health and Social Care.

Conflicts of Interest

AC has received research and consultancy fees from the INCiPiT (Italian Network for Paediatric Trials), CARIPLO Foundation, and Angelini Pharma, outside the submitted work. EGO has received research and consultancy fees from Angelini Pharma. All other authors declare no conflicts.

Multimedia Appendix 1
Search strategy.
[DOCX File , 21 KB - mental_v9i8e38600_app1.docx]

Multimedia Appendix 2
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart.
[DOCX File , 165 KB - mental_v9i8e38600_app2.docx]

Multimedia Appendix 3
Impact of COVID-19 on mental health services: summary of key findings.
[DOCX File , 71 KB - mental_v9i8e38600_app3.docx]

Multimedia Appendix 4
World distribution of 101 included papers.
[DOCX File , 235 KB - mental_v9i8e38600_app4.docx]

Multimedia Appendix 5
Characteristics of 72 papers on face-to-face services.
[DOCX File , 27 KB - mental_v9i8e38600_app5.docx]

References

1. Hong J, Sheriff R, Smith K, Tomlinson A, Saad F, Smith T, et al. Impact of COVID-19 on telepsychiatry at the service and individual patient level across two UK NHS mental health Trusts. Evid Based Ment Health 2021 Nov;24(4):161-166 [FREE Full text] [doi: 10.1136/ebmental-2021-300287] [Medline: 34583940]
2. Adhanom Ghebreyesus T. Addressing mental health needs: an integral part of COVID-19 response. World Psychiatry 2020 Jun;19(2):129-130 [FREE Full text] [doi: 10.1002/wps.20768] [Medline: 32394569]
3. Taquet M, Luciano S, Geddes JR, Harrison PJ. Bidirectional associations between COVID-19 and psychiatric disorder: retrospective cohort studies of 62 354 COVID-19 cases in the USA. The Lancet Psychiatry 2021 Feb;8(2):130-140 [FREE Full text] [doi: 10.1016/s2215-0366(20)30462-4]
4. Taquet M, Geddes JR, Husain M, Luciano S, Harrison PJ. 6-month neurological and psychiatric outcomes in 236 379 survivors of COVID-19: a retrospective cohort study using electronic health records. The Lancet Psychiatry 2021 May;8(5):416-427 [FREE Full text] [doi: 10.1016/s2215-0366(21)00084-5]
5. Helms J, Kremer S, Merdji H, Clere-Jehl R, Schenck M, Kummerlen C, et al. Neurologic Features in Severe SARS-CoV-2 Infection. N Engl J Med 2020 Jun 04;382(23):2268-2270 [FREE Full text] [doi: 10.1056/nejmc2008597]
6. Rogers JP, Chesney E, Oliver D, Pollak TA, McGuire P, Fusar-Poli P, et al. Psychiatric and neuropsychiatric presentations associated with severe coronavirus infections: a systematic review and meta-analysis with comparison to the COVID-19 pandemic. The Lancet Psychiatry 2020 Jul;7(7):611-627 [FREE Full text] [doi: 10.1016/s2215-0366(20)30203-0]
7. Varatharaj A, Thomas N, Ellul M, Davies NW, Pollak T, Tenorio E, et al. Neurological and neuropsychiatric complications of COVID-19 in 153 patients: a UK-wide surveillance study. The Lancet Psychiatry 2020 Oct;7(10):875-882 [FREE Full text] [doi: 10.1016/S2215-0366(20)30287-X]
8. 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. The Lancet 2020 Mar;395(10227):912-920 [FREE Full text] [doi: 10.1016/s0140-6736(20)30460-8]
9. Botchway S, Fazel S. Remaining vigilant about COVID-19 and suicide. The Lancet Psychiatry 2021 Jul;8(7):552-553 [FREE Full text] [doi: 10.1016/s2215-0366(21)00117-6]
10. Baraniuk C. How to vaccinate the world against covid-19. BMJ 2021 Feb 05;372:n211 [FREE Full text] [doi: 10.1136/bmj.n211] [Medline: 33547042]
11. Hale T, Angrist N, Goldszmidt R, Kira B, Petherick A, Phillips T, et al. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). Nat Hum Behav 2021 Apr;5(4):529-538 [FREE Full text] [doi: 10.1038/s41562-021-01079-8] [Medline: 33686204]

12. 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. The Lancet Psychiatry 2020 Sep;7(9):813-824 [FREE Full text] [doi: 10.1016/s2215-0366(20)30307-2]
13. Policy Brief: COVID-19 and the Need for Action on Mental Health. United Nations. 2020. URL: https://unsdg.un.org/sites/default/files/2020-05/UN-Policy-Brief-COVID-19-and-mental-health.pdf [accessed 2022-07-30]
14. The impact of COVID-19 on mental, neurological and substance use services. World Health Organization. URL: https://www.who.int/publications/i/item/978924012455 [accessed 2022-07-30]
15. Torous J, Wykes T. Opportunities From the Coronavirus Disease 2019 Pandemic for Transforming Psychiatric Care With Telehealth. JAMA Psychiatry 2020 Dec 01;77(12):1205-1206. [doi: 10.1001/jamapsychiatry.2020.1640] [Medline: 32391857]
16. Kinoshita S, Cortright K, Crawford A, Mizuno Y, Yoshida K, Hilty D, et al. Changes in telepsychiatry regulations during the COVID-19 pandemic: 17 countries and regions' approaches to an evolving healthcare landscape. Psychol. Med 2020 Nov 27:1-8. [doi: 10.1017/s0033291720004584]
17. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. World Health Organization. 2020. URL: https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020 [accessed 2022-07-30]
18. Cipriani A, Ostinelli E, Smith K, Macdonald O, Sheriff R, Harrison P, et al. COVID-19 and its impact on mental health patients, staff and services: systematic review with recommendations for clinical practice and research. PROSPERO International prospective register of systematic reviews. URL: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42020178819 [accessed 2022-07-30]
19. COAP (COVID-19 Open Access Project) COVID-19 Living Evidence. URL: https://zika.ispm.unibe.ch/assets/data/pub/search_beta/ [accessed 2022-07-30]
20. COVID-19: Stringency Index. Our World in Data. URL: https://ourworldindata.org/grapher/covid-stringency-index [accessed 2022-07-30]
21. Fredwall M, Terry D, Enciso L, Burch MM, Trott K, Albert DVF. Short-term outcomes in pediatric and adolescent patients with psychogenic nonepileptic events seen by telemedicine during the COVID-19 pandemic. Epilepsy Behav 2021 Apr;117:107739 [FREE Full text] [doi: 10.1016/j.yebeh.2020.107739] [Medline: 33601287]
22. Stewart R, Broadbent M. Using past and current data to estimate potential crisis service use in mental healthcare after the COVID-19 lockdown: South London and Maudsley data. medRxiv. 2020. URL: https://www.medrxiv.org/content/10.1101/2020.06.29.20142448v1 [accessed 2022-07-30]
23. Baugh JJ, White BA, McEvoy D, Yun BJ, Brown DF, Raja AS, et al. The cases not seen: Patterns of emergency department visits and procedures in the era of COVID-19. Am J Emerg Med 2021 Aug;46:476-481 [FREE Full text] [doi: 10.1016/j.ajem.2020.10.081] [Medline: 33189517]
24. Amad A, Magnat M, Quilès C, Yrondi A, Sauvaget A, Bulteau S, et al. [Evolution of electro-convulsive therapy activity in France since the beginning of the COVID-19 pandemic]. Encephale 2020 Jun;46(3S):S40-S42 [FREE Full text] [doi: 10.1016/j.encep.2020.04.004] [Medline: 32370981]
25. Casano K, Capone E. Unexpected cancellation on a catatonic patient's electroconvulsive therapy due to the coronavirus pandemic. Gen Psychiatr 2020 Sep 24;33(6):e100271 [FREE Full text] [doi: 10.1136/gpsych-2020-100271] [Medline: 33083693]
26. Grover S, Mehra A, Sahoo S, Avasthi A, Tripathi A, D'Souza A, et al. State of mental health services in various training centers in India during the lockdown and COVID-19 pandemic. Indian J Psychiatry 2020;62(4):363. [doi: 10.4103/psychiatry.indianjpsychiatry_567_20]
27. Grover S, Mehra A, Sahoo S, Avasthi A, Tripathi A, D'Souza A, et al. Impact of COVID-19 pandemic and lockdown on the state of mental health services in the private sector in India. Indian J Psychiatry 2020;62(5):488. [doi: 10.4103/psychiatry.indianjpsychiatry_568_20]
28. Bollmann A, Hohenstein S, Pellissier V, Stengler K, Reichardt P, Ritz JP, et al. Utilization of in- and outpatient hospital care in Germany during the Covid-19 pandemic insights from the German-wide Helios hospital network. PLoS One 2021;16(3):e0249251 [FREE Full text] [doi: 10.1371/journal.pone.0249251] [Medline: 33765096]
29. Clerici M, Durbano F, Spinogatti F, Vita A, de Girolamo G, Micciolo R. Psychiatric hospitalization rates in Italy before and during COVID-19: did they change? An analysis of register data. Ir J Psychol Med 2020 Dec 05;37(4):283-290 [FREE Full text] [doi: 10.1017/ipm.2020.29] [Medline: 32368994]
30. Lee A, Mo FYM, Lam LCW. Higher psychogeriatric admissions in COVID-19 than in severe acute respiratory syndrome. Int J Geriatr Psychiatry 2020 Dec;35(12):1449-1457 [FREE Full text] [doi: 10.1002/gps.5422] [Medline: 32892460]
31. Itrat A, Jagadheesan K, Danivas V, Lakra V. A comparative study of access to inpatient psychiatric treatment in a public mental health service in Melbourne during COVID-19. Indian J Psychiatry 2020;62(9):454 [FREE Full text] [doi: 10.4103/psychiatry.indianjpsychiatry_852_20]
32. Jagadheesan K, Danivas V, Itrat Q, Sekharan L, Lakra APV. COVID-19 and psychiatric admissions: An observational study of the first six months of lockdown in Melbourne. Psychiatry Res 2021 Jun;300:113902 [FREE Full text] [doi: 10.1016/j.psychres.2021.113902] [Medline: 33799197]

33. Arendt F, Markiewitz A, Mestas M, Scherr S. COVID-19 pandemic, government responses, and public mental health: Investigating consequences through crisis hotline calls in two countries. Soc Sci Med 2020 Nov;265:113532 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113532] [Medline: 33223385]
34. Beran C, Sowa NA. Adaptation of an Academic Inpatient Consultation-Liaison Psychiatry Service During the SARS-CoV-2 Pandemic: Effects on Clinical Practice and Trainee Supervision. J Acad Consult Liaison Psychiatry 2021;62(2):186-192 [FREE Full text] [doi: 10.1016/j.psym.2020.11.002] [Medline: 33288272]
35. Berthaut E, Chamignon T. Le fil institutionnel en période Covid : utilisation de la plateforme Discord au CMPA. Neuropsychiatrie de l'Enfance et de l'Adolescence 2021 May;69(3):142-144 [FREE Full text] [doi: 10.1016/j.neurenf.2021.01.002]
36. Cantini S, Caponigro G, Fusco O, Deligiannis P, Pastorini A, Cubeddu A, et al. Psychoncology in front of Covid-19. Tumori 2020 Oct 20;106(2_suppl):95-96 [FREE Full text] [doi: 10.1177/0300891620953388]
37. Carpiniello B, Tusconi M, Zanalda E, Di Sciascio G, Di Giannantonio M, Executive Committee of The Italian Society of Psychiatry. Psychiatry during the Covid-19 pandemic: a survey on mental health departments in Italy. BMC Psychiatry 2020 Dec 16;20(1):593 [FREE Full text] [doi: 10.1186/s12888-020-02997-z] [Medline: 33327940]
38. Connolly SL, Stolzmann KL, Heyworth L, Weaver KR, Bauer MS, Miller CJ. Rapid Increase in Telemental Health Within the Department of Veterans Affairs During the COVID-19 Pandemic. Telemed J E Health 2021 Apr 01;27(4):454-458 [FREE Full text] [doi: 10.1089/tmj.2020.0233] [Medline: 32926664]
39. Datta N, Derenne J, Sanders M, Lock JD. Telehealth transition in a comprehensive care unit for eating disorders: Challenges and long-term benefits. Int J Eat Disord 2020 Nov 27;53(11):1774-1779 [FREE Full text] [doi: 10.1002/eat.23348] [Medline: 32715512]
40. Garcia-Huidobro D, Rivera S, Valderrama Chang S, Bravo P, Capurro D. System-Wide Accelerated Implementation of Telemedicine in Response to COVID-19: Mixed Methods Evaluation. J Med Internet Res 2020 Oct 06;22(10):e22146 [FREE Full text] [doi: 10.2196/22146] [Medline: 32903195]
41. Ghiretti F, Gildoni G, Grassi GM, Torricelli L, Benassi E, Bonaretti E, et al. Psychological Support to the Community During the COVID-19 Pandemic: Field Experience in Reggio Emilia, Northern Italy. Front Psychol 2020 Nov 17;11:561742 [FREE Full text] [doi: 10.3389/fpsyg.2020.561742] [Medline: 33281663]
42. Ghosh A, Mahintamani T, Pillai RR, Mattoo SK, Basu D. Telemedicine-assisted stepwise approach of service delivery for substance use disorders in India. Asian J Psychiatr 2021 Apr;58:102582 [FREE Full text] [doi: 10.1016/j.ajp.2021.102582] [Medline: 33607350]
43. Graell M, Morón-Nozaleda MG, Camarneiro R, Villaseñor Á, Yáñez S, Muñoz R, et al. Children and adolescents with eating disorders during COVID-19 confinement: Difficulties and future challenges. Eur Eat Disord Rev 2020 Nov 29;28(6):864-870 [FREE Full text] [doi: 10.1002/erv.2763] [Medline: 32729139]
44. Guinart D, Marcy P, Hauser M, Dwyer M, Kane JM. Mental Health Care Providers' Attitudes Toward Telepsychiatry: A Systemwide, Multisite Survey During the COVID-19 Pandemic. Psychiatr Serv 2021 Jun;72(6):704-707 [FREE Full text] [doi: 10.1176/appi.ps.202000441] [Medline: 33593104]
45. Hall P, Sukhera J. Emergency Diversion Clinic: Lessons from Southwestern Ontario's Pandemic Response. Journal of the Canadian Academy of Child and Adolescent Psychiatry 2020;29(4):271 [FREE Full text]
46. Hames JL, Bell DJ, Perez-Lima LM, Holm-Denoma JM, Rooney T, Charles NE, et al. Navigating uncharted waters: Considerations for training clinics in the rapid transition to telepsychology and telesupervision during COVID-19. Journal of Psychotherapy Integration 2020 Jun;30(2):348-365 [FREE Full text] [doi: 10.1037/int0000224]
47. Harrison L, DeMasi L, Butkus S, O'Neill K, Kwon N, D'Angelo J, et al. 163 Implementation of a Telephonic-Based Model to Continue to Address Substance Use as Part of Usual Care in Emergency Departments during COVID-19. Annals of Emergency Medicine 2020 Oct;76(4):S64 [FREE Full text] [doi: 10.1016/j.annemergmed.2020.09.175]
48. Hazarika M, Das B, Das S, Baruah A, Sharma N, Barua C, et al. Profile of distress callers and service utilisation of tele-counselling among the population of Assam, India: an exploratory study during COVID-19. Open J Psychiatry Allied Sci 2021;12(1):7-12 [FREE Full text] [doi: 10.5958/2394-2061.2021.00001.x] [Medline: 33163629]
49. Hoffnung G, Feigenbaum E, Schechter A, Guttman D, Zemon V, Schechter I. Children and Telehealth in Mental Healthcare: What We Have Learned From COVID-19 and 40,000+ Sessions. Psychiatr Res Clin Pract 2021 Sep;3(3):106-114 [FREE Full text] [doi: 10.1176/appi.prcp.20200035] [Medline: 33821240]
50. Humer E, Pieh C, Kuska M, Barke A, Doering BK, Gossmann K, et al. Provision of Psychotherapy during the COVID-19 Pandemic among Czech, German and Slovak Psychotherapists. Int J Environ Res Public Health 2020 Jul 04;17(13):4811 [FREE Full text] [doi: 10.3390/ijerph17134811] [Medline: 32635422]
51. Johnson S, Dalton-Locke C, Vera San Juan N, Foye U, Oram S, Papamichail A, COVID-19 Mental Health Policy Research Unit Group. Impact on mental health care and on mental health service users of the COVID-19 pandemic: a mixed methods survey of UK mental health care staff. Soc Psychiatry Psychiatr Epidemiol 2021 Jan;56(1):25-37 [FREE Full text] [doi: 10.1007/s00127-020-01927-4] [Medline: 32857218]
52. Karim M, Wadoo O, Reagu SM, Amro R, Abdulla MA. Telepsychiatry in the Arabian Gulf region - Implications beyond the Covid-19 pandemic. Asian J Psychiatr 2020 Dec;54:102397 [FREE Full text] [doi: 10.1016/j.ajp.2020.102397] [Medline: 33271697]

53. Khanra S, Mukherjee A, Goyal N, Das B, Munda SK. Service utilization and saved travel cost in telepsychiatry consultation by outpatients at a psychiatric hospital in India during COVID-19 pandemic. Asian J Psychiatr 2021 Mar;57:102568 [FREE Full text] [doi: 10.1016/j.ajp.2021.102568] [Medline: 33535135]
54. Lian W, Wen L, Zhou Q, Zhu W, Duan W, Xiao X, et al. Digital Health Technologies Respond to the COVID-19 Pandemic In a Tertiary Hospital in China: Development and Usability Study. J Med Internet Res 2020 Nov 24;22(11):e24505 [FREE Full text] [doi: 10.2196/24505] [Medline: 33141679]
55. Looi JC, Allison S, Bastiampillai T, Pring W, Reay R. Australian private practice metropolitan telepsychiatry during the COVID-19 pandemic: analysis of Quarter-2, 2020 usage of new MBS-telehealth item psychiatrist services. Australas Psychiatry 2021 Apr 06;29(2):183-188 [FREE Full text] [doi: 10.1177/1039856220975294] [Medline: 33280401]
56. Looi JC, Allison S, Bastiampillai T, Pring W. Private practice metropolitan telepsychiatry in smaller Australian jurisdictions during the COVID-19 pandemic: preliminary analysis of the introduction of new Medicare Benefits Schedule items. Australas Psychiatry 2020 Dec 05;28(6):639-643 [FREE Full text] [doi: 10.1177/1039856220960381] [Medline: 33016772]
57. Looi JC, Allison S, Bastiampillai T, Pring W. Private practice metropolitan telepsychiatry in larger Australian states during the COVID-19 pandemic: an analysis of the first 2 months of new MBS telehealth item psychiatrist services. Australas Psychiatry 2020 Dec 30;28(6):644-648 [FREE Full text] [doi: 10.1177/1039856220961906] [Medline: 32997521]
58. Looi JC, Allison S, Bastiampillai T, Pring W, Reay R, Kisely SR. Increased Australian outpatient private practice psychiatric care during the COVID-19 pandemic: usage of new MBS-telehealth item and face-to-face psychiatrist office-based services in Quarter 3, 2020. Australas Psychiatry 2021 Apr 24;29(2):194-199 [FREE Full text] [doi: 10.1177/1039856221992634] [Medline: 33626304]
59. Lunsky Y, Bobbette N, Selick A, Jiwa M. "The doctor will see you now": Direct support professionals' perspectives on supporting adults with intellectual and developmental disabilities accessing health care during COVID-19. Disabil Health J 2021 Jul;14(3):101066 [FREE Full text] [doi: 10.1016/j.dhjo.2021.101066] [Medline: 33531290]
60. Mehtani NJ, Ristau JT, Snyder H, Surlyn C, Eveland J, Smith-Bernardin S, et al. COVID-19: A catalyst for change in telehealth service delivery for opioid use disorder management. Subst Abus 2021 Mar 08;42(2):205-212 [FREE Full text] [doi: 10.1080/08897077.2021.1890676] [Medline: 33684331]
61. Mishkind MC, Shore JH, Bishop K, D'Amato K, Brame A, Thomas M, et al. Rapid Conversion to Telemental Health Services in Response to COVID-19: Experiences of Two Outpatient Mental Health Clinics. Telemed J E Health 2021 Jul 01;27(7):778-784 [FREE Full text] [doi: 10.1089/tmj.2020.0304] [Medline: 33393857]
62. Moreland A, Guille C, McCauley JL. Increased availability of telehealth mental health and substance abuse treatment for peripartum and postpartum women: A unique opportunity to increase telehealth treatment. J Subst Abuse Treat 2021 Apr;123:108268 [FREE Full text] [doi: 10.1016/j.jsat.2020.108268] [Medline: 33612199]
63. Myers Virtue S, Howrey HL, Duffy NM, Wallace M. Shifting psychosocial oncology care to telepsychology during the COVID-19 pandemic. J Psychosoc Oncol 2021;39(3):416-427 [FREE Full text] [doi: 10.1080/07347332.2021.1894526] [Medline: 33792503]
64. Naik S, Rahul P, Harihara S, Pahuja E, Chithra NK, Ramachandraiah S, et al. Telephonic follow-up during COVID-19 to maintain continuity of care for persons with psychiatric disorders. Asian J Psychiatr 2021 Mar;57:102564 [FREE Full text] [doi: 10.1016/j.ajp.2021.102564] [Medline: 33516153]
65. Peppou L, Economou M, Skali T, Papageorgiou C. From economic crisis to the COVID-19 pandemic crisis: evidence from a mental health helpline in Greece. Eur Arch Psychiatry Clin Neurosci 2021 Mar;271(2):407-409 [FREE Full text] [doi: 10.1007/s00406-020-01165-4] [Medline: 32666279]
66. Peralta E, Taveras M. Effectiveness of teleconsultation use in access to mental health services during the coronavirus disease 2019 pandemic in the Dominican Republic. Indian J Psychiatry 2020;62(9):492 [FREE Full text] [doi: 10.4103/psychiatry.indianjpsychiatry_1047_20]
67. Perricone G, Rotolo I, Beninati V, Billeci N, Ilarda V, Polizzi C. The Lègami/Legàmi Service-An Experience of Psychological Intervention in Maternal and Child Care during COVID-19. Pediatr Rep 2021 Mar 22;13(1):142-161 [FREE Full text] [doi: 10.3390/pediatric13010021] [Medline: 33810190]
68. Pierce BS, Perrin PB, Tyler CM, McKee GB, Watson JD. The COVID-19 telepsychology revolution: A national study of pandemic-based changes in U.S. mental health care delivery. Am Psychol 2021 Jan;76(1):14-25 [FREE Full text] [doi: 10.1037/amp0000722] [Medline: 32816503]
69. Prior M, Bellè M, Burelli M, Favaretto A. On-line psychological treatment with oncology patients during COVID-19 emergency. Tumori 2020 Oct 20;106(2_suppl):81 [FREE Full text] [doi: 10.1177/0300891620953388]
70. Probst T, Stippl P, Pieh C. Changes in Provision of Psychotherapy in the Early Weeks of the COVID-19 Lockdown in Austria. Int J Environ Res Public Health 2020 May 27;17(11):3815 [FREE Full text] [doi: 10.3390/ijerph17113815] [Medline: 32471295]
71. Rainwater L, Lichiello S, Duckworth K, Tolbert A, Moskop R, McQuellon R. Psychosocial oncology care during a pandemic: Effect of offering telemental health in an academic medical center. Psychooncology 2021 Mar;30(S1):18.1 [FREE Full text] [doi: 10.1002/pon.5637] [Medline: 33656764]

72. Ravindran S, Channaveerachari NK, Seshadri SP, Kasi S, Manikappa SK, Cherian AV, et al. Crossing barriers: Role of a tele-outreach program addressing psychosocial needs in the midst of COVID-19 pandemic. Asian J Psychiatr 2020 Oct;53:102351 [FREE Full text] [doi: 10.1016/j.ajp.2020.102351] [Medline: 32871399]
73. Rosen C, Morland LA, Glassman LH, Marx BP, Weaver K, Smith CA, et al. Virtual mental health care in the Veterans Health Administration's immediate response to coronavirus disease-19. Am Psychol 2021 Jan;76(1):26-38 [FREE Full text] [doi: 10.1037/amp0000751] [Medline: 33119331]
74. Salum GA, Rehmenklau JF, Csordas MC, Pereira FP, Castan JU, Ferreira AB, et al. Supporting people with severe mental health conditions during the COVID-19 pandemic: considerations for low- and middle-income countries using telehealth case management. Braz J Psychiatry 2020 Aug;42(4):451-452 [FREE Full text] [doi: 10.1590/1516-4446-2020-1078] [Medline: 32578690]
75. Sampaio M, Haro MVN, De Sousa B, Melo WV, Hoffman HG. Therapists Make the Switch to Telepsychology to Safely Continue Treating Their Patients During the COVID-19 Pandemic. Virtual Reality Telepsychology May Be Next. Front Virtual Real 2021 Jan;1:576421 [FREE Full text] [doi: 10.3389/frvir.2020.576421] [Medline: 33585834]
76. Sharma A, Sasser T, Schoenfelder Gonzalez E, Vander Stoep A, Myers K. Implementation of Home-Based Telemental Health in a Large Child Psychiatry Department During the COVID-19 Crisis. J Child Adolesc Psychopharmacol 2020 Sep;30(7):404-413 [FREE Full text] [doi: 10.1089/cap.2020.0062] [Medline: 32639849]
77. Singh Bhandari S, Joseph SJ, Udayasankaran JG, Konthoujam B, Shoib S, Dutta S. Telepsychiatry: A Feasible Means to Bridge the Demand-Supply Gaps in Mental Health Services During and After the COVID-19 Pandemic: Preliminary Experiences from Sikkim State of India. Indian J Psychol Med 2020 Oct 27;42(5):500-502 [FREE Full text] [doi: 10.1177/0253717620951282] [Medline: 33414609]
78. Staples L, Nielssen O, Kayrouz R, Cross S, Karin E, Ryan K, et al. Rapid report 2: Symptoms of anxiety and depression during the first 12 weeks of the Coronavirus (COVID-19) pandemic in Australia. Internet Interv 2020 Dec;22:100351 [FREE Full text] [doi: 10.1016/j.invent.2020.100351] [Medline: 33110762]
79. Steeg S, Carr MJ, Trefan L, Ashcroft DM, Kapur N, Nielsen E, et al. Primary care clinical management following self-harm during the first wave of COVID-19 in the UK. medRxiv. URL: https://www.medrxiv.org/content/10.1101/2021.03.19.21253969v1 [accessed 2022-07-30]
80. Stewart R, Martin E, Broadbent M. Mental health service activity during COVID-19 lockdown: South London and Maudsley data on working age community and home treatment team services and mortality from February to mid-May 2020. medRxiv. URL: https://www.medrxiv.org/content/10.1101/2020.06.13.20130419v1 [accessed 2022-07-30]
81. Vonderlin R, Biermann M, Konrad M, Klett M, Kleindienst N, Bailer J, et al. [Implementation and evaluation of a telephone hotline for professional mental health first aid during the COVID-19 pandemic in Germany]. Nervenarzt 2022 Jan;93(1):24-33 [FREE Full text] [doi: 10.1007/s00115-021-01089-x] [Medline: 33725184]
82. Yaffa S, Adi E, Itai P, Marit J, Doron G, Daniel S. Treatment of eating disorders in adolescents during the COVID-19 pandemic: a case series. J Eat Disord 2021 Mar 10;9(1):17 [FREE Full text] [doi: 10.1186/s40337-021-00374-z] [Medline: 33568207]
83. Zhong B, Zhou DY, He MF, Li Y, Li WT, Ng CH, et al. Mental health problems, needs, and service use among people living within and outside Wuhan during the COVID-19 epidemic in China. Ann Transl Med 2020 Dec;8(21):1392 [FREE Full text] [doi: 10.21037/atm-20-4145] [Medline: 33313137]
84. Zulfic Z, Liu D, Lloyd C, Rowan J, Schubert KO. Is telepsychiatry care a realistic option for community mental health services during the COVID-19 pandemic? Aust N Z J Psychiatry 2020 Dec;54(12):1228 [FREE Full text] [doi: 10.1177/0004867420937788] [Medline: 32571079]
85. Merchant R, Torous J, Rodriguez-Villa E, Naslund JA. Digital technology for management of severe mental disorders in low-income and middle-income countries. Curr Opin Psychiatry 2020 Sep;33(5):501-507 [FREE Full text] [doi: 10.1097/YCO.0000000000000626] [Medline: 32520747]
86. Walker P, Whittaker C, Watson OJ, Baguelin M, Winskill P, Hamlet A, et al. The impact of COVID-19 and strategies for mitigation and suppression in low- and middle-income countries. Science 2020 Jul 24;369(6502):413-422 [FREE Full text] [doi: 10.1126/science.abc0035] [Medline: 32532802]
87. Adepoju P. Africa turns to telemedicine to close mental health gap. The Lancet Digital Health 2020 Nov;2(11):e571-e572 [FREE Full text] [doi: 10.1016/s2589-7500(20)30252-1]
88. Arango C, Wykes T, Moreno C. Mental health care and COVID-19. The Lancet Psychiatry 2020 Dec;7(12):1013 [FREE Full text] [doi: 10.1016/s2215-0366(20)30480-6]
89. Kola L, Kohrt BA, Hanlon C, Naslund JA, Sikander S, Balaji M, et al. COVID-19 mental health impact and responses in low-income and middle-income countries: reimagining global mental health. The Lancet Psychiatry 2021 Jun;8(6):535-550 [FREE Full text] [doi: 10.1016/s2215-0366(21)00025-0]
90. Thomas N, McDonald C, de Boer K, Brand RM, Nedeljkovic M, Seabrook L. Review of the current empirical literature on using videoconferencing to deliver individual psychotherapies to adults with mental health problems. Psychol Psychother 2021 Sep 23;94(3):854-883 [FREE Full text] [doi: 10.1111/papt.12332] [Medline: 33620133]

91. Appleton R, Williams J, Vera San Juan N, Needle JJ, Schlief M, Jordan H, et al. Implementation, Adoption, and Perceptions of Telemental Health During the COVID-19 Pandemic: Systematic Review. J Med Internet Res 2021 Dec 09;23(12):e31746 [FREE Full text] [doi: 10.2196/31746] [Medline: 34709179]
92. Cowan K, McKean AJ, Gentry MT, Hilty DM. Barriers to Use of Telepsychiatry: Clinicians as Gatekeepers. Mayo Clin Proc 2019 Dec;94(12):2510-2523 [FREE Full text] [doi: 10.1016/j.mayocp.2019.04.018] [Medline: 31806104]
93. Yellowlees P. Impact of COVID-19 on Mental Health Care Practitioners. Psychiatr Clin North Am 2022 Mar;45(1):109-121 [FREE Full text] [doi: 10.1016/j.psc.2021.11.007] [Medline: 35219432]
94. Fong R, Tsai CF, Yiu OY. The Implementation of Telepractice in Speech Language Pathology in Hong Kong During the COVID-19 Pandemic. Telemed J E Health 2021 Jan 01;27(1):30-38 [FREE Full text] [doi: 10.1089/tmj.2020.0223] [Medline: 32667859]
95. Wahezi SE, Duarte RA, Yerra S, Thomas MA, Pujar B, Sehgal N, et al. Telemedicine During COVID-19 and Beyond: A Practical Guide and Best Practices Multidisciplinary Approach for the Orthopedic and Neurologic Pain Physical Examination. Pain Phys 2020;23:S205-S237. [doi: 10.36076/ppj.2020/23/s205]
96. Waller G, Pugh M, Mulkens S, Moore E, Mountford VA, Carter J, et al. Cognitive-behavioral therapy in the time of coronavirus: Clinician tips for working with eating disorders via telehealth when face-to-face meetings are not possible. Int J Eat Disord 2020 Jul 08;53(7):1132-1141 [FREE Full text] [doi: 10.1002/eat.23289] [Medline: 32383530]
97. Smith K, Ostinelli E, Macdonald O, Cipriani A. COVID-19 and Telepsychiatry: Development of Evidence-Based Guidance for Clinicians. JMIR Ment Health 2020 Aug 28;7(8):e21108 [FREE Full text] [doi: 10.2196/21108] [Medline: 32658857]
98. Digital technologies and telepsychiatry. Oxford Health Biomedical Research Centre. URL: https://oxfordhealthbrc.nihr.ac.uk/our-work/oxppl/table-5-digital-technologies-and-telepsychiatry/ [accessed 2022-07-30]
99. Domestic violence and abuse. Oxford Health Biomedical Research Centre. URL: https://oxfordhealthbrc.nihr.ac.uk/our-work/oxppl/domestic-violence-and-abuse/ [accessed 2022-07-30]
100. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital Mental Health and COVID-19: Using Technology Today to Accelerate the Curve on Access and Quality Tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
101. Clozapine treatment. Oxford Health Biomedical Research Centre. URL: https://oxfordhealthbrc.nihr.ac.uk/our-work/oxppl/table-1-clozapine-treatment/ [accessed 2022-07-30]
102. Krzyzaniak N, Greenwood H, Scott AM, Peiris R, Cardona M, Clark J, et al. The effectiveness of telehealth versus face-to face interventions for anxiety disorders: A systematic review and meta-analysis. J Telemed Telecare 2021 Dec 03:1357633X2110537 [FREE Full text] [doi: 10.1177/1357633x211053738]
103. Loucas C, Fairburn CG, Whittington C, Pennant ME, Stockton S, Kendall T. E-therapy in the treatment and prevention of eating disorders: A systematic review and meta-analysis. Behav Res Ther 2014 Dec;63:122-131 [FREE Full text] [doi: 10.1016/j.brat.2014.09.011] [Medline: 25461787]
104. Hoffman L, Wisniewski H, Hays R, Henson P, Vaidyam A, Hendel V, et al. Digital Opportunities for Outcomes in Recovery Services (DOORS): A Pragmatic Hands-On Group Approach Toward Increasing Digital Health and Smartphone Competencies, Autonomy, Relatedness, and Alliance for Those With Serious Mental Illness. J Psychiatr Pract 2020 Mar;26(2):80-88 [FREE Full text] [doi: 10.1097/PRA.0000000000000450] [Medline: 32134881]
105. Triana AJ, Gusdorf RE, Shah KP, Horst SN. Technology Literacy as a Barrier to Telehealth During COVID-19. Telemed J E Health 2020 Sep 01;26(9):1118-1119 [FREE Full text] [doi: 10.1089/tmj.2020.0155] [Medline: 32429770]
106. Ibragimov K, Palma M, Keane G, Ousley J, Crowe M, Carreño C, M. S. F. Mental Health Working Group. Shifting to Tele-Mental Health in humanitarian and crisis settings: an evaluation of Médecins Sans Frontières experience during the COVID-19 pandemic. Confl Health 2022 Mar 14;16(1):6 [FREE Full text] [doi: 10.1186/s13031-022-00437-1] [Medline: 35164807]
107. Dimitris MC, Gittings M, King NB. How global is global health research? A large-scale analysis of trends in authorship. BMJ Glob Health 2021 Jan 26;6(1):e003758 [FREE Full text] [doi: 10.1136/bmjgh-2020-003758] [Medline: 33500263]
108. Shore J. Managing Virtual Hybrid Psychiatrist-Patient Relationships in a Digital World. JAMA Psychiatry 2020 May 01;77(5):541-542 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.0139] [Medline: 32159756]
109. Shore J, Schneck CD, Mishkind MC. Telepsychiatry and the Coronavirus Disease 2019 Pandemic-Current and Future Outcomes of the Rapid Virtualization of Psychiatric Care. JAMA Psychiatry 2020 Dec 01;77(12):1211-1212 [FREE Full text] [doi: 10.1001/jamapsychiatry.2020.1643] [Medline: 32391861]

Abbreviations

MERS: Middle East respiratory syndrome
TMH: telemental health
WHO: World Health Organization

Edited by J Torous; submitted 08.04.22; peer-reviewed by N Kozloff, D Hilty; comments to author 14.06.22; revised version received 05.07.22; accepted 21.07.22; published 22.08.22.

Please cite as:
Zangani C, Ostinelli EG, Smith KA, Hong JSW, Macdonald O, Reen G, Reid K, Vincent C, Syed Sheriff R, Harrison PJ, Hawton K, Pitman A, Bale R, Fazel S, Geddes JR, Cipriani A
Impact of the COVID-19 Pandemic on the Global Delivery of Mental Health Services and Telemental Health: Systematic Review
JMIR Ment Health 2022;9(8):e38600
URL: https://mental.jmir.org/2022/8/e38600
doi:10.2196/38600
PMID:35994310

Original Paper

Predicting Patient Wait Times by Using Highly Deidentified Data in Mental Health Care: Enhanced Machine Learning Approach

Amir Rastpour[1], BSc, MSc, PhD; Carolyn McGregor[1,2], BSc, PhD

[1]Faculty of Business and Information Technology, Ontario Tech University, Oshawa, ON, Canada

[2]Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, Australia

Corresponding Author:
Amir Rastpour, BSc, MSc, PhD
Faculty of Business and Information Technology
Ontario Tech University
2000 Simcoe St N
Oshawa, ON, L1G 0C5
Canada
Phone: 1 905 721 8668 ext 2830
Fax: 1 905 721 3167
Email: amir.rastpour@ontariotechu.ca

Abstract

Background: Wait times impact patient satisfaction, treatment effectiveness, and the efficiency of care that the patients receive. Wait time prediction in mental health is a complex task and is affected by the difficulty in predicting the required number of treatment sessions for outpatients, high no-show rates, and the possibility of using group treatment sessions. The task of wait time analysis becomes even more challenging if the input data has low utility, which happens when the data is highly deidentified by removing both direct and quasi identifiers.

Objective: The first aim of this study was to develop machine learning models to predict the wait time from referral to the first appointment for psychiatric outpatients by using real-time data. The second aim was to enhance the performance of these predictive models by utilizing the system's knowledge while the input data were highly deidentified. The third aim was to identify the factors that drove long wait times, and the fourth aim was to build these models such that they were practical and easy-to-implement (and therefore, attractive to care providers).

Methods: We analyzed retrospective highly deidentified administrative data from 8 outpatient clinics at Ontario Shores Centre for Mental Health Sciences in Canada by using 6 machine learning methods to predict the first appointment wait time for new outpatients. We used the system's knowledge to mitigate the low utility of our data. The data included 4187 patients who received care through 30,342 appointments.

Results: The average wait time varied widely between different types of mental health clinics. For more than half of the clinics, the average wait time was longer than 3 months. The number of scheduled appointments and the rate of no-shows varied widely among clinics. Despite these variations, the random forest method provided the minimum root mean square error values for 4 of the 8 clinics, and the second minimum root mean square error for the other 4 clinics. Utilizing the system's knowledge increased the utility of our highly deidentified data and improved the predictive power of the models.

Conclusions: The random forest method, enhanced with the system's knowledge, provided reliable wait time predictions for new outpatients, regardless of low utility of the highly deidentified input data and the high variation in wait times across different clinics and patient types. The priority system was identified as a factor that contributed to long wait times, and a fast-track system was suggested as a potential solution.

(JMIR Ment Health 2022;9(8):e38428) doi:10.2196/38428

KEYWORDS

mental health care; outpatient clinics; wait time prediction; machine learning; random forest; system's knowledge

Introduction

The length and predictability of wait times are important factors that impact patient satisfaction, treatment effectiveness, and the efficiency of care that the patients receive. Providing patients with accurate wait time predictions and informing them about potential appointment delays increase the patients' satisfaction level and enable care providers and staff members to manage the patient flow more effectively and efficiently [1-3]. Lengthy wait times are significantly associated with prognosis deterioration in mental health care [4] and are associated with higher rates of no-shows that adversely impact wait time management. The issue of long wait times is worse for children and youth with mental health problems, with some waiting as long as 2.5 years [5].

A great deal of research has been conducted on wait time prediction and identification of factors that drive lengthy wait times in physical health care sectors, including emergency departments [6], maternity emergency rooms [7], radiology departments [3], and oncology departments [8]. These wait time prediction models are usually developed for systems in which care is provided during a single visit to a care provider and the care is provided to patients individually. In contrast, mental health care is offered in a different context: the care is usually provided through multiple consecutive visits, the number of which is not necessarily known at the beginning of treatment, and the care can be provided to a group of patients, as in group consultation sessions. In addition, the psychiatric clinics also face a high rate of no-shows that make the task of wait time prediction even more difficult. Because of these intrinsic differences in the care provided for patients with physical problems and psychiatric patients, the wait time models developed for physical care cannot be readily used in the context of psychiatric care.

The task of predicting wait times becomes even more challenging when the available data are highly deidentified, that is, all direct identifiers (such as name, address, license plate) and quasi identifiers (such as gender, date of birth, zip code) are removed. Although it is a common practice to deidentify research data by removing the direct identifiers, the removal of quasi identifiers makes highly deidentified data more attractive from the privacy point of view, but at the same time, it compromises the utility of the information in the data that can otherwise be used to analyze and improve the system [9,10].

We use state-of-the-art machine learning (ML) methods to predict outpatient wait times at a tertiary mental health hospital by using real-time highly deidentified data. In addition, we use these models to identify key factors that drive the wait times. ML methods are sophisticated tools that can capture hidden patterns in large and imperfect data more effectively than conventional linear regression methods. ML methods are resistant to noise in the data and they quickly adapt to operational changes in wait time management processes without human supervision [11-14]. ML methods have been widely used in the mental health care sector for diagnosis [15-18], prognosis [19-22], treatment [23-25], and other medical purposes. A few literature review papers provide systematic surveys of these papers [26-29]. However, to the best of our knowledge, ML methods have not yet been applied to provide wait time prediction in the mental health care sector, although other health care sectors have benefitted from these sophisticated methods in their waiting list management.

A system's knowledge is obtained through systems thinking, which is defined as seeing the relationship among components (rather than seeing the components individually) and observing the patterns of change (rather than static "snapshots") [30]. In the context of an emergency department, it has been shown that ML models provide more accurate wait time predictions when they are enhanced with the system's knowledge in the *presence* of quasi identifiers such as age and gender [31]. We obtain and use the system's knowledge to enhance the predictive power of our ML models in the *absence* of quasi identifiers.

The first objective of this study was to develop 6 ML methods (namely, linear regression, random forest, weighted k-nearest neighbors, support vector machine, neural network, and decision tree) for real-time prediction of wait time for new outpatients in 8 outpatient clinics in Ontario Shores Centre for Mental Health Sciences (Ontario Shores) in Ontario, Canada. The second objective was to enhance the predictive power of ML models by using the system's knowledge while having highly deidentified input data. The third objective was to assess variable importance to understand what factors drove long wait times. The fourth objective was to develop models such that care providers could understand and use them relatively easily without the need for background knowledge on ML models and their implementation.

Methods

Data Source and Data Preparation

In this research, we used highly deidentified retrospective administrative data from Ontario Shores to build ML models for predicting new outpatients' wait times. Our focus was on 8 outpatient clinics, namely, Anxiety and Mood Disorders (AMD) Clinic, Traumatic Stress Clinic, Borderline Personality Self-Regulation Clinic, Women's Clinic, Prompt Care Clinic, Prompt AMD, Prompt Transitional Aged Youth, and Prompt Adolescent Consultation. Our data included 4998 patients whose first appointment was between April 1, 2017 and September 30, 2019 (both days inclusive). We excluded 30 (0.6%) patients because of missing referral dates. Selection of patients based on the date of their first appointment, rather than their referral date, had caused a selection bias in favor of 2 groups of patients: (1) those who had longer wait times among patients whose referral date was just before April 1, 2017 and (2) those who had shorter wait times among patients whose referral date was just before September 30, 2019. To address this selection bias, we removed all patients whose referral date was before April 1, 2017, and for each clinic, we removed all patients whose referral date was after September 30, 2019 minus the 80th percentile of the wait times in that specific clinic. After removing the biased data, we were left with 4187 referral entries. Table 1 shows the breakdown of patient count by clinic.

Table 1. Summary statistics of patients' wait time across different clinics (N=4187).

Clinic	Patients (n)	Mean (SD) days	Median days
Anxiety and Mood Disorders Clinic	298	98.08 (105.71)	64
Traumatic Stress Clinic	203	173.85 (113.58)	165
Borderline Personality Self-Regulation Clinic	181	107.42 (60.33)	112
Women's Clinic	155	80.19 (47.08)	75
Prompt Care Clinic	2338	29.05 (23.90)	21
Prompt Anxiety and Mood Disorders Consultation	436	186.42 (138.16)	205.5
Prompt Transitional Aged Youth Consultation	402	54.5 (45.77)	37
Prompt Adolescent Consultation	174	97.14 (57.17)	90.5

Variables

Outcome Variable

We aimed to predict the wait time, defined as the time from referral to the first appointment.

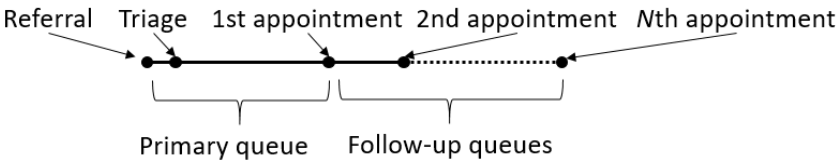
Predictor Variables Included in Our Data Set

Relevant variables for each patient were selected from the electronic health record. The medical variables included referral date, triage date, priority level (low, medium, and high) designated at triage, all appointment dates, status of each appointment (attended, no-show, and cancelled), and possible status changes while waiting or receiving care.

Engineered Predictor Variables

To better understand how our 8 clinics provided care to their patients, we had monthly feedback sessions with the care providers. In these sessions, components of the care system were identified and their interactions were outlined. Figure 1 presents a schematic view of the care processes at our clinics of interest. At first, care providers (usually family doctors) sent referrals to Ontario Shores. Then, a triage clinician assessed the patients and made intake decisions (accept or decline). Accepted patients entered the wait list associated with each clinic and remained there until the first appointment with their clinician. The outcome variable, wait time, was actually the wait time of the primary queue, as presented in Figure 1. Primary queue patients who were ahead of the new arrival directly impacted the new patient's wait time. Understanding how the system worked led us to understand that although follow-up queues were downstream to the primary queue, they also indirectly impacted the wait time of the new patients. This indirect impact happened because the follow-up queues utilized the same resources (clinicians) that the primary queue utilized. Therefore, the new patient's wait time depended on how much of the resource capacity was utilized by the follow-up queues. Another important output from our systems analysis discussions was that the clinics had utilizations close to 100%, which meant all of the care capacity offered by the clinics were assigned to patients. This helped us to approximate the offered care capacity by adding up the provided care. Obtaining the system's knowledge led us to define and measure the following predictor variables: (1) number of patients from each priority level in primary and follow-up queues, (2) the accumulative wait time of patients from each priority level who were in follow-up queues, (3) the accumulative amount of service (treatment) that patients from each priority level in follow-up queues (note that the amount of service received in the primary queue is zero) had already received, (4) the accumulative amount of time patients from each priority level had already spent in follow-up queues, and (5) the total care capacity during 30-, 60-, and 90-day time windows just before the referral date.

Figure 1. A schematic view of an outpatient receiving mental health care.



Missing Values

Some patients were missing their designated priority level at triage. However, those patients had a priority designated to them at a later date (possibly owing to re-evaluations while waiting). If a priority level at triage was missing, we replaced that with the priority level designated at the closest date after triage.

Dimensionality Reduction

High dimensionality, that is, having too many variables in a model, may cause many complications, including overfitting and producing a higher sampling variance (ie, sensitivity to small fluctuations in the training set) [32-34]. We reduced the dimensionality of our data by selecting a subset of variables (and discarding the rest) while retaining as much information as possible from all variables. We kept all of the medical variables, and among variables obtained from systems analysis, we calculated the pairwise Pearson correlation and removed the redundant information if the correlation was larger than 90%, as in [35].

Outliers

We used the generalized extreme studentized deviate method to identify the outliers [36]. This method iteratively applies the generalized extreme studentized deviate test and progressively evaluates anomalies by removing potential outliers and recalculating the test statistic and the associated critical value. The procedure continues until all outliers are identified.

ML Methods

Implementation

We examined 6 different ML methods, namely, linear regression, random forest, weighted k-nearest neighbors, support vector machine, neural network, and decision tree [11]. We used R version 4.0.2 (2020-06-22) and developed all our predictive models in the tidymodels ecosystem of packages [37]. The tidymodels ecosystem was used to streamline the modeling procedure and to avoid coding variations caused by using separate packages for each ML tool. Streamlining the modeling procedure simplified the implementation and debugging steps and therefore made the models more likely to be used by Ontario Shores.

Tuning and Evaluation

For each of the ML modeling approaches, there were multiple hyperparameters that we needed to tune to make sure that the obtained output was the best (or close-to-best) possible from that model. To obtain good models and to avoid overfitting, data for each clinic were randomly divided into training (75% of the data) and testing (the remaining 25%) sets. First, we applied the Latin hypercube sampling method [38] to create the search grid within the range of values of each hyperparameter. Then, we selected the best value for each hyperparameter by conducting an exhaustive grid search using the 10-fold cross-validation method. As our outcome variable, wait time, was continuous, we used the root mean square error (RMSE) to compare the performance of different models.

Ethics Approval

This study was approved by the Research Ethics Boards at Ontario Tech University (15596) and Ontario Shores (19-009-D).

Results

Summary Statistics

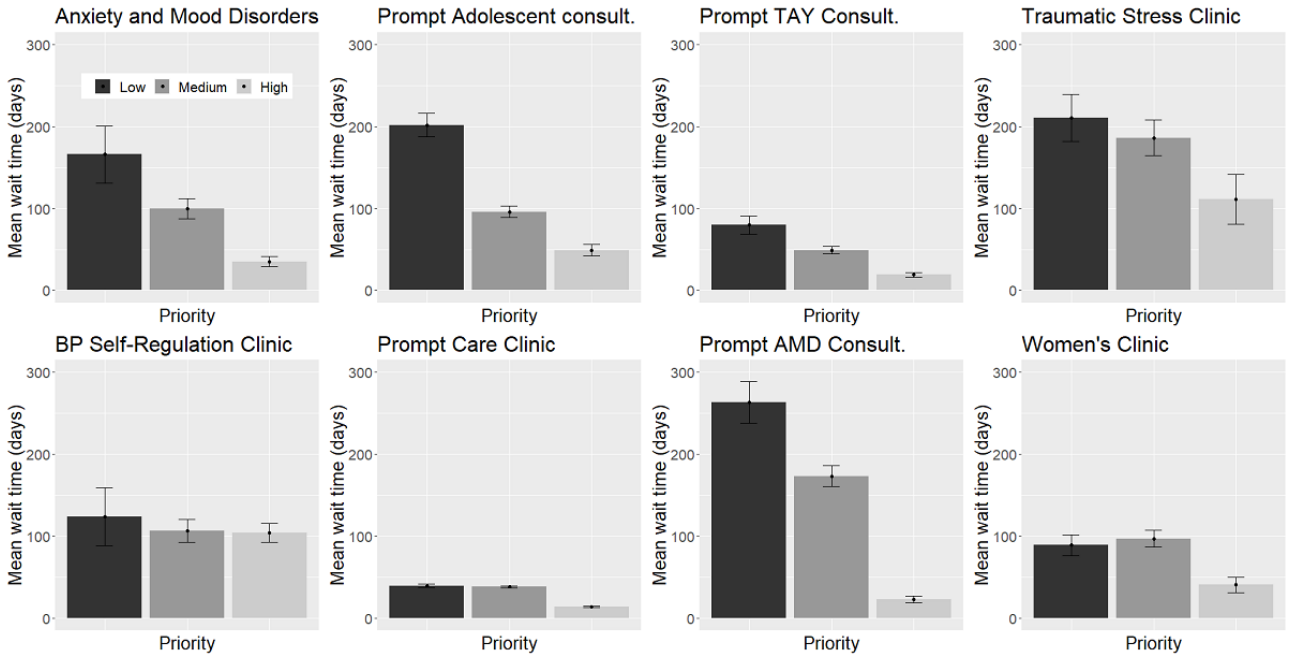
The summary statistics of our data for different clinics are shown in Table 1. The average wait time widely varied between the clinics from 29 days for Prompt Care Clinic to 186 days for Prompt AMD Consultation. For more than half of the clinics, the average wait time was longer than 3 months. We also calculated the median wait times, which varied between 21 days for Prompt Care Clinic and 205.5 days for Prompt AMD consultation, and confirmed long wait times. Our data also showed large standard deviation values, which varied between 23.9 days for Prompt Care Clinic and 138.16 days for Prompt AMD consultation. Having a high variation among patients' wait times made it difficult for both care providers and patients to be able to plan ahead.

The summary statistics of the appointments across different clinics are shown in Table 2. In total, the data set included 30,342 appointments out of which 4862 (16%) were no-shows. The number of scheduled appointments widely varied among clinics, from 307 appointments for Prompt Adolescent Consultation to 10,506 appointments for Borderline Personality Self-Regulation Clinic. The proportion of no-shows also widely varied across different clinics, from 2.9% (17/584) for Prompt AMD and Transitional Aged Youth Clinics to 22.2% (623/2804) for Women's Clinic. Figure 2 shows the average wait time and 95% CI of all patients stratified by the priority level and clinic. This figure illustrates that low-priority patients had the longest wait time in all clinics, except for the Women's Clinic, where the medium-priority patients had the longest average wait time.

Table 2. Summary statistics of all the appointments and no-show appointments per patient across clinics.

Clinic	All appointments (N=30,342)			No-show appointments (n=4862)		
	Patients (n)	Mean (SD)	Median	Patients, n (%)	Mean (SD)	Median
Anxiety and Mood Disorders Clinic	6830	21.68 (23.03)	17	1431 (20.9)	4.54 (6.84)	2
Traumatic Stress Clinic	5617	27.27 (16.08)	26	1148 (20.4)	5.57 (5.30)	4
Borderline Personality Self-Regulation Clinic	10,506	57.73 (50.18)	42.5	1453 (13.8)	7.98 (7.71)	5
Women's Clinic	2804	17.97 (14.53)	17.5	623 (22.2)	3.99 (5.16)	2
Prompt Care Clinic	3167	1.34 (0.8)	1	158 (4.9)	0.07 (0.32)	0
Prompt Anxiety and Mood Disorders Consultation	584	1.08 (0.28)	1	17 (2.9)	0.03 (0.19)	0
Prompt Transitional Aged Youth Consultation	527	1.05 (0.22)	1	14 (2.6)	0.03 (0.18)	0
Prompt Adolescent Consultation	307	1.74 (2.3)	1	18 (5.8)	0.10 (0.37)	0

Figure 2. The mean wait time with 95% confidence interval by priority level and clinic. AMD: Anxiety and Mood Disorders; BP: Borderline Personality; Consult.: Consultation; TAY: Transitional Aged Youth.



Model Performance

We first applied our dimensionality reduction approach to the training set. Variables introduced in the section "Engineered Predictor Variables" appeared to contain similar information. We dropped all of them except for "Number of patients from each priority level in primary and follow-up queues." We used these variables along with variables introduced in the section "Predictor Variables Included in Our Data Set" to build our ML models. Table 3 displays the best RMSE values obtained from each of the models for each of the clinics when we included the engineered predictors. For each clinic (ie, each row), the best performing method is italicized. As different clinics followed different operational schemes and their wait times had different profiles, there was not a single ML model that outperformed all of the rest across all clinics. However, the random forest method appeared to be the most promising method as it provided the minimum RMSE values for 4 (out of the 8 clinics) and the second minimum RMSE for the other 4 clinics. It is notable that the linear regression method, regardless of its simplicity, outperformed other ML methods at some clinics such as Women's Clinic. This can be attributed to the existence of linear patterns in the data [39], having small sample sizes [40], and the fact that the grid search method provides an optimal combination of the selected subset of hyperparameter values, but it cannot guarantee the global optimality of the output [13].

Table 3. Comparison of the root mean square error of different machine learning methods[a].

Clinic	Linear regression	Random forest	K-nearest neighbors	Support vector machine	Neural network	Decision tree
Anxiety and Mood Disorders Clinic	66.88	*49.89*	70.37	52.64	83.44	50.65
Traumatic Stress Clinic	94.02	93.54	98.95	*86.6*	108.36	102.71
Borderline Personality Self-Regulation Clinic	50	*49.47*	51.94	50.61	61.98	56.16
Women's Clinic	*33.45*	36.16	42	39.63	46.93	56.73
Prompt Care Clinic	19.04	*16.49*	16.83	17.13	16.87	17.13
Prompt Anxiety and Mood Disorders Consultation	121.25	119.6	125.64	*117.83*	142.53	131.28
Prompt Transitional Aged Youth Consultation	29.82	26.19	26.3	*23.56*	28.92	28.34
Prompt Adolescent Consultation	20.4	*18.04*	31.8	19.13	26.6	20.84

[a]For each clinic (ie, each row), the best performing method is italicized.

Hyperparameter Tuning

Table 4 displays the list of hyperparameters that we used to tune each of the ML methods, the range of values for each hyperparameter, and the selected values for the AMD clinic. The selected values varied across clinics.

Table 5 displays the selected values of the random forest hyperparameters across different clinics. For some settings, the neural network method with 1 hidden layer could be the same as the linear regression method [12]; to avoid duplications, we did not consider such settings for the neural network method.

Table 4. Hyperparameters used for tuning the machine learning methods and their selected values for the Anxiety and Mood Disorders Clinic.

Machine learning method, parameter	Range	Selected value	Explanation
Linear regression	N/A[a]	N/A	N/A
Random forest			
mtry	1 to 20	16	Number of predictors at each split
min_n	2 to 40	14	Minimum node size
K-nearest neighbors			
neighbors	1 to 15	13	Number of neighbors to consider
dist_power	0.1 to 2	0.21	Minkowski distance parameter
weight_func	__[b]	Rectangular	Kernel function for weighting sample distribution
Support vector machine			
cost	2^{-10} to 2^{5}	22.31	The cost of wrong predictions
rbf_sigma	10^{-10} to 10^{0}	$10^{-1.76}$	Radial basis function parameter
margin	0 to 0.2	0.11	Epsilon for support vector machine insensitive loss function
Neural network			
hidden_units	1 to 10	9	Number of units in the hidden model
penalty	10^{-10} to 10^{0}	$10^{-0.39}$	Amount of weight decay
epochs	10^{1} to 10^{3}	993	Number of training iterations
Decision tree			
cost_complexity	10^{-10} to 10^{-1}	$10^{-8.02}$	Cost/complexity parameters
tree_depth	1 to 15	3	Maximum depth of the tree
min_n	2 to 40	17	Minimum node size

[a]N/A: not applicable.

[b]triweight, triangular, rectangular, rank, optimal, inv, gaussian, epanechnikov, cos, biweight.

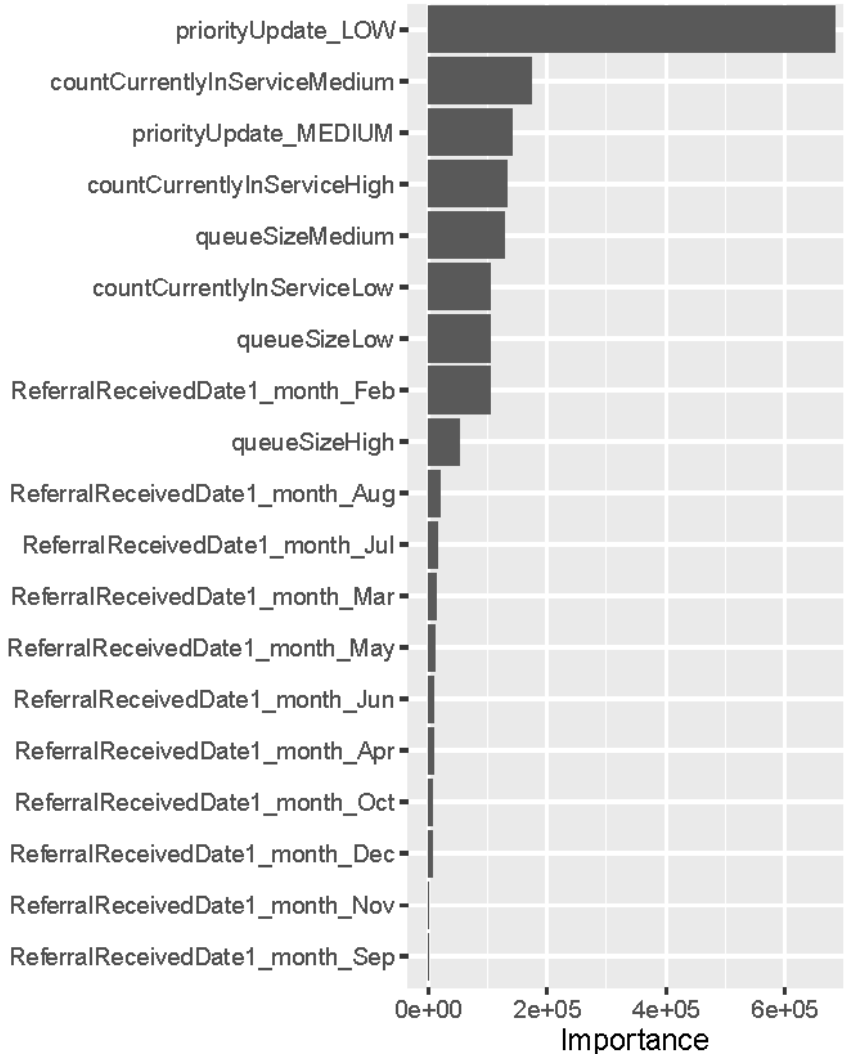
Table 5. Hyperparameters of the random forest method across different clinics.

Clinic	Count of splitting variables	Count of trees	Minimal node size
Anxiety and Mood Disorders Clinic	16	1000	14
Traumatic Stress Clinic	17	1000	29
Borderline Personality Self-Regulation Clinic	2	1000	31
Women's Clinic	17	1000	39
Prompt Care Clinic	17	1000	39
Prompt Anxiety and Mood Disorders Consultation	19	1000	30
Prompt Transitional Aged Youth Consultation	11	1000	10
Prompt Adolescent Consultation	11	1000	10

Variable Importance

The random forest method provides measures of importance for predictor variables. These measures of importance help the user to identify variables that have the most and the least impacts on the outcome variable. Figure 3 displays the importance of predictor variables, measured by impurity (variance of the responses) at the AMD Clinic. The rankings of the importance of predictor variables were similar to those shown in Figure 3 at other clinics. According to Figure 3, *priorityUpdate*, *countCurrentlyInService*, and *queueSize* were the most influential variables. In our models, *priorityUpdate* variables denoted the last priority assigned to each patient, *countCurrentlyInService* variables denoted how many patients of different priority levels were currently receiving service (ie, were in the follow-up queues) at the referral time, and *queueSize* variables denoted how many patients of different priority levels were waiting in the primary queue at the referral time. The seasonality variables did not play important roles in wait time prediction.

Figure 3. Importance of the predictor variables, measured by impurity (variance of the responses), at the Anxiety and Mood Disorders Clinic.



Discussion

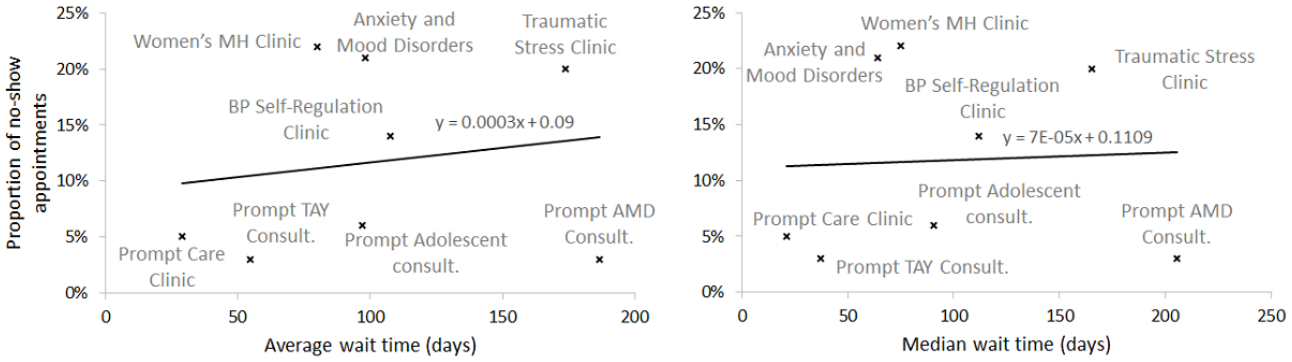
Current State

Although operations management and ML tools have been widely used in different sectors of care for physical diseases to improve the waiting list management, there has not been such studies in the mental health care sector. Previous research has demonstrated positive effects of operational and policy improvements that have taken place to improve the wait list management in physical health care, for example, cancer care [41]. In 2015, the Canadian Wait Time Alliance [42] reported that although there had been significant progress in the wait time management in 5 areas of focus in the 2004 Health Accord (hip and knee replacement, cataract, bypass surgery, radiation therapy, and diagnostic imaging), mental health care struggled with long waiting times and required immediate attention nationwide. That report also outlined that universal measures did not even exist to track access to psychiatric care across the country. Loebach and Ayoubzadeh [43] compared the waiting times for psychiatric patients and patients with physical problems in the province of Ontario, Canada, and concluded that while the former group often ended up waiting beyond the target waiting times specified by the province, the latter group often received their treatments within the target time window.

No-show Rates

In addition to adverse impacts that long wait times have on patients' health conditions, they also cause higher no-show rates that cause operational complications for health care managers [44,45]. The no-show rate may depend on factors such as wait time and quality of care and may vary between 5% and 80% in different health care sectors [44,45]. Figure 4 illustrates the positive correlation between longer wait times and higher no-show rates in our data, which indicates that shortening wait times may decrease the no-show rates as well.

Figure 4. Correlation between no-show appointments and wait times. AMD: Anxiety and Mood Disorders; BP: Borderline Personality; Consult.: Consultation; MH: Mental Health; TAY: Transitional Aged Youth.



Random Forest Method

In this research, we applied 6 different ML methods to predict wait times in mental health care and to identify factors that drive the long wait times. The input data were highly deidentified, which limited the data utility. The random forest method enhanced by the system's knowledge turned out to be the most promising method. The good performance of this model could be attributed to some appealing computational features of the random forest method, including its low sensitivity to outliers and its ability in capturing complex interactions between predictor variables [46]. From a practical point of view, another appealing feature of the random forest method is its relatively low sensitivity to parameter tuning [46]. Identifying the random forest method as a superior ML method to predict wait times is consistent with findings of [47] that identify this method as the most accurate method to calculate the probability of waiting more than 1 day before receiving treatment for patients with opioid use disorder.

Managerial Insights

The impurity measure of variable importance, which was the basis for ranking predictive variables in Figure 3, indicates that the total amount of increase in the mean square error of the model output resulted from a random permutation of each variable in the test set. According to Figure 3, the long wait times can be attributed to the usage of the priority system in assigning the care resources to patients. In priority systems, low-priority patients are preceded by patients from higher priority levels and may end up waiting an extended period of time to receive a relatively simple treatment. It is likely that during their long wait times, low-priority patients are changed to higher-priority patients owing to condition deterioration. This phenomenon has been observed in other health care sectors, and using the "fast-track" system for low-priority patients has been suggested as a potential solution [48-50]. In a fast-track system, the waiting line is broken into 2 separate lines: one for low-priority patients and one for patients with higher priorities. One potential advantage of this approach is that because of the simpler nature of care required by low-priority patients, they can be attended by "less-trained" clinicians, freeing up the "more-trained" clinicians for patients with more complex needs. One potential disadvantage of the fast-track system is that the improvement in waiting time of low-priority patients may come at the cost of longer wait times for patients with higher priorities.

Limitations

Small sample sizes coupled with very large variations within wait times of each clinic imposed the main limitation in this study. There was also a significant difference between the wait time profiles of clinics such that the generalized models for all clinics performed poorly in comparison to models for individual clinics. In addition, the following approximations also impacted the accuracy of model predictions.

1. Care Resource Capacity Limitation: There was no access to the real capacity offered to patients at each clinic at a given day and therefore, we created proxy variables to approximate the capacity.
2. Group Meeting Limitation: Of the clinics that we reviewed within Ontario Shores, the AMD, Traumatic Stress, Borderline Personality Self-Regulation, and Women's clinics provided care through group meetings where multiple patients attend at the same time. The dynamics of group visits in clinics that provide group treatments were not clear and therefore, we could not explicitly capture the potential impacts of these treatments on wait times.

Conclusion

In this study, we used retrospective highly deidentified administrative data from 8 clinics at Ontario Shores to build 6 different ML models to predict wait times. We enhanced our models by system knowledge to mitigate the limiting impact of deidentification on our data utility. The data included 4187 patients who received care through 30,342 appointments. The random forest method provided the minimum RMSE values for 4 of the 8 clinics and the second minimum RMSE for the other 4 clinics. The priority system was identified as a factor that contributed to long wait times, and a fast-track system was suggested as a potential solution. Despite the challenges with the wait time source data, this research provided Ontario Shores with a deeper understanding of the extent of and contributors to their wait times on a clinic-by-clinic basis. This research provided Ontario Shores with information and knowledge to pursue quality improvement initiatives to reduce wait times.

Acknowledgments

We are grateful to Ontario Shores and its members for providing the data, partial funding, and helping us with understanding the data set. Ontario Shores provided partial funding for this research. Ontario Shores had no role in the study design, analysis, interpretation of the findings, writing the manuscript, and the decision to submit the paper for publication. The authors thank the editorial team for their thoughtful comments that led to considerable improvements in this paper.

Authors' Contributions

AR was involved with the conception and design of the study, and he conducted all the statistical and predictive analyses. He jointly drafted the initial and revision documents with CMG and created intellectual contents. CMG provided the clinical context information for mental health and Ontario Shores. She assisted with the design of the analysis with regard to inclusion and exclusion criteria for clinics and patient types based on the various challenges with the data for use by the machine learning techniques. Both authors gave their approval of the final version to be submitted for peer review.

Conflicts of Interest

None declared.

References

1. Camacho F, Anderson R, Safrit A, Jones A, Hoffmann P. The Relationship between Patient's Perceived Waiting Time and Office-Based Practice Satisfaction. North Carolina Medical Journal 2006 Nov 01;67(6):409-413 [FREE Full text] [doi: 10.18043/ncm.67.6.409]
2. Jaworsky C, Pianykh O, Oglevee C. Patient Feedback on Waiting Time Displays. Am J Med Qual 2017;32(1):108. [doi: 10.1177/1062860616658974] [Medline: 27418619]
3. Curtis C, Liu C, Bollerman T, Pianykh O. Machine Learning for Predicting Patient Wait Times and Appointment Delays. J Am Coll Radiol 2018 Sep;15(9):1310-1316. [doi: 10.1016/j.jacr.2017.08.021] [Medline: 29079248]
4. Reichert A, Jacobs R. The impact of waiting time on patient outcomes: Evidence from early intervention in psychosis services in England. Health Econ 2018 Nov;27(11):1772-1787 [FREE Full text] [doi: 10.1002/hec.3800] [Medline: 30014544]
5. Kids can't wait: 2020 provincial budget recommendations for high-quality, accessible child and youth mental health and addictions care for all Ontario families. Children's Mental Health Ontario. URL: https://cmho.org/wp-content/uploads/CMHO-Report-WaitTimes-2020.pdf [accessed 2022-07-30]
6. Ang E, Kwasnick S, Bayati M, Plambeck EL, Aratow M. Accurate Emergency Department Wait Time Prediction. MSOM 2016 Feb;18(1):141-156 [FREE Full text] [doi: 10.1287/msom.2015.0560]
7. Pereira S, Portela F, Santos M, Machado J, Abelha A. Predicting Pre-triage Waiting Time in a Maternity Emergency Room Through Data Mining. Smart Health 2016:105 [FREE Full text] [doi: 10.1007/978-3-319-29175-8_10]
8. Joseph A, Hijal T, Kildea J, Hendren L, Herrera D. Predicting waiting times in radiation oncology using machine learning. 2018 Presented at: 16th IEEE International Conference on Machine Learning and Applications; January; Cancun, Mexico. [doi: 10.1109/icmla.2017.00-16]
9. Deidentification guidelines for structured data. Information and Privacy Commissioner of Ontario. 2016. URL: https://www.ipc.on.ca/wp-content/uploads/2016/08/Deidentification-Guidelines-for-Structured-Data.pdf [accessed 2022-07-30]
10. Zaman ANK, Obimbo C, Dara R. An improved differential privacy algorithm to protect re-identification of data. 2017 Presented at: IEEE Canada International Humanitarian Technology Conference (IHTC); July 21-22; Toronto, Canada. [doi: 10.1109/IHTC40297.2017]
11. Murphy K. Machine Learning: A Probabilistic Perspective. Cambridge, MA, USA: MIT Press; 2012.
12. James G, Witten D, Hastie T, Tibshirani R. Chapter 10. In: An Introduction to Statistical Learning: With Applications in R. New York, NY, USA: Springer; 2013.
13. Kuhn M, Johnson K. Applied Predictive Modeling. NY, USA: Springer; 2013.
14. Machine learning: what it is and why it matters. SAS Institute. URL: https://www.sas.com/en_us/insights/analytics/machine-learning.html [accessed 2022-07-30]
15. Hueniken K, Somé NH, Abdelhack M, Taylor G, Elton Marshall T, Wickens C, et al. 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 Nov 17;8(11):e32876 [FREE Full text] [doi: 10.2196/32876] [Medline: 34705663]
16. Sonkurt H, Altınöz AE, Çimen E, Köşger F, Öztürk G. The role of cognitive functions in the diagnosis of bipolar disorder: A machine learning model. Int J Med Inform 2021 Jan;145:104311. [doi: 10.1016/j.ijmedinf.2020.104311] [Medline: 33202371]
17. Mumtaz W, Qayyum A. A deep learning framework for automatic diagnosis of unipolar depression. Int J Med Inform 2019 Dec;132:103983. [doi: 10.1016/j.ijmedinf.2019.103983] [Medline: 31586827]
18. Wshah S, Skalka C, Price M. Predicting Posttraumatic Stress Disorder Risk: A Machine Learning Approach. JMIR Ment Health 2019 Jul 22;6(7):e13946 [FREE Full text] [doi: 10.2196/13946] [Medline: 31333201]

19. Cos H, Li D, Williams G, Chininis J, Dai R, Zhang J, et al. Predicting Outcomes in Patients Undergoing Pancreatectomy Using Wearable Technology and Machine Learning: Prospective Cohort Study. J Med Internet Res 2021 Mar 18;23(3):e23595 [FREE Full text] [doi: 10.2196/23595] [Medline: 33734096]
20. Wang M, Ge W, Apthorp D, Suominen H. Robust Feature Engineering for Parkinson Disease Diagnosis: New Machine Learning Techniques. JMIR Biomed Eng 2020 Jul 27;5(1):e13611 [FREE Full text] [doi: 10.2196/13611]
21. Morel D, Yu K, Liu-Ferrara A, Caceres-Suriel A, Kurtz S, Tabak Y. Predicting hospital readmission in patients with mental or substance use disorders: A machine learning approach. Int J Med Inform 2020 Jul;139:104136 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104136] [Medline: 32353752]
22. Hatton C, Paton L, McMillan D, Cussens J, Gilbody S, Tiffin P. Predicting persistent depressive symptoms in older adults: A machine learning approach to personalised mental healthcare. J Affect Disord 2019 Mar 01;246:857-860. [doi: 10.1016/j.jad.2018.12.095] [Medline: 30795491]
23. Crane N, Jenkins L, Bhaumik R, Dion C, Gowins J, Mickey B, et al. Multidimensional prediction of treatment response to antidepressants with cognitive control and functional MRI. Brain 2017 Feb;140(2):472-486 [FREE Full text] [doi: 10.1093/brain/aww326] [Medline: 28122876]
24. Hahn T, Kircher T, Straube B, Wittchen H, Konrad C, Ströhle A, et al. Predicting treatment response to cognitive behavioral therapy in panic disorder with agoraphobia by integrating local neural information. JAMA Psychiatry 2015 Jan;72(1):68-74. [doi: 10.1001/jamapsychiatry.2014.1741] [Medline: 25409415]
25. Whitfield-Gabrieli S, Ghosh S, Nieto-Castanon A, Saygin Z, Doehrmann O, Chai X, et al. Brain connectomics predict response to treatment in social anxiety disorder. Mol Psychiatry 2016 May;21(5):680-685. [doi: 10.1038/mp.2015.109] [Medline: 26260493]
26. Shatte A, Hutchinson D, Teague S. Machine learning in mental health: a scoping review of methods and applications. Psychol. Med 2019 Feb 12;49(09):1426-1448 [FREE Full text] [doi: 10.1017/s0033291719000151]
27. Thieme A, Belgrave D, Doherty G. Machine Learning in Mental Health. ACM Trans. Comput.-Hum. Interact 2020 Oct 31;27(5):1-53 [FREE Full text] [doi: 10.1145/3398069]
28. Cecula P, Yu J, Dawoodbhoy F, Delaney J, Tan J, Peacock I, et al. Applications of artificial intelligence to improve patient flow on mental health inpatient units - Narrative literature review. Heliyon 2021 Apr;7(4):e06626 [FREE Full text] [doi: 10.1016/j.heliyon.2021.e06626] [Medline: 33898804]
29. 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 [FREE Full text] [doi: 10.1038/s41746-020-0233-7] [Medline: 32219184]
30. Senge P. Chapter 5. In: The Fifth Discipline: The Art & Practice of the Learning Organization. Redfern, Sydney, Australia: Currency; 2006.
31. Kuo Y, Chan N, Leung J, Meng H, So A, Tsoi K, et al. An Integrated Approach of Machine Learning and Systems Thinking for Waiting Time Prediction in an Emergency Department. Int J Med Inform 2020 Jul;139:104143. [doi: 10.1016/j.ijmedinf.2020.104143] [Medline: 32330853]
32. Greene W. Econometric Analysis (6th Edition). USA: Pearson; 2008.
33. Witten IH, Frank E, Hall M. Data Mining: Practical Machine Learning Tools and Techniques (Morgan Kaufmann Series in Data Management Systems). San Francisco, USA: Morgan Kaufmann; 4th edition; Sep 30, 2011.
34. Shmueli G. To Explain or to Predict? Statist Sci 2010 Aug 1;25(3):289 [FREE Full text] [doi: 10.1214/10-sts330]
35. Mowbray F, Zargoush M, Jones A, de Wit K, Costa A. Predicting hospital admission for older emergency department patients: Insights from machine learning. Int J Med Inform 2020 Aug;140:104163. [doi: 10.1016/j.ijmedinf.2020.104163] [Medline: 32474393]
36. Rosner B. Percentage Points for a Generalized ESD Many-Outlier Procedure. Technometrics 1983 May;25(2):165-172 [FREE Full text] [doi: 10.1080/00401706.1983.10487848]
37. Kuhn M, Julia S. Tidy Modeling With R: A Framework for Modeling in the Tidyverse. 2021. URL: https://www.tmwr.org/ [accessed 2021-10-22]
38. McKay MD, Beckman RJ, Conover WJ. A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics 1979 May;21(2):239 [FREE Full text] [doi: 10.2307/1268522]
39. Kim Y. Comparison of the decision tree, artificial neural network, and linear regression methods based on the number and types of independent variables and sample size. Expert Systems with Applications 2008 Feb;34(2):1227-1234 [FREE Full text] [doi: 10.1016/j.eswa.2006.12.017]
40. Jiao S, Gao Y, Feng J, Lei T, Yuan X. Does deep learning always outperform simple linear regression in optical imaging? Opt. Express 2020 Jan 27;28(3):3717 [FREE Full text] [doi: 10.1364/oe.382319]
41. Rastpour A, Begen M, Louie A, Zaric G. Variability of waiting times for the 4 most prevalent cancer types in Ontario: a retrospective population-based analysis. CMAJ Open 2018;6(2):E227-E234 [FREE Full text] [doi: 10.9778/cmajo.20170118] [Medline: 29880658]
42. Eliminating code gridlock in Canada's health care system: 2015 wait time alliance report card. Wait Time Alliance. URL: https://www.waittimealliance.ca/wta-reports/2015-wta-report-card/ [accessed 2022-07-30]
43. Loebach R, Ayoubzadeh S. Wait times for psychiatric care in Ontario. UWOMJ 2017 Dec 03;86(2):48-50 [FREE Full text] [doi: 10.5206/uwomj.v86i2.2027]

44. Kheirkhah P, Feng Q, Travis L, Tavakoli-Tabasi S, Sharafkhaneh A. Prevalence, predictors and economic consequences of no-shows. BMC Health Serv Res 2016 Jan 14;16:13 [FREE Full text] [doi: 10.1186/s12913-015-1243-z] [Medline: 26769153]
45. Marbouh D, Khaleel I, Al Shanqiti K, Al Tamimi M, Simsekler M, Ellahham S, et al. Evaluating the Impact of Patient No-Shows on Service Quality. RMHP 2020 Jun;Volume 13:509-517 [FREE Full text] [doi: 10.2147/rmhp.s232114]
46. Zhang C, Ma Y. Ensemble Machine Learning: Methods and Applications. Boston, MA, USA: Springer; 2012.
47. Kong Y, Zhou J, Zheng Z, Amaro H, Guerrero E. Using machine learning to advance disparities research: Subgroup analyses of access to opioid treatment. Health Serv Res 2022 Apr;57(2):411-421 [FREE Full text] [doi: 10.1111/1475-6773.13896] [Medline: 34657287]
48. Cooke M, Wilson S, Pearson S. The effect of a separate stream for minor injuries on accident and emergency department waiting times. Emerg Med J 2002 Jan;19(1):28-30 [FREE Full text] [doi: 10.1136/emj.19.1.28] [Medline: 11777867]
49. Sanchez M, Smally A, Grant R, Jacobs L. Effects of a fast-track area on emergency department performance. J Emerg Med 2006 Jul;31(1):117-120. [doi: 10.1016/j.jemermed.2005.08.019] [Medline: 16798173]
50. Lin D, Patrick J, Labeau F. Estimating the waiting time of multi-priority emergency patients with downstream blocking. Health Care Manag Sci 2014 Mar;17(1):88-99 [FREE Full text] [doi: 10.1007/s10729-013-9241-3] [Medline: 23690253]

Abbreviations

AMD: Anxiety and Mood Disorders
ML: machine learning
RMSE: root mean square error

Edited by J Torous; submitted 31.03.22; peer-reviewed by M Zargoush, D Carvalho, T Sagi; comments to author 03.05.22; revised version received 18.06.22; accepted 18.07.22; published 09.08.22.

Please cite as:
Rastpour A, McGregor C
Predicting Patient Wait Times by Using Highly Deidentified Data in Mental Health Care: Enhanced Machine Learning Approach
JMIR Ment Health 2022;9(8):e38428
URL: https://mental.jmir.org/2022/8/e38428
doi:10.2196/38428
PMID:35943774

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org