Contents

Original Papers

Smartphone-Based Self-Reports of Depressive Symptoms Using the Remote Monitoring Application in Psychiatry (ReMAP): Interformat Validation Study (e24333)
Janik Goltermann, Daniel Emden, Elisabeth Leehr, Katharina Dohm, Ronny Redlich, Udo Dannlowski, Tim Hahn, Nils Opel. ............................. 3

Participant Engagement in a Transmedia Storytelling Web-Based App Intervention for Mental Health of Latina Women: Qualitative Analysis (e22575)
Patricia Soderlund, Adrienne Martinez Hollingsworth, MarySue Heilemann. .......................................................... 13

Considerations in Designing Digital Peer Support for Mental Health: Interview Study Among Users of a Digital Support System (Buddy Project) (e21819)
Nazanin Andalibi, Madison Flood. .......................................................... 27

Effects of ACT Out! Social Issue Theater on Social-Emotional Competence and Bullying in Youth and Adolescents: Cluster Randomized Controlled Trial (e25860)
Jon Agle y, Mi Kyoun g Jun, Lor i Eldridge, Daniel Agley, Yunyu Xiao, Steve Sussman, Lilian Golzani-Arroyo, Stephanie Dickinson, W a santh a Jayawardene, Ruth Gassman. ............................. 44

Psychiatric Profiles of eHealth Users Evaluated Using Data Mining Techniques: Cohort Study (e17116)
Jorge Lopez-Castroman, Diana Abad-Tortosa, Aurora Cobo Aguilera, Philippe Courtet, Maria Barrigón, Antonio Ar tés, Enrique Baca-García. 6

A Mindfulness-Based Intervention for Student Depression, Anxiety, and Stress: Randomized Controlled Trial (e23491)
Paul Ritvo, Farah Ahmad, Christo El Morr, Meysam Pirbaglou, Rahim Moineddin, MVC Team. .......................................................... 90

Comparing Effectiveness Between a Mobile App Program and Traditional Cognitive Behavior Therapy in Obsessive-Compulsive Disorder: Evaluation Study (e23778)
Hyunchan Hwang, Su Jin Bae, Ji Hong, Doug Hart. .......................................................... 107

A Blended Electronic Illness Management and Recovery Program for People With Severe Mental Illness: Qualitative Process Evaluation Alongside a Randomized Controlled Trial (e20860)
Titus Beentjes, Betsie van Gaal, Hester Vermeulen, Maria Nijhuis-van der Sanden, Peter Goossens. ......................................................... 121

The Effects of Downloading a Government-Issued COVID-19 Contact Tracing App on Psychological Distress During the Pandemic Among Employed Adults: Prospective Study (e23699)
Norito Kawakami, Natsu Sasaki, Reiko Kuroda, Kanami Tsuno, Kotaro Imamura. .......................................................... 133
Psychological Impact of the COVID-19 Pandemic on Chinese Health Care Workers: Cross-Sectional Survey Study (e23125)
Jie Ni, Fang Wang, Yihai Liu, Mingyue Wu, Yan Jiang, Yujie Zhou, Yujie Zhou, Dujuan Sha. ................................................................. 142

Review

Text Message Interventions in Adolescent Mental Health and Addiction Services: Scoping Review (e16508)
Sarah MacDougall, Susan Jerrott, Sharon Clark, Leslie Campbell, Andrea Murphy, Lori Wozney. ............................................................. 76

Corrigenda and Addenda

Correction: Evaluation of a Mobile App to Enhance Relational Awareness and Change During Cognitive Analytic Therapy: Mixed Methods Case Series (e27159)
Stephen Kellett, Katherine Easton, Martin Cooper, Abigail Millings, Melanie Simmonds-Buckley, Glenys Parry. ........................................ 131
Smartphone-Based Self-Reports of Depressive Symptoms Using the Remote Monitoring Application in Psychiatry (ReMAP): Interformat Validation Study

Janik Goltermann1*, MSc; Daniel Emden1*, Dipl-Inf; Elisabeth Johanna Leehr1, PhD; Katharina Dohm1, PhD; Ronny Redlich1,2, PhD; Udo Dannlowski1, MD, PhD; Tim Hahn1, PhD; Nils Opel1,3, MD

1Department of Psychiatry, University of Münster, Münster, Germany
2Institute of Psychology, University of Halle, Halle, Germany
3Interdisciplinary Centre for Clinical Research Münster, University of Münster, Münster, Germany
*these authors contributed equally

Abstract

Background: Smartphone-based symptom monitoring has gained increased attention in psychiatric research as a cost-efficient tool for prospective and ecologically valid assessments based on participants’ self-reports. However, a meaningful interpretation of smartphone-based assessments requires knowledge about their psychometric properties, especially their validity.

Objective: The goal of this study is to systematically investigate the validity of smartphone-administered assessments of self-reported affective symptoms using the Remote Monitoring Application in Psychiatry (ReMAP).

Methods: The ReMAP app was distributed to 173 adult participants of ongoing, longitudinal psychiatric phenotyping studies, including healthy control participants, as well as patients with affective disorders and anxiety disorders; the mean age of the sample was 30.14 years (SD 11.92). The Beck Depression Inventory (BDI) and single-item mood and sleep information were assessed via the ReMAP app and validated with non–smartphone-based BDI scores and clinician-rated depression severity using the Hamilton Depression Rating Scale (HDRS).

Results: We found overall high comparability between smartphone-based and non–smartphone-based BDI scores (intraclass correlation coefficient=0.921; \(P<.001\)). Smartphone-based BDI scores further correlated with non–smartphone-based HDRS ratings of depression severity in a subsample (\(r=0.783; \ P<.001; \ n=51\)). Higher agreement between smartphone-based and non–smartphone-based assessments was found among affective disorder patients as compared to healthy controls and anxiety disorder patients. Highly comparable agreement between delivery formats was found across age and gender groups. Similarly, smartphone-based single-item self-ratings of mood correlated with BDI sum scores (\(r=-0.538; \ P<.001; \ n=168\)), while smartphone-based single-item sleep duration correlated with the sleep item of the BDI (\(r=-0.310; \ P<.001; \ n=166\)).

Conclusions: These findings demonstrate that smartphone-based monitoring of depressive symptoms via the ReMAP app provides valid assessments of depressive symptomatology and, therefore, represents a useful tool for prospective digital phenotyping in affective disorder patients in clinical and research applications.

(JMIR Ment Health 2021;8(1):e24333) doi:10.2196/24333

KEYWORDS
mobile monitoring; smartphone; digital biomarkers; digital phenotyping; course of illness; psychometric quality; mood disorders; depression; affective disorders; mobile phone
Introduction

The phasic development of symptoms over time in the form of disease episodes is one of the key characteristics of affective disorders. These disease trajectories can be used as an informative predictor as well as an outcome measure in psychiatric research and personalized medicine. However, the assessment of the development of symptoms over time is challenging. The value of cross-sectional assessments is limited as they can only capture an excerpt of the symptom history and it is unclear whether this excerpt reflects, for example, the peak of an affective episode or a fully or partially remitted state and whether episodes are recurrent. Collecting this information retrospectively from the patients is one approach to gaining insights into their former symptom history, which is likely to be biased by their current depressive state [1]. Thus, multiple prospective assessments of symptoms are needed for a valid interpolation of the underlying disease trajectory. Although such prospective instruments based on a paper-and-pencil format exist [2], their use is limited due to low cost-efficiency as well as low patient compliance [3]. In recent years, the utilization of smartphone apps for psychological and psychiatric assessment has increased considerably due to the cost-efficiency and practicability of these apps [4-6].

Several proof-of-concept studies have pointed to the utility of smartphone-based data in affective disorder research [7]. Smartphone-based measures can be categorized into passive sensor data (eg, geolocation, distance, steps, acceleration, and app activities) and active self-report. The latter, which entails daily diaries, reiterated questionnaires, and ecological momentary assessments, utilizes multiple assessments per day, thereby acquiring different micro- or macrolevels of affective symptomatology [8]. The focus of this paper is the assessment and validation of active self-report data.

The potential of continuous monitoring of psychomotor activity based on acceleration and location for a differentiation of unipolar and bipolar patients has been demonstrated [9,10]. Recent studies have also indicated that smartphone-based movement parameters allow for a prediction of intradinidividual, daily mood state changes [4,11-14]. However, such prospective investigations require in-depth knowledge of the psychometric properties of the acquired data especially when it comes to the validity of smartphone-based measurements. This point appears particularly important in study designs that entirely rely on smartphone-based data.

Consequently, the comparability between smartphone-based and non–smartphone-based versions (ie, conventional paper-and-pencil or stationary computer-based versions) of psychometric instruments has also received increasing attention [15]. Besides the obvious difference in the format in which content is presented, differences in the assessment setting (ie, laboratory or clinical setting vs variable situations in real life) as well as technical reservations could lead to different assessment results. Particularly when using smartphones, potential distractions may become more likely, with the environments of reporting participants being less controllable. Initial evidence suggests that scores derived from digital and paper-and-pencil psychometric instruments seem to be generally comparable, however, with considerable variance in the agreement [16-18]. Yet, a considerable number of previous studies investigating the reliability and validity of digital phenotyping methods have focused on computer-based assessments that might differ from mobile assessments via the participants’ smartphones as outlined above. For the Beck Depression Inventory (BDI), interformat reliability between non–smartphone-based paper-and-pencil versions and computer-based versions has been demonstrated across several studies [16], while large-scale validation reports of agreement between smartphone-based and non–smartphone-based versions are currently lacking.

Data from pilot studies indicate agreement between smartphone-delivered, daily self-rated mood and clinician-rated mood via Hamilton Depression Rating Scale (HDRS) scores among bipolar patients [19]; in addition, Juengst et al demonstrated high comparability between mood-related symptoms among traumatic brain injury patients assessed either via smartphone self-reports or via telephone interview [20]. In a systematic review of the literature including data from three studies and a total of 89 bipolar outpatients, significant medium-sized correlations between daily, smartphone-based self-report assessments of depressive symptoms and established clinical rating scales were reported [21]. Regarding smartphone-based monitoring in major depression, Torous et al reported high agreement between daily, smartphone-based self-reports and paper-and-pencil assessments using the Patient Health Questionnaire-9 (PHQ-9) among 13 adult patients with major depressive disorder (MDD) [22]; similarly, Cao et al reported agreement between daily, smartphone-based self-reported mood and the PHQ-9 among 13 adolescent participants [23]. One systematic review that investigated the psychometric properties of mobile mood monitoring among young people concluded that there is enormous heterogeneity in the validity of smartphone-based delivery formats and more high-quality studies are needed [15].

In sum, while the aforementioned findings of overall agreement between smartphone-based self-reported depressive symptoms and established clinical scales is encouraging, it appears important to denote that limited sample sizes in previous reports as well as systematic differences, including sample properties, technical properties, and assessment type, currently limit our understanding of the reliability and validity of smartphone-based assessments of depressive symptoms. It thus remains unclear to what degree validation reports of smartphone-based self-reports are generalizable across assessment instruments, cohorts, and applications; hence, app- or study-specific validation of measurements remains the gold standard.

Therefore, the aim of this study is to assess the validity of smartphone-based assessments of depressive symptoms using the Remote Monitoring Application in Psychiatry (ReMAP) app. To this end, we use smartphone-based depression self-reports using single-item and BDI questionnaire data and investigate their comparability with non–smartphone-based versions of the BDI, a well-established and standardized self-report instrument used among psychiatric patients and healthy control participants. We test the hypotheses that both
delivery formats—smartphone-based and non–smartphone-based assessments—yield comparable results and, therefore, that smartphone-based monitoring of depressive symptoms via the ReMAP app provides valid assessments of depressive symptomatology. Furthermore, we aim to investigate potential differences in the agreement between smartphone-based and non–smartphone-based assessments of depressive symptoms across diagnostic groups as well as across age and gender.

Methods

Participants

The ReMAP study was designed as a prospective, naturalistic observational study. An overall sample of 173 participants was included in the analyses; participants had a mean age of 30.14 years (SD 11.92). The single inclusion criterion for this study was availability of a smartphone-based BDI that was completed within 4 weeks of a non–smartphone-based BDI. The sample included adults that were either healthy controls (n=101) or belonged to one of the following diagnostic groups: MDD (n=43), bipolar disorder (n=5), MDD with comorbid social anxiety disorder (SAD) (n=9), SAD only (n=2), or specific phobia (SP), spider subtype (n=13). Participants were recruited for ReMAP participation in the context of ongoing longitudinal cohort studies over which assessments were parallelized; details on subsamples from all cohorts are provided in the Multimedia Appendix.

Participants were informed about the possibility of voluntary additional participation in the ReMAP study in a face-to-face meeting at the time they presented at the Department of Psychiatry, University of Münster, Germany, in the context of ongoing, longitudinal cohort assessments. Interested subjects were extensively briefed about aims; methods, especially type and amount of collected data; details on data security (ie, details on data transfer and storage); and financial compensation. The study was approved by the local Institutional Review Board, and written informed consent was obtained before participation.

Non–Smartphone-Based Measures and Procedures

All measures that were not assessed via smartphone (ie, conventionally administered in interviews or via paper-and-pencil or tablet questionnaires) will be referred to as non–smartphone-based assessments and are described below. Presence or absence of a psychiatric diagnosis was assessed in all participants via the Structured Clinical Interview for DSM-IV (Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition) Axis I Disorders (SCID-I) [24,25] prior to participation in the ReMAP study. All healthy control participants were free from any history of a psychiatric disorder. As part of the original study assessments, participants from all cohorts provided self-reports of depressive symptoms via the BDI-I [26] or the BDI-II [27]. Both versions of the BDI are standardized and valid instruments for the assessments of depressive symptoms and represent well-established assessment tools in research and clinical routines for assessing the presence and extent of depressive symptoms. Additional assessments of clinician-rated depression severity via the HDRS [28] were available for a subset of 51 participants.

The ReMAP Smartphone App

Development of ReMAP began in mid-2018 at the Institute for Translational Psychiatry in Münster. It is a native app for iOS and Android, based on Apple ResearchKit, Apple Health, and Google Fit. After an anonymous log-in with a provided subject ID, the app works in background mode and monitors the number of steps taken by the user, the distance walked, the accelerometer, and GPS position data. The data are encrypted on the smartphone and sent regularly via REST-API (REpresentational State Transfer application programming interface) to a back end specifically developed for ReMAP, which is provided on university servers. In addition, the app regularly enables the user to fill out various questionnaires regarding sleep and mood as well as to create short voice recordings. Measures used in this study’s analyses are described below.

Smartphone-Based Measures and Procedures

After written informed consent was obtained, each participant was provided an individual subject ID (ie, subject code). The participant was then asked to download the developed ReMAP smartphone app and to start the app. At this time, subjects were asked to confirm participation in the study again and to enter their individual subject IDs.

In addition to the continuous assessment of passive data, all participants were asked to provide self-reported ratings of depressive symptoms. To this end, participants filled out a digital version of the BDI-I that was integrated into ReMAP every 2 weeks. Moreover, participants rated their mood and sleep duration by answering single items every 3 days. For the single mood question (ie, “How is your mood today?”), participants provided their responses via touch screen on a scale from 1 (very bad) to 10 (very good). For the single sleep question (ie, “How many hours did you sleep last night?”), participants provided their response on a scale from 0 to 13 hours. For all self-reported data, the app sent out weekly push notifications on a random basis during the daytime with a variance of 2 days or every 2 weeks in case of the BDI. The time of the day when notifications were sent was systematically varied in order to avoid bias from systematically assessing symptom self-reports (eg, only during the morning). Participants were instructed that answering all questions was optional and they were free to choose their time of answering whenever items were made available.

Again, for this study, smartphone-based and non–smartphone-based data were only included if the time interval between completion of the ratings between both delivery formats was less than 4 weeks, in order to minimize potential bias due to temporal change in depressive symptoms. Further, for each participant, the respective BDI, mood, and sleep assessments from the time point with the shortest interval between smartphone-based and non–smartphone-based assessments were included for this study.

Statistical Analyses

Agreement between non–smartphone-based and smartphone-based BDI scores was assessed by absolute agreement using a two-way, mixed-effects intraclass correlation
To this end, the non–smartphone-based measures were compared with the temporally closest smartphone-based BDI scores available, resulting in the shortest interval possible. This analysis was further repeated for the over-1-week-interval and the under-1-week-interval groups separately in order to assess the influence of the test-retest interval on the agreement between measurements. In addition, the analysis was repeated separately among healthy controls, affective disorder (ie, MDD, SAD + MDD, and bipolar disorder) patients, and anxiety disorder (ie, SP, spider subtype; and SAD) patients, as well as for the two non–smartphone-based BDI versions (ie, BDI-I and BDI-II). The internal consistency of the smartphone-based BDI was assessed via Cronbach $\alpha$ and compared with the internal consistency of the non–smartphone-based BDIs.

The smartphone-based single mood item was correlated with the non–smartphone-based and smartphone-based BDI scores. Although it covers different levels of symptomatology (ie, the BDI assesses complex symptoms over time, while the single mood item assesses only the current subjective mood [8]), the BDI questionnaire was used for validation based on the assumption that both measures are sensitive for current mood.

For validation of the smartphone-based single sleep item, it was correlated with the smartphone-based and non–smartphone-based BDI item assessing sleeping disturbance. Analogous to the BDI analysis, one mood and one sleep assessment were used for analysis based on the shortest interval to the non–smartphone-based measures. For further validation, the ReMAP BDI and the ReMAP single mood item were both correlated with clinician-rated depression severity using the HDRS.

All analyses were conducted using SPSS, version 26 (IBM Corp). A multiple test correction was undertaken across all significance tests (n=34) in order to avoid $\alpha$ error accumulation using a false-discovery-rate (FDR) correction following the Benjamini-Hochberg procedure [30]. Assuming an FDR q value of .05, this approach yielded a corrected significance threshold of $P<.04$.

**Results**

**Descriptive Statistics**

Mean BDI scores and their range across all participants were similar for ReMAP (mean 5.35, SD 8.63; range 0-44) and non–smartphone-based BDI (mean 6.46, SD 9.06; range 0-47). Absolute differences between both measurements were, on average, 3.02 points (SD 3.76) with a considerable range covering 0 to 26 points. The mean test-retest interval was 5.84 days (SD 7.29), ranging from 0.20 to 28.70 days. Detailed descriptive statistics across subgroups of the sample are provided in Table S1 in Multimedia Appendix 1. Among the included participants who completed a smartphone-based BDI within 4 weeks of completing non–smartphone-based measures, the percentages of participants who also provided single items for mood and sleep within a maximum interval of 4 weeks were 97.11% and 95.95%, respectively.

**Validity of Affective Symptom Assessment via ReMAP**

The overall agreement between ReMAP and the non–smartphone-based BDI was very high (ICC 0.921, 95% CI 0.890-0.942). Separate investigations of the BDI agreement in several subgroups yielded highly comparable ICCs across both BDI versions (ie, BDI-I and BDI-II), across different test-retest intervals, across different age groups, and across males and females—the ICC was over 0.888 for all subgroups. Separate investigations across different diagnostic statuses yielded the highest BDI agreement between delivery formats in the subgroup with affective disorders (ICC 0.912), while healthy controls and participants with anxiety disorders (ie, SP, spider subtype; and SAD) showed moderate agreement between BDIs (ICC 0.639 and ICC 0.736, respectively). Similarly, higher agreement was found among acutely depressed as compared to remitted MDD patients (see Multimedia Appendix 1). ICC statistics for the full sample and all subgroups are presented in Table 1. Scatterplots of ReMAP BDI scores over non–smartphone-based BDI scores are provided in Figure 1.

The internal consistency of the ReMAP BDI (Cronbach $\alpha=.944$, n=174) was virtually identical to both non–smartphone-based BDI versions (BDI-I: $\alpha=.945$, n=54; BDI-II: $\alpha=.944$, n=108). For further validation, the ReMAP BDI was correlated with clinician-rated depression severity using the HDRS in a subset of the sample (n=51). The analysis yielded a strong significant correlation ($r=0.783; P.<.001$) that was comparable to the association between the HDRS score and the score of the non–smartphone-based BDI ($r=0.682; P.<.001$).
Table 1. Intraclass correlation agreement of the Remote Monitoring Application in Psychiatry (ReMAP) Beck Depression Inventory-I (BDI-I) with the full sample and stratified subsamples.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Number of participants (N=173), n (%)</th>
<th>Intraclass correlation coefficient</th>
<th>95% CI</th>
<th>P valuea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full sample</td>
<td>173 (100)</td>
<td>0.921</td>
<td>0.890-0.942</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>BDI-I non–smartphone based</td>
<td>64 (37.0)</td>
<td>0.921</td>
<td>0.870-0.952</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>BDI-II non–smartphone based</td>
<td>109 (63.0)</td>
<td>0.919</td>
<td>0.863-0.850</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>≤1-week intervalb</td>
<td>126 (72.8)</td>
<td>0.934</td>
<td>0.890-0.958</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>&gt;1-week intervalc</td>
<td>47 (27.2)</td>
<td>0.888</td>
<td>0.799-0.938</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Healthy controls</td>
<td>101 (58.4)</td>
<td>0.639</td>
<td>0.454-0.760</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Affective disorders</td>
<td>57 (32.9)</td>
<td>0.912</td>
<td>0.851-0.948</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Anxiety disorders</td>
<td>15 (8.7)</td>
<td>0.736</td>
<td>0.252-0.910</td>
<td>.008</td>
</tr>
<tr>
<td>Age ≤35 years</td>
<td>131 (75.7)</td>
<td>0.899</td>
<td>0.851-0.931</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Age &gt;35 years</td>
<td>42 (24.3)</td>
<td>0.962</td>
<td>0.930-0.980</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Male</td>
<td>41 (23.7)</td>
<td>0.969</td>
<td>0.919-0.986</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Female</td>
<td>132 (76.3)</td>
<td>0.904</td>
<td>0.864-0.933</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

a All P values below a false discovery rate–corrected significance threshold of P<.04 are considered statistically significant.
b Participants completed the smartphone-based BDI within 1 week of completing non–smartphone-based measures.
c Participants completed the smartphone-based BDI and non–smartphone-based measures more than 1 week apart.

Figure 1. Beck Depression Inventory (BDI) scores via the Remote Monitoring Application in Psychiatry (ReMAP) smartphone app over non–smartphone-based BDI scores across diagnostic groups.

After including all data points with a test-retest interval of up to 4 weeks, the single item for mood assessed via ReMAP correlated moderately with the sum scores of the ReMAP BDI ($r=0.538; P<.001; n=168$) and with both non–smartphone-based BDI versions (BDI-I: $r=0.485, P<.001, n=61$; BDI-II: $r=0.504, P<.001, n=107$). Further, a significant negative correlation between the ReMAP single mood item and the HDRS score was observed ($r=0.369; P=.008; n=51$). Correlations of the single mood item across subsamples are provided in Table S2 in Multimedia Appendix 1.

The single item for sleep assessed via ReMAP was correlated with the BDI item assessing sleeping disturbance. After including all data points with test-retest intervals of up to 4 weeks, this analysis yielded significant negative associations.
with the sleep item from the ReMAP BDI \((r=-0.310; \ P<.001; n=166)\) and with the sleep item of both non–smartphone-based BDI versions (BDI-I: \(r=-0.279, \ P=0.03, n=63\); BDI-II: \(r=-0.202, \ P=0.04, n=102\)). Separate correlation analyses of the single mood and sleep ReMAP items across disorder subgroups are presented in Table S3 in Multimedia Appendix 1. The general pattern of results yielded the strongest associations in the affective disorder group.

**Discussion**

With this study, we demonstrate that smartphone-based monitoring of depressive symptoms via the ReMAP app provides valid assessments of depressive symptomatology. The overall high agreement between the non–smartphone-based and smartphone-based versions of the BDI confirm that digital assessments via the ReMAP app using the participants’ smartphones have the potential to offer valid estimates of the trajectory of participants’ moods. This notion is additionally supported by the observed correlation of smartphone-administered single-item ratings regarding mood and sleep with corresponding non–smartphone-based assessments. Importantly, the validity of smartphone-based assessments could furthermore be demonstrated by using clinical rating scales as a criterion with a strong correlation of smartphone-based BDI and non–smartphone-based HDRS scores.

The observation of high agreement between self-reported smartphone-based assessments of depressive symptoms and classic non–smartphone-based assessments in this study is supported by previous findings from pilot studies among MDD patients [22] and from a systematic review among bipolar patients [21]. Furthermore, the comparability of the non–smartphone-based and smartphone-based versions of the BDI in our study matches similar results of agreement between paper-and-pencil and computer versions of the BDI [16].

Our findings of overall high validity of smartphone-based and conventional non–smartphone-based assessments of depressive symptoms in a relatively large and heterogeneous sample critically underscores the potential of mobile assessment tools in psychiatric research. Considering that smartphone-based assessments offer valid data on patients’ mood states, an expansion of mobile data acquisition in the clinical and research context appears desirable. The cost-efficiency of smartphone-based data might thus allow the acquisition of valid data on patients’ long-term disease trajectories at an unprecedented scale. Together with previous studies investigating the comparability of the BDI versions (ie, BDI-I and BDI-II) [27,31] as well as delivery formats [16], our findings add to an increasing evidence base of high comparability of smartphone-based and conventional non–smartphone-based assessments of depressive symptoms.

We furthermore demonstrate that agreement between smartphone-based and non–smartphone-based assessments of depressive symptoms does not depend on the age or gender of participants, which supports the generalizability of smartphone-based assessments of depressive symptoms. This notion appears especially noteworthy considering the relatively large sample size, in comparison with previous reports, as well as the age range of participants included in this study (ie, 18-68 years of age). The inclusion of older participant groups seems relevant, as previous studies have emphasized that smartphone apps for mental health monitoring should meet the needs (eg, easy handling) of older and potentially less technically proficient individuals in order to assure adherence among these group members [32].

An important observation of this study was that higher agreement between smartphone-based and non–smartphone-based assessments of depressive symptoms was found among affective disorder patients compared to anxiety disorder patients or healthy controls. Notably, while the agreement in the affective disorder sample can be estimated as excellent, intraclass correlations indicate a lower, but still moderate to good, agreement in the healthy control and anxiety disorder samples [33]. This might partly be traced back to the much higher variance of depression severity in the affective disorder group. Lower variance in depression scores in nonaffective clinical samples has previously been suggested to account for findings of low reliability among substance addiction patients [34]. Further, small sample sizes of some participant subgroups limit the weight of this finding, particularly for the anxiety disorder subgroup (n=15). These findings may call for a cautious interpretation of findings based on self-reported symptom data in healthy or nonaffective disorder populations. However, they also seem to contradict previous findings. The authors of a meta-analysis investigating BDI reliability concluded that nonclinical samples show a very good test-retest reliability, while only very limited data are available for test-retest reliability in clinical samples [31]. Sporadic reports of lower retest reliabilities as found by one study [35] were explained by the authors as natural changes in depression severity over time [31]. The difference in reliabilities across samples may, in part, stem from differences in statistical analyses, as traditional Pearson correlations that were used by the cited studies can produce substantially different results than ICC agreement estimates, which are now often recommended for retest analysis [36].

Besides validation of a smartphone version of the BDI, this study found moderate to high agreement between mood ratings via smartphone-based single-item assessments and established clinical scores using the BDI, regardless of the delivery format of the BDI. This finding is of particular importance considering that completion of an entire questionnaire is time-consuming and, hence, the usage of single items might provide a valid possibility of assessing mood on a frequent basis. Importantly, these findings are tentative and limited by the fact that the single mood item and the BDI questionnaires may systematically assess differing concepts in regard to the symptom level as suggested by previous scholars [8]. However, although this distinction may account for agreement between both measures, the high agreement also points to substantial overlap between the macrolevel BDI questionnaire and the more microlevel single mood item.

In sum, the associations between questionnaire data (ie, the BDI) and single-item mood self-reports pose the following question: Which measure may be better suited for specific
research contexts and could one of the two be omitted completely? One may argue that single mood items seem to provide a sufficient proxy for the assessment of mood fluctuations that is more time-efficient and could, therefore, be assessed more frequently as compared to a more exhaustive BDI questionnaire. On the other hand, it may be more beneficial to have a more elaborate symptom profile as obtained, for example, via the BDI in exchange for assessment frequency. It remains to be investigated what temporal and content-related resolution is most beneficial for the investigation of the development of depressive symptoms and for specific feature engineering using machine learning algorithms. Likely, the most beneficial trade-off between the two highly depends on the specific research question.

Compared with the single mood item, the single sleep item showed a lower correlation with corresponding non–smartphone-based assessments in the form of sleep disturbance items within the BDI questionnaire. One possible explanation for this low association is that both items measure slightly different aspects of sleep: while the BDI sleep item assesses increased and decreased sleep duration and, depending on the BDI version, also a combination with subjective sleep quality, the single sleep item assesses purely the duration of sleep during the last night. Further, sleep quality or disturbance may be a more heterogeneous construct and, thus, more difficult to assess via a single item. Another possible explanation for this finding could be that variability in the sleep quality, as well as sleep duration, may be less temporally stable as compared to mood changes. Thus, the test interval of up to 4 weeks may be too long in order to validate the smartphone-based assessment of sleep duration. Considering that smartphone-based and non–smartphone-based assessment methods lie several days or weeks apart, the association between them seems to be reasonably high.

Strengths of this study include the relatively large sample of participants and the availability of smartphone-based data along with conventional psychometric and clinical data. Furthermore, this study included participants with differing psychiatric diagnoses and a high variability in age, thus allowing the assessment of generalizability across such participant groups. Further, a wide variety of assessment forms were used for validation, considering multiple sources of information. The application of non–smartphone-based BDI versions (ie, self-report), as well as clinical ratings (ie, HDRS), underlines the validity of the smartphone-based assessments via the ReMAP app. Limitations include the lack of prospective clinical follow-up data. Future large-scale studies are warranted to assess the prognostic validity of smartphone-based self-reports in affective disorder patients.

Smartphone-based monitoring of depressive symptoms remains a timely matter of critical relevance for translational psychiatry. These results demonstrate overall high validity of smartphone-based assessments of depressive symptoms and should, thus, encourage researchers to apply mobile apps toward continuous prospective assessments of depressive symptoms.

Acknowledgments
Funding was provided by the German Research Foundation (DFG) (project No. 44541416-TRR 58); project grants C09 and Z02 of the Collaborative Research Center Transregio (CRC-TRR) 58 were awarded to UD. Funding was also provided by the Interdisciplinary Center for Clinical Research (IZKF) of the Faculty of Medicine, University of Münster (grant Dan3/012/17 was awarded to UD and grant SEED 11/19 was awarded to NO). In addition, funding was provided by the Innovative Medizinische Forschung (IMF) of the Faculty of Medicine, University of Münster (grant OP121710 was awarded to NO and TH and grants LE121703 and LE121904 were awarded to EL). Further, we are deeply indebted to all personnel involved in recruitment and data management as well as to all participants of this study.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Descriptive statistics of subsamples, correlations of the Remote Monitoring Application in Psychiatry (ReMAP) single mood item with Beck Depression Inventory (BDI) scores across subsamples, and correlations of the ReMAP single sleep item with the BDI sleep item across subsamples, correlations of the ReMAP single mood item with BDI sleep item across subsamples, and correlations of the ReMAP single sleep item with the BDI sleep item across subsamples.

[DOCX File, 32 KB - mental_v8i1e24333_app1.docx ]

References


Abbreviations

BDI: Beck Depression Inventory
CRC-TRR: Collaborative Research Center Transregio
DFG: German Research Foundation
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition
FDR: false discovery rate
HDRS: Hamilton Depression Rating Scale
ICC: intraclass correlation coefficient
IMF: Innovative Medizinische Forschung
IZKF: Interdisciplinary Center for Clinical Research
MDD: major depressive disorder
PHQ-9: Patient Health Questionnaire-9
ReMAP: Remote Monitoring Application in Psychiatry
REST-API: REpresentational State Transfer application programming interface
SAD: social anxiety disorder
SCID-I: Structured Clinical Interview for DSM-IV Axis I Disorders
SP: specific phobia

Edited by J Torous; submitted 15.09.20; peer-reviewed by S Steingrimsson, F Bos; comments to author 08.10.20; revised version received 28.10.20; accepted 06.11.20; published 12.01.21.

Please cite as:
Smartphone-Based Self-Reports of Depressive Symptoms Using the Remote Monitoring Application in Psychiatry (ReMAP): Interformat Validation Study
JMIR Ment Health 2021;8(1):e24333
URL: https://mental.jmir.org/2021/1/e24333
doi:10.2196/24333
PMID:33433392

©Janik Goltermann, Daniel Emden, Elisabeth Johanna Leehr, Katharina Dohm, Ronny Redlich, Udo Dannlowski, Tim Hahn, Nils Opel. Originally published in JMIR Mental Health (http://mental.jmir.org), 12.01.2021. This is an open-access article
distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Mental Health, is properly cited. The complete bibliographic information, a link to the original publication on http://mental.jmir.org/, as well as this copyright and license information must be included.
Participant Engagement in a Transmedia Storytelling Web-Based App Intervention for Mental Health of Latina Women: Qualitative Analysis

Patricia D Soderlund¹, PhD, RN; Adrienne S Martinez Hollingsworth², PhD, RN; MarySue V Heilemann³, PhD, RN, FAAN

¹School of Medicine, University of Minnesota, Duluth, Duluth, MN, United States
²Department of Internal Medicine, UCLA School of Medicine, National Clinician Scholars Program, University of California, Los Angeles, Los Angeles, CA, United States
³School of Nursing, University of California, Los Angeles, Los Angeles, CA, United States

Corresponding Author:
MarySue V Heilemann, PhD, RN, FAAN
School of Nursing
University of California, Los Angeles
Factor Building, Box 6919
Los Angeles, CA, 90095-6919
United States
Phone: 1 310 206 4735
Email: mheilema@sonnet.ucla.edu

Abstract

Background: Stigma, fear, and lack of knowledge regarding treatment options or where to get help create delays for Latina women in accessing needed mental health help. Story-based media interventions hold appeal for Latina women. Thus, we drew upon the Social Cognitive Theory by Bandura to create an evidence-based, transmedia storytelling web-based app for mental health called Catalina: Confronting My Emotions to connect Latina women to a curated set of mental health resources. Understanding how Latina women perceive various aspects of the web-based app will help design future expansions.

Objective: A previously published analysis led to the development of a category on how participants related to the lead character (Catalina) in the story line of the web-based app as a real person. However, the purpose of this analysis was to gain an understanding of participants’ experiences with the extension of the dramatic story line of the web-based app beyond Catalina to a Latina nurse-therapist character named Veronica, who was featured prominently in the app’s interactive content and bonus videos.

Methods: Qualitative analyses were conducted with interview data from a community-based sample of 28 English-speaking Latina women aged between 21 and 50 years who scored above the threshold for anxiety (Generalized Anxiety Disorder-7) and/or depression (Patient Health Questionnaire-9) but were not suicidal at screening. Data were collected 72 hours after participants engaged with our transmedia storytelling web-based app for mental health. Grounded theory methodology guided the analysis and interpretation of data that had been collected telephonically, recorded, and transcribed with identifiers removed. Analyses included initial and focused coding using process codes (gerund form of verbs in codes focused on action), informed by symbolic interactionism, and the development of categories with properties through constant comparison, memo writing, and the use of charts and diagrams.

Results: Our participants experienced a multiphase process that was most heavily related to Veronica, the Latina nurse-therapist character in our web-based app, who led them through a process to a place of action. We conceptualized this process as moving from passive viewer to active participant of a transmedia storytelling web-based app intervention. Overall, 3 new conceptual categories provided insight into women’s experiences, including encountering a trustworthy nurse-therapist character, taking in messages that dispel old beliefs, and preparing when and how to take action. Each category has nuanced properties that reflect participants’ experiences.

Conclusions: Active engagement with our web-based app led our sample to successfully transition from the viewpoint of the observer to the viewpoint of the experiencer, moving from a passive position of watching to active engagement that involved imagining, thinking, reflecting, and acting. Careful development of dramatic material for health-related web-based apps using
Background

The median delays in seeking mental health care for depression and anxiety in the United States have been estimated to be 9 and 8 years, respectively, even if the person knows that their symptoms are due to a mental health problem; these delays were significantly associated with being Latinx or Black [1]. A meta-analysis on the duration of untreated unipolar depression showed that participants who failed to initiate care within 8 weeks of becoming aware of depression had a lower probability of responding well to treatment or achieving remission [2]. In another study, 70% of the 287 Latina women in a community-based sample delayed seeking needed physical health care; furthermore, having depression or anxiety was associated with 3.1 times greater odds of delaying needed health care (95% CI 1.6-5.9) [3]. Therefore, for Latina women, the consequences of untreated depression and anxiety threaten both physical and mental health.

The prevalence of depression is higher among Latina women than among Latino men in the United States [4,5], as it is globally [6]. Compounding the problem, Latina women have been reluctant to seek care [7] due to fear, guilt, shame, privacy concerns, lack of family support, or distrust of professionals [8]. This includes English-speaking Latinx men and women [9], who have reported higher levels of self-stigma and have more often said they would conceal a potential mental health problem compared with non-Latinx White individuals [10]. Latinx men and women are approximately half as likely as people of other racial or ethnic groups to obtain mental health care [11]. Thus, a dynamic intervention that helps this group overcome barriers to care and identify available resources is crucial.

As mental health literacy is a segue to treatment engagement [12], it is an asset that would benefit Latina women with symptoms who wish to obtain mental health help. Key features include having the ability to recognize a mental health disorder as it develops, being clear about professional and self-help treatment options, and knowing where to obtain resources and information, as well as how to help someone else who is struggling with their mental health [13]. However, the question arises of how to most effectively reach Latina women with this content. This is especially challenging with Latina women, who avoid even talking about mental health due to fear of stigmatizing labels [12]. Nonthreatening strategies are needed to increase the confidence of Latina women and encourage them to seek help and make contact with a provider. A discreet approach to introducing and delving into the topic of mental health is likely to be more acceptable to Latina women; however, for the approach to be effective in helping Latina women achieve the goal of mental health and wellness [14], it should be uncomplicated, engaging, and directly related to getting treatment.

One strategy is to use technology that most Latina women have in their pocket, purse, or desktop, that is, their smartphone, tablet, or computer. Latinx men and women report high levels of internet use both for entertainment [15,16] and for accessing health information [17]. Latinx men and women have high rates of smartphone ownership [18], which supports the idea of a mental health web-based app intervention that is accessible on smartphones or other mobile devices and can be ultimately personalized as well as scalable to maximize reach [19]. However, any endeavor to create a web-based app needs to be guided by theory and based on scientific evidence, as Torous and Roberts [20] have explained. They cautioned that only very few of the 10,000 mental health apps that were available for the public to download in 2017 were actually based on data [21]. With such a large number of available apps, it can be quite challenging for users to find the few evidence-based apps among those that were not developed based on data [20]. In addition to being evidence-based, developers need to take extra steps to ensure that their apps are user friendly, designed with user input [22,23], and respectful of users’ privacy [21,23]. Ultimately, if the app is useful for meeting users’ needs, it holds promise for raising the otherwise low rates of engagement with mental health apps [21].

When considering a target group of symptomatic Latina women, apps also need to be culturally acceptable, desirable, and enjoyable to them, so that they will be attracted to using the apps. Latinx men and women are considered a mobile-first community with some of the highest levels of smartphone engagement. For example, 90% of all Latinx consumers use their smartphones to perform video streaming. In particular, Latina women report high levels of television watching and movie going [16], spending more than 30 hours per week watching television and more than 22 hours per week using their smartphones while exploring apps, downloading and viewing videos, or surfing the net. Latina women also report using YouTube, Google+, Instagram, Snapchat, and Twitter at higher rates than non-Hispanic White women [24]. For these reasons, we created an evidence-based, storytelling mental health intervention that could link untreated Latina women to the needed care and resources via a smartphone, tablet, or computer using transmedia. Transmedia involves telling stories that unfold and extend across multiple digital platforms, including smartphones, tablets, or computers, which offer users and viewers a media experience that has the potential to go deeper as they gain different points of view related to the story world, plot, or characters through extra scenes or bonus videos [25].
Previous Work

Fueled by input from Latina women, informed by Social Cognitive Theory by Bandura related to media, vicarious learning, and behavior change [26], and inspired by the 1970s edutainment productions by Miguel Sabido on Televisa in Mexico [27], we created a Hollywood-quality web-based app with a responsive design that can be used on smartphones, tablets, or computers. We partnered with a programmer and Latinx media and film professionals, as described elsewhere [28,29]. Our web-based app is character driven in that the momentum of the web-based app is carried forward by the characters themselves. In addition, our characters are messengers of health-related content embedded in the story and bonus videos. This means that the story line is extended by featuring characters in additional bonus videos that are psychoeducational or therapeutic. In this way, the characters present helpful information from within the story world while also enhancing the participants’ media experience. For these reasons, we developed the characters with special attention to deidentified input from Latina women struggling with depression who participated in previous research studies [30-38]. In addition, during the design phase, we received critique and input from Latina therapists and non-Latina women therapists who worked with symptomatic Latina women to maximize usefulness from a provider point of view. These sources of input influenced all scripts, directing, and acting. In theater testing, Latina women of the target demographic gave feedback useful for editing and culturally tailoring the content. Latina focus group participants named it Catalina: Confronting My Emotions. Our approach heeds advice voiced by Torous et al [22,23] about the importance of a collaborative approach in app development that includes patients and providers.

A mixed methods study with 2 major aims was conducted. The first was to quantitatively analyze the feasibility, acceptability, and efficacy of the web-based app intervention, and the second was to qualitatively explore the experiences and perceptions of 28 English-speaking Latina women in a community sample that engaged with the web-based app. All participants were aged between 21 and 50 years and scored above the threshold for anxiety [39] and/or depression [40] but were not found to be suicidal at screening. Interviews were conducted by phone approximately 72 hours after participants engaged with media in the web app.

Due to the design of the home webpage of our web-based app, participants were led to engage with the videos by starting with the story-based videos about a character, Catalina, portrayed by actress Yareli Arizmendi. These videos included a 11-minute webisode (online television episode) about Catalina struggling with feelings of sadness, frustration, and worry in her daily life. This was followed by a 3-minute video log of Catalina, making a recording for her best friend in which she confides in her best friend that she is thinking of seeking therapy. Users then watch a 3-minute extra scene of Catalina coming out of a community counseling center and follow her as she continues walking down the street while leaving a voice message for her best friend about her positive experience in therapy with a character named Veronica. At this point, participants have not yet seen Veronica, but they hear Catalina talk about the therapy session with Veronica as something she values.

The goal of the early videos was to attract and pique the interest of participants in the dramatic story so they would be motivated to click to open the bonus and interactive videos that extended the story through Veronica. Thus, the story involved some drama, some romance, and some tense scenes depicting difficulties Catalina had involving an old boyfriend, her mother, and her emotions. Our strategy was successful, and all 28 participants clicked to open and watch all features of the web-based app [28]. They got their first glimpse of Veronica, portrayed by actress Yareli Arizmendi, in a 4-minute bonus video in which she speaks from her point of view as a Latina and a nurse who is a therapist. Veronica looks directly into the camera as she speaks to the participants. The script allows Veronica to talk about Catalina’s situation in an accessible way and to build upon it to share how common depression and anxiety are among women and to provide various points of psychoeduction. She then invites participants to click to open a sequence of interactive videos. Over the course of 5 short 1- to 2-minute interactive videos, Veronica again speaks to the audience directly, posing questions about what is holding them back from getting care [41] to help them consider their own situation, emotional needs, and desires. Finally, Veronica invites participants to click to open her blog, written from her point of view, which contains carefully selected links to hotlines, local clinic webpages, telephone numbers, and websites that provide free or low-cost resources and information about mental health [28,29].

Our team analyzed and published an analysis of qualitative data from our sample about their perceptions of the fictional lead character of the story, Catalina, whom they related to as a real person with a past, present, and future [29]. However, as a major portion of the transmedia web-based app involved Veronica, the nurse-therapist character, it was imperative for us to analyze the qualitative data from participants about Veronica. She was the character who guided participants through all other aspects of the web-based app experience, including the transmedia bonus videos, interactive features, and resource-rich blog.

Objectives

The purpose of this qualitative analysis is to explore, describe, analyze, and interpret the perceptions of English-speaking Latina women with elevated levels of depression and/or anxiety related to the nurse-therapist character in our transmedia web-based app, named Veronica. Through this analysis, we aim to provide knowledge for consideration in the future when developing nurse characters or other health provider characters in mental health apps designed especially for Latina women and other users struggling with depression or anxiety symptoms.

Methods

Design

For this qualitative analysis, we used the techniques of grounded theory methodology [42,43] to analyze interview data that were collected as part of a mixed methods intervention study. As was described elsewhere [28,29], approval was obtained from the...
UCLA Institutional Review Board. Symbolic interactionism [44,45] informed our use of grounded theory techniques; symbolic interactionism holds that people make sense of reality through the social aspects of their daily lives and meaning emerges through social interactions. This sensitized us during analysis to focus on what was meaningful to participants about their web-based app experience, with special attention to interactions with the characters in the context of their own lives.

**Sampling and Recruitment Procedures**

To attract participants, flyers were distributed at 9 community-based sites located in a metropolitan area of Southern California. Interested women called the study phone and were screened. Purposive sampling was used with the goal of recruiting English-speaking Latina women who scored above the threshold for anxiety [39] or depression but were not suicidal [40]. Those who met the inclusion criteria and gave web-based informed consent were aged 21-50 years, could speak and read English, and had access to the internet via a tablet, computer, or smartphone.

**Data Collection**

Participants completed a web-based baseline survey before engaging with the web-based app. A telephone interview was conducted up to 3 days later. A semistructured interview guide was used [29], which included open-ended questions about the Latina women’s experiences with each aspect of the transmedia web-based app, their perceptions of Veronica, and attitudes about help-seeking. Being semistructured, the interview was designed to allow participants to expand on topics as they desired. As is recommended by grounded theory methodology, participants’ answers in early interviews influenced questions posed in later interviews. Thus, additional questions that emerged in early interviews about what Latina participants found pertinent were added to subsequent interviews as data collection proceeded. All telephone interviews were conducted by a nurse scientist who is also a mental health nurse practitioner (first author). After completing several steps in the study, including the audio-recorded telephone interview (lasting an average of 45 minutes), participants received a US $60 gift card (via US mail, text message, or email). All digital recordings were password protected and were stored on a server with a firewall. Audio recordings were transcribed verbatim using a professional and secure transcription service. All transcripts were deidentified, checked for accuracy, and uploaded into Atlas.ti (qualitative data analysis; Scientific Software Development GmbH) [46]. All participants who enrolled in the study completed the entire study.

**Data Analysis**

We engaged in a rigorous qualitative analysis of the data using the techniques of grounded theory methodology. Initial coding of each line of data was performed using process codes (gerunds) to maximize our focus on the actions of each participant (including their thinking and self-talk) and their point of view as an individual. This reduces tendencies to project beyond the data or to make an interpretive leap prematurely based on one particular line of data. Process coding also directed the analysis to take into account multiple participants’ perspectives. Initial codes were created for each line of 10 of the 28 transcripts. Then, we identified the most significant and most frequent codes (called focused codes, which are similar to themes based on grounded theory methodology) and used them to sort codes with similar meanings into groups. Symbolic interactionism enhanced our analysis because it provided guidance for asking ourselves questions during analysis of the data about women’s interactions with the characters and with themselves as they experienced the web-based app. This helped us to identify what aspects participants found meaningful. Then, we developed the focused codes into categories. The qualitative software Atlas.ti (qualitative data analysis; Scientific Software Development GmbH) [46] facilitated the analytic process by serving as a platform to organize data so that it could be easily shared between researchers. Excel charts facilitated our use of a constant comparison of data with data and codes with codes. We were limited to only one interview per participant, so we could not perform theoretical sampling. However, the techniques of grounded theory methodology allowed us to scrutinize every line of the very rich data we had collected to perform a systematic and rigorous analysis. Memo writing deepened the analysis, facilitated discussions about patterns in the data, enhanced researcher reflexivity, and led to checking potential biases throughout the analysis. From this, we were able to develop a robust description of categories with nuanced properties. Diagrams were instrumental in illustrating processes and identifying relationships between and within categories.

**Results**

**Participant Details**

Demographics and levels of depression and anxiety symptoms at screening for our sample of 28 Latina participants are reported in Table 1 [28]. More than half of our sample reported watching story-based dramas or comedies on television once or more each week and watching videos, movies, or story-based shows on the internet at least weekly, using a smartphone, tablet, or computer.
Table 1. Sample demographics (N=28).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Highest education level</strong></td>
<td></td>
</tr>
<tr>
<td>Some high school (but not all)</td>
<td>3 (11)</td>
</tr>
<tr>
<td>Graduated from high school or earned a general educational development certificate</td>
<td>7 (25)</td>
</tr>
<tr>
<td>Some technical, trade, or vocational school after high school</td>
<td>3 (11)</td>
</tr>
<tr>
<td>Attended at least one course in college</td>
<td>8 (29)</td>
</tr>
<tr>
<td>Graduated with an associate degree</td>
<td>1 (4)</td>
</tr>
<tr>
<td>Graduated with a bachelor’s degree</td>
<td>3 (11)</td>
</tr>
<tr>
<td>Graduated with a master’s degree</td>
<td>2 (7)</td>
</tr>
<tr>
<td>Chose not to answer the question</td>
<td>1 (4)</td>
</tr>
<tr>
<td><strong>Finances</strong></td>
<td></td>
</tr>
<tr>
<td>Ability to meet weekly financial needs</td>
<td></td>
</tr>
<tr>
<td>Not difficult</td>
<td>3 (11)</td>
</tr>
<tr>
<td>Somewhat difficult</td>
<td>6 (21)</td>
</tr>
<tr>
<td>Very difficult</td>
<td>19 (68)</td>
</tr>
<tr>
<td>Chose not to answer</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Mental health symptoms</strong></td>
<td></td>
</tr>
<tr>
<td>Depression and anxiety</td>
<td></td>
</tr>
<tr>
<td>Depression: PHQ-9\textsuperscript{a} score ≥10</td>
<td>3 (11)</td>
</tr>
<tr>
<td>Anxiety: GAD-7\textsuperscript{b} score ≥10</td>
<td>4 (14)</td>
</tr>
<tr>
<td>Depression and anxiety scores each ≥10</td>
<td>21 (75)</td>
</tr>
</tbody>
</table>

\textsuperscript{a}PHQ-9: Patient Health Questionnaire 9-item.

\textsuperscript{b}GAD-7: Generalized Anxiety Disorder scale 7-item.

**Moving From Passive to Active Participant in a Transmedia Storytelling Web-Based App Intervention**

The Latina women in our sample engaged in a multi-phase process that we conceptualized as moving from passive viewer to active participant of a transmedia storytelling web-based app intervention. This process is depicted using the metaphor of interlocking gears. The process begins with the movement of the first gear, which, in this depiction, is the symbolic connection to the lead character in the web-based app. Our sample of symptomatic Latina women found Catalina to be highly relatable (Figure 1). The analysis and description of this conceptual category was previously published [29]. We called it relating to the lead character (Catalina) as a real person. The turning of the first gear causes the next gear to turn, which then engages the next gear until all the gears are turning.
This analysis picks up after the first gear is in motion. In other words, having found Catalina to be relatable, the Latina women in our sample moved on to an active engagement with the second character, Veronica, a nurse therapist. This analysis provides a nuanced description of the second, third, and fourth phases of the process that Latina women went through upon meeting Veronica in the media of the web-based app. These 3 phases are depicted as interlocking gears and are representative of 3 new categories. These 3 categories give clarity to the process of moving from passive viewer to active participant. Each new category has 2 or 3 properties. The categories are as follows: encountering a trustworthy nurse-therapist character, taking in messages that dispel old beliefs, and preparing when and how to take action.

This analysis picks up after the first gear is in motion. In other words, having found Catalina to be relatable, the Latina women in our sample moved on to an active engagement with the second character, Veronica, a nurse therapist. This analysis provides a nuanced description of the second, third, and fourth phases of the process that Latina women went through upon meeting Veronica in the media of the web-based app. These 3 phases are depicted as interlocking gears and are representative of 3 new categories. These 3 categories give clarity to the process of moving from passive viewer to active participant. Each new category has 2 or 3 properties. The categories are as follows: encountering a trustworthy nurse-therapist character, taking in messages that dispel old beliefs, and preparing when and how to take action.

**Encountering a Trustworthy Nurse-Therapist Character**

All participants clicked the links on the web-based app that extended the story from a focus on Catalina to bonus videos wherein Veronica, the nurse-therapist character, spoke directly to the audience. Our participants spoke at length about their experience of Veronica in interviews. They recognized Veronica as a trustworthy nurse who was also a capable therapist. As participants described their perceptions and reactions to Veronica, it was the element of trust that mattered most to them. This category had 2 properties: (1) perceiving her to be earnest, sincere, and nonjudgmental and (2) feeling confident due to her competence (Figure 2).

**Perceiving Her to be Earnest, Sincere, and Nonjudgmental**

Sincerity was perceived as a crucial aspect of Veronica’s trustworthiness. One participant described her as “very genuine” and indicated that Veronica’s choice of words and style of speaking conveyed an honest intention to help and a deep sense of caring. This made the media interaction feel enjoyable, comfortable, and safe. She stated:

> “I get that... warm vibe from her, and, that’s someone that I would feel comfortable opening up to and being able to talk to and not feeling judged.”

Participants did not focus on Veronica as a fictional character. Rather, they described how earnest she was, which impressed them. One woman said, “regardless if she really was a nurse, she seemed very sincere.” Participants explained that they felt certain that Veronica would understand instead of judging them,

**Feeling confident due to her competence**

She [Veronica] felt very eager to help. So, her tone of voice was soft but very straightforward but also like there was some love in it. So, I really liked hearing from her. I felt safe.

Participants did not focus on Veronica as a fictional character. Rather, they described how earnest she was, which impressed them. One woman said, “regardless if she really was a nurse, she seemed very sincere.” Participants explained that they felt certain that Veronica would understand instead of judging them,
that they trusted her, and perceived her to be the type of therapist who would be really open to listening. In the context of their recent emotional struggles, the participants valued her lack of judgment, implying that it was not just important but necessary for them to feel comfortable.

Women described Veronica’s manner, how she looked, and how they felt in response to her demeanor. For example, one woman explained that Veronica seemed “warm and approachable,” that “her demeanor was very welcoming and very caring,” and although this participant said she typically would have preferred a male counselor, Veronica “felt like she was like someone that I could go and just share how I’m feeling” as she seemed to “be neutral and impartial.”

Putting the emphasis on the effect Veronica had on her, another participant said, “she seems to calm my nerves.” After recounting how she perceived warmth from Veronica, another woman emphasized her personal expectations and then projected feelings she anticipated having in the presence of Veronica:

*I get, that - for me - warm vibe from her [Veronica], and, that’s someone that I would feel comfortable opening up to and being able to talk to and not feeling judged.*

**Feeling Confident Due to Her Competence**

The second property of trustworthiness in the nurse character was her competence, which inspired confidence. Participants believed that Veronica embodied different facets of competence, which made them feel confident about her. One woman said, “and, that’s why I could feel comfortable—the way she sounded confident and [like] someone you can talk to, feel comfortable with.” Another said she was “very empathetic, very professional.”

Many were impressed by her expertise as a nurse therapist saying, “she’s very knowledgeable” and “it looked like she knew what she’s doing and talking about.” Another summed up Veronica’s proficiency by declaring firmly, “she’s great at what she does.” Participants commented on her presence as a skillful therapist and regarded her as a master:

*Veronica seemed to be a person that has studied, I guess, human behavior in difficult situations. She seems like she is a person that is capable of helping a person in need and she seems to be trustworthy.*

Participants also viewed her as a capable therapist who had a wide range of skills, able to help “no matter what” type of problem was presented. Her capability engendered confidence as one woman exclaimed, “I feel like she would give me a lot of good advice for my life, and how to deal with situations.” Several women expressed assurance that Veronica could handle any sticky or complex situation they might encounter. One woman stated:

*. . . because Veronica knows how to deal with problems like that, so I know she would be someone that could definitely help me over the problems, whatever I’m going through.*

Others expressed themselves by describing a hypothetical scenario and then explaining how Veronica could tailor her skills to match the needs of such a person with such a problem at such a moment. One woman imagined confiding in Veronica and said “…you can tell you can trust her. You could tell her something and she would respond in a way that’s in words you need to hear.”

**Taking in Relevant Messages That Dispel Old Beliefs**

Our participants reported that the messages in the videos that involved Veronica were meaningful to them. How they described this content indicated that the messages were effective in bringing them new insight and deeper understanding of mental health. Thus, the second category is taking in relevant messages that dispel old beliefs. The 2 steps of this process are represented by 2 properties of this category: (1) taking in new messages because they were personally relevant and meaningful and (2) using these messages to dispel old myths and erroneous beliefs about mental health (Figure 3).

Figure 3. Properties of the category of taking in relevant messages that dispel old beliefs.

**Messages Were Relevant and Personally Meaningful**

When the interviewer asked general questions about participants’ thoughts related to seeking help for emotions in general, participants spontaneously answered in a way that integrated Veronica’s messages. They personalized the content of Veronica’s messages to fit their own situation. Remarkably, they remembered her messages even though the web-based app had not prompted them at any point to memorize or recite any content in the app. Furthermore, when asked how they felt about
Veronica, they responded as if Veronica and her messages were so intertwined that they were the same thing. For example, when asked how she felt about Veronica, one participant did not answer by describing a feeling. Rather, she answered by citing one of Veronica’s messages that she found relevant in her life and related it to her own family, saying:

I liked the point that she [Veronica] brought up. I believe she had said, “First,” you know, “you have to make sure you’re okay with yourself, and you need to seek help if you want to,” you know, “have a better life and also help your family.”

The content delivered by Veronica in the web-based app seemed to be taken in and absorbed by the women. During the interviews, women transposed what Veronica said by putting it into the first person as if it was their own idea, never mentioning that they were using Veronica’s words. Sometimes during an interview, women shared something Veronica had said as if they were educating the interviewer. For example, one woman who was pregnant at the time of the interview gave a brief description of Veronica and then quickly went on to paraphrase one of Veronica’s messages saying, “…because I matter, and in order for me to take care of bringing a life into this world, you need to make sure you are ready to help others.” With conviction, this participant explained the reasons why seeking therapy would be crucial for her and her family. She shared her own view but never mentioned that Veronica had said very similar words in the video.

Another participant described how she “wished” she could talk to a person like Veronica when “problems in the home” occurred. She reiterated Veronica’s advice from the video and expanded upon it, bringing in how emotional struggles can take a toll on one’s body, which she said was why taking action felt helpful, as one woman said:

She gave really good advice. Like, getting therapy is good. It’s good to get help. It helps. You read, you take it out, you take it out of your body. It’s not good to be holding things in. It’s good to talk to somebody, you know? It helps. It makes things better. You feel it helps you more be less stressful, less nervous, you feel like a different person.

Using Messages to Dispel Old Beliefs

Through the process of taking in new messages, the women experienced a shift in thinking that spurred new insight. In the process, it dispelled old beliefs and erroneous myths about mental health and mental health care. Women said that Veronica clarified how common mental health challenges were and in doing so “put things into perspective.” They explained how Veronica helped them see their symptoms differently, which made them feel more comfortable and less odd or unusual. One woman said, “maybe it’s more normal than we think!” and then added, “I like that she [Veronica] made me feel like, OK, this is a thing that happens to a lot of people.” Her experience with Veronica dispelled a previous belief that she was the only one, as it did for many participants. Another woman with the same realization summed it up saying, “…we’re not alone. There’s a lot of other women out there that are going through the same thing.”

One woman dispelled an old myth that had prevented her from seeking help “until” she felt it was finally time. She shared her feeling that she had to wait until the right time, but she never seemed to know when the time was right. Referring to how, over time, she quietly endured emotional struggles that seemed to multiply, she said, “I guess I feel like we take it in, we take it in until we finally burst.” She then explained how Veronica’s messages helped change her perspective about the right time to seek help. She said:

I guess after seeing these videos and seeing the way the therapist talks, that’s what, made me feel comfortable to know that you don’t have to wait until you can’t take it anymore. You could get help before so that you won’t get to that point.

Participants described how Veronica’s messages brought awareness, saying that she “made me realize” things. In the context of the overall story line in the web-based app, another participant shared how she realized a key point about therapy. Previously, she thought that if she went to a therapy appointment, she would be doing all the talking. Now she realized that actually, “maybe it’s more of a combination where we’re both talking.” This and the other details that Veronica shared about getting mental health care were practical and helpful, as one woman said:

...she explained how it would help to get therapy, what the benefits are to talk to someone, how it would make you feel better and possibly change your life or even your lifestyle.

These details, coming from Veronica, dispelled the idea that getting help was not acceptable, and that, in fact, it could lead to a positive change in their personal lives.

Preparing When and How to Take Action

After the women decided that Veronica was trustworthy and engaged in the process of integrating messages they found to be relevant, they entered a part of the process that involved preparing when and how to take action. This category has 3 properties: (1) being prompted to re-examine the need for help, (2) feeling ready to access help, and (3) having a vision for the kind of therapist to seek (Figure 4).
Being Prompted to Re-examine the Need for Help

After finding Veronica to be trustworthy, participants found that the content of her messages helped them reflect to gain a more enlightened view of themselves. They felt prompted to apply what they learned from Veronica and to step back, take an honest look at their ability to cope with their emotional struggles, and examine it in the context of their life circumstances. One woman with 5 children, who never sought therapy before and always put her children’s needs before her own, was moved by Veronica’s message. This woman said, “She [Veronica] gives good advice, and I just feel everything she said is true.” She went on to discuss the possibility of finding a therapist saying, “I get to thinking, I am thinking about it... They [the videos] give you good information, like the counselor [Veronica], and it’s true, it’s good to talk to people.” Another participant reflected on how she felt before engaging with the transmedia web-based app intervention saying, “I’ve never been, ‘OK, I’m going to go’.” However, she continued to say that now she found herself open in a way that was new for her. Admitting that, thanks to Veronica, she had re-evaluated her need for mental health care, she said, “She makes me feel like, ‘Oh wait, do I really have a problem?’, or ‘Should I go talk to someone?’” Her willingness to be open and inquisitive about her own need for help was key to moving forward in the help-seeking process.

Feeling Ready to Access Help

The second property of preparing to get help involved feeling ready. Some participants, who announced their unequivocal desire to seek help after engaging with the web-based app, attributed their decision to the character, Veronica. This included her presence overall and her insightful messages. One woman, struggling to cope with her life as a single mother said, “I’ve been having a lot of anxiety due to all my stress, just overall in my life.” She expressed her enthusiasm and enhanced readiness to seek therapy by saying:

I guess just seeing Veronica and the way that she is and how she brought up good points. So, it’s something that I would definitely be interested in doing.

After coming to terms with Veronica’s messages, another participant indicated that she was more prepared to take the risk of seeking help. She said, “I feel like I would give it a try, yeah. I would be interested to talk to someone.” Another woman gained hope and optimism that therapy had the potential to improve her life:

Actually, all the stuff that she said, it’s true. You shouldn’t feel alone. You should always get help, especially when you’re feeling down or depressed. If you get help, it’s going to help you a lot, it’s going to help you a lot. I don’t know yet, because I haven’t gone to get help, but I am pretty sure that if I could get help, I know it’s going to help, and I know that’s what I need.

Having a Vision for the Kind of Therapist to Seek

The third property of preparing to get help involved desiring to find a therapist like Veronica. Participants were so moved by Veronica that some not only wanted to seek professional help but also wanted to find a therapist who was like her. One woman, who revealed that Veronica changed her thinking about seeking mental health care, shared that feeling comfortable with a therapist was a priority. She felt that it would maximize her ability to open up and talk more freely. She stated:

I was thinking about, I wish I could get in contact or help from a person like her. It makes me feel kind of, you know, when you feel fine about someone, that you can talk?

Another participant expressed how the web-based app’s depiction of a therapist like Veronica actually gave her hope. It provided the impetus to take the initiative to seek help for herself. She stated:

I wanted to believe that there will be another therapist like her as well. You know, like, that understands you. So, it just made me feel comfortable to actually take that step.

Another woman who had thought about seeking help “plenty of times” but never followed through said, “I just worry so much, like, if they’re really going to understand me.” However, watching Veronica, she said, “If I had a therapist like her, I would not have a problem meeting with a therapist, honestly.” This helped generate the motivation to plan for her course of action to enter therapy. As a result, she planned to “talk to my doctor. He’ll recommend someone nearby my house.”
Discussion

Principal Findings

Our web-based app, Catalina: Confronting My Emotions, offered our Latina participants a culturally acceptable digital bridge that ultimately led them to resources for getting the needed mental health care. Although the first step of the process of engagement with our web-based app involved participants identifying with Catalina, it was Veronica, the fictional nurse-therapist character, who served as the catalyst through the remaining steps of the process that led participants to options for action. Our sample became actively engaged with the web-based app and transitioned from the point of view of the observer to that of the experiencer, moving from a passive position of watching to an active engagement, guided by Veronica. This process included imagining, thinking, reflecting, and acting. No doubt the process of engagement started when participants recognized themselves in the relatable lead character of the story line, Catalina [29]; this put the metaphorical gears of the dynamic process in motion (Figure 1). However, it was engagement with Veronica, the nurse-therapist character, that soon became the focus of their experience as they clicked to open transmedia bonus videos and other features. Through this, participants shifted from a focus on how Veronica helped Catalina in the story to personally experience how Veronica was actually helping them in real time. As such, Bandura’s dynamic of social modeling [26,47] was in motion first through Catalina and then through Veronica. By increasing engagement with the nurse-therapist character (Veronica), participants were able to apply her messages to themselves.

As Veronica’s messages mattered to them, they digested them. They came from someone they saw as a knowledgeable, competent health professional, who they wanted to listen to and engage with. This messenger, Veronica, also seemed to really care. She was a nurse and a therapist, but, importantly, she was also a Latina woman who understood them. Veronica’s messages prompted participants to replace previous assumptions with new perspectives on the healing potential of therapy and other mental health resources. As the old misconceptions about mental health were defused, the gears turned onward. The next phase of the dynamic process led participants to begin preparing themselves to take action. They were able to focus on when, how, and what they could do to make a change in their lives. The potency of Veronica’s messages added to the power of our web-based app to become a conduit to care for our sample, over a third of whom took action to use a resource, get information, or make an appointment within one week of engagement [28].

As participants already found the Catalina character to be highly relatable, the bar was set high for any characters who followed in the web-based app’s transmedia bonus videos. The results of our analysis showed that Veronica attracted and held the participants’ attention, deepening engagement and investment in the web-based app experience. There is no doubt that this is in part due to the talent of the Hollywood actress who portrayed Veronica (Yarel Arizmendi). Participants saw Veronica as welcoming not only because of how she looked or the sound of her voice but also because of the way she conveyed a sense of loving care to participants. From a film-directing point of view, Veronica was meant to be a knowledgeable character who was also compassionate [48]. Data from participants indicated that Arizmendi’s performance met the mark.

As we analyzed the transcripts from interviews with Latina women in this study, we found it impossible to tease apart participants’ views of the character (Veronica) from the content of the messages she delivered. Participants seemed to perceive them as being integrated as one and the same. Many women incorporated the actual language used by Veronica into their own self-expression. Women voiced her words as their own, which suggests that the things she communicated rang true to them, most likely because they were relevant to their lives. This underscores the importance of partnering with members of the target group to allow their concerns to be central to the design of a storytelling mental health web-based app [21,22,49]. It also indicates that casting a talented actress who could powerfully and poignantly embody a character was vitally important.

Our scripts were based on the priorities of Latina women who had struggled with depression that we had identified in previous research [30,31,36-38]. This included Latina women’s own past experiences with depression when alone or in therapy sessions with a therapist as well as lessons they learned and found valuable. As noted earlier, well before filming, the scripts were written by a Latinx script writer from Hollywood and revised with input from therapists. After filming, the first cut of the videos was vetted in theater testing by a sample of Latina women who fit the demographics of the target group [28]. This was done before our mixed methods study was implemented with symptomatic Latina women. The input from our participants for this analysis implies that the script indeed reflected what they needed to hear and that content was delivered in a way that they found easy to receive and absorb. Taking the messages from the script into account was not a passive act for them. Rather, the participants took an active role in the web-based app experience, considering its video content, as the story unfolded, clicking to watch each of the bonus videos and engaging in the interactive questions on screen as the story extended from a focus on Catalina to a focus on Veronica to a focus on them.

Participants became aware that distressing symptoms warranted professional help and their engagement with the web-based app expanded their vision of both the kind of help they wanted to find and where to get it. Interestingly, these are similar to the elements of mental health literacy that Jorn [13] described. It is possible that what helped facilitate this process was the compassion that the character, Veronica, expressed in her portrayal of a nurse therapist. Kemp et al [48] performed an analysis of compassionate mental health care delivered in digital form, such as through web-based apps. They defined compassion as having several elements that seem to pertain to the way that Veronica was perceived by Latina women in our sample. For example, through her gestures and words, Veronica conveyed awareness that many women have emotional struggles and need help. Participants saw Veronica as focused on helping by reducing suffering. Veronica also encouraged them to click the next link, engage with the interactive sequence, and think about what was holding them back. Participants could have chosen
to stop the web-based app engagement at any point in time and just turn it off. However, they continued, likely because of their connection to Veronica and because engagement felt useful to their own journey [21]. Although simultaneously being knowledgeable and caring in a way that Latina women welcomed, Veronica also modeled mental health literacy; she shared information about depression and anxiety, options for treatment, and where to get help [13]. Women responded with high interest, trust, and, ultimately, action.

Engagement with this web-based app created a unique space for participants, one that helped optimize mental health contemplation and led to action. It provided our Latina participants with an opportunity to experience what it might be like to have Catalina’s therapist interact with them. In the interactive sequence of 5 short 2-minute videos [28], Veronica posed questions from the confidence ruler and the importance ruler based on motivational interviewing [41]. Participants could move through this sequence at their own pace, replay a video, or return to it later. This gave participants the choice to engage with Veronica through a back-and-forth experience of videos and questions. It included ample space for personal introspection, possibly contributing to their feeling that this might be what an actual therapeutic interaction felt like. Participants valued this experience, which further enhanced engagement; it is noteworthy that we had no attrition in this study [28].

Torus et al [14] have called for the creation of digital technologies that lead to increased access to care while fostering a therapeutic relationship. Our character-driven web-based app holds promise for moving in that direction. For example, although our purpose was to engage symptomatic Latina women and then connect them to sources of therapy and help, our results showed that engagement in our web-based app led to statistically significant drops in depression and anxiety scores, both 1 and 6 weeks after engagement [28]. It is possible that this was partially due to a kind of therapeutic alliance that women developed with Veronica. Although she was not providing live therapy to participants, she was interacting therapeutically with them to enhance their self-awareness, confidence, and sense of the importance of getting needed care. Veronica also normalized mental health symptoms and helped women feel less odd, thereby challenging the stigma associated with depression and anxiety. Overall, Veronica met the participants’ needs.

Limitations and Next Steps
The findings of this analysis are not meant to be generalizable to all Latina women or to all Latina women with mental health symptoms. In addition, it is not generalizable to users of story-based apps, character-driven apps, or transmedia apps. Nonetheless, the knowledge generated here advances an understanding about important elements of mental health provider characters featured in storytelling web-based apps for underserved groups, specifically, English-speaking Latina women with depression and/or anxiety. Future research with a Spanish version of our web-based app is needed to test feasibility, acceptability, and efficacy with a sample of Spanish-speaking Latina women. In addition, a future randomized controlled trial with a much larger sample that is broader is needed to validate the study with evidence from symptomatic but untreated Latina women, both Spanish and English speakers, in comparison with an attention control. Potential extensions of this work include using a theory-driven approach to create additional character-driven webisodes or bonus videos that extend the story while supporting participants who are engaged in needed therapy (actual or virtual). Other interactive modules that are character-led and therapeutic could be created and tested in English and Spanish. Further exploration, development, and experimentation hold promise for gaining needed insight into the most potent uses of story-based features in mental health web-based apps. Finally, testing the web-based app with women of other ethnic groups or translation of the web-based app into languages other than English and Spanish would provide valuable insight related to helping other populations. Diverse, multidisciplinary teams will be crucial as we advance the science of web-based apps for health that use transmedia storytelling.

Conclusions
As researchers and developers continue to produce mental health apps, research and testing are crucial to identify the areas where we need to focus our efforts to provide effective web-based or app-related tools [23]. Our web-based app was both a useful conduit for mental health resources and an experience that was dynamically experienced by symptomatic Latina participants. Our positive results to date are likely due to many things including the theory-informed [26,45] process of development that we undertook and the multi-layered approach of drawing upon input from Latinx collaborators at every step. This analysis of qualitative data collected after use of the web-based app by the target group about their experience is crucial for informing the next step of app development. Although a theory-driven, collaborative, data-based approach is time consuming, it is key to creating web-based apps that are desirable and acceptable to a target group such as symptomatic Latina women and, therefore, more likely to be successful in meeting their needs.

Authors’ Contributions
PS collected all data and led data analysis and interpretation of results; AM collaborated in data analysis and interpretation; MH designed the study and collaborated in data analysis and interpretation. MH was primary in overseeing the entire study. All authors collaborated in writing and editing the manuscript.

Conflicts of Interest
None declared.


Considerations in Designing Digital Peer Support for Mental Health: Interview Study Among Users of a Digital Support System (Buddy Project)

Nazanin Andalibi¹, BSc, MSc, PhD; Madison K Flood², AS

¹University of Michigan School of Information, Ann Arbor, MI, United States
²Department of History of Art, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Nazanin Andalibi, BSc, MSc, PhD
University of Michigan School of Information
105 S State St
Ann Arbor, MI, 48109
United States
Phone: 1 734 763 1460
Email: andalibi@umich.edu

Abstract

Background: Peer support is an approach to cope with mental illness, and technology provides a way to facilitate peer support. However, there are barriers to seeking support in offline and technology-mediated contexts.

Objective: This study aims to uncover potential ways to design digital mental health peer support systems and to outline a set of principles for future designers to consider as they embark on designing these systems. By learning how existing systems are used by people in daily life and by centering their experiences, we can better understand how to design mental health peer support technologies that foreground people's needs. One existing digital peer support system is Buddy Project, the case study in this paper.

Methods: This paper reports on an interview study with Buddy Project users (N=13). Data were analyzed using the constant comparative approach.

Results: Individuals matched through Buddy Project developed supportive friendships with one another, leading them to become each other's peer supporters in their respective journeys. It was not only the mental health peer support that was important to participants but also being able to connect over other parts of their lives and identities. The design of Buddy Project provided a sense of anonymity and separation from pre-existing ties, making it easier for participants to disclose struggles; moreover, the pairs appreciated being able to browse each other's social media pages before connecting. Buddy Project has an explicit mission to prevent suicide and demonstrates this mission across its online platforms, which helps reduce the stigma around mental health within the peer support space. Pairs were matched based on shared interests and identities. This choice aided the pairs in developing meaningful, compatible, and supportive relationships with each other, where they felt seen and understood. However, the pairs were concerned that matching based on a shared mental health diagnosis may lead to sharing unhealthy coping mechanisms or comparing themselves and the severity of their experiences with their peers.

Conclusions: The results of this study shed light on desirable features of a digital mental health peer support system: matching peers based on interests and identities that they self-identify with; having an explicit mental health–related mission coupled with social media and other web-based presences to signal that discussing mental health is safe within the peer support ecosystem; and not matching peers based on a broad mental health diagnosis. However, if the diagnosis is important, this matching should account for illness severity and educate peers on how to provide support while avoiding suggesting unhelpful coping mechanisms; allowing for some degree of anonymity and control over how peers present themselves to each other; and providing relevant information and tools to potential peers to help them decide if they would like to embark on a relationship with their matched peer before connecting with them.

(JMIR Ment Health 2021;8(1):e21819) doi:10.2196/21819
Mental health is a topic of utmost importance to public health. Nearly 20% of US adults live with a mental illness, with the majority never accessing care [1]. Every day, an average of 129 Americans die by suicide [2]. In 2017, it was estimated that 18.9% of all US adults had a mental, behavioral, or emotional disorder [1]. More specifically, young adults aged from 18 to 25 years had the highest prevalence of mental illness (25.8%) among other age groups; however, the percentage of young adults who received mental health services was lower (38.4%) than adults of other age groups [1]. Although young adults tend to be one of the most vulnerable groups with respect to mental illnesses, their mental health care needs often remain unmet [3]. As a result of mental illness, suicidal ideations are estimated to affect 25% of young adults, with helplessness being the most reported motive [4].

The major barriers to accessing and seeking mental health care include stigma associated with mental illness, shortage of trained professionals, treatment costs, concerns about confidentiality, lack of knowledge of resources, and inaccessibility of services [5,6]. Furthermore, the United States is projected to have shortages in most types of mental health care providers by 2025 [7]. As such, new approaches that can complement or expand the capacity of mental health care are needed. One such approach is to facilitate seeking and finding social and peer support. Social support and peer support are well-established approaches to cope with illnesses, including mental illness [8]; however, when struggling with mental illness, seeking and finding appropriate support is difficult [9]. One key mechanism through which social and peer support may be accessed is through technology and the internet [10]. Peer support, typically exchanged between individuals who share an experience, can facilitate accessing social support by providing information support regarding professional help [11] and emotional and esteem support by boosting one’s self-esteem [11], increasing hope [12], reducing the feeling of isolation [13], and allowing peers to feel that they belong to a community [14].

Previous work [10,11,14-16] has established the benefits of creating technology-mediated support systems, including the option to use pseudonyms, an opportunity for social interaction, increased accessibility to a wide range of coping mechanisms and support tools, an increased likelihood of seeking help, a better chance of finding a destigmatized conversation space, and being part of a positive and supportive community of similar individuals. In addition, individuals experiencing mental illness often face unique barriers to build friendships or other social connections [15,17]. An example is social anxiety, which may lead to lower friendship quality or the fear of being rejected or embarrassed [18]. Therefore, using the internet as a way to develop these social connections has proved to be helpful. More broadly, young adults are reluctant to seek professional face-to-face help for their mental health conditions [19]. Therefore, examining other mechanisms that might be helpful to their coping and engagement with peer support is important.

Reports show that approximately 90% of American adults use the internet [20]. In fact, a growing portion of young adults use the internet to access mental health resources [21]. The moment an individual takes the first step to find support and connect to others is a critical point in their illness journey [22], whether that is seeking peer support or other types of care. However, with some exceptions [15], design guidelines for mental health technologies have mostly focused on systems that bring individuals with mental illness in contact with their providers rather than peers [23]. In fact, technology design for peers to support each other has lagged behind other technological innovations for mental health [24]. Although we know that peer support is helpful in coping with and managing mental illness [16] and that technology has the potential to facilitate this coping process [13], there remains a need to interrogate existing peer support systems for their success or lack thereof and learn from them to contribute to our knowledge about how technology-mediated mental health peer support may be designed. By learning how existing systems are used by people in their natural settings and by centering on their experiences, we can better understand how to design peer support technologies that center people’s needs. We address this gap by taking Buddy Project [25], an online peer support system for mental health and a nonprofit organization, as a case study. Buddy Project is a peer support system interested in fostering friendship, peer support, and connection, which is one aspect of mental health care. Buddy Project is not a replacement for formal treatment or a place to monitor mental health.

Buddy Project

For this study, we turned to Buddy Project, an online peer-to-peer support system that aims at fostering relationships that provide support to those struggling with mental illness. Buddy Project was founded in 2015 by Gabby Frost, a young person who wanted to prevent suicide and self-harm while advocating for mental health. Buddy Project’s mission to prevent suicide is now displayed on the website’s home page and in its Twitter and Instagram biographies (Figure 1). So far, the nonprofit movement has paired more than 236,000 adolescents aged between 12 and 25 years. Participants are matched based on age difference and shared interests and connected using their social media accounts (Twitter or Instagram). The sign-up process is shown in Figure 2, where users choose the social media account they want to use. Peers are then redirected to a Google form. On the first page, they provide their email address, first name, and Twitter or Instagram username, depending on the sign-up process they selected. On the second page, peers rank their first to fifth interests. A complete list of the interest options is included in Multimedia Appendix 1. We developed higher-level themes for these interests to gain a better sense of what they include. Interests categories include arts and entertainment (eg, musicians, television shows, books), identity (eg, religion, political ideologies, gender, and sexuality), and time zones. The third page prompts peers to provide their age.
The last page informs peers what to expect once they submit the Google form. It states that they will receive an email once paired with their buddy, that their selection is not automatic, and that peers are welcome to sign up for multiple buddies, along with the links to Buddy Project’s website, Instagram, Twitter, and Facebook accounts.

Buddy Project’s founder manually pairs all of the buddies using a Microsoft Excel spreadsheet. When pairing buddies, they take into consideration the interests they selected when signing up and their age. If someone wants to refuse a buddy, they can simply do so by not connecting with them once paired, and they can sign up to be paired with someone else. There is currently no formal training offered to buddies because buddies sign up to make friends who understand them (and thus can provide them with peer support), not provide or receive any type of professional help or crisis counseling.

We conducted in-depth semistructured interviews with 13 Buddy Project users.

Figure 1. Buddy Project mission statements on Instagram biography.

Figure 2. The sign-up page on the Buddy Project website.
Literature Review

Social and Peer Support for Mental Health

Social support is crucial for maintaining well-being [8]. Broadly defined, social support is characterized by meaningful interactions that provide some sort of support [26-28]. Social support can take many forms, such as emotional (communicating care and compassion), esteem (communicating confidence in one’s ability), informational (providing information, advice, and tips), network (communicating that one is not alone and that there are many others who understand them), and tangible or instrumental support (providing tangible help and services) [29]. Receiving adequate and appropriate social support improves mental health and well-being [30]. For example, emotional and esteem support are successful in reducing the impact of a negative, unhealthy view of oneself by providing individuals with a higher sense of self-worth [31].

Gaining mental health support poses challenges because it can be time consuming, draining, and costly, and those affected by mental illness may have additional personal or social barriers to build friendships or connections [22]. For example, some may feel uncomfortable or unequipped to approach people or communicate in person to receive support or may have cognitive or social impairments, hindering their ability to seek and receive social support [22]. A main challenge in coping with mental illness is the stigma surrounding it, which is the idea that if one seeks or receives help, they are weak, incapable of taking care of themselves, or are inferior to people who can cope with mental illness on their own [11]. Stigma creates a barrier that prevents people from accessing or acquiring information about mental health resources [13]. Increasing communication about mental health–related topics, coping mechanisms, and common emotions reduces the perceived embarrassment associated with these conversations [32]. Although seeking help online can be beneficial in part because it can limit the consequences of stigma, there is still space to remove barriers to seeking help.

The dominant method of promoting mental health is providing clinical mental health services via professionals such as psychologists, psychotherapists, or social workers. However, most people with mental health disorders receive no treatment [33]. A complementary approach to seeking formal treatments is engaging in peer support. Peer support refers to the support that people with lived experience of an illness or condition (eg, mental illness) can provide to one another [34]. The key principles of peer support include respect, shared responsibility, and an agreement of what is helpful [35]. Peer support’s importance as a key recovery service for people with mental illnesses has been established globally [36].

Digital Peer Support and Mental Health

Technology is increasingly applied to deliver peer support to individuals with mental health conditions [37]. Digital peer support is defined as peer support mediated through technology [38]. Technologies such as support forums and groups, mental health–focused mobile apps, and more broadly social media, have the potential to facilitate finding social and peer support and coping with mental illness. Along with increasing the amount of social support one receives [39], peer support can complement other resources (eg, professional therapy) one may use to cope with mental illness. The support one can obtain from a trained professional is qualitatively different from the type of support that one can receive from similar others with comparable lived experiences [40].

Research has focused on developing and evaluating mental health apps. Mental health apps provide features such as information, monitoring medications and symptoms, telepsychiatry, cognitive behavioral therapy, and support groups [41]. However, interest in using mental health apps does not mean actual high usage of these apps [42-47].

Extant research also shows how people turn to online communities to seek and exchange social support with many similar others [48,49] and express themselves [50]. Some online spaces also afford one-on-one peer-to-peer support, which allows peers to connect over personal experiences [51]. Regardless of whether conversations are one-to-many (ie, one post from one person to many recipients) as in the case of forums or one-on-one as in the case of private chats, because of this shared experience, individuals feel more connected, less alone, and less ashamed of what they are going through [13]. Moreover, finding support online can be crucial for some more than others; for example, individuals with intersecting marginalized identities often find safe online communities necessary to receive social support [48]. Some degree of anonymity present in some peer support technologies (eg, allowing the use of pseudonyms rather than enforcing the use of physical world identities) removes barriers to seek support when facing stigma, making social support a more attainable coping resource for some in comparison with nonanonymously settings (eg, in other online spaces) [22].

Online and offline, peer-to-peer support unites individuals with mental health conditions, providing them the opportunity to engage in sharing experiences, feelings, coping mechanisms, advice, and support to improve their mental health condition [15]. Although traditional face-to-face communication cannot be replaced with technology [49], connecting with peers via the internet allows for widespread accessibility and often an easier way for users to engage with one another [13].

In summary, digital peer support for mental health is an emerging and promising research space that has the potential to help improve mental health conditions, self-management skill development, social functioning, hope, and empowerment [38]. A systematic literature review on digital peer support interventions found that digital peer support interventions are feasible and acceptable, with high potential for clinical effectiveness [52]. Furthermore, peer support has the potential to improve and change not just how we approach mental health but also social change more broadly [53].

It is important to learn from existing services and platforms that employ technology for mental health peer support in settings outside of controlled research studies (in the real world) to examine users’ perceptions and inform the design of future digital peer support systems. We turn to Buddy Project (described earlier) as a case study to do exactly that.
Methods

Recruitment

The recruitment process began with a screening survey that the Buddy Project organization shared on their Twitter and Instagram accounts. To participate in the survey, participants needed to (1) be a social media user, (2) live in the United States, (3) be at least 18 years old, and (4) be a current or previous user of Buddy Project. Being a social media user was important for us to situate Buddy Project within other online support systems that participants may have used. If they met these qualifications, participants were asked to complete the second portion of the survey that asked questions about demographics, social media use, length of longest relationship with buddies, number of buddies they had matched with, month and year when they first used Buddy Project, and overall experience with their buddies.

We did not screen participants on the basis of mental health status or diagnosis, as we were interested in learning experiences with Buddy Project, and Buddy Project does not match buddies based on the diagnosis. The survey was open from June 12 to 25, 2019, and received 123 responses. From the responses, 63 participants met the initial criteria, 38 were invited to participate in the study, and 13 were interviewed. We sent an interview invitation to all qualified participants who had consistent survey responses (e.g., the time of their longest relationship with a buddy did not extend past the date of when they started using Buddy Project) and had used Buddy Project within the last 2 years. We also purposefully recruited participants from diverse demographics to the extent possible. The invitation included study details and an online consent form. A total of 13 individuals completed these forms and participated in the interview. Table 1 includes the details of participants. We continued recruiting participants through the interview process and stopped when no new themes emerged. We offered a US $15 gift card to interview participants. This study was approved by our institutional review board.

Apart from assisting with recruitment, Buddy Project had no other role in any part of this study. We plan to share this paper with the founder after peer review.
<table>
<thead>
<tr>
<th>Participant</th>
<th>Age (years)</th>
<th>Gender</th>
<th>Race</th>
<th>Education</th>
<th>Living area</th>
<th>Social media</th>
<th>User status</th>
<th>When they joined BP</th>
<th>Total buddy count</th>
<th>Longest relationship</th>
<th>Overall experience</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>19</td>
<td>Woman</td>
<td>Latina</td>
<td>Some college</td>
<td>Urban</td>
<td>TW, SC, TB</td>
<td>Previous</td>
<td>2015</td>
<td>1</td>
<td>4 years</td>
<td>Both positive and negative</td>
</tr>
<tr>
<td>P2</td>
<td>19</td>
<td>Woman</td>
<td>Black</td>
<td>Some college</td>
<td>Urban</td>
<td>FB, TW, IG</td>
<td>Previous</td>
<td>2017</td>
<td>4</td>
<td>A few months</td>
<td>Mostly positive</td>
</tr>
<tr>
<td>P3</td>
<td>25</td>
<td>Man</td>
<td>White</td>
<td>Some graduate school</td>
<td>Urban</td>
<td>FB, TW, IG</td>
<td>Current</td>
<td>Mid-2018</td>
<td>1</td>
<td>Almost a year</td>
<td>Mostly positive</td>
</tr>
<tr>
<td>P4</td>
<td>20</td>
<td>Woman</td>
<td>South Asian</td>
<td>Some college</td>
<td>Urban</td>
<td>FB, TW, IG, SC</td>
<td>Current</td>
<td>2018</td>
<td>2</td>
<td>6 months</td>
<td>Mostly positive</td>
</tr>
<tr>
<td>P5</td>
<td>23</td>
<td>Man</td>
<td>White</td>
<td>Some graduate school</td>
<td>Urban</td>
<td>FB, TW, IG, SC</td>
<td>Previous</td>
<td>Mid-2014</td>
<td>5 to 10</td>
<td>4 years</td>
<td>Mostly positive</td>
</tr>
<tr>
<td>P6</td>
<td>24</td>
<td>Woman</td>
<td>White</td>
<td>College</td>
<td>Rural</td>
<td>FB, TW, IG, SC</td>
<td>Current</td>
<td>Mid-2018</td>
<td>1</td>
<td>1 year</td>
<td>Mostly positive</td>
</tr>
<tr>
<td>P7</td>
<td>20</td>
<td>Woman</td>
<td>White</td>
<td>Some college</td>
<td>Urban</td>
<td>FB, TW, IG, SC</td>
<td>Previous</td>
<td>2018</td>
<td>1</td>
<td>A few months</td>
<td>Mostly positive</td>
</tr>
<tr>
<td>P8</td>
<td>18</td>
<td>Woman</td>
<td>White</td>
<td>High school</td>
<td>Urban</td>
<td>FB, TW, IG, TB</td>
<td>Current</td>
<td>April 2019</td>
<td>1</td>
<td>1 month</td>
<td>Mostly positive</td>
</tr>
<tr>
<td>P9</td>
<td>18</td>
<td>Nonbinary</td>
<td>White</td>
<td>High school</td>
<td>Rural</td>
<td>TW, SC, TB</td>
<td>Current</td>
<td>2014</td>
<td>4</td>
<td>2 years</td>
<td>Mostly positive</td>
</tr>
<tr>
<td>P10</td>
<td>20</td>
<td>Woman</td>
<td>White</td>
<td>Some college</td>
<td>Rural</td>
<td>FB, TW, IG, SC</td>
<td>Previous</td>
<td>2017</td>
<td>2</td>
<td>4-5 months (still check in once in a while)</td>
<td>Mostly positive</td>
</tr>
<tr>
<td>P11</td>
<td>19</td>
<td>Woman</td>
<td>White</td>
<td>Some college</td>
<td>Rural</td>
<td>FB, IG, SC</td>
<td>Previous</td>
<td>Mid-2018</td>
<td>3</td>
<td>6 months</td>
<td>Mostly positive</td>
</tr>
<tr>
<td>P12</td>
<td>19</td>
<td>Gender-fluid</td>
<td>White</td>
<td>Some college</td>
<td>Rural</td>
<td>TW, IG, SC</td>
<td>Current</td>
<td>2017</td>
<td>5</td>
<td>9 months</td>
<td>Mostly positive</td>
</tr>
<tr>
<td>P13</td>
<td>20</td>
<td>Woman</td>
<td>Asian</td>
<td>Some college</td>
<td>Rural</td>
<td>FB, SC, TB, RD</td>
<td>Previous</td>
<td>2015</td>
<td>6</td>
<td>A couple of months</td>
<td>Both positive and negative</td>
</tr>
</tbody>
</table>

Participants typed their gender and race, and we report the terms they used to describe themselves.

aBP: Buddy Project.
bTW: Twitter.
cSC: Snapchat.
dTB: Tumblr.
eFB: Facebook.
fIG: Instagram.
RD: Reddit.

Interviews and Data Collection

Participants included 10 women, 1 man, 1 nonbinary, and 1 gender-fluid person. The average age of participants was 20.3 (SD 2.25) years (range 18-25 years). In total, 6 participants lived in rural areas, and 7 lived in urban areas; 9 participants were White, 1 was Black, 1 was Latina, 1 was Asian, and 1 was South Asian. A total of 7 participants had previously used Buddy Project but were no longer active users and 6 were current users. The number of buddies ranged from 1 to 10, and the length of contact with a buddy ranged from 1 month to 4 years. A total of 11 participants described their experience with Buddy Project.
as being positive overall, whereas 2 described it as both positive and negative.

We conducted in-depth semistructured interviews that allowed participants to provide details about their experiences with their buddies while still covering topics regarding their experience with Buddy Project. The interviews were conducted via participants’ preferred method of voice or video call, and the audio was transcribed for analysis. The average length of the interviews was 47 min (range 28-56 min).

The interviews began with the interviewer sharing the goals of the study and asking for permission to record the conversation. The participants were then asked about internet and social media use before diving into their stories and experiences with Buddy Project. We asked to hear the story of how the participant found Buddy Project, why they started participating, why they continued or stopped connecting with their buddies, what the relationship with their buddy was like, and what topics they discussed. We asked questions about perceptions of the shared interest feature and their best and worst buddy experiences. We continued with questions regarding how their buddies compared with other friends, online and offline, and when relevant, they were probed about how they used Buddy Project to cope with mental illness or distress and how this compared with their other coping mechanisms (if any). It was important to us that participants only discussed experiences they felt absolutely comfortable with; therefore, we only probed on specific mental health experiences only if they mentioned the topic organically. We paid particular attention to a unique feature of Buddy Project, that is, the way buddies are matched based on the interests they select when they sign up. The interview protocol is available in Multimedia Appendix 2.

Analysis

We started coding the data with an open coding procedure using Dedoose—a software for analyzing qualitative data. We followed the constant comparative approach [54], where we looked for patterns, consistencies, and differences in the data in an iterative manner. Specifically, one author coded one interview. Then, the 2 authors met with each other to discuss and refine the codes. The same author then coded the next 8 interviews with the constant comparative approach where we looked for patterns, consistencies, and differences in the data. Specifically, one author coded one interview. Then, the 2 authors met with each other to discuss and refine the codes. The same author then coded 4 more interviews, for a total of 5, and the authors met to sort codes into themes and to further discuss and refine each code and theme. The same author then coded the next 8 interviews with these codes in mind, and the 2 authors met frequently to continue refining the themes. No further codes or themes emerged in the latter process.

Results

We provide an overview of how participants connected with their buddies. Overall, participants used a variety of communication tools to connect with their buddies, including social media, text messages, video calls, and phone calls, with the most common being messaging and social media, which was most convenient for long-distance buddies so they did not have to pay data and messaging rates. These channels also protected buddies from releasing their phone numbers. One participant (P5) spent 2 weeks in person visiting her buddy. Communication frequency varied from multiple times a day to just enough interaction to stay in contact.

In the remainder of this section, we report 4 key themes that were derived from our analysis of participants’ experiences that contribute to Buddy Project’s success: building a support system for buddies, a stated and visible mission to prevent suicide in Buddy Project’s official online presence, matching buddies based on shared interests and identities, and not matching buddies based on shared diagnosis.

Building a Peer Support System for Buddies

Buddies Providing Peer Support

Participants shared reflections on how buddies provided peer support. For example, P11 emphasized the importance of peer support and how Buddy Project contributes to making meaningful supportive connections:

I think [Buddy Project is] one of the most amazing things that there is, especially in terms of promoting your mental health...and having friends [that] actually are there for you and [Buddy Project is] trying to connect you to people with those interests [so that] you have those friends.

P11 gave an example of what the friendship with her buddy is like and how they exchange social support when she said the following:

They [the buddy] were definitely always trying to make sure the other person was happy no matter what the situation they just went through was...they were actually just very good about everything no matter how difficult a situation. They wouldn't go to bed unless you were okay and they were always there. That's just, I think, something everyone needs.

P8 touched on how her friendship with her buddy counters loneliness when she said: “I did feel more lonely before [having the buddy] ... But now I feel, I think, like I'm friends with someone.”

Referring to buddies as friends and these connections as something everyone needs are important demonstrations of what important gaps buddies fill in participants’ lives.

Buddies found that having the option to talk about mental health was important but not a necessary part of daily conversations with their buddies. Having someone who is simply there to talk about their struggles as needed is something that P5 found beneficial. She described the supportive relationship that she had with her buddy as follows:

Being there for each other when we need it, when we're both low. Knowing how to help each other feel better but just knowing that the other person is there if you just need to vent completely. They'll listen and understand and if they have anything to say, being able to say something that might help.

Similarly, P6 said:

I think it's important to have someone that doesn't have a bias, that shares in your issues. Her and I...
Participants did not talk to their buddies about mental health challenges all the time nor did they want to; however, knowing that they could do so should they wish, was comforting and a unique aspect of buddies’ relationships with one another compared with their other relationships. These examples illustrate buddies providing emotional support (ie, communicating care and compassion) to each other.

Participants also noted sharing advice and tips, a type of informational support, to their buddies. For example, P9 described their relationship with their buddy as follows:

They got from the very basic, where do you live and whatnot, to asking about family life and being able to talk about like, “Oh, this is going on in my mental health region,” and giving each other advice on how to deal with it. And if things got too bad, they would stop me and be like, “Hey, you should talk to your parents about this or someone around you.”

Similarly, P13 noted the following:

I was giving the type of advice or comfort that I was trying to seek from other people… it’s how I help my friends… I did the same for my buddies.

Sense of Anonymity and Separation From Existing Known Ties

Participants noted that difficulties in disclosing mental health or other sensitive information with peers or friends (not buddies) included privacy concerns and not trusting that the other person would keep their information confidential. In contrast, with buddies, a sense of anonymity and separation of buddies from networks of known ties eased difficulties in feeling safe to share intimate information with others. P4 described it as follows:

If I were to come out to a friend, even if they’re a friend I trust, there’s still a part of this bigger circle that I’m in. They could always tell other friends, or they could accidentally slip up and tell a classmate, or a coworker, things like that. Whereas, my buddy, we’re geographically in a different realm. I know it’ll stay between us and if it doesn’t I don’t have to worry about who they’re sharing things with.

Peers may be unwilling to disclose their struggles with mental illness with known (not anonymous) peers in their social circles for a variety of reasons, such as the stigma surrounding mental illness (or talking about it), lack of helpful skills to effectively do so [55], or privacy concerns [56]. Buddies were often desired, in part, because they were perceived to likely be outside of one’s network of existing ties and without connections to one’s existing social network.

Although for some the separation from one’s network of known ties was helpful and provided a sense of desired anonymity, there were still other cases where the amount of afforded anonymity with buddies was insufficient. For example, P13 described how she used a completely anonymous forum as a mental health resource:

If I had a whole long thing that I needed to just rant about or get advice on but I wasn’t comfortable talking to anyone in my life or even a buddy, I would go on the forum.

We see how although buddies were helpful, participants still felt the need to seek other resources to find the support they needed sometimes. Nevertheless, in this example, we see the unique position the buddy had in participants’ lives such that they could not talk about certain things even with a buddy. In this sense, a forum provided more anonymity than the buddy, whereas a buddy provided more anonymity compared with one’s in-person friends. As such, following previous work [57], we conceptualized anonymity as a continuum rather than a binary.

Although human connections can be supportive and helpful, they can also cause harm. Buddy Project provides participants with the Twitter or Instagram username of their peer match, which participants often used to ease worry about connecting with a total stranger. P10 noted the following:

It’s an easy way to connect to people that’s safer [than connecting with a random stranger] because you’re connecting me to social media so you can see who the person is and you can know a little about them just by scrolling through profiles and it’s not you’re just talking to a random stranger:

Connecting with strangers on the internet is often a concern for those who seek online support [15]. When participants were able to gather cues on social media to assess the buddy, they felt safer to begin a conversation with them.

A Stated Mission to Prevent Suicide on Official Buddy Project’s Online Presence

Mental illness can prevent people from feeling included and connected with peers, often categorizing those diagnosed as minorities, creating even more disconnection from one’s broader social environment. Overcoming stigma is often the first step to seeking and receiving support for mental illness [58]. Creating a safe space for discussing mental illness often requires a commitment to destigmatization, and Buddy Project is doing exactly that with their explicit mission to prevent suicide and commitment to raise awareness for mental health via their social media outlets. When asked what she thinks about the positive messages on Buddy Project’s Twitter and Instagram, P10 responded as follows:

I think it’s great that they do that because not many people on social media are willing to just openly post about [mental health] so seeing the Wallpaper Wednesdays and all the positive messages... I think it’s great that there is an organization out there that’s willing to do that.

Relatedly, P3 described Wallpaper Wednesday as “weekly wallpapers for your phone and your mobile devices.” He continued as follows:
Shared Interests as Conversation and Compatibility Aid

Matching Buddies Based on Shared Interests and Identities

The explicit framing of Buddy Project allowed less perceived stigma between buddies to discuss mental health, resulting in more discussions of mental health compared with participants' other social settings online and offline. We see how having a mission to raise mental health awareness and demonstrating that in action (eg, through providing resources or social media posts) is an important characteristic for technology-mediated peer-to-peer support systems that want to encourage participants to discuss mental health and exchange support.

Matching Buddies Based on Shared Interests and Identities

We found that another feature that participants deemed helpful with Buddy Project was how peers are matched on the basis of shared interests and identities (eg, music, television shows, identity). Our analysis suggests that shared interests act to promote conversation between new buddies who are otherwise technically 2 strangers at the time of being paired. In addition, salient shared identities help with ensuring that one will be understood when discussing mental health and intersections of their identities with mental health.

Shared Interests as Conversation and Compatibility Aid

Matching buddies on the basis of shared interests had several perceived benefits and functions for participants, which stem from the central theme that shared interests act as support for better conversations, leading to deeper connections at a higher pace. This way of being matched acted as a means for starting conversations, as P3 described:

I think the shared interest is good because it gives you something to talk about. If you didn't have a shared interest I think it would be a little harder to start up conversations whereas if you have that common interest you can be like, 'What did you think about this part of our common interest?' And then you can start a conversation from there and then it can drift off into other things about your life.

Similarly, P10 described her experience with using the shared interest to start conversation:

It made it easier to talk to them because you went in with this baseline interest that you can both just start talking about and it wasn't awkward icebreaker conversations...you get that instant connection and you understand already why you are talking to this person as opposed to if you're meeting somebody in person for the first time you don't always know what to talk about.

She continued to describe the process:

If I message them first, I'll just go in with, “Hey, our interest is this. What do you like about it or what do you know about it?” And just to start the conversation where it's not awkward back and forth like, “Hi, how are you?” Because those conversations can get tedious and boring.

In this way, shared interests mitigated the risk of potentially losing out on relationships that have the potential to be helpful because buddies can begin their relationship with discussing topics that are meaningful to them (and are not necessarily related to mental health) rather than topics that they do not share an interest in.

Participants also noted that the shared interest feature improves the likelihood that one will find a compatible companion. P1 described this as follows:

If I saw something that I thought was cool I could send it to her to also see, or if she saw something she could send it to me and then that would just keep our conversation going.

The shared interests made it more likely that participants were matched with a compatible buddy. They acted as optional conversation prompts and facilitated the formation of deeper connections between buddies.

Shared Identity to Ensure Being Understood When Discussing Mental Health

Participants noted that it was more helpful to connect with someone who shared a salient identity with them—identities that they perceived to shape their mental health experiences. For example, P6, a self-identified neurodiverse person, said as follows:

I think it's definitely easier [to connect to another neuro-diverse individual to discuss mental health]...

I have one friend in real life [also neuro-diverse] that...
her and I say people who are neuro-typical ... just don’t understand necessarily sometimes the issues that we have or that we’re going through.

She continued:

I think it’s easier to talk to someone who kind of gets it a little bit, given I would talk to anyone that had any issues... I think it’s really helpful to talk to someone, and have someone else just get it.

P6 found it easier to talk to neurodiverse peers, especially about mental health, compared with others who did not share this identity facet with her. In fact, many neurodiverse individuals consider neurodiversity as part of their identity, not an illness [59].

As another example, P12 described how Buddy Project “helps you find people ...Especially with pan[sexual], it helped me find people that are like myself.” As a result of having a buddy with similar identities, participants felt that they would be understood and seen within their buddy relationship.

Not Matching Buddies Based on a Shared Mental Health Diagnosis

Participants’ accounts suggest that shared mental health diagnosis may not always be the best way to match peers who need social support to cope with mental illness. We found that buddies typically base their first conversations off their stated shared interests. As previously described, buddies’ experiences have been positive with the shared interest feature, whereas the mental health orientation of Buddy Project creates a shared understanding that discussing mental health with buddies is not taboo. Participants noted that when the shared diagnosis replaces the shared interest, the dynamic of the relationship changes. The friendship aspect is lost, as discussing mental health diagnosis is not how people begin to form friendships with one another; of course, once a relationship exists, discussing mental health is something participants feel more comfortable with. In this section, we describe how participants reflected on the perceived benefits and drawbacks of being matched with someone that shares their same mental illness diagnosis. These include concerns about differences in the severity of buddies’ experiences, engaging in self-comparison, and sharing unhealthy coping mechanisms.

Comparing Self and Severity

According to participants, one of the perceived drawbacks of being matched with someone based on a diagnosis is that mental illnesses can vary in severity. P11 described this drawback as follows:

The cons would be just the different levels of each different type of mental illness. Anxiety has so many different levels that sometimes you really can’t connect because someone’s [level of severity] could be so low and then someone else’s [level of severity] could be so high.

As a result of different severities of anxiety, participants’ symptoms can be so varied that it could hardly look like the same condition. It is likely that if 2 people are paired with the same mental illness, they are expected to be able to connect over the diagnosis and experiences. When this does not happen because of different severity levels, participants may become discouraged from using peer-to-peer support or not find the validation and support that they need, making mental health support resources scarce.

Participants also noted that matching based on diagnosis might lead to self-comparison on coping and wellness between buddies. For example, P12 described the competitiveness that self-comparison may cause:

I guess if you’re talking to someone who’s had a similar diagnosis as me but their experience was still different. I feel like sometimes it would become a game of “Oh, mine is worse” or that kind of game.

When individuals compare the severity of their mental illness with that of someone who seems to be coping more effectively than they are, one is subject to negative self-talk [60], which is not helpful in coping.

On the other hand, when participants share the same diagnosis and the same severity level, they will often share similar experiences that allow for a deeper understanding of each other. P6 spoke to the point that anxiety varies in severities and that a benefit of connecting based on shared diagnosis and severity is that one is more likely to find someone who truly understands their experience. P6 said that she would rather connect with someone that has the same diagnosis and severity level than someone who just has the same diagnosis broadly:

I think it’s better to be able to just connect with someone on that [severity] level, just because they do truly understand what you’re going through or how you’re feeling...

Variations in severity can lead one to compare their mental state with their peers, which is not always a helpful coping strategy. However, sharing the same severity of mental health diagnosis can result in a deeper understanding for one another and less competitiveness. Overall, for participants wanting to connect with others over a similar diagnosis, severity level was a criterion to ensure the connection would be helpful. Further research is required to examine how the design of peer support technologies would account for fluctuating severities. In this study, we provide preliminary evidence that is a relevant criterion to consider.

Sharing Unhealthy Coping Mechanisms

Another potential downfall to matching buddies based on shared diagnosis is the risk of sharing unhealthy coping mechanisms. P10 gave an example of sharing unhealthy coping mechanisms when asked about being matched with a buddy with the same condition:

If you sleep all day to help you cope with something, telling the other person that they might try sleeping more to see if that helps them. Not the healthiest option, but it’s what you do.

Other unhealthy coping mechanisms may include self-harm, substance abuse, changes in diet, and a disconnection from reality. These negative changes can result in decreased mental...
health [61], which defeats the purpose of any peer support system.

Overall, participants’ accounts highlight the nuances that shared diagnosis matching may have, sometimes leading to more understanding and sometimes leading to unhealthy coping mechanisms or otherwise unhelpful behaviors such as self-comparison.

Discussion

Principal Findings

Although Buddy Project is not the most technically sophisticated peer support system, we found that its design principles are effective in facilitating relationships that help young adults cope with mental illness by engaging in peer support. By taking Buddy Project as a case study that uses technology to pair up young people to cope with mental illness, we provide insights into the qualities possessed by effective technology-mediated peer support systems. We found that designing a technology-mediated peer-to-peer support system with a commitment to mental health awareness is an effective way to create a coping tool for young adults facing mental health challenges. This clear commitment lowers the stigma attached to mental health discussions between buddies and provides a shared understanding that sharing mental health–related experiences is a legitimate need that can be met within this peer support system.

Future digital peer support tools for mental health can use a combination of the features that Buddy Project users found helpful: (1) semiguided chat spaces made possible through matching based on interests and identities that peers self-identify with and having an explicit mental health–related mission coupled with social media and other online presence that makes it clear that discussing mental health is safe within the peer support ecosystem; (2) not matching based on broad mental health diagnosis; however, if diagnosis is important to peers to incorporate such matching with educating peers on how to provide helpful support and how to avoid unhelpful coping mechanisms and accounting for symptom severity in matching; (3) allowing for some degree of anonymity and control over how peers present themselves to each other; (4) allowing potential peers to assess the fit and whether they would like to embark on a relationship with their peers based on browsing public social media profiles of one another or innovating privacy-preserving means of assessing initial fit and trust. We describe these in detail in the remainder of this section.

Semiguided Chat Spaces for Mental Health Coping and Developing Meaningful and Supportive Relationships

Mental Health–Related Mission

We found that having an explicit message such as Buddy Project’s mission to prevent suicide creates a space where peers feel safe to talk about mental health; however, mental health is not all that they feel compelled to talk about. Removing people’s fear of being judged by their illness not only contributes to more fluid communication but is an essential element of a trustworthy relationship [62]. Once people recognize their condition and are able to communicate their experiences with others, they are taking the first step to recovery and are more likely to seek further help [16]. We recommend that future peer support systems and organizations aiming at facilitating mental health peer support make it an explicit point to advocate for mental health destigmatization through all their outlets in practice.

Shared Interests and Identities

Although the mental health mission provides a stigma-free space to discuss mental health among peers, buddies found that being matched based on shared interests and identities (and not diagnosis) provided complementary context for them to develop deep meaningful friendships in which discussing mental health and other topics were welcome.

Semiguided Conversation Space

Taken together, shared interests and identities and the mental health mission statement create what we call a semiguided conversation space: an open and flexible yet somewhat guided conversation space, ideal for connections that are both deep and supportive. We describe how we came to this notion in the remainder of this section.

O’Leary et al [24] designed a chat system in Google Docs to study the impact of guided and unguided chats for technology-mediated mental health peer support. In their study, the guided chats followed a script, whereas the unguided chats had no prompts at all. They found that guided chats resulted in deep connections, where peers provided solutions to problems and new perspectives. Unguided chats resulted in smooth conversations that offered personal connections. Their participants referred to unguided chats as pleasant and relaxing acting as a distraction, or temporary relief, from one’s problems. We found that chats with buddies fell somewhere in between what O’Leary et al [24] referred to as guided and unguided chats, which we call a semiguided conversation space.

We describe semiguided chats as those that have the characteristics of both guided and unguided chats to some extent. Semiguided chats are similar to guided chats because of the way the shared interests or identity and mission statement guide conversations among buddies; they are different in that they do not use explicit prompts for buddies to engage in conversations about as guided chats do. Unlike a fully unguided chat space where this shared understanding is missing, the sociotechnical space with semiguided chats is not completely void of possible topics that peers would know they can safely speak to their buddies about (eg, shared interests or identities, mental health). We found that the semiguided conversation space provided by Buddy Project facilitates the development of meaningful friendships and personal connections and a supportive context within which discussing mental health was safe and comfortable and where buddies would exchange stories and perspectives to help each other. However, buddies did not feel that they needed to discuss mental health all the time nor did they feel that they absolutely needed to provide actionable advice to their peers in contrast to what a guided chat space would require. We suggest that a semiguided chat space will allow peers to not only gain the benefits of both guided and unguided chats but will also
likely not evoke unwanted feelings for buddies (eg, an unwanted sense of responsibility to help their peers, as was the case in the study by O’Leary et al [24]).

Semiguided chats offer buddies the option to fluidly switch between deep and light conversations, whereas unguided and guided chats [24] were focused on one or the other. Semiguided chats allow peers to find the middle ground between exchanging social support and problem solving, which could allow peers to use both mechanisms in one resource (ie, relationship with the buddy). We suggest that semiguided chats are a good design principle when creating a sociotechnical space for mental health discussions and peer support. They allow peers to have freedom in what they talk about so the conversation does not feel forced, yet they are free flowing. Combined with a message about mental health, the conversation will still likely gravitate toward becoming an effective coping tool, as it did for Buddy Project users participating in our study.

Previous research has recommended connecting peers in online communities on the basis of similar features other than diagnosis in contexts such as breast cancer [63], various cancer diagnoses [64], and caregivers of individuals with cognitive illnesses [65]. Specifically related to mental illness, recent work [24] suggests that technology can enhance peer support for mental health by matching peers based on similarities that go beyond diagnosis. However, this previous research begs the questions of what this peer support might look like in practice, what peers’ attitudes toward it might be in practice, or what other features are important to make peer support matching helpful to peers. Buddy Project’s design, as we found, fills this gap in our knowledge.

**Striking a Balance: Matching Peers Based on the Mental Health Diagnosis or Not**

The participants in this study expressed concern about the potential costs of being matched based on shared diagnosis—comparing one’s self and severity of illness and sharing unhealthy coping mechanisms. In fact, individuals with mental illnesses do not always prefer for their peers to have the same diagnosis as theirs [15]. This finding resonates with previous studies, suggesting that sharing unhealthy coping mechanisms is one of the most likely disadvantages of online support groups and peer support [38,55,60,66]. This can include suicidal ideations being shared among peers, and even the possibility of a suicidal pact, and an overall diminished self-esteem and well-being [55]. In the case of depression, which affects an estimated 25% of young adults [67], hearing a peer’s depressive thoughts can cause a downward spiral [56]. Being matched based on the same diagnosis can also lead to lower self-esteem and hope when people engage in self-comparison and feel as though others are doing better [16], which counteracts the peer support that one should be receiving [17].

However, with the right system and the right peer match (eg, facilitated through shared interests and identities), sharing healthy coping mechanisms could be a positive experience for participants as it acts as an outlet to seek and provide helpful mental health resources and provide a sense of solidarity and connection to peers [22]. For example, if a participant has been seeking help from a therapist, they could suggest their peer to do the same and support their claim with personal experience, making it more likely that their peer seeks additional mental health support. For this to be possible, we argue that individuals would have to be educated on what healthy and unhealthy coping mechanisms are. In addition, when matched based on diagnosis, peers can provide support to each other around common stressors, symptoms, stigma, and other challenges, resulting in them feeling less alone [58].

A useful approach that addresses the concerns and allows users to receive the benefits of shared diagnosis pairing could be offering educational material to peers that explains to them the potential dangers of these unhealthy actions and instead educating them on healthy coping mechanisms. There is further evidence for an unmet demand for such educational programs, as many individuals express interest in training to become a peer counselor [68]. For instance, the 7 Cups of Tea website educates users on active listening and support provision; similarly, Crisis Hotlines trains their volunteers.

Technological spaces such as Buddy Project or other services dedicated to mental health advocacy and awareness can use their online presence (eg, social media posts) to provide such educational material. They can also provide a set of resources to buddies once they are matched, educating them on healthy coping mechanisms and how to be a supportive, compassionate buddy. The effectiveness of any such approach should be investigated in future research. Our study provides initial insights that there is room for improvement when it comes to engaging in healthy behavior in technology-mediated peer support systems for mental health when a shared diagnosis is a factor in matching peers.

**Anonymity, Privacy, Intimacy, and Safety in Relationships Between Buddies**

Privacy concerns and the risk associated with sharing information are key challenges for internet users because they can make them more reluctant to sharing information, asking for support, and building connections with other users [69]. It can also be difficult to reach out to strangers with whom one has no pre-existing ties to find support. One of the reasons Buddy Project is a popular choice among young people seeking online peer support is the sense of anonymity and separation from participants’ existing and known networks (eg, family, classmates). They are not completely anonymous, as buddies have access to their peers’ Twitter or Instagram account and share varying levels of personal information with each other; however, there is a sense that they are unlikely to have a pre-existing tie with their buddy. Therefore, their buddy would also not know anyone in their existing social network. There are also no requirements for sharing physical world names or identities, as they are in platforms such as Facebook. As a result, buddies feel they can safely share their mental health struggles in confidence, without worrying about others’ undesired access to their personal information and struggles.

A sense of anonymity facilitates openness in self-expression, support seeking, and support provision around sensitive topics [70-72]. Our findings suggest that control over how to present oneself (eg, using a pseudonym or just the first name) coupled with separation from networks of known ties (ie, existing social
connections such as friends and family) and some contextual information through access to the prospective buddy’s social media accounts, or shared interests, helped peers to take the first steps in a relationship with a buddy. Once there was more trust developed in the relationship, participants revealed more and deeper information about themselves, consistent with the social penetration theory [73]. This theory suggests that relationships develop and become more intimate over time as people share more intimate information with each other and move away from solely shallow information [73]. Here, we see how access to prospective buddies’ social media profiles before conversing with them for the first time in tandem with the mental health mission and shared interests matching provide a fruitful sociotechnical space for developing deep supportive relationships with buddies.

Giving individuals who sign up to be paired an option to add their social media information is how Buddy Project conducts its matching process; however, users do not need to have public profiles. We learned that being able to assess their proposed buddy before connecting with them through the availability of some signals on their social media was helpful to participants. However, this would be harder for those whose online profiles are private or do not include much information. We suggest that Buddy Project or similar systems provide ways for their users to assess their buddies before deciding to connect; this does not have to occur using social media platforms. For example, an alternative would be creating a profile within the system that has an about me page, how long the person has been a user, buddies paired with, and any other information that users choose to share about themselves.

Some drawbacks of connecting online to find peer-to-peer social support are privacy concerns and cyberbullying [31]. Our findings demonstrate privacy concerns but not much about cyberbullying concerns—an important area for future research. In summary, our findings suggest that peer support systems should consider allowing some degree of anonymity, separation from existing social networks, and control over presentation, along with providing tools to assess and maintain safety and initial fit with prospective buddies before connecting with them.

Conclusions
We contribute to an understanding of desirable features for digital mental health peer support systems: (1) matching peers on the basis of interests and identities they identify with; (2) having an explicit mental health mission coupled with other online presence to signal that discussing mental health is not a taboo within the digital peer support system; and (3) not matching peers based on broad mental health diagnosis; however, if diagnosis is crucial to account for, accounting for illness severity and educating peers on how to provide support while avoiding suggesting unhelpful coping mechanisms; (4) allowing for some anonymity and control over how peers present themselves to each other; and (5) providing relevant information and tools to potential peers to aid in their decision in connecting with the proposed peers before connecting with them.

Limitations and Future Work
As is common in interview studies, this study’s sample was not representative of Buddy Project users or their experiences with the service, even though we sought a diverse participant group in demographics and experiences. Future work is needed to evaluate our findings with a larger and representative population, possibly through other methods (eg, survey of Buddy Project users). Nonetheless, following best practices in interview research, our goal is not generalizability [74], rather generating conceptual insights. The challenges faced by individuals with multiple intersecting marginalized identities are worthy of future exploration. For example, although we note that connecting over identities is helpful, future research should explore what an ideal connection would look like when multiple identity facets are concerned.

Our study does not account for all kinds of mental illnesses, and further research is needed to examine the similarities and differences in designing for various mental health peer support systems. Nevertheless, we identify the factors that such designs should consider.

Within our interview participants, we did not encounter anyone who mentioned that they had encountered fake profiles or those with malicious intentions. However, this does not mean that such harmful interactions do not happen. It is possible that those willing to interview with us had more positive experiences. For example, Buddy Project shared our study link on their social media accounts. It is possible that those who follow Buddy Project on social media have had better experiences with the platform (although we attempted to recruit individuals with both positive and negative experiences). However, examining how to design technologies to facilitate trust and safety while reducing harmful behavior is an ongoing area of research that is beyond the scope of this study.

Acknowledgments
The authors wish to thank the participants in the study and the founder of Buddy Project, Gabby Frost, for assisting with recruiting participants. The second author was partly supported through the University of Michigan’s Undergraduate Research Opportunity Program, to which the authors are thankful.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Buddy Project's interests and identity options.

References


25. Buddy Project. URL: http://www.buddy-project.org/ [accessed 2020-12-03]


30. Wright KB, Bell SB, Wright KB, Bell SB. Health-related support groups on the internet: linking empirical findings to social support and computer-mediated communication theory. J Health Psychol 2003 Jan;8(1):39-54. [doi: 10.1080/13591050300801429] [Medline: 22113899]


53. Mead S. Peer Support as a Socio-Political Response to Trauma and Abuse. Intentional Peer Support. URL: https://docs.google.com/document/d/1trJ35i4dXX5AIW8nbg7804t7-RIPE9_DbPm5kSST9_Q/edit [accessed 2020-06-24]


©Nazanin Andalibi, Madison K Flood. Originally published in JMIR Mental Health (http://mental.jmir.org), 04.01.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Mental Health, is properly cited. The complete bibliographic information, a link to the original publication on http://mental.jmir.org/, as well as this copyright and license information must be included.
Original Paper

Effects of ACT Out! Social Issue Theater on Social-Emotional Competence and Bullying in Youth and Adolescents: Cluster Randomized Controlled Trial

Jon Agley¹, MPH, PhD; Mikyoung Jun¹, MPH, PhD; Lori Eldridge¹, MS; Daniel L Agley¹, EdD; Yunyu Xiao², MPhil, PhD; Steve Sussman³,⁴,⁵, PhD; Lilian Golzarri-Arroyo⁶, MS; Stephanie L Dickinson⁶, MS; Wasantha Jayawardene¹, MD, PhD; Ruth Gassman¹, PhD

¹Prevention Insights, Department of Applied Health Science, School of Public Health, Indiana University Bloomington, Bloomington, IN, United States
²School of Social Work, Indiana University Bloomington and Indiana University-Purdue University Indianapolis, Indianapolis, IN, United States
³Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States
⁴Department of Psychology, Dana and David Dornsife College of Letters, Arts and Sciences, University of Southern California, Los Angeles, CA, United States
⁵School of Social Work, University of Southern California, Los Angeles, CA, United States
⁶Biostatistics Consulting Center, School of Public Health Bloomington, Indiana University Bloomington, Bloomington, IN, United States

Corresponding Author:
Jon Agley, MPH, PhD
Prevention Insights, Department of Applied Health Science
School of Public Health
Indiana University Bloomington
501 N Morton St
Suite 110
Bloomington, IN, 47404
United States
Phone: 1 812 855 3123
Email: jagle@indiana.edu

Abstract

Background: Schools increasingly prioritize social-emotional competence and bullying and cyberbullying prevention, so the development of novel, low-cost, and high-yield programs addressing these topics is important. Further, rigorous assessment of interventions prior to widespread dissemination is crucial.

Objective: This study assesses the effectiveness and implementation fidelity of the ACT Out! Social Issue Theater program, a 1-hour psychodramatic intervention by professional actors; it also measures students’ receptiveness to the intervention.

Methods: This study is a 2-arm cluster randomized control trial with 1:1 allocation that randomized either to the ACT Out! intervention or control (treatment as usual) at the classroom level (n=76 classrooms in 12 schools across 5 counties in Indiana, comprised of 1571 students at pretest in fourth, seventh, and tenth grades). The primary outcomes were self-reported social-emotional competence, bullying perpetration, and bullying victimization; the secondary outcomes were receptiveness to the intervention, implementation fidelity (independent observer observation), and prespecified subanalyses of social-emotional competence for seventh- and tenth-grade students. All outcomes were collected at baseline and 2-week posttest, with planned 3-months postrtest data collection prevented due to the COVID-19 pandemic.

Results: Intervention fidelity was uniformly excellent (>96% adherence), and students were highly receptive to the program. However, trial results did not support the hypothesis that the intervention would increase participants’ social-emotional competence. The intervention’s impact on bullying was complicated to interpret and included some evidence of small interaction effects (reduced cyberbullying victimization and increased physical bullying perpetration). Additionally, pooled within-group reductions were also observed and discussed but were not appropriate for causal attribution.

Conclusions: This study found no superiority for a 1-hour ACT Out! intervention compared to treatment as usual for social-emotional competence or offline bullying, but some evidence of a small effect for cyberbullying. On the basis of these results and the within-group effects, as a next step, we encourage research into whether the ACT Out! intervention may engender
a bystander effect not amenable to randomization by classroom. Therefore, we recommend a larger trial of the ACT Out! intervention that focuses specifically on cyberbullying, measures bystander behavior, is randomized by school, and is controlled for extant bullying prevention efforts at each school.

**Trial Registration:** Clinicaltrials.gov NCT04097496; https://clinicaltrials.gov/ct2/show/NCT04097496

**International Registered Report Identifier (IRRID):** RR2-10.2196/17900

doi:10.2196/25860

**KEYWORDS**
cyberbullying; bullying; social-emotional learning; SEL; social-emotional competence; RCT; randomized controlled trial; outcome; emotion; bully; prevention; school; intervention; assessment; effectiveness; implementation; fidelity; reception; children; young adults; adolescents

**Introduction**

**ACT Out! Social Issue Theater**
The ACT Out! Ensemble was founded in 1995 and is currently operated by Claude McNeal Productions (CMP), a professional theater troupe incorporated as a not-for-profit [1]. The ensemble uses scripted content (scenarios) generated to meet an audience’s needs and transforms it into improvisational, interactive theater performances. The actors deliver performances focused on a variety of topics salient to youth and adolescents, including bullying, diversity, inclusion, and substance use [1]. A facilitator pauses the action at the end of each scenario and enables discussion between attendees and the actors, who remain in character for the discussion [1]. For example, a scenario about bullying may instruct the group to create a scene with “an example of a male student making a female student physically uncomfortable.” After the actors present an interpretation of that scene, the facilitator pauses the action and asks the audience “whether the male character meant to make the female character uncomfortable,” with the actors participating, in character, in the audience discussion.

The premise behind the ACT Out! Social Issue Theater intervention, outlined in the trial protocol, is that dramatic performances can enable emotional catharsis, thereby allowing new ways of feeling and thinking about behaviors and attitudes [2]. In other words, there is likely a difference between students discussing or attending a lecture on an issue like bullying in their own lives (where their own identity has weight and affects perceptions) and students’ emotional responses to a scenario that seems real and familiar (eg, bullying) but that is occurring with characters rather than with themselves or their peers. The latter case would theoretically enable students to process their reactions to bullying separately from their own or their peers’ identities. This is facilitated by the high caliber of talent involved in the intervention performances; shows by CMP have received positive reviews from, among other venues, the New York Times, NBC, and Time magazine [3]. For this trial, CMP developed and revised 15 vignettes (5 per participating grade level) addressing bullying and cyberbullying using principles of social-emotional learning (SEL) and reviewed the content with the research team. A presentation of the 5 psychodramatic vignettes was planned to last approximately 1 hour, including student interaction with the characters, and these performances constituted the intervention for this trial. This paper describes a rigorous evaluation, through a cluster randomized controlled trial, of the ACT Out! Ensemble’s theater performance, addressing bullying and SEL.

**Social and Emotional Learning**
In the United States and internationally, schools, school-based professionals, and policymakers have begun focusing on positive development models as a means of addressing the numerous, complex, and detrimental behavioral patterns and associated outcomes (eg, bullying, mental health problems, self-harm, and substance use) observed among youth and adolescents [4]. Such approaches deliberately avoid a deficit approach (“fixing what is wrong”) and emphasize the development of assets or protective factors in youth. Among the most common and conceptually similar positive development models are positive youth development [5], which emphasizes skills development, healthy relationship development, supportive community systems, and SEL. SEL focuses on the instruction of skills such as social problem-solving, recognizing emotions in others, and emotional self-regulation [6]. There have been numerous SEL programs implemented and evaluated in schools in recent years; a summary of over 300 studies contained in 4 meta-analyses identified generally positive short-term outcomes [7] across multiple domains (eg, substance use). In general, performances by the ACT Out! Ensemble are structured to model aspects of SEL, such as healthy relationships, regardless of the additional topic being addressed (eg, bullying). We have summarized additional content related to SEL and social-emotional competence (SEC) as it pertains to this study in our published protocol [2].

**Bullying and Victimization**
School bullying is frequently mentioned among the detrimental behaviors addressed by SEL programs [8,9]. Bullying is an unfortunate reality for youth attending US schools; a meta-analysis of 80 studies (youth aged 12 to 18 years) found a 35% student-level prevalence of traditional bullying and a 15% prevalence of cyberbullying [10]. Being bullied in childhood and adolescence has been associated with long-term, negative consequences that persist into midlife in areas such as mental and general physical health and lower socioeconomic status attainment [11]. Further, bullying victimization appears likely to cause notable increases in anxiety and depression among those victimized [12].

A recent meta-analysis of traditional bullying identified 65 school-based bullying prevention programs, but only 8 had been
evaluated more than once [13]. In general, such programs tend to be slightly more effective in reducing bullying perpetration and less effective in reducing victimization [14]. Mean values from meta-analyses have been somewhat consistent in terms of victimization, reporting reductions of 15-16% [14] and 17-20% [15]. A separate meta-analysis, focused only on cyberbullying, reported a mean reduction in victimization of 14-15% [16]. However, a meta-analysis of victimization studies limited to randomized controlled trials with a high level of rigor reported a rather small effect size (standard mean difference of -.09) [17]. Each of the cited meta-analyses note high levels of heterogeneity in the types of programs and outcomes across included programs and studies; importantly, these differences extended to study design and rigor, with some scholars noting that beneficial effects appear to be weaker when measured as part of randomized controlled trials [18].

As one might expect, reductions in victimization tend to be larger for more intensive and multifaceted programs, but implementing such programs can be both expensive and complicated [11,19,20]. For example, a summary of nontargeted (general population), relatively efficacious bullying prevention programs for US elementary schools found that program durations ranged from 11 weeks to 3 years [21]. Even efficacious programs described as “brief” prevention curricula can last 1 week or more and involve multiple interlocking components [22]. Further, several programs that have reported favorable results in efficacy trials have not always produced the same results in effectiveness (“real world”) trials, potentially due to issues with implementation fidelity and existing, confounding antibullying programming [23,24]. In concluding their report from a recent effectiveness trial, Rapee et al [24] noted, “clearly, producing a sizeable impact on school-based victimization is extremely difficult.” Therefore, there is a demonstrated need for inexpensive, simple-to-implement bullying prevention programming, but achieving positive outcomes from such interventions is likely to be especially challenging. For this reason, we believe that innovative or out-of-the-box strategies to address bullying merit serious consideration.

Psychodrama and Professional Acting as Innovation

Given the difficulty in addressing school-based bullying with lengthy and multipartite curricula, one might wonder why a short (1-hour) dramatic performance would be hypothesized to have even a short-term effect on SEC or bullying. A small body of literature has examined psychodrama as a prevention or behavior-change mechanism in youth, but these studies have covered diverse behaviors [25], have involved multiple, separate components such as teacher training [26], or have used students or school employees rather than professional actors as dramatis personae [27]. ACT Out! Social Issue Theater is different than each of these examples because it uses trained, professional actors and requires no involvement from schools outside of planning the visit (when implemented outside of a study). We were unable to find a precedent for this intervention structure in the literature.

Our decision to analyze this intervention was based on our a priori understanding of the value this brief intervention might yield as well as the remarkable community- and school-level support for the program. Prior to this study, more than 500,000 individuals had viewed a performance by the ACT Out! Ensemble [1], providing a notable depth of informal, qualitative evidence supporting the program. Uncontrolled evaluations of the program from 2015 also suggested substantive behavioral benefits [1]. Thus, given the importance of both SEL and bullying prevention in schools, and the unique position occupied by the ACT Out! Ensemble, we determined that an independently conducted, randomized controlled trial of this intervention was a valuable contribution to the prevention literature.

Study Objectives

This study primarily aims to assess whether a 1-hour exposure to ACT Out! Social Issue Theater is superior to treatment as usual for developing SEC and reducing bullying (both bullying behavior and victimization) in elementary, middle, and high school students at a 2-week posttest. Secondly, the study aims to determine whether the same intervention is superior to treatment as usual in developing specific subdomains of SEC (social awareness, emotion regulation, relationship skills, and responsible decision making) among middle and high school students at a 2-week posttest. Finally, the study also aims to assess student receptivity to ACT Out! Social Issue Theater using previously validated measures indicating student agreement with positive (eg, “enjoyable”) and negative (eg, “boring”) adjectives. Additional details are available in the trial protocol [2]. All outcomes were measured at the individual participant level, but randomization occurred at the cluster level (classroom) because performances are intended to be delivered to groups and because research literature [23,24] has indicated that pre-existing school-level programs addressing bullying and SEC often vary between schools and may contribute to statistical noise in randomized trials using school as the cluster (eg, treatment as usual may not be consistent between schools).

Methods

Trial Design

The ACT Out! trial was a proof-of-concept cluster randomized superiority trial with 2 groups and 1:1 allocation. The unit of measurement was individual students, but the unit of randomization was the classroom, stratified by school (with 1 exception, Multimedia Appendix 1). For each school, half of the classrooms were randomly assigned to the intervention arm and half to the control arm. Schools with an odd number of classrooms had a single classroom randomly selected for exclusion (though if the school requested, that classroom was permitted to complete the survey for appearance’s sake, and the results were then discarded by the study team).

Participants and Recruitment

The ACT Out! trial was conducted among 12 public and charter schools in Indiana: 4 schools in Marion County, 3 in Ripley County, 2 in Boone County, 2 in Lawrence County, and 1 in Monroe County. For reasons described in the protocol, clusters were selected only from grades 4, 7, and 10 [2]. All students in the selected classrooms and schools were eligible to participate. As planned, we recruited schools until meeting a threshold of
approximately 80 participating classrooms (around 1594 students) across both conditions.

Schools were selected based on their willingness to participate in the project as described, which included classroom-level randomization and inclusion of all eligible classrooms in the study’s allocation processes. Authorizing officials for schools or school corporations were required to provide a signed letter of agreement prior to participating in the study. At the individual level, the project used a waiver of parental consent (opt out), as approved by the institutional review board. Parents and legal guardians were permitted to review study procedures and were provided with a description of the study a minimum of 2 weeks prior to any individual-level interaction with subjects, along with instructions for how to opt out; students, their parents, and their guardians all had the ability to opt a student out from participating either formally or by survey noncompletion. The rationale for this approach was a combination of the low risk posed by the study as well as the desire to avoid unintentional exclusion of underrepresented minorities and high-risk populations, as described in the protocol [2]. This study and all consent procedures were carried out according to, and approved by, the Indiana University Institutional Review Board.

**Intervention**

**ACT Out! Social Issue Theater**

The intervention was a psychodramatic, improvisational performance that was delivered to classrooms (separately, except in 1 case where 2 small classrooms attended together) by members of the CMP professional theater company. Interventions were scheduled to last approximately 1 hour and were delivered during the school day. Each 1-hour performance consisted of 5 vignettes focused on bullying and cyberbullying and was designed to be interactive; after each scenario, the student audience was invited to converse with the performers, who remained in character. In each case, a moderator from CMP also managed the overall performance (eg, calling on students to ask questions of the characters). While the scenarios were improvisational in nature, they were designed to remain true to core concepts that were prespecified and agreed upon by CMP and the research team (eg, the identity of the characters, the nature of the conflict, and methods of bullying). To ensure this, fidelity data were captured from all performances (described in the Quality Control). The written specifications for each vignette, by grade level, are provided in Multimedia Appendix 2.

**Treatment as Usual (Control)**

Classrooms randomized to treatment as usual were provided with the preparation materials for the survey (see Data Collection) and completed the survey tools in the classroom during the school day (both at baseline and 2-week posttest). Students were not otherwise informed about the ACT Out! intervention by study personnel. Within schools, we were not aware of any systematic differences between intervention and control classrooms aside from the ACT Out! intervention itself, though schools themselves likely had different SEL and bullying programs at the school level (our statistical models incorporated random effects at both the classroom and school level).

**Outcomes**

All measured outcomes were prespecified in the clinical trial registration and the published protocol [2], along with the rationale for their selection, and have been validated. Unfortunately, certain outcomes were not possible to collect due to the COVID-19 pandemic, which led to mandated school closures in the state during part of the data collection period (Multimedia Appendix 1). However, we do not have reason to suspect that the data we collected prior to closures were substantively affected.

Thus, this study collected the following 2 primary outcomes: (1) social-emotional competence and (2) bullying behavior and bullying victimization. Overall social-emotional competence was measured at baseline and at 2-week posttest using the Delaware Social-Emotional Competency Scale (DSECS-S) [28]. This scale prompted students to “Please read each statement and mark the response that best shows how much it is like you,” with response options of 1=Not like me at all, 2=Not much like me, 3=Somewhat like me, and 4=Very much like me. An example statement from the scale is, “I can control how I behave.” This scale demonstrated good internal consistency [intention-to-treat (ITT) pretest α=.78] for the study sample. Bullying behavior and experiences of being bullied (victimization) were measured at baseline and at 2-week posttest using the Bullying and Cyberbullying Scale for Adolescents (BCS-A) [29]. These questions measured the number of times (between 0 and 4+) in the past 2 weeks that students “bullied another school student” or, separately, “had been bullied.” These sections were further separated into subsections for “online/on the internet or mobile phones” (eg, cyberbullying behaviors) and “offline/faceto-face” (including physical, verbal, and relational behaviors).

Further, the study collected the following 2 secondary outcomes: (1) receptiveness to the intervention and (2) prespecified subanalyses of social-emotional competence for seventh- and tenth-grade students. Student receptivity to the intervention was measured at 2-week posttest (intervention arm only) using questions to assess the degree to which they found the intervention to be enjoyable, interesting, a waste of time, boring, understandable, difficult to understand, believable, important, and helpful [30]. Social-emotional competence subdomains (social awareness, emotion regulation, relationship skills, and responsible decision-making) were measured at baseline and 2-week posttest (seventh and tenth grades only) using scales from the Washoe County School District Social-Emotional Competency Assessment (WCSD-SECA) [31]. These scales prompted students to “Please tell us how easy or difficult each of the following are for you,” with response options of 1=Very difficult, 2=Difficult, 3=Easy, and 4=Very easy. An example item from one scale is, “Getting along with my teachers.” Though the scales were previously developed to be reliable and valid [31], the items are relatively heterogeneous, perhaps contributing to their rather mediocre internal consistency with this sample [ITT pretest α=.57 (social awareness), .65 (emotion regulation), .68 (relationship skills), and .66 (responsible decision-making)].
Sample Size

The rationale for choices made in preparing the sample size calculation is provided in the protocol [2]. We estimated the sample size required to detect a moderate effect (Cohen $d=0.30$) with a 2-sided significance of .05 and a power of .80 to be 340 participants. We estimated an intraclass correlation of 0.153 based on a prior school-based cluster study on cigarette smoking with a similar methodology [2] and assumed approximately 20 students per classroom, yielding a design effect of 3.907. We took the resultant estimate of 1328 students and multiplied it by 1.2 to account for attrition and potential loss of matched pairs due to survey matching procedures, producing the final sample size target of 1594 students across approximately 80 classrooms.

Randomization and Allocation

Sequence Generation and Type of Randomization

Simple randomization occurred at the cluster level using a smartphone app produced by Random.org [32]. Randomization of clusters occurred within schools with a 1:1 allocation. In the specific instance where clusters first had to be created (a single school), the website version of Random.org was used to randomly assign students to evenly sized clusters and then to assign clusters to a study arm.

Concealment, Implementation, and Blinding

Since the generation of the allocation sequence was computerized, it was concealed to all members of the research team until the moment of assignment. Because of needs driven by school planning, there was some variability in the generation and assignment process. Decisions were made as follows: Schools that agreed to participate were asked to identify all clusters within the selected grade level (eg, fourth, seventh, or tenth). If schools were willing and able to accommodate it, allocation sequences were generated by the fidelity checker immediately prior to the intervention (eg, on-site, in the schools). However, most schools (10/12) were unable to accommodate this method. Subsequently, most schools were asked to list classrooms by a fixed characteristic (eg, the time the homeroom met, teacher’s name) at the time of school enrollment. One of the researchers generated a random sequence, applied it directly to classrooms, and shared the sequence with school administrators, identifying which classrooms would be allocated to which arm; the researcher asked the administrators not to share this information with teachers until necessary for planning efforts.

Consent was obtained from an administrative authority at participating schools at the time of enrollment and prior to randomization. Students and their parents or legal guardians were notified at least 2 weeks in advance of the intervention and provided an opportunity to opt out of the study but were not informed about their classroom’s allocation. Due to the nature of this study, blinding of participants, school officials, and researchers was not feasible. However, multiple independent statisticians were involved in conducting and reviewing analyses, and some were blinded to the meaning of study arm coding.

Data Collection

Survey Administration

Once a school enrolled in the study, each classroom was provided with a study packet containing surveys and response forms, a manila envelope, a white envelope, and an administrator checklist [2]. Each classroom was also assigned a unique code consisting of the grade level, study arm, and a randomly generated cluster ID. This code was prefilled on the back of each survey form and on the front of each envelope to facilitate data quality control.

Classroom teachers administered the surveys by following the step-by-step administrator checklist. Surveys that were handed out to students were placed back in the manila envelope, regardless of whether they were completed, while extra surveys were placed in the white envelope (unused). The pretest was completed 0-3 days prior to the intervention, depending on school schedules and availability. The posttest was completed 14-27 days after the intervention, with most classrooms completing the posttest within 14 to 17 days, depending on school schedules and the ability to facilitate the posttest.

Quality Control

Data were collected using a customized form created with Scantron DesignExpert (Scantron) and scanned directly into a database with an Insight 700c scanner (Scantron) to avoid data entry errors. However, one of the survey matching elements required a handwritten response; these were typed manually into the database. To verify intervention fidelity, at least one individual who was not a member of the ACT Out! Ensemble attended every performance and documented the concordance between a prespecified checklist of elements for the intervention and the performance itself. These checklists were developed separately for each grade (since the scenarios vary) and are available in Multimedia Appendices 3-5. To establish coding reliability, a second individual attended performances for 6 clusters to conduct fidelity checks, and interrater reliability was computed (Multimedia Appendix 1).

Survey Matching Procedures

This study required an anonymous procedure to match students’ surveys between pretest and posttest while remaining compliant with the requirements of the institutional review board. As described in the protocol [2], even recent meta-analyses had not identified a best-practice solution to such a dilemma [33]. Thus, the project team developed and used a novel anonymous matching procedure based on unique self-generated identification code elements and machine-assisted weighted matching. This approach was sufficiently complex that it required a separate full-length manuscript to articulate [34], and a complete description would extend well beyond the scope of this paper.

Analytic Methods

Missing Data

Multiple imputation (MI) using the Markov Chain Monte Carlo approach was completed using SAS (version 9.4; SAS Institute), utilizing PROC MI and MIANALYZE with the assumption that
data were missing at random. All variables that were collected were imputed for all analyses. Numbers of iterations were based on missingness in the per-protocol (PP) analysis, which had 1184 pretests and posttests, and thus 2368 surveys. Percent missingness ranged from a low of 1.22% to a high of 8.57%. SEC items not asked of fourth-grade students were present on 2078 surveys (removing 145 fourth-grade participants). Percent missingness within those variables ranged from 2.84% to 4.96%. Given this information, we selected 10 imputations for our analyses (integer greater than the missingness in the variable with the highest level of missingness, 10 > 8.57 [35]). Bias was also mitigated by presenting outcomes from 4 approaches.

**Statistical Analyses**

All outcomes were continuous, so linear mixed models using restricted maximum likelihood were fitted for each analysis (SAS PROC MIXED) with repeated measures for each participant. The time of survey administration (pretest or posttest), study arm, and time and study arm interaction were treated as fixed effects. The interaction of time by study arm was the hypothesis test for causal effects (eg, intervention group improved significantly more than the control group). All analyses allowed for clustering of students within schools and classrooms as random intercepts to alleviate the issue of inflated standard errors and used Kenward-Rogers degrees of freedom approximation to account for the cluster randomized trial design. P values were 2-sided and treated as significant at .05 or less; however, in keeping with recommendations from the American Statistical Society, we did not use P values as the sole determinant of outcome importance. Instead, we provided the full dataset and analytic code, and interpreted the output based on a combination of effect size, clinical significance, standard errors, and significance [36]. Similarly, we produced 4 sets of output, the ITT analysis in which all data were analyzed in the arm to which they were randomized (with and without MI; 1537 pretest and 1209 posttest), and the PP analysis in which only data resulting from a completed protocol were analyzed (with and without MI; 1184 pretest and posttest), in accordance with reporting recommendations [37] (Figure 1). Notably, the 931 cases that were matched do not represent the totality of students who complied with the protocol but rather are those who provided internally consistent information for the variables used to match anonymous surveys. The PP analyses included an additional 253 individuals (1184 in total) who were likely to have completed both surveys, even though their specific surveys could not be reliably matched between time points. The 931 individuals were included as repeated measures in the analysis, and the remaining 253 were included with surveys that were unmatched between time points.
ITT has been suggested to more closely represent effectiveness while PP represents efficacy [38]. Thus, superiority randomized controlled trials typically emphasize ITT analyses with imputation. In this case, however, the preponderance of excluded data resulted predominantly from an unexpected global event (COVID-19) rather than intervention nonadherence. Being able to attribute attrition directly to an external factor unrelated to the study is rare in controlled studies. As a result, it is less clear to us that ITT better reflects the true findings than PP in this specific instance. Thus, we interpreted ITT and PP analyses in tandem, providing all data from each analysis. We felt that this approach, while more complex in terms of preparing written text, facilitated transparency in explicating the findings.

**Results**

**Recruitment**

The formal study start date was October 16, 2019, and interventions were delivered from November 6, 2019, to February 28, 2020. School recruitment was terminated in February once the anticipated numbers of clusters and participants reached the planned total. Initially, a total of 13 classrooms was allocated to each arm, resulting in a total of 26 classrooms (13 intervention and 13 control) for data collection. However, due to unforeseen circumstances, such as the COVID-19 pandemic, several classrooms were unable to participate. Ultimately, 39 classrooms (774 students) were included in the intervention group and 37 classrooms (763 students) were included in the control group. The consort flow diagram illustrates the recruitment process and the exclusion of classrooms due to various reasons.

**Intention-to-Treat Analysis:** 1,537 pretests, 1,209 posttests, of which 931 pairs matched.

**Per Protocol Analysis:** 1,184 pretests and 1,184 posttests, of which 931 pairs matched.
schools from across Indiana participated in the trial, comprising 84 classrooms for the eligible grade levels. Of those 84 classrooms, 42 were randomized to the intervention arm and 42 were randomized to the control arm. One school did not follow protocol and failed to correctly administer pretest surveys to either arm prior to intervention delivery, so it (4 classrooms) was summarily removed from the trial. In addition, 3 schools provided incorrect counts of classrooms to be randomized (3 control classrooms and 1 intervention classroom), so sequences were generated that included classrooms that did not exist. Upon discovery of this discrepancy, sequences and assignments were not altered because it would have affected allocation concealment. Thus, despite 1:1 allocation, the number of baseline classrooms was 76 (37 control and 39 intervention classrooms). Finally, an additional 11 classrooms (5 control and 6 intervention classrooms) at a single school were slightly delayed in completing posttests, and then schools were shut down for the academic year due to COVID-19 prior to data collection. Thus, the number of classrooms that completed posttests was 65 (32 control and 33 intervention). Table 1 shows the baseline characteristics of the 2 trial arms for all students who provided data, excluding blank surveys.
Table 1. Baseline sample characteristics of the study participants (n=1537).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Control group (n=763)</th>
<th>Intervention group (n=774)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>386 (51.8)</td>
<td>404 (52.9)</td>
</tr>
<tr>
<td>Female</td>
<td>359 (48.2)</td>
<td>360 (47.1)</td>
</tr>
<tr>
<td>Missing</td>
<td>18 (—)</td>
<td>10 (—)</td>
</tr>
<tr>
<td><strong>Grade, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fourth</td>
<td>81 (10.6)</td>
<td>73 (9.4)</td>
</tr>
<tr>
<td>Seventh</td>
<td>307 (40.2)</td>
<td>293 (37.9)</td>
</tr>
<tr>
<td>Tenth</td>
<td>375 (49.2)</td>
<td>408 (52.7)</td>
</tr>
<tr>
<td><strong>Race, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>526 (72.2)</td>
<td>548 (73.1)</td>
</tr>
<tr>
<td>African-American or Black</td>
<td>67 (9.2)</td>
<td>77 (10.3)</td>
</tr>
<tr>
<td>Asian</td>
<td>14 (1.9)</td>
<td>11 (1.5)</td>
</tr>
<tr>
<td>Native American or Alaskan Native</td>
<td>10 (1.4)</td>
<td>7 (0.9)</td>
</tr>
<tr>
<td>Hawaiian or Pacific Islander</td>
<td>0 (0.0)</td>
<td>2 (0.3)</td>
</tr>
<tr>
<td>Multiracial</td>
<td>74 (10.2)</td>
<td>61 (8.1)</td>
</tr>
<tr>
<td>Other</td>
<td>38 (5.2)</td>
<td>44 (5.9)</td>
</tr>
<tr>
<td>Missing</td>
<td>34 (—)</td>
<td>24 (—)</td>
</tr>
<tr>
<td><strong>Hispanic/Latino, n (%)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>92 (12.7)</td>
<td>87 (11.8)</td>
</tr>
<tr>
<td>No</td>
<td>634 (87.3)</td>
<td>653 (88.2)</td>
</tr>
<tr>
<td>Missing</td>
<td>37 (—)</td>
<td>34 (—)</td>
</tr>
<tr>
<td><strong>Bullying victimization</strong>&lt;sup&gt;b&lt;/sup&gt;, mean (SD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Traditional (physical)</td>
<td>0.57 (0.88)</td>
<td>0.57 (0.93)</td>
</tr>
<tr>
<td>Traditional (verbal)</td>
<td>1.13 (1.49)</td>
<td>1.20 (1.55)</td>
</tr>
<tr>
<td>Traditional (relational)</td>
<td>0.78 (1.20)</td>
<td>0.82 (1.23)</td>
</tr>
<tr>
<td>Cyber</td>
<td>0.48 (0.83)</td>
<td>0.56 (0.93)</td>
</tr>
<tr>
<td><strong>Bullying perpetration</strong>&lt;sup&gt;b&lt;/sup&gt;, mean (SD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Traditional (physical)</td>
<td>0.25 (0.59)</td>
<td>0.24 (0.60)</td>
</tr>
<tr>
<td>Traditional (verbal)</td>
<td>0.56 (1.05)</td>
<td>0.61 (1.13)</td>
</tr>
<tr>
<td>Traditional (relational)</td>
<td>0.26 (0.74)</td>
<td>0.25 (0.71)</td>
</tr>
<tr>
<td>Cyber</td>
<td>0.25 (0.61)</td>
<td>0.24 (0.59)</td>
</tr>
<tr>
<td><strong>Social-emotional competence</strong>&lt;sup&gt;c&lt;/sup&gt;, mean (SD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-awareness&lt;sup&gt;d&lt;/sup&gt;, mean (SD)</td>
<td>2.94 (0.48)</td>
<td>2.90 (0.50)</td>
</tr>
<tr>
<td>Emotion regulation&lt;sup&gt;d&lt;/sup&gt;, mean (SD)</td>
<td>2.37 (0.61)</td>
<td>2.44 (0.62)</td>
</tr>
<tr>
<td>Relationship skills&lt;sup&gt;d&lt;/sup&gt;, mean (SD)</td>
<td>2.79 (0.52)</td>
<td>2.76 (0.55)</td>
</tr>
<tr>
<td>Responsible decision-making&lt;sup&gt;d&lt;/sup&gt;, mean (SD)</td>
<td>2.92 (0.54)</td>
<td>2.93 (0.55)</td>
</tr>
</tbody>
</table>

<sup>a</sup>—: not available.

<sup>b</sup>Count variable scored from 0 (no instances of any exemplars of bullying in the category) to 4 (4 or more instances of every exemplar of bullying in the category) [29].

<sup>c</sup>Scored from 1 to 4, where 4 is the optimal score [31].

<sup>d</sup>Seventh- and tenth-grade students only; scored from 0 to 4, where 4 is the optimal score [28].
Fidelity

Interrater reliability for fidelity coding was excellent (96.5% concordance). Fidelity to the intervention was high for all grades, though some seventh- and tenth-grade classrooms were not able to proceed through all 5 scenarios due to timing constraints. These instances were noted distinctly from fidelity because they resulted from longer-than-expected time completing the pretest survey immediately prior to the intervention, a factor that would not exist outside of the study. Fidelity was computed as the sum of all completed checkpoints for all clusters divided by the sum of all possible checkpoints for all clusters, within each grade. For the 5 fourth-grade classrooms, intervention fidelity was 100%. For the 15 seventh-grade classrooms, fidelity was 96.9%, but the fifth scenario was excluded for all clusters due to time constraints. For the 14 tenth-grade classrooms, fidelity was 98.4%, but 3 checkpoints for 1 classroom were excluded (fidelity checker was meeting with administrators), the third scenario was excluded for 2 clusters, the fourth scenario was excluded for 3 clusters, the fifth scenario was excluded for 3 clusters, and both the fourth and fifth scenarios were excluded for 1 cluster, all due to time constraints.

Outcomes

Sociodemographic Characteristics

All baseline data for the study are provided in Table 1, sorted by study arm. No significant sociodemographic differences were observed between the control and intervention arms based on chi-square tests with unadjusted alpha (.05) and pairwise exclusion of cases with missing sociodemographic values. The overall study sample was mostly male (790/1509, 52.4%), non-Hispanic/Latino (1287/1466, 87.8%), and White (1074/1479, 72.6%). More participants were in seventh (600/1537, 39.0%) and tenth (782/1537, 50.9%) grades than in fourth grade (154/1537, 10.0%).

Primary Objective 1

In our protocol [2], we hypothesized that ACT Out! Social Issue Theater was superior to treatment as usual for the development of overall SEC in students enrolled in elementary, middle, and high schools, measured approximately 2 weeks postintervention using the DSECS-S. The data did not support this hypothesis; no clinically or statistically significant interactions were observed (although this was not a “clinical” study, we use the term “clinical significance” to indicate findings where the magnitude, in our opinion, might reasonably be inferred to be of interest or value to potential stakeholders).

Primary Objective 2

In our protocol [2], we hypothesized that ACT Out! Social Issue Theater was superior to treatment as usual for reducing frequency (count) of perpetration of and victimization from traditional bullying (physical, verbal, and relational) and cyberbullying, measured approximately 2 weeks postintervention. These findings were complex. To interpret clinical significance, it is important to know that bullying scores are mean values based on count data of multiple bullying behaviors within a category, and so interpreting outcomes is different than for an attitudinal scale. For example, an individual who scored 1 for physical bullying victimization would need to have reported 1 instance of each of the 4 types of physical bullying victimization that compose that scale [2,29]. Similar interpretation applies to verbal (2 types), relational (2 types), and cyber (5 types) bullying. For example, a baseline bullying score of 0.568 (physical victimization) means that the average student reported experiencing more than 2 instances of physical bullying behaviors in the past 2 weeks (computed as .568*4). This should be taken into account when interpreting bullying outcomes.

There was limited causal, clinically significant evidence of small reductions (assessed via interaction effects of time by study arm) for cyberbullying victimization (favoring the intervention arm) and physical bullying victimization (favoring the control arm). For cyberbullying, reductions in the intervention arm victimization score ranged from -0.08 (P=.011, ITT with MI) to -0.13 (P<.001, PP without MI). This corresponded to mean cyberbullying victimization reductions of 0.40 to 0.65 instances/2 weeks, where the interaction term comparing the reduction in the intervention arm to the control arm was marginally significant in the PP analysis without MI (P=.067) but was increasingly nonsignificant in other models, ranging up to P=.301 for ITT with MI. For physical bullying, reductions in the control arm victimization score ranged from -0.13 (P<.001, ITT with MI; interaction P=.013) to -0.14 (P<.001, PP without MI; interaction P=.062). This corresponded to mean physical bullying victimization reductions of 0.52 to 0.56 instances/2 weeks. There was also limited evidence of a small effect of similar magnitude for increased physical bullying perpetration via the interaction effects. Increases in the intervention arm perpetration score ranged from 0.06 (P=.013, interaction P=.060; ITT with MI) to 0.08 (P=.005, interaction P=.032; ITT without MI), corresponding to increased perpetration of 0.24 to 0.32 instances/2 weeks, while the control arm did not have significant increases. However, only the ITT without MI model showed significance.

Finally, we observed an overall decrease in bullying victimization across pooled study participants (both arms). Findings were fairly uniform across models, so we provide only the most conservative (ITT with MI) outcomes in this paper. This included small-to-moderate, clinically significant overall reductions in physical (as above), verbal (control: -0.29, P<.001; intervention: -0.29, P<.001), and relational (control: -0.15, P=.001; intervention: -0.15, P=.003) bullying victimization. These corresponded to reductions of 0.58 (verbal) and 0.30 (relational) mean victimization instances/2 weeks. There was also some evidence of a small or moderate overall decrease in verbal bullying perpetration (control: -0.10, P=.023; intervention: -0.12, P=.005), corresponding to a reduction of 0.20-0.24 mean verbal bullying perpetration instances/2 weeks. These findings were not causal (ie, they were not observed differentially for the intervention clusters) and do not directly support the original hypothesis for this objective (except, potentially, for cyberbullying victimization); however, we believe that they do provide some favorable evidence for the program. Table 2 shows the ITT study outcomes, and Table 3 shows the PP study outcomes.
Table 2. Intention-to-treat study outcomes with standard errors.

<table>
<thead>
<tr>
<th>Variable (# observations)</th>
<th>Control</th>
<th>Intervention</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pretest, mean (SE), n=576</td>
<td>Posttest, mean (SE), n=614</td>
<td>Posttest, mean (SE), n=595</td>
</tr>
<tr>
<td></td>
<td>P value</td>
<td>P value of difference</td>
<td>P value</td>
</tr>
<tr>
<td>Bullying victimization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical (2623)</td>
<td>0.60 (.05)</td>
<td>0.46 (0.05) &lt;.001</td>
<td>0.58 (0.05)</td>
</tr>
<tr>
<td>Verbal (2585)</td>
<td>1.16 (.08)</td>
<td>0.85 (0.08) &lt;.001</td>
<td>1.24 (0.08)</td>
</tr>
<tr>
<td>Relational (2569)</td>
<td>0.82 (0.06)</td>
<td>0.65 (0.07) &lt;.001</td>
<td>0.83 (0.06)</td>
</tr>
<tr>
<td>Cyber (2622)</td>
<td>0.49 (0.05)</td>
<td>0.45 (0.05) .210</td>
<td>0.58 (0.05)</td>
</tr>
<tr>
<td>Bullying perpetration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical (2604)</td>
<td>0.28 (0.04)</td>
<td>0.28 (0.04) .860</td>
<td>0.26 (0.04)</td>
</tr>
<tr>
<td>Verbal (2576)</td>
<td>0.59 (0.05)</td>
<td>0.50 (0.06) .022</td>
<td>0.62 (0.05)</td>
</tr>
<tr>
<td>Relational (2570)</td>
<td>0.28 (0.04)</td>
<td>0.30 (0.04) .516</td>
<td>0.28 (0.04)</td>
</tr>
<tr>
<td>Cyber (2595)</td>
<td>0.27 (0.03)</td>
<td>0.28 (0.03) .623</td>
<td>0.26 (0.03)</td>
</tr>
<tr>
<td>Social-emotional competence (2699)</td>
<td>3.19 (0.02)</td>
<td>3.17 (0.03) .182</td>
<td>3.17 (0.02)</td>
</tr>
<tr>
<td>Social awareness (2384)</td>
<td>2.94 (0.02)</td>
<td>2.93 (0.03) .847</td>
<td>2.91 (0.02)</td>
</tr>
<tr>
<td>Emotion regulation (2374)</td>
<td>2.38 (0.03)</td>
<td>2.51 (0.03) &lt;.001</td>
<td>2.43 (0.03)</td>
</tr>
<tr>
<td>Relationship skills (2366)</td>
<td>2.78 (0.03)</td>
<td>2.80 (0.03) .315</td>
<td>2.75 (0.03)</td>
</tr>
<tr>
<td>Responsible decision-making (2355)</td>
<td>2.91 (0.03)</td>
<td>2.95 (0.03) .148</td>
<td>2.92 (0.03)</td>
</tr>
<tr>
<td>Multiple imputation analyses (2746)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bullying victimization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical</td>
<td>0.61 (.05)</td>
<td>0.48 (.05) &lt;.001</td>
<td>0.61 (.05)</td>
</tr>
<tr>
<td>Verbal</td>
<td>1.18 (.08)</td>
<td>0.89 (.08) &lt;.001</td>
<td>1.28 (.07)</td>
</tr>
<tr>
<td>Relational</td>
<td>0.84 (.06)</td>
<td>0.69 (.06) .001</td>
<td>0.89 (.06)</td>
</tr>
<tr>
<td>Cyber</td>
<td>0.51 (.05)</td>
<td>0.47 (.05) .270</td>
<td>0.60 (.05)</td>
</tr>
<tr>
<td>Bullying perpetration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical</td>
<td>0.30 (.03)</td>
<td>0.29 (.04) .928</td>
<td>0.30 (.03)</td>
</tr>
<tr>
<td>Verbal</td>
<td>0.61 (.05)</td>
<td>0.51 (.06) .233</td>
<td>0.67 (.05)</td>
</tr>
<tr>
<td>Relational</td>
<td>0.30 (.04)</td>
<td>0.33 (.04) .553</td>
<td>0.31 (.04)</td>
</tr>
<tr>
<td>Cyber</td>
<td>0.28 (.03)</td>
<td>0.29 (.03) .726</td>
<td>0.29 (.03)</td>
</tr>
<tr>
<td>Social-emotional competence</td>
<td>3.19 (.02)</td>
<td>3.17 (.03) .209</td>
<td>3.17 (.02)</td>
</tr>
<tr>
<td>Social awareness</td>
<td>2.94 (.02)</td>
<td>2.93 (.02) .732</td>
<td>2.91 (.02)</td>
</tr>
<tr>
<td>Emotion regulation</td>
<td>2.38 (.03)</td>
<td>2.52 (.03) &lt;.001</td>
<td>2.43 (.03)</td>
</tr>
</tbody>
</table>

Linear Mixed Models (no imputation)

Bullying victimization

Social-emotional competence (2699)

Multiple imputation analyses (2746)
<table>
<thead>
<tr>
<th>Variable (# observations)</th>
<th>Control</th>
<th></th>
<th>Intervention</th>
<th></th>
<th>Interaction</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pretest, mean (SE), n=763</td>
<td>Posttest, mean (SE), n=614</td>
<td>Posttest, mean (SE), n=595</td>
<td>Difference in differences (SE)</td>
<td>Pretest, mean (SE), n=774</td>
<td>Posttest, mean (SE), n=774</td>
</tr>
<tr>
<td>Relationship skills</td>
<td>2.78 (.03)</td>
<td>2.80 (.03)</td>
<td>.295</td>
<td>2.75 (.03)</td>
<td>2.78 (.03)</td>
<td>.182</td>
</tr>
<tr>
<td>Responsible decision-making</td>
<td>2.91 (.03)</td>
<td>2.95 (.03)</td>
<td>.142</td>
<td>2.91 (.03)</td>
<td>2.98 (.03)</td>
<td>.005</td>
</tr>
</tbody>
</table>
Table 3. Per-protocol study outcomes with standard errors.

<table>
<thead>
<tr>
<th>Variable (Observations)</th>
<th>Control</th>
<th>Intervention</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pretest, mean (SE), n=603</td>
<td>Posttest, mean (SE), n=603</td>
<td>P value of difference</td>
</tr>
<tr>
<td>Bullying victimization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical (2285)</td>
<td>0.61 (.05)</td>
<td>0.47 (.05)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Verbal (2252)</td>
<td>1.19 (.08)</td>
<td>0.87 (.08)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Relational (2239)</td>
<td>0.82 (.07)</td>
<td>0.65 (.07)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Cyber (2285)</td>
<td>0.47 (.05)</td>
<td>0.44 (.05)</td>
<td>.292</td>
</tr>
<tr>
<td>Bullying perpetration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical (2272)</td>
<td>0.27 (.04)</td>
<td>0.27 (.04)</td>
<td>.954</td>
</tr>
<tr>
<td>Verbal (2245)</td>
<td>0.57 (.06)</td>
<td>0.49 (.06)</td>
<td>.060</td>
</tr>
<tr>
<td>Relational (2244)</td>
<td>0.26 (.04)</td>
<td>0.29 (.04)</td>
<td>.314</td>
</tr>
<tr>
<td>Cyber (2263)</td>
<td>0.25 (.04)</td>
<td>0.27 (.04)</td>
<td>.388</td>
</tr>
<tr>
<td>Social-emotional competence (2336)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social awareness (2043)</td>
<td>2.94 (.03)</td>
<td>2.93 (.03)</td>
<td>.640</td>
</tr>
<tr>
<td>Emotion regulation (2034)</td>
<td>2.40 (.03)</td>
<td>2.52 (.03)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Relationship skills (2029)</td>
<td>2.77 (.03)</td>
<td>2.79 (.03)</td>
<td>.348</td>
</tr>
<tr>
<td>Responsible decision-making (2024)</td>
<td>2.91 (.04)</td>
<td>2.94 (.04)</td>
<td>.175</td>
</tr>
<tr>
<td>Multiple imputation analyses (2368)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bullying victimization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical</td>
<td>0.61 (.05)</td>
<td>0.48 (.05)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Verbal</td>
<td>1.20 (.08)</td>
<td>0.91 (.08)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Relational</td>
<td>0.83 (.07)</td>
<td>0.68 (.07)</td>
<td>.002</td>
</tr>
<tr>
<td>Cyber</td>
<td>0.48 (.05)</td>
<td>0.46 (.05)</td>
<td>.453</td>
</tr>
<tr>
<td>Bullying perpetration</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical</td>
<td>0.27 (.04)</td>
<td>0.28 (.04)</td>
<td>.856</td>
</tr>
<tr>
<td>Verbal</td>
<td>0.57 (.06)</td>
<td>0.49 (.06)</td>
<td>.072</td>
</tr>
<tr>
<td>Relational</td>
<td>0.27 (.04)</td>
<td>0.31 (.04)</td>
<td>.180</td>
</tr>
<tr>
<td>Cyber</td>
<td>0.26 (.03)</td>
<td>0.28 (.03)</td>
<td>.400</td>
</tr>
<tr>
<td>Social-emotional competence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social awareness</td>
<td>2.94 (.03)</td>
<td>2.93 (.03)</td>
<td>.609</td>
</tr>
<tr>
<td>Emotion regulation</td>
<td>2.40 (.03)</td>
<td>2.53 (.03)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Relationship skills</td>
<td>2.77 (.03)</td>
<td>2.79 (.03)</td>
<td>.354</td>
</tr>
<tr>
<td>Responsible decision-making</td>
<td>2.91 (.03)</td>
<td>2.94 (.03)</td>
<td>.171</td>
</tr>
</tbody>
</table>
Secondary Objective 1

Table 4 shows student perception of the ACT Out! intervention. Students were highly receptive to the ACT Out! performance. Merging affirmative responses (both “Yes” and “YES!” or, for negative questions, “No” and “NO!”), students found that the intervention was enjoyable (443/537, 82.5%), interesting (429/512, 83.8%), understandable (433/521, 83.1%), believable (433/517, 83.8%), important (417/506, 82.4%), and helpful (401/513, 78.2%). They also found that it was not a waste of time (424/518, 81.9%), boring (406/511, 79.5%), or difficult to understand (444/511, 86.9%).
Table 4. Perceptions of ACT Out! intervention. (Percentages may not add exactly to 100 due to rounding.)

<table>
<thead>
<tr>
<th>Variable (# observations)</th>
<th>Values, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enjoyable (537)</td>
<td></td>
</tr>
<tr>
<td>NO!</td>
<td>46 (8.6)</td>
</tr>
<tr>
<td>No</td>
<td>48 (8.9)</td>
</tr>
<tr>
<td>Yes</td>
<td>235 (43.8)</td>
</tr>
<tr>
<td>YES!</td>
<td>208 (38.7)</td>
</tr>
<tr>
<td>Interesting (512)</td>
<td></td>
</tr>
<tr>
<td>NO!</td>
<td>34 (6.6)</td>
</tr>
<tr>
<td>No</td>
<td>49 (9.6)</td>
</tr>
<tr>
<td>Yes</td>
<td>236 (46.1)</td>
</tr>
<tr>
<td>YES!</td>
<td>193 (37.7)</td>
</tr>
<tr>
<td>Waste of time (518)</td>
<td></td>
</tr>
<tr>
<td>NO!</td>
<td>230 (44.4)</td>
</tr>
<tr>
<td>No</td>
<td>194 (37.5)</td>
</tr>
<tr>
<td>Yes</td>
<td>56 (10.8)</td>
</tr>
<tr>
<td>YES!</td>
<td>38 (7.3)</td>
</tr>
<tr>
<td>Boring (511)</td>
<td></td>
</tr>
<tr>
<td>NO!</td>
<td>227 (44.4)</td>
</tr>
<tr>
<td>No</td>
<td>179 (35.0)</td>
</tr>
<tr>
<td>Yes</td>
<td>64 (12.5)</td>
</tr>
<tr>
<td>YES!</td>
<td>41 (8.0)</td>
</tr>
<tr>
<td>Understandable (521)</td>
<td></td>
</tr>
<tr>
<td>NO!</td>
<td>36 (6.9)</td>
</tr>
<tr>
<td>No</td>
<td>52 (10.0)</td>
</tr>
<tr>
<td>Yes</td>
<td>241 (46.3)</td>
</tr>
<tr>
<td>YES!</td>
<td>192 (36.9)</td>
</tr>
<tr>
<td>Difficult to understand (511)</td>
<td></td>
</tr>
<tr>
<td>NO!</td>
<td>245 (48.0)</td>
</tr>
<tr>
<td>No</td>
<td>199 (38.9)</td>
</tr>
<tr>
<td>Yes</td>
<td>43 (8.4)</td>
</tr>
<tr>
<td>YES!</td>
<td>24 (4.7)</td>
</tr>
<tr>
<td>Believable (517)</td>
<td></td>
</tr>
<tr>
<td>NO!</td>
<td>45 (8.7)</td>
</tr>
<tr>
<td>No</td>
<td>39 (7.5)</td>
</tr>
<tr>
<td>Yes</td>
<td>234 (45.3)</td>
</tr>
<tr>
<td>YES!</td>
<td>199 (38.5)</td>
</tr>
<tr>
<td>Important (506)</td>
<td></td>
</tr>
<tr>
<td>NO!</td>
<td>36 (7.1)</td>
</tr>
<tr>
<td>No</td>
<td>53 (10.5)</td>
</tr>
<tr>
<td>Yes</td>
<td>202 (39.9)</td>
</tr>
<tr>
<td>YES!</td>
<td>215 (42.5)</td>
</tr>
<tr>
<td>Helpful (513)</td>
<td></td>
</tr>
<tr>
<td>NO!</td>
<td>47 (9.2)</td>
</tr>
</tbody>
</table>
Secondary Objective 2

As planned, we conducted secondary analyses to see whether the intervention was superior to treatment as usual for specific subdomains of SEC measured for seventh- and tenth-grade students only (social awareness, emotion regulation, relationship skills, and responsible decision-making) using the WCSD-SECA. No notable or significant interaction effects were observed. However, all pooled students (across conditions) reported a small, statistically significant increase in emotion regulation. In the ITT model with MI, this was 0.14 (P<.001) for the control arm and 0.12 (P<.001) for the intervention arm. As above, such effects do not support causal attribution to the intervention.

Discussion

Brief Summary

The ACT Out! Social Issue Theater trial was a prespecified and preregistered cluster randomized controlled trial conducted by a research team external to the program developers and supported by additional analysts who both were independent of the research team and the developers. Study findings were mixed.

Implementation Fidelity

For school-based programs, implementation fidelity tends to be inconsistently documented and highly variable [39]. There is some evidence that achieving good fidelity may require intensive training of those delivering the intervention (eg, teachers, [40]) and may depend on school- and teacher-level variables [41]. Since the ACT Out! intervention was an improvisational and interactive intervention by actors, there was little precedent as to whether implementation fidelity would be achievable. We found that when provided with guidelines for core content elements, the professional actors in CMP were able to deliver nearly all (96.9% to 100%) prespecified content (Multimedia Appendices 3-5), even accounting for variance in student responses and different actors playing different roles, distinguishing this intervention from many school-based interventions. However, ACT Out! performances were short in duration and the actors had the delivery of the performance as their primary purpose; in contrast, teachers and schools must balance many different requirements simultaneously, involving the delivery of multiple sessions, and so high fidelity is conceptually reasonable to expect and may be a benefit of programs of this type.

Student Receptivity

The degree to which students report enjoying a program or finding it to be realistic and engaging can be interpreted as an indicator of program quality and is a common component of process evaluation [30]. The ACT Out! intervention was very positively received by students across all specified metrics. Although not an indicator of emotional competence, we infer that this may be interpreted as a partial measure of emotional response to the intervention.

Assessment of Effects on SEC

One of our primary hypotheses was that the ACT Out! intervention would improve students’ SEC in the short-term; this was suspected to be the mechanism through which the proposed emotional catharsis of psychodrama [2] could be measured. This study did not support that hypothesis, with analyses demonstrating neither statistical nor clinical significance, except improvement in emotion regulation regardless of treatment condition for seventh- and tenth-grade students.

During the project kickoff meeting, individuals at CMP expressed concern about quantitatively measuring SEC, at one point asking the research team, “How do you measure the sunrise?” We collected SEC data using 2 different tools that approached SEC in complementary ways [28,31]; however, optimal measurement of SEC remains a topic of debate, even among the national SEL workgroup [42]. Thus, we note (as with all such measurement) that the trial did not definitively find that the ACT Out! intervention had not engendered SEC development; rather, it found that SEC, as measured by the DSECS-S and WCSD-SECA (2 validated tools), was not affected by the intervention. It does not necessarily follow that those tools are the optimal or correct ways to measure students’ responses to the ACT Out! intervention (indeed, WCSD-SECA subscore reliability for this sample was suboptimal). It may also be the case that there was a partial ceiling effect [43] on overall SEC, as the baseline scores were relatively high: 3.202 and 3.176 for control and intervention groups, respectively, on a scale from 1-4, potentially making improvement from any source more difficult to achieve and measure. Given that all schools in the state already are required to offer SEL programming, this may also suggest that an intervention such as this would more appropriately be tested with subgroups of individuals who have lower baseline SEC (eg, not as a universal SEL program, but as an indicated program).

Assessment of Effects on Bullying

We hypothesized that the classrooms viewing the ACT Out! intervention would report reduced bullying victimization and perpetration [both traditional (physical, verbal, relational) bullying and cyberbullying] relative to the control classrooms. There was little evidence for an effect on perpetration, though 1 of the 4 models indicated the potential for slight increases for physical bullying in the intervention arm. There was also some evidence of a small reduction in cyberbullying victimization attributable to the intervention, though not for the traditional forms of bullying. That cyberbullying might be influenced...
separately from traditional bullying is reasonable, as cyberbullying victimization is unique in several ways, such as where it occurs (in a digital space “outside” of school) and its ubiquity [44]. Further, some of the dramatic scenarios (2 of 5 for both seventh and tenth grades) explicitly focused on cyberbullying as opposed to traditional bullying.

There was also evidence that physical bullying victimization was lower among the control group than the intervention group at posttest, though this was not an iatrogenic effect since neither group reported increased victimization scores. This finding may have been related to the issue we discuss subsequently.

**Interpretation of Pooled Victimization Effects**

Students reported clinically and statistically significant reductions in physical, verbal, and relational bullying victimization at posttest relative to pretest in aggregate (pooled across both study arms). This does not directly address the study hypotheses, as improvements were seen in students who did not participate in the intervention. However, this study examined students across 3 grade levels (fourth, seventh, and tenth), with interventions offered at different times over the course of nearly 6 months, in 12 schools and 65 classrooms, with a relatively short timeframe between pretest and posttest (14 to 27 days). Thus, it was implausible (though not impossible) that an external factor outside of the study was responsible for this finding.

We therefore considered effects that may have resulted from the study procedures. It was possible that participants’ responses at posttest were affected by completing the same items at pretest, either due to guessing and wanting to affect the study hypotheses or simple item-related biases introduced by familiarity with the questions [45]. However, some evidence suggests that questionnaire items do not affect student behavior [46]. Further, if this were the case, we would expect to have seen similar effects for cyberbullying victimization, bullying perpetration, and SEC, which we did not. We were also hesitant to ascribe these findings to regression to the mean [47]. The decrease was observed only for a specific subset of variables; as was already noted, bullying is common in school-based settings, so the likelihood that a large group of students from disparate settings and grades would significantly deviate from the population mean for bullying frequency is not conceptually strong. In addition, the impact of outlier cases on bullying frequency was minimized by the design, since the instance count in the questionnaire terminates at 4 (“4 or more times”). We also carefully avoided biases at the level of trial design by developing a protocol according to SPIRIT 2013 guidelines [48], preregistering the study, reporting even minor deviations, and attending to common sources of bias in clinical and prospective studies [49], though the possibility of unexplored confounding bias always must be considered [50]. Finally, 110 students in the control group indicated that they had seen a play or presentation by ACT Out! Ensemble before, though the degree of confounding influence, or the topic of prior performances, is unknown.

One such potential source of variance may have been the decision to randomize at the level of the classroom rather than at the level of the school. As has long been established, rigorous studies and evaluation of school-based programs is very methodologically difficult [51]. Prior research has indicated that randomizing at the school level can be problematic because schools often implement various SEL, SEC, and bullying programs, and so confounding variance can be introduced [23]. We were especially concerned about this when developing the protocol because SEL is written into expectations for Indiana schools, but not prescriptively (eg, schools can address it in different ways) [2]. By randomizing classrooms, we attempted to evade this problem by ensuring that a school’s other activities outside of this study were relatively equally represented among intervention and control clusters. However, it is important to consider that bullying does not occur within pre-existing clusters; that is, one is not limited to bullying or being bullied by students in the same homeroom period or English class. Thus, it is possible that an intervention affecting a bullying-related behavior, delivered to a random half of clusters within a grade and school, would have an effect on all clusters.

If this were the core mechanism explaining our data, then both bullying perpetration and victimization would theoretically be reduced; however, only victimization was reported to have been lowered. Thus, after careful consideration of the findings, including lack of SEC effect, high student receptivity and intervention fidelity, and significant time effects for traditional bullying victimization only, we hypothesize that the ACT Out! intervention may engender a heightened, defensive bystander response in participants. Bystander intervention in bullying has been associated with self-efficacy (belief that an intervention can be successful) [52], perceived knowledge about how one might successfully intervene [53], and the degree to which a bullying event is interpreted as serious [54]. Many of the scenarios for each grade level emphasized the roles of individuals other than the bully and the victim within the vignette and illustrated ways that others could intervene in a situation. They also identified potentially serious consequences that could emerge from bullying, including self-harm, so we infer surface-level plausibility of this explanation.

It is important to emphasize that while the data from this study were consistent with this explanation, the study itself does not provide causal evidence that the intervention engenders an increased likelihood for bystanders to intervene in bullying, nor was this an initial study hypothesis. Rather, we only know definitively that reported victimization decreased over time, in aggregate, among all participants. Additional research will be needed to determine the mechanism(s) by which this occurred. However, it is important to interpret and acknowledge all study findings to promote transparent research literature, and we have attempted to do so here.

**Limitations and Strengths**

This study was truncated unexpectedly by the COVID-19 pandemic, which had the effect of preventing planned 3-month outcome data collection and moderately affecting the number of clusters available for analysis for short-term outcomes (loss of 11 clusters). There were also several unplanned deviations from the study protocol, each of which has been documented in **Multimedia Appendix 1**. Participating schools were from both urban and rural counties in Indiana, and student participants were generally more diverse than the population of Indiana as
a whole. However, some caution should be used when generalizing these findings outside of the participating schools, especially since participating schools were those that volunteered to participate in a randomized trial. Further, given sample proportions, the results can be generalized more readily to middle and high school students than elementary school students. The study also had several notable strengths, including prespecification of all analyses; use of multiple objective, external consultants to the research team; thorough documentation of all protocol deviations; use of validated measures; and provision of all student documents and data in an open-source format.

Conclusions and Next Steps
This study found no superiority for a 1-hour ACT Out! intervention compared to treatment as usual for SEC or offline bullying, but some evidence of a small effect for cyberbullying. As was already indicated, SEL and bullying interventions in schools tend to be lengthy and involved, and bullying interventions in particular may struggle to demonstrate effectiveness in randomized trials [24]. Since ACT Out! is much shorter and highly scalable, we interpreted the findings in this study, though few and small in magnitude, with interest.

We suggest several next steps for research in this area. First, a rigorous follow-up study with a new sample would be valuable, which addresses issues related to the interpretability of bullying victimization data, including measures of possible bystander effects and randomization at the school level while, if feasible, controlling for ongoing and recent bullying prevention programs. In doing so, scenario emphasis might also be reasonably shifted toward cyberbullying and away from physical bullying, for which potential iatrogenic effects in perpetration were computed in one of the models, though overall physical victimization declined. Second, additional data might also be collected on the sustainability of the effects beyond 2 weeks as well as on whether there is a dose-response relationship (eg, “Would 2 performances within a semester more strongly reduce victimization?”). This could also be extended by collecting measures related more broadly to student mental health, academic performance, and perceived school climate. Finally, on the practical side, given the high intervention fidelity, high student receptiveness, and preliminary evidence related to cyberbullying victimization, it would not be unreasonable for CMP to offer a performance of scenarios focused on cyberbullying prevention to supplement, rather than replace, extant bullying prevention programming. In practice (eg, outside of a trial), this intervention has comparatively low fiscal cost, only 1 hour of time is utilized, it requires no teacher time or preparation, and it may have some benefits.

Additional Resources
To facilitate replication and transparent research processes, we have included supplemental files that may be valuable to researchers. These include a table of intracluster correlations (Multimedia Appendix 6), as well as the analysis syntax and datasets for per-protocol and intention-to-treat analyses (Multimedia Appendices 7-9, respectively).

Acknowledgments
Funding for this study was provided by Lilly Endowment Inc, grant no. 2019 0543, to Claude McNeal Productions. Funding was provided to Prevention Insights via a subaward from that grant. Claude McNeal Productions and their representatives own the rights to the ACT Out! Social Issue Theater program. No one from that organization was involved in preparing the study protocol, interpreting findings, conducting analyses, or writing this manuscript, both as a matter of practice and per written agreement in the subaward to Prevention Insights.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Deviations from protocol.
[DOCX File , 14 KB - mental_v8i1e25860_app1.docx ]

Multimedia Appendix 2
ACT Out! scenario guide.
[DOCX File , 24 KB - mental_v8i1e25860_app2.docx ]

Multimedia Appendix 3
Fourth-grade fidelity instrument.
[DOCX File , 16 KB - mental_v8i1e25860_app3.docx ]

Multimedia Appendix 4
Seventh-grade fidelity instrument.
[DOCX File , 16 KB - mental_v8i1e25860_app4.docx ]
Multimedia Appendix 5
Tenth-grade fidelity instrument.

Multimedia Appendix 6
Intracluster correlation table.

Multimedia Appendix 7
Analysis syntax.

Multimedia Appendix 8
Dataset for per-protocol analysis.

Multimedia Appendix 9
Dataset for intent-to-treat analysis.

Multimedia Appendix 10
CONSORT E-HEALTH checklist (V 1.6.1).

References
20. Serwacki M, Nickerson A, Schrantz M. Guide to School-Wide Bullying Prevention Programs. Alberti Center for Bullying Abuse Prevention: University at Buffalo 2017 [FREE Full text]


Abbreviations

BCS-A: Bullying and Cyberbullying Scale for Adolescents  
CMP: Claude McNeal Productions  
DSECS-S: Delaware Social-Emotional Competency Scale  
ITT: intention to treat  
MI: multiple imputation  
PP: per-protocol  
SEC: social-emotional competence  
SEL: social-emotional learning  
WCSD-SECA: Washoe County School District Social-Emotional Competency Assessment
Effects of ACT Out! Social Issue Theater on Social-Emotional Competence and Bullying in Youth and Adolescents: Cluster Randomized Controlled Trial

JMIR Ment Health 2021;8(1):e25860
URL: http://mental.jmir.org/2021/1/e25860/
doi:10.2196/25860
PMID:33338986

©Jon Agley, Mikyoung Jun, Lori Eldridge, Daniel L Agley, Yunyu Xiao, Steve Sussman, Lilian Golzarri-Arroyo, Stephanie L Dickinson, Wasantha Jayawardene, Ruth Gassman. Originally published in JMIR Mental Health (http://mental.jmir.org), 06.01.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Mental Health, is properly cited. The complete bibliographic information, a link to the original publication on http://mental.jmir.org/, as well as this copyright and license information must be included.
Psychiatric Profiles of eHealth Users Evaluated Using Data Mining Techniques: Cohort Study

Jorge Lopez-Castroman1,2,3,4*, MD, PhD; Diana Abad-Tortosa2*, PhD; Aurora Cobol Aguilera5,6, MS; Philippe Courtet1,4,7, MD, PhD; Maria Luisa Barrigón8,9, MD, PhD; Antonio Artés3,5,6, PhD; Enrique Baca-García2,3,8,9,10,11,12,13, MD, PhD

1Institute of Functional Genomics, CNRS-INSERM, Montpellier, France
2Department of Psychiatry, Nîmes University Hospital, Nîmes, France
3CIBERSAM, Madrid, Spain
4University of Montpellier, Montpellier, France
5Department of Signal Theory, Universidad Carlos III de Madrid, Madrid, Spain
6Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain
7Department of Psychiatric Emergency and Acute Care, Lapeyronie Hospital, University of Montpellier, Montpellier, France
8Universidad Autónoma de Madrid, Madrid, Spain
9Department of Psychiatry, Hospital Universitario Fundación Jiménez Díaz, Madrid, Spain
10Department of Psychiatry, University Hospital Villalba, Villalba, Madrid, Spain
11Department of Psychiatry, University Hospital Infanta Elena, Valdemoros, Madrid, Spain
12Department of Psychiatry, University Hospital Rey Juan Carlos, Mostoles, Madrid, Spain
13Universidad Católica del Maule, Talca, Chile

*these authors contributed equally

Corresponding Author:
Jorge Lopez-Castroman, MD, PhD
Institute of Functional Genomics, CNRS-INSERM
39 Ave Charles Flauhault
Montpellier, 34093
France
Phone: 33 4996145
Email: jorgecastroman@gmail.com

Abstract

Background: New technologies are changing access to medical records and the relationship between physicians and patients. Professionals can now use e-mental health tools to provide prompt and personalized responses to patients with mental illness. However, there is a lack of knowledge about the digital phenotypes of patients who use e-mental health apps.

Objective: This study aimed to reveal the profiles of users of a mental health app through machine learning techniques.

Methods: We applied a nonparametric model, the Sparse Poisson Factorization Model, to discover latent features in the response patterns of 2254 psychiatric outpatients to a short self-assessment on general health. The assessment was completed through a mental health app after the first login.

Results: The results showed the following four different profiles of patients: (1) all patients had feelings of worthlessness, aggressiveness, and suicidal ideas; (2) one in four reported low energy and difficulties to cope with problems; (3) less than a quarter described depressive symptoms with extremely high scores in suicidal thoughts and aggressiveness; and (4) a small number, possibly with the most severe conditions, reported a combination of all these features.

Conclusions: User profiles did not overlap with clinician-made diagnoses. Since each profile seems to be associated with a different level of severity, the profiles could be useful for the prediction of behavioral risks among users of e-mental health apps.

(JMIR Ment Health 2021;8(1):e17116) doi:10.2196/17116

KEYWORDS
mental disorders; suicide prevention; suicidal ideation; data mining; digital phenotyping
Introduction

The development of new technologies shows promise for causing a revolution in the way chronic diseases are followed and treated [1]. In the past few decades, we have seen the following two major technological changes directly connected to the availability of medical information: (1) introduction of electronic health records in most health care facilities, and (2) accessibility to portable devices capable of acquiring information about their users. Both systems are already being used to enhance communication between health providers and final users and to improve the overall performance of health care. Indeed, public and private entities are massively investing in the development of web-based platforms or smartphone apps through which patients can organize their medical agenda, have access to all or part of their medical records, provide their input, and join their medical referents [2].

It seems reasonable to believe that the follow-up of persons with mental illness will be improved if eHealth systems lead to an increased interaction with health care providers. Our group and others have shown that electronic assessment is feasible with proper adaptation [3]. Efficient monitoring may prompt health responses in cases of emergency [4], inform accurately about real-life behaviors between medical appointments, reduce unnecessary visits, and sustain therapeutic decisions [5]. Other parameters, such as biomarkers and input from close relatives, can be added to the monitoring system. This kind of ecosystem already exists [6], and a growing body of evidence has shown that eHealth tools improve treatment outcomes in terms of engagement, symptom improvement, well-being, and self-care [7-13]. Their combination with machine learning techniques has also shown positive impacts on the diagnosis, prediction, and prevention of several diseases, such as cancer [14-16].

Despite these advances, the majority of potential users, such as elderly persons with low educational levels [17], seem to be unenthusiastic about e-mental health tools [18]. Utilization rates have been associated with the characteristics of health professionals [19], but there is little knowledge about the kinds of patients who use e-mental health, how they become users, and what are their patterns of use. Young age, high education, and dissatisfaction with the health care system might be common features among eHealth users [20]. We do not know if this profile also applies to patients with mental disorders, but their digital phenotype [21] is likely to contain valuable information for clinicians and providers alike [22,23]. Alterations in the patterns of use could help clinicians to detect pathological or risky behaviors and individual needs, and increase treatment efficiency.

In this article, a nonparametric latent feature model based on the Indian Buffet Process (IBP) explores the response patterns of 2254 psychiatric outpatients to a web-based questionnaire. We aimed to define specific profiles according to response patterns and then link profiles with psychiatric disorders. The study will thus specifically describe patients with psychiatric diagnoses who have used an eHealth application at least once. We hypothesize that data mining techniques, such as IBP, can be used to associate information from different questionnaires and assessments in a plausible model that could serve ultimately to plan health care delivery.

Methods

Sample

Participants were recruited from psychiatric outpatient facilities in the catchment area of Fundacion Jimenez Diaz, a University Hospital in Madrid, Spain. This hospital is part of the National Health Service and provides medical coverage to about 850,000 people. From May 2014 onwards, all clinicians working at the six mental health centers of the catchment area received specific training and were encouraged to use the MEmind Wellness Tracker systematically in their clinical activity. A total of 2254 patients signed up on the MEmind platform and completed the assessment, and they were subsequently included in the study. The assessment comprised the collection of information about sociodemographic features and diagnoses. Participants also filled up a short questionnaire. For this study, we used broad inclusion criteria. Every patient attending psychiatric consultations independent of diagnosis was considered. Thus, all clinicians in the catchment area were instructed to propose the use of the web application to every outpatient they saw with no restriction whatsoever regarding their diagnoses or their clinical statuses. The total number of outpatients who consulted during the study period was 30,808.

For the purpose of this study, we included only participants who voluntarily accessed the application and responded to an open-text field. We made this choice to select proactive participants who completed most of the questions at the user end. We noted a missing data rate of 12%, which resulted from the sum of clinical missing data and a lack of completeness of the questionnaires at the user end of the application.

This study was performed in agreement with the ethic requirements of the Declaration of Helsinki (World Medical Association, 2013) and was approved by the Institutional Review Board of the University Hospital Fundación Jiménez Díaz (Madrid, Spain). All participants provided written informed consent to participate in the study.

Assessment

Questionnaires

The data set consists of 23 questions from the following three different questionnaires: (1) a brief day assessment related to sleep quality, appetite, medication intake, aggressiveness, and suicidal behavior (six items); (2) the Who-5 Well-Being Index [24] (five items); and (3) the ninth version of the General Health Questionnaire [25] (12 items). All these questionnaires are short self-reported measures of current mental well-being. All items are yes-or-no questions, followed by the degree of agreement reported on a Likert scale (0 to 100 points). Although participants could repeat the assessment, only data from the first report were included in the model.

Clinical Diagnoses

Diagnostic coding was based on the International Statistical Classification of Diseases and Related Health Problems (ICD-10) [26]. Thus, diagnoses of mental disorders were
classified into 10 groups (F0 to F9) according to ICD-10 (Multimedia Appendix 1). The corresponding physician coded the diagnosis for each patient and completed the clinical global impression (CGI) scale [27], which reflects the global functioning of a patient according to the view of the clinician on a scale (0 to 7 points). The CGI scale provides a summary measure accounting for patient history, psychosocial factors, behavior, and the impact of symptoms on the patient’s ability to function (Multimedia Appendix 2).

**Data Processing**

First, the scores for the items with a positive valence in the questionnaire data set (items 1 to 15) were inversed. In this way, a higher score for any item of the questionnaire indicated poorer mental health. Second, we dichotomized every item score using a specific threshold in order to code the top 10% scores with the value “1” and the remaining 90% with the value “0” (Multimedia Appendix 3). The use of a centesimal scale increases the sensitivity of the questionnaire. Responders tend to avoid extreme values unless they identify completely with them [28], but extreme responders do not seem to be affected by the length of the response scale [29]. By using the highest scores, we made sure that only the extreme responders were separated. The histograms of dichotomized scores for the items are shown in Figure 1.

![Histograms of scores provided by eHealth users to each of the 23 questions of the mental well-being questionnaire.](image)

**Figure 1.** Histograms of scores provided by eHealth users to each of the 23 questions of the mental well-being questionnaire. Scores range from 0 to 100. The last histogram presents the average score for all questions.

The clinical records of the participants provided a second source of data. These records included sex, age, clinical diagnoses, and CGI values. CGI values presented missing data, so subsequent analyses including this variable were carried out with a total sample of 2000 participants. For analyses involving clinical diagnoses, the total sample was 1787 participants. All patients with missing data were excluded from this part of the data modeling. Comorbid diagnoses were also examined when present.

**Data Modeling**

We applied the **Sparse Poisson Factorization Model (SPFM)** to model the data. The SPFM is based on the IBP [30], a nonparametric probabilistic method that proposes a sparse analysis of the variables. The input data for the model must be binary or categorical. The SPFM decomposes the input matrix into the following two nonnegative and disperse matrixes: matrix Z and matrix β. Latent factor sets can be calculated from them. The binary Z matrix represents the number of active factor sets for each patient (Multimedia Appendix 4). The β matrix weights the contribution of each factor set to each item of the questionnaire. Each factor set is characterized by precise values on the 23 questions. A higher weight (β) of a factor set for an item is associated with a greater probability to find a high score in that item when that particular factor set is active. The SPFM also estimates a **bias term**, a factor set that is present in all the patients of the sample [31]. The bias term is the “default” situation of an eHealth user and represents a profile shared by all patients that is independent of any additional feature. In that sense, the bias term allows the machine learning function to be shifted to better fit the data in a similar way as done by the y-intercept in an equation.

Different profiles have been obtained by applying the basic clustering method K-means [32]. The procedure classifies a given data set through a certain number of clusters fixed a priori. This method, applied on the Z matrix, associates data with similar characteristics into different clusters by using centroids. Thus, it allows clustering patients who show similar activation of their factor sets (Figure 2).
**Results**

**Sample Description**

The sample involved 2254 patients, including 1184 (52.53%) women, 795 (35.27%) men, and 275 (12.20%) patients with missing data on sex. The mean age was 52.0 years (SD 15.1). Medical reports about the patients showed a CGI mean score of 2.95 (SD 1.95) (Multimedia Appendix 5) with a high percentage of participants scoring 3 (mildly ill; 844/2000, 42.20%) or 4 (moderately ill; 632/2000, 31.60%) and only a few scoring 6 or 7 (severely ill or extremely ill; 10/2000, 0.50%). According to the ICD-10 criteria, participants with mood disorders (F3; 347/1787, 19.43%), stress-related disorders (F4; 962/1787, 53.82%), and adult personality disorders (F6; 178/1787, 9.96%) represented most of the sample (Multimedia Appendix 6).

**Data Classification**

The SPFM latent model analysis found the following three components in the assessment: one bias term and two factor sets. Both the bias term and factor sets involved groups of items of the questionnaire that are particularly informative. The bias term is present for all patients and reflects a common behavioral pattern. On the other hand, factor sets 1 and 2 are based on subsets of answers with high informational value to discriminate patients. Factor sets 1 and 2 can be present or absent for a particular patient. They are present if the corresponding subset of responses has a high score (value “1”: highest 10% scores) and are absent if the corresponding subset of responses has a low score (value “0”: remaining 90% of the scores).

The K-means algorithm applied on the Z matrix established four different patient profiles according to the presence of none, one, or both factor sets. Profile 0 presents only the bias term, profile 1 presents the bias term plus factor set 1, profile 2 presents the bias term plus factor set 2, and profile 3 presents the bias term and each factor set. The number of patients in each profile is shown in Table 1. In Table 2 and Multimedia Appendix 7, we can appreciate how each profile responds differently to the questionnaire.

**Table 1. Number of patients in each profile.**

<table>
<thead>
<tr>
<th>Profile</th>
<th>Number of patients</th>
<th>Factor set</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1113</td>
<td>Bias term (0)</td>
</tr>
<tr>
<td>1</td>
<td>480</td>
<td>0+1</td>
</tr>
<tr>
<td>2</td>
<td>616</td>
<td>0+2</td>
</tr>
<tr>
<td>3</td>
<td>45</td>
<td>0+1+2</td>
</tr>
</tbody>
</table>
Table 2. Average score for each item in the self-reported questionnaire of current mental well-being according to the β matrix of factor sets.

<table>
<thead>
<tr>
<th>Questions</th>
<th>Profile 0</th>
<th>Profile 1</th>
<th>Profile 2</th>
<th>Profile 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. How many hours did you sleep today? (from 0 to 12)(^a)</td>
<td>0.0689</td>
<td>0.1525</td>
<td>0.0041</td>
<td>0.1567</td>
</tr>
<tr>
<td>2. Quality of sleep(^a)</td>
<td>0.0390</td>
<td>0.2089</td>
<td>0.0040</td>
<td>0.2129</td>
</tr>
<tr>
<td>3. Do you have appetite?(^a)</td>
<td>0.0446</td>
<td>0.1980</td>
<td>0.0009</td>
<td>0.1990</td>
</tr>
<tr>
<td>4. Do you take your medication?(^a)</td>
<td>0.0795</td>
<td>0.0566</td>
<td>0.0225</td>
<td>0.0791</td>
</tr>
<tr>
<td>5. I felt joyful and with good mood(^a)</td>
<td>0.0133</td>
<td>0.3228</td>
<td>0.0006</td>
<td>0.3235</td>
</tr>
<tr>
<td>6. I felt peaceful and relaxed(^a)</td>
<td>0.0223</td>
<td>0.3181</td>
<td>0.0012</td>
<td>0.3193</td>
</tr>
<tr>
<td>7. I felt active and robust(^a)</td>
<td>0.0130</td>
<td>0.3811</td>
<td>0.0103</td>
<td>0.3915</td>
</tr>
<tr>
<td>8. I felt awake, fresh, and rested(^a)</td>
<td>0.0496</td>
<td>0.3894</td>
<td>0.0041</td>
<td>0.3936</td>
</tr>
<tr>
<td>9. My daily life has many interesting things(^a)</td>
<td>0.0199</td>
<td>0.4236</td>
<td>0.0000</td>
<td>0.4236</td>
</tr>
<tr>
<td>10. Have you been able to keep focus on the tasks you did?(^a)</td>
<td>0.0276</td>
<td>0.3467</td>
<td>0.0013</td>
<td>0.3481</td>
</tr>
<tr>
<td>11. Have you felt that you have a useful role in life?(^a)</td>
<td>0.0187</td>
<td>0.3756</td>
<td>0.0018</td>
<td>0.3774</td>
</tr>
<tr>
<td>12. Have you felt able to make decisions?(^a)</td>
<td>0.0173</td>
<td>0.3529</td>
<td>0.0002</td>
<td>0.3531</td>
</tr>
<tr>
<td>13. Have you enjoyed regular activities from daily life?(^a)</td>
<td>0.0100</td>
<td>0.3739</td>
<td>0.0016</td>
<td>0.3756</td>
</tr>
<tr>
<td>14. Have you felt able to cope with your issues?(^a)</td>
<td>0.0149</td>
<td>0.3674</td>
<td>0.0016</td>
<td>0.3691</td>
</tr>
<tr>
<td>15. Do you feel reasonably happy taking into account the circumstances?(^a)</td>
<td>0.0174</td>
<td>0.3591</td>
<td>0.0030</td>
<td>0.3622</td>
</tr>
<tr>
<td>16. Do you feel aggressiveness?</td>
<td>0.1745</td>
<td>0.0053</td>
<td>0.2362</td>
<td>0.2415</td>
</tr>
<tr>
<td>17. Do you have suicidal thoughts?</td>
<td>0.3177</td>
<td>0.0012</td>
<td>0.4573</td>
<td>0.4585</td>
</tr>
<tr>
<td>18. Have you had worries interfering with your sleep?</td>
<td>0.0422</td>
<td>0.0210</td>
<td>0.2931</td>
<td>0.3142</td>
</tr>
<tr>
<td>19. Have you felt constantly overwhelmed or tense?</td>
<td>0.0109</td>
<td>0.0006</td>
<td>0.3108</td>
<td>0.3115</td>
</tr>
<tr>
<td>20. Have you felt unable to overcome your troubles?</td>
<td>0.0060</td>
<td>0.0015</td>
<td>0.3005</td>
<td>0.3021</td>
</tr>
<tr>
<td>21. Have you felt unhappy or depressed?</td>
<td>0.0001</td>
<td>0.0037</td>
<td>0.4213</td>
<td>0.4251</td>
</tr>
<tr>
<td>22. Have you lost self-confidence?</td>
<td>0.0083</td>
<td>0.0071</td>
<td>0.4784</td>
<td>0.4855</td>
</tr>
<tr>
<td>23. Have you felt worthlessness?</td>
<td>0.1097</td>
<td>0.0010</td>
<td>0.6198</td>
<td>0.6208</td>
</tr>
</tbody>
</table>

\(^a\)The scores from these items were inversed during data processing.

### Patient Profiles

The bias term was associated with high scores in suicide thoughts and aggressiveness (items 5 and 6), as well as feelings of worthlessness (item 23). All patients in our sample shared the features of the bias term, but about half of them (n=1141) also presented one or two different factor sets. Those presenting factor set 1 were included in profile 1, which was characterized by the absence of positive mood, low sleep quality, low energy, and feelings of loss of control (items 1-3 and 5-15). Patients presenting factor set 2 were included in profile 2, which was characterized by intense suicidal thoughts, aggressiveness, intense feelings of depression and worthlessness, low self-confidence, and worries interfering with sleep (items 16-23). All these characteristics were simultaneously active in patients with profile 3, who presented simultaneously with both factors. No statistical differences were found between the profiles regarding the distribution of age or sex (\(F_3=1.391, P=.24\) and \(X^2_3=0.56, P=.90\)).

### Association Between Patient Profile and Medical Evaluation

After modeling the data, we compared CGI scores and clinical diagnoses between profiles. The results showed that the CGI scores were higher than the mean in profile 1 (3.2, SD 1.27), with the largest percentage of participants evaluated with a score of 4 (192/453, 42.4%). For profile 2, the CGI scores were lower than the mean (2.78, SD 1.28), with a high percentage of participants evaluated with a score of 3 (256/557, 45.9%). Results in profile 3 were not compared given the low number of patients.

Most diagnoses fell within the F4, F3, and F6 categories in each profile and in the total sample, corresponding with affective disorders, neurotic and stress-related disorders, and disorders...
of adult personality and behavior. The distribution of participants with profile 1 was similar for all the types of diagnoses. However, profile 2 seemed to be more frequent among patients with diagnoses of schizophrenia and psychological, behavioral, and emotional disorders with onset in childhood/adolescence (F2: 43/90, 48%; F8: 3/5, 60%; and F9: 14/38, 37%; Table 3).

### Table 3. Distribution of patient profiles according to the main ICD-10 diagnostic categories for psychiatric disorders (F0-F9).

<table>
<thead>
<tr>
<th>Category</th>
<th>Profile, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
</tr>
<tr>
<td>F0</td>
<td>7 (50.0)</td>
</tr>
<tr>
<td>F1</td>
<td>26 (45.6)</td>
</tr>
<tr>
<td>F2</td>
<td>37 (41.1)</td>
</tr>
<tr>
<td>F3</td>
<td>215 (49.6)</td>
</tr>
<tr>
<td>F4</td>
<td>614 (51.2)</td>
</tr>
<tr>
<td>F5</td>
<td>51 (48.1)</td>
</tr>
<tr>
<td>F6</td>
<td>109 (49.1)</td>
</tr>
<tr>
<td>F7</td>
<td>6 (66.6)</td>
</tr>
<tr>
<td>F8</td>
<td>2 (40.0)</td>
</tr>
<tr>
<td>F9</td>
<td>20 (52.6)</td>
</tr>
</tbody>
</table>

### Discussion

The data modeling approach we applied was able to discriminate four different profiles of patients based on the answers to a brief electronic questionnaire. All profiles shared a component associated with feelings of aggressiveness, worthlessness, and suicidal thoughts (bias term or profile 0), which seemed to be common among patients who used e-mental health tools [33], such as the MEmind Wellness Tracker.

Sex and age distributions showed very little variability across the profiles, facilitating comparisons between them. In addition to profile 0 (bias term, default pattern), three profiles where found based on the scores of different sets of questions. It is important to bear in mind that a factor set was classified as active only when the scores were in the top 10% of the corresponding items. For example, even if the item of low sleep quality is absent from profile 2, a patient with that profile could still have high scores in that item compared with the general population and thus have relatively low sleep quality.

Patients in profile 1 reported a lack of positive mood, low quality of sleep, low energy, feelings of loss of control, and difficulties to face problems. These symptoms could be reactive to life difficulties and partly due to a lack of coping skills. Patients in profile 2 presented high scores in depressive feelings, worries interfering with their sleep, feelings of being overwhelmed and unable to overcome troubles, low self-confidence, and feelings of worthlessness. This pattern seems to be related with a greater inward focus and depressive-like symptomatology. Interestingly, patients in profile 2 also reported the highest scores for suicidal thoughts and feelings of aggressiveness. Indeed, patients in profile 2 reported five of the 10 ICD-10 diagnostic criteria for a depressive episode, including disturbed sleep, depressive feelings, reduced self-confidence, ideas of worthlessness, and ideas of suicide [26]. Surprisingly, those in profile 2 were evaluated by their physicians as having a higher level of functionality (CGI) than those in profile 1, despite higher levels of suicidal thoughts, aggressiveness, and depression in profile 2. This points to discordances between the medical assessment and the self-reported momentary assessment. Finally, profile 3 involved a small group of patients with high scores in all the items of the questionnaire. They shared the features of profile 1 and profile 2, and reported the most severely affected psychological state in our sample (the highest levels of distress).

Our study suggests that the analysis of data from electronic self-assessments can discriminate profiles or clusters of patients sharing similar clinical characteristics. These features do not seem to overlap with usual clinical diagnoses, since no differences were found in the prevalence of previous psychiatric diagnoses between profiles. Most patients in each profile received diagnoses in F4 (anxiety disorders) and F3 (mood disorders) ICD-10 categories, which were numerically the most common diagnoses in the sample. However, diagnoses of disorders with an onset during childhood and adolescence (eg, F8 and F9) and schizophrenia (F2) were overrepresented in profile 2. Profile 3 was particularly overrepresented among the small group of patients with organic mental disorders (F0) in the sample, which could implicate a more complex disease course. Interestingly, in a previous paper, we found that the assessments made by clinicians did not correlate well with patients’ self-reports within 24 h of a clinical evaluation [34].

The presence of sporadic suicide thoughts can be relatively frequent in psychiatric patients, but eHealth apps could help to identify profiles with higher suicide risk, such as profile 2. Previous literature has suggested improvements in mood, well-being, anxiety, and self-awareness, as well as a higher adherence to treatments among users of eHealth apps [5,35-37]. Electronic assessment tools, such as the one used in our study, may support physicians to discriminate patients with high suicide risk in order to adjust their interventions.

Among the limitations of our study, we note the use of only baseline assessments and incomplete clinical information. The
described profiles might not be reflective of eHealth users who continue to use the app regularly. Besides, our intention was not to map the participants onto Diagnostic and Statistical Manual of Mental Disorders (DSM) or ICD categories but rather to identify symptomatic profiles that are not necessarily reflected in psychiatric diagnoses. This study was designed to explore the utility of a new method to classify e-mental health users, and it needs to be completed with follow-up data. Nonetheless, once the SPFM is trained, it will be possible to analyze changes in patient profiles during continuous assessment with several time points. It will also be possible to link the electronic assessment with medical records. Our results could help to select the most performing questions according to mental disorders or patient profiles, which, in turn, could be used to create shorter and more efficient questionnaires. We can see in our study that the question about medication intake had very low informative value.

There are still many concerns regarding e-mental health that need to be addressed. One of the main concerns reported by both professionals and users is related to the privacy, ownership, and responsible use of medical information [38,39]. This is one of the major challenges that eHealth needs to address by means of privacy-preserving technologies [40]. Accessibility and difficulties to find reliable sources of medical information are also important concerns in the population, especially among older adults [41]. Medical professionals also have doubts about the capacity of online information to improve the knowledge of patients and have reported concerns regarding the capacity of telemedicine to enhance physician-patient bond [39]. All these concerns must be addressed in order to improve the acceptability and use of eHealth tools. A recent study suggested that there is still a low preference for the use of eHealth tools among the adult general population [17]. However, those who have already used eHealth apps usually feel confident to continue using them. Some studies have reported a sense of security and the existence of a relational bond between eHealth apps and patients with psychiatric diseases [42,43]. Our analyses show that machine learning can help to classify e-mental health users and provide clues for their diagnoses and, importantly, their needs in terms of treatment. If machine learning helps physicians to take clinical treatment decisions based on data, the social perception about available eHealth tools will certainly improve.

Acknowledgments
This study received grant support from Instituto de Salud Carlos III (ISCIII PI13/02200; PI16/01852), Delegación del Gobierno para el Plan Nacional de Drogas (20151073), the American Foundation for Suicide Prevention (LSRG-1-005-16), the Ministerio de Ciencia, Innovación y Universidades (RTI2018-099655-B-I00; TEC2017-92552-EXP), and the Comunidad de Madrid (Y2018/TCS-4705, PRACTICO-CM).

Conflicts of Interest
EBG and AAR designed the MEmind application. The authors declare no other conflicts of interest.

Multimedia Appendix 1
List of categories from the ICD-10 Classification of Mental and Behavioral Disorders.
[DOCX File, 13 KB - mental_v8i1e17116_app1.docx]

Multimedia Appendix 2
Clinical global impression (CGI) scores: severity of illness.
[DOCX File, 12 KB - mental_v8i1e17116_app2.docx]

Multimedia Appendix 3
Dichotomized processing of data from the questionnaire scores.
[DOCX File, 14 KB - mental_v8i1e17116_app3.docx]

Multimedia Appendix 4
A binary Z matrix, presenting the number of active factor sets for each patient. Each line corresponds to a single patient. All patients present the bias term or factor set 0.
[DOCX File, 48 KB - mental_v8i1e17116_app4.docx]

Multimedia Appendix 5
Number of patients by clinical global impression (CGI) scores.
[DOCX File, 28 KB - mental_v8i1e17116_app5.docx]

Multimedia Appendix 6
Percentage of patients according to ICD-10 diagnoses.
Multimedia Appendix 7

Average scores for each item of the self-reported questionnaire of current mental well-being according to the gamma distribution. The model is based on a β matrix. A higher weight (β) of a factor set for an item is associated with a greater probability to find a high score in that item when that particular factor set is active.

References


Abbreviations

CGI: clinical global impression
IBP: Indian Buffet Process
ICD: International Statistical Classification of Diseases and Related Health Problems
SPFM: Sparse Poisson Factorization Model
Text Message Interventions in Adolescent Mental Health and Addiction Services: Scoping Review

Sarah MacDougall1*, MPH; Susan Jerrott2*, PhD; Sharon Clark3*, PhD; Leslie Anne Campbell4*, PhD; Andrea Murphy5*, PharmD; Lori Wozney6*, PhD

1Maritime SPOR SUPPORT Unit, Halifax, NS, Canada
2Nova Scotia Health, Halifax, NS, Canada
3IWK Health, Halifax, NS, Canada
4Department of Community Health and Epidemiology, Dalhousie University, Halifax, NS, Canada
5College of Pharmacy, Dalhousie University, Halifax, NS, Canada
6Mental Health and Addictions, Policy and Planning, Nova Scotia Health, Dartmouth, NS, Canada

*all authors contributed equally

Abstract

Background: The vast majority of adolescent mental health and substance use disorders go undiagnosed and undertreated. SMS text messaging is increasingly used as a method to deliver adolescent health services that promote psychological well-being and aim to protect adolescents from adverse experiences and risk factors critical for their current and future mental health. To date, there has been no comprehensive synthesis of the existing literature on the extent, range, and implementation contexts of these SMS text message interventions.

Objective: The objective of this scoping review was to map and categorize gaps in the current body of peer-reviewed research around the use of SMS text messaging–based interventions for mental health and addiction services among adolescents.

Methods: A scoping review was conducted according to Levac’s adaptation of Arksey and O’Malley’s methodological framework for scoping reviews in six iterative stages. A search strategy was cocreated and adapted for five unique databases. Studies were screened using Covidence software. The PICO (patient, intervention, comparator, outcome) framework and input from multiple stakeholder groups were used to structure and pilot a data extraction codebook. Data were extracted on study methodology and measures, intervention design, and implementation characteristics, as well as policy, practice, and research implications.

Results: We screened 1142 abstracts. Of these, 31 articles published between 2013 and 2020 were eligible for inclusion. Intervention engagement was the most common type of outcome measured (18/31), followed by changes in cognitions (16/31; eg, disease knowledge, self-awareness) and acceptability (16/31). Interventions were typically delivered in less than 12 weeks, and adolescents received 1-3 messages per week. Bidirectional messaging was involved in 65% (20/31) of the studies. Limited descriptions of implementation features (eg, cost, policy implications, technology performance) were reported.

Conclusions: The use of SMS text messaging interventions is a rapidly expanding area of research. However, lack of large-scale controlled trials and theoretically driven intervention designs limits generalizability. Significant gaps in the literature were observed in relation to implementation considerations, cost, clinical workflow, bidirectionality of texting, and level of personalization and tailoring of the interventions. Given the growth of mobile phone–based interventions for this population, a rigorous program of large-scale, well-designed trials is urgently required.

(JMIR Ment Health 2021;8(1):e16508) doi:10.2196/16508
KEYWORDS
adolescent; mental health; eHealth; text messaging; SMS; information science; cell phone; implementation; review

Introduction

Limits of Face-to-face Mental Health and Addiction Care for Adolescents

Many mental health disorders emerge in adolescence, which contribute to the existing burden of disease among young people and later in life [1]. More than 50% of adult mental disorders have their onset before the mid-teen years [2,3]. Furthermore, adolescents experiencing depressive symptoms more than two-standard-deviations above the mean predicts a twofold to three-fold greater risk for an adult major depressive episode [4]. Notably, a survey of 10,123 adolescents aged 13 to 18 years in the continental United States showed that 40% of participants with one disorder also met criteria for another lifetime disorder [5]. Substance use and mental health disorders, for example, commonly co-occur [6] and are closely related to increased morbidity and mortality [7]. A recent meta-analysis of 41 studies conducted between 1985 and 2012 in 27 countries estimated a global point prevalence of mental disorders in children and adolescents of 13% [8]. From a global perspective, neuropsychiatric disorders are the leading cause of years lost due to disability among 10- to 24-year-olds [9].

The persistent lack of available services to identify and meet these needs is concerning. While several face-to-face psychological therapies have demonstrated effectiveness in the treatment of mental health and addictions, many children and adolescents do not receive or have access to these treatments [10,11]. Despite decades of effort to improve access, demand continues to outstrip provider capacity for face-to-face services [12,13]. In particular, children experiencing poverty [14], children in rural areas [15], and youth and families who experience self-stigma due to prejudice and stereotyping [16] consistently face unequal access or barriers to care. Even when face-to-face services are available, treatment engagement is challenging. For example, high levels of missed appointments and premature termination of therapy are significant for adolescents, with no-show and attrition rates of 40%–60% commonly reported for this population [17]. Given this global burden and the unsustainability of traditional intervention approaches to meet the rising demand, it is vital for new lines of research to generate innovative strategies and interventions for adolescent mental health and addictions.

Mental health systems facing these challenges increasingly look to advancements in information and communications technologies to augment provider capacity, promote healthy behavior and lifestyle changes, and overcome barriers that limit help-seeking [18,19]. However, the technology marketplace is continuously evolving into more dynamic, diverse, and sophisticated functionalities. As technologies are expected to increase in scope and impact, rapid updating and analysis of emerging evidence is needed to inform future research and signal new interventions to policy makers in order to accelerate implementation of high-quality services once effectiveness of these interventions has been established.

Leveraging Trends in Text Messaging for Adolescent Mental Health Research

One rapidly growing field of study is the use of SMS text messaging for delivering mental health and addiction interventions. Seven billion people, or 95% of the global population, live in an area covered by a mobile-cellular network [20], making SMS text messaging one of the most widely used information and communication technologies. Texting is used by most adolescent cell phone owners and has surpassed phone calls, instant messaging, social network messaging, and face-to-face talking as the preferred mode of communication for this age group [21]. Adolescents report convenience, discreetness, increased communication effectiveness, and reduced anxiety associated with texting or talking on the phone in comparison to in-person evaluations with a physician [22] as reasons for preferring this modality. The immediacy of reaching adolescents through texting is apparent, with 91% of texts read within the first 3 minutes of receipt [23]. Cell phone ownership among 12- to 17-year-olds has been steadily rising over the past several years and is consistent across race and gender groups [24]. Reports from 2019 indicated that among US teens, 69% have a smartphone by the age of 12 [25]; however, teens from lower-income families are slightly less likely to own cell phones than teens from higher-earning families [24]. Although smartphone ownership is rapidly growing, only about a third of the world’s population (approximately 2.6 billion) used a smartphone in 2017 [26], compared to over 5 billion mobile phone subscribers. Even using conservative estimates, simple SMS text messaging that does not require smartphone capabilities will remain an important tool to reach adolescents for some time to come. Consequently, researchers and decision makers need strategic guidance on where new investments in SMS text messaging intervention development and testing are best positioned in the future.

Related Work and the Need for Mapping the Current Evidence Base

Texting is among the most frequently used technologies for low-intensity behavioral health interventions [27] and has demonstrated effectiveness in supporting a range of healthy behavior changes among adolescents with diabetes [28] and obesity [29] and for many other topics including sexual health [30] and contraception [31]. Interventions in these other health domains typically make use of multiple persuasive system design features (eg, personalization, reminders, feedback, branching/tailoring) [32], but there are significant gaps in the knowledge base about breadth of features used to deliver service through this modality and the contexts and populations in which SMS text messaging interventions focused on mental health or addictions have been tested.

The most recent synthesis of SMS text messaging intervention research for mental health or addictions present several limitations for addressing adolescent population needs. A 2016 review of SMS text messaging mental health interventions by Watson et al [33] excluded studies involving children and
adolescents, and reviews by Berrouiguet et al [34] and Rathbone and Prescott [35] did not provide information on the target age range of studies in their review, making it challenging to conduct any subanalysis of adolescent interventions specifically. Importantly, findings from these adult-focused reviews may not be generalizable to adolescent populations due to the unique developmental changes adolescents experience and their different technology preferences. A 2014 meta-analysis of 14 adolescent-focused intervention studies using SMS text messaging revealed a summary effect size of 0.25 on measures of substance use reduction [36]. However, 11 of the studies focused on tobacco use alone, making the generalization to the broader mental health and addictions continuum limited. A review by Badawy and Kuhns [37] of mobile phone interventions targeting adolescents published between 1995 and 2015 limited inclusion to studies where the primary or secondary outcome related to preventive behavior adherence, meaning that the review provides only a partial synthesis of SMS text messaging research across the full continuum of care. The most recently published relevant review was by Garrido et al in 2019 [38] and synthesized data on a broad range of mobile phone interventions for adolescent anxiety and depression, identifying only 4 text messaging interventions of any study design in their search strategy. Cursory scans of recently published literature suggest there are significantly more primary studies on mental health and addiction interventions for adolescents using text messages than have been previously reviewed. Further, limited discussion in previous reviews on the implementation characteristics of interventions under study (eg, costs, information and communications technology infrastructure required, provider training requirements) provide decision makers with few insights to inform real-world requirements of offering these interventions as sustainable services.

It is clear from the gaps identified that the foundational understanding of how these interventions have been implemented and designed, the measures used to evaluate impact, the contexts and mechanisms identified by researchers to explain intervention impacts, and recommendations being offered for policy, practice, and future research requires updating. The objective of this scoping review was to understand the current state of peer-reviewed research around the use of text message–based interventions for mental health and addiction services among adolescents.

Specifically, the review proposed to answer the following questions: What outcomes are measured to determine effectiveness and engagement? What are the technological and clinical design features of these interventions and services? What implementation contexts, mechanisms, barriers, and facilitators are described? What are the recommendations for practice, policy, and research reported by study authors?

### Methods

#### Overview

A scoping review design is ideal for broad mapping and characterizing of existing research [39]. Scoping reviews share a similar process to systematic reviews, since they both are rigorous and transparent in identifying eligible literature but are divergent in purpose. Using a scoping review framework allows for broad exploration of the research to map key concepts, evidence types, and gaps in research in a defined field [40].

The Levac adaptation to Arksey and O’Malley’s [41] methodological framework for scoping reviews was applied in six iterative stages: (1) identifying the research question, (2) identifying relevant studies, (3) selecting studies, (4) charting the data, (5) collating, summarizing, and reporting on the articles and (6) consulting with stakeholders. The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guideline extension for minimum reporting standards in scoping reviews [42] and Joanna Briggs Institute recommendations for scoping reviews [43] were also followed.

#### Search Strategy

The search strategy was developed in collaboration with an evidence synthesis specialist at the Maritime SPOR SUPPORT Unit (MSSU), which is an organization that provides support to researchers and brings together key stakeholders to work on government priority projects. MEDLINE, Embase, PsycINFO, CINAHL, and Scopus databases were searched to identify a broad range of articles in October 2018. A search conducted in the Ovid MEDLINE “In-Process & Other Non-Indexed Citations” database was used to reduce the chance of omitting articles not included in PubMed and to capture the most recent literature possible. An example of the search terms and search strategy is available in Multimedia Appendix 1. Reference lists of related reviews were hand-searched for any additional citations. The search was updated by an evidence synthesis specialist at the MSSU on June 1, 2020.

#### Study Selection

After removing duplicates, titles and abstracts were uploaded into Covidence software (Covidence). The extensive time and effort requirements for conducting an unfunded review [44] and stakeholders on the overarching project team promoting rapid review approaches to inform decision making shaped the review protocol. To expedite the process while maintaining rigor, the primary author, with experience in eMental Health–related systematic reviews and text messaging intervention design (LW), completed screening at the title and abstract level, following defined eligibility criteria (see Multimedia Appendix 2). Any abstracts considered questionable were moved to the full-text review phase. For titles and abstracts identified in our updated search in June 2020, two reviewers (LW and SM) independently screened titles and abstracts. No test of agreement between reviewers was conducted; instead, all discrepancies were discussed until a decision on moving to full-text review was made.

#### Eligibility Criteria

The screener followed the “excluded terms” approach outlined by Carter [45] to improve accuracy and efficiency of single-screener processes without compromising the quality. To this end, several “exclude” terms were generated to quickly reject studies at the title/abstract level that had a high likelihood of being excluded. Exclude terms included HIV, neurodevelopmental, domestic violence, infant, pregnancy, cancer, sexual health, and contraception.
As several meta-analyses and systematic reviews of tobacco and smoking cessation interventions, including a 2016 smoking cessation Cochrane review [46], have already established high-quality evidence for SMS text messaging interventions in that area, intervention studies that focused only on smoking cessation were excluded in order to focus on mapping and characterizing the less well-established literature.

The inclusion criteria were as follows: the article was a primary study or abstract reporting outcome data; the study targeted children or adolescents (included in the age range ≤18 years) by design; text messages were one of the primary delivery mechanisms; the study was published in English; and the intervention was for mental health and addiction care (eg, symptom management, appointment reminders).

The exclusion criteria were as follows: the article was a systematic review, commentary, editorial, or protocol; the study targeted parents of children and adolescents receiving mental health care; the study targeted only smoking cessation or smoking prevention; SMS text messaging was only used for research data collection and had no therapeutic/educational purpose; or the intervention targeted a subgroup of patients with mental health as a secondary issue relative to an overarching medical condition (eg, cancer, HIV, pregnancy).

Full-text review was completed independently by one reviewer (Swati Rathore) at the outset. A validity check on a randomly selected set of 20 full texts was undertaken independently by a second reviewer (SM) to explore potential bias or inconsistency. Excellent interrater agreement (κ=0.90) was established. Questionable studies were discussed and resolved through consensus with a third reviewer (LW). In the case of a study that had a published protocol as well as a study outcome paper, only the outcome paper was included.

Data Extraction

Suggestions on synthesizing evidence from complex interventions were followed [47]. Briefly, such frameworks emphasize that decision makers are more interested in knowing when the intervention works (delivered by whom, how, how often, and in which setting) than in answering the simple question of “does it work?”. A codebook was drafted (LW and Swati Rathore) based on the PICO (patient, intervention, comparator, outcome) framework. Through consultation and discussion with a range of stakeholders (psychologists/clinicians, academic researchers, patient advisors, policy and planning experts, and mental health system administrators at the local and provincial level), the codebook was refined to highlight interoperability, communication; and (7) recommendations (ie, recommendations for practice, policy, and research made by the primary study authors).

The form was piloted independently by two reviewers (Swati Rathore and SM) for 3 articles to determine if both parties were in agreement and to see if any clarifications were needed. This enabled an early check on consistency and relevance. Data extraction was then split between the two coders, and discrepancies or questions were resolved by consensus with input from another reviewer (LW).

Analysis

All data for this scoping review were entered in Microsoft Excel (Microsoft Corporation). After data cleaning, a descriptive, analytical approach was used to generate summary statistics (counts, percentages, etc) of the data extracted for key sections 1-6. For section 7, key themes and issues from each study were identified by scrutinizing the results and discussion sections using thematic content analysis.

Results

Study Characteristics

We screened 1142 abstracts for possible inclusion (Figure 1). After title and abstract screening, 102 full-text articles were screened for eligibility, with 71 articles excluded at this stage. Thirty-one studies were included, with 28 published in peer-review journals, 2 dissertations, and 1 abstract that included outcome data [48-78].

Included articles were published between 2013 and 2020. Of the 31 included studies, 18 (58%) were conducted in the United States. With respect to study design, 29% (9/31) incorporated randomization, 19% (6/31) were qualitative designs, and 52% (16/31) were observational or cross-sectional. In 68% (21/31) of the studies, fewer than 100 study participants were exposed to the intervention or service (see Multimedia Appendix 2).
Measures and Outcomes

In addition to bespoke measures, over forty unique standardized outcome measures were used across the studies, measuring a wide range of symptom, global functioning, and therapeutic experience self-report scales and inventories. The measures categories with the widest use were intervention engagement (18/31; eg, number of texts sent, number of URL links clicked), cognitions (16/31; eg, disease knowledge, self-awareness), and acceptability (16/31; eg, satisfaction). Adherence to a predefined treatment protocol was measured in 7 studies, typically for substance use-related interventions (eg, reduction in number of drinks consumed). Fewer studies tracked social/relational outcomes (5/31; eg, quality of social relationships) or physical changes (2/31; eg, blood alcohol level). Forty-two percent of the studies (13/31) explicitly reported evidence that the measures used were validated or reliable (eg, test-retest, internal consistency, and content validity checks). Twenty-six percent (8/31) reported outcomes that were measured immediately following exposure to the intervention. An additional 32% (10/31) measured outcomes within 3 months of intervention completion, but not immediately following, and 3 studies completed outcome tracking at 6 months postintervention or longer. Studies by Gonzales et al [59] and Whittaker et al [78] included a measure 12 months postintervention. Eighty-seven percent of studies (27/31) reported on at least one positive outcome including, for example, lowered stress [49], cost efficiency [52], less substance use [57], and treatment protocol adherence [51]. In the only randomized controlled trial of an intervention for adolescent depression that enrolled more than 500 participants, researchers found no evidence of benefit in depressive symptoms from a cognitive behavioral therapy–based SMS text message intervention compared to a control program [78]. However, youth reported finding the program helpful. One study reported not being able to complete outcome measurement due to major recruitment and retention issues during implementation [71]. Another study reported an adverse event/safety issue related to the intervention [71]. Across the 16 studies reporting on acceptability measures, the majority of adolescent participants (ie, >70%) in each case study reported being satisfied with the intervention.
Intervention Characteristics

Interventions ranged considerably in length/duration and intensity (see Multimedia Appendix 3). The models or frameworks most frequently cited for guiding intervention development were social cognitive theory (5/31) and cognitive behavioral therapy (3/31), with the health belief model (1/31) and normalization process theory (1/31) also explicitly identified. Nine studies reported that the intervention had been co-designed with adolescents. Texting frequency ranged from several times per day for one depression-focused study [56] to once every two weeks or month for another depression-focused study [54]. Most frequently, interventions were delivered in less than 12 weeks, and adolescents received 1-3 messages per week over that time. Thirty-five percent (11/31) of studies specifically indicated the intervention was for substance use or problem drinking, and 32% (10/31) focused on adolescents with depression. Almost two-thirds of studies (20/31) had bidirectional texting, either automated bounce-back prompts (eg, [57]) or person-driven (eg, [65]). The 2020 study by Haug et al [65] incorporated multiple uses of automated text messages where messages were used to push weblinks, video clips, and pictures in some instances but also used to help assess and deliver individualized feedback through quizzes (eg, “reply to this SMS with ‘yes’ or ‘no’… are you meeting friends or going out today?”). Based on responses, a tailored but automated prompt would be sent (eg, “Hey Mike, great plan! Take a moment and imagine exactly how you could implement this plan…. Have a nice evening!”). Text messages across all studies were typically designed to convey empathy and encouragement (eg, “Stop… Breathe… And think about how you got through difficult times before. You got this!” [56]) and to support coping. The content and authenticity of the messages were cited by adolescent respondents as key elements of their engagement. As one respondent in the Duan et al study [58] reflected, “Don’t include messages that make me feel like my parents are teaching me something. I need encouragement from a friend, not from a teacher or parent”. Among studies utilizing bidirectional texting, 60% (12/20) required the researcher to respond to at least some of the texts. In 29% of studies (9/31), adolescents had the option to self-refer to the intervention. All other instances required provider or researcher referral or recruitment pathways. Two studies reported offering the intervention in more than one language.

Implementation Contexts and Features

Community characteristics of adolescents enrolled in the interventions were explicitly described in only 10 of 31 studies and included limited details on community socioeconomic status, urban/rural geographic location, or generic statements related to culturally diverse community samples. Demographic information on the adolescent study participants revealed that most of the interventions were designed to target 13- to 17-year-olds. In 5 studies, the intended age range of end users covered more than 10 years (eg, Antiss and Davies, 2015 [49], ranged from 12 to 24 years of age). Forty-eight percent of the studies (15/31) reported providing adolescents with an honorarium or compensation for study participation. At least some level of training for the provider was noted in 35% (11/31) of studies. Twenty-nine percent of studies (9/31) reported on the technical quality or performance (eg, all messages were received). Sustainability (plans for scale-up or long-term operational planning) was discussed in 2 studies. Eight studies (26%) provided details on privacy and security of electronic data, which included things like requiring password-protected phones [61], limiting collection and storage of personal identifying health information [53], and confidentiality approval [74]. None of the studies reported on the direct costs of delivering the intervention. In 32% (10/31) of studies, the technical infrastructure used to deliver the text messages required some level of interoperability between interfaces (eg, text redirecting to a survey tool housed on another platform). As retention of adolescents in the studies was quite high, overall there was limited discussion of barriers to uptake or nonadherence. One author noted that even though only 38% of texts (eg, asking a participant to rate their mood from 1-10) were responded to by adolescents in the study, participants still reported improved outcomes [69].

Practice, Policy, and Research Recommendations

Forty-two percent (13/31) of study authors reported that ease of use and broad accessibility were strengths of the services that made them feasible to roll out and resulted in high levels of uptake. Increased ability to tailor the scheduling of messages (eg, [67]; messages before or after school) or the content of messages (eg, [72]; including youth’s name in the texts) was noted by 35% (11/31) of authors as a way to practically improve interventions further. Multiple study authors noted the importance of working with youth to co-design and shape the messages. The intensity of exposure to texts (ie, time of day, number sent, and frequency) was pointed to as an important practical consideration by 19% (6/31) of authors. No studies explicitly addressed or identified specific policy-related implications of their findings. For example, there were no explicit recommendations around ethics of sending youth text messages, electronic communication policies, risk, privacy, or safety monitoring guidelines that would need to be in place for the service to operate beyond the research. Two primary sub-themes were identified in future research directions proposed by primary authors: (1) opportunities for comparative component studies that look at whether and which features (eg, tailoring, different messaging schedules, and different media) improve adherence or outcomes (15/31; 48%) and (2) need for well-powered randomized controlled trials (13/31; 42%).

Discussion

Principal Findings

To our knowledge, this review provides the first comprehensive mapping of the current literature on use of SMS text messaging for delivery of mental health and addiction interventions to adolescents. The aim was to detail outcomes being measured, clinical and technical features of interventions, implementation contexts in which they have been studied, and the ensuing recommendations made by primary authors to support innovative research in the field moving forward. Findings suggest a growing evidence base regarding text message interventions for adolescent mental health and addictions. These interventions appear to be highly acceptable and accessible to adolescents,
easy to use, and have at least some positive impacts according to most studies. While these findings are consistent with the literature among adults receiving physical and mental health text message–delivered services [35], the research literature remains limited in several important ways. Overall, there are significant limitations in the trial designs. These include issues with small sample sizes (eg, pilot studies of less than 100 participants are typical), limited intervention duration, lack of randomized designs, nonstandard outcome measures and definitions, and few repeated measures in studies with longitudinal designs.

Limited grounding in, or at least reporting on, theoretical frameworks that drive intervention development was observed in this review. While most interventions tended to follow a basic architecture of less than 12 weeks and 2-3 messages per week, the content, levels of interactivity and personalization, use of media, and underpinning aims (eg, behavioral activation, acquisition of new knowledge, emotional regulation) were vastly different. Significant heterogeneity in intervention features was noted in our review, such as fixed message content or personalized content; fixed-frequency or real-time support; standardized versus personalized messages; and unidirectional versus bidirectional communication. Without theoretical and therapeutic models to define and articulate how these intervention elements work to produce which outcomes (ie, satisfaction versus clinical change versus changes in thoughts and feelings), there is limited ability to generate and replicate rigorous hypothesis testing or clarify which mental health and addiction conditions at which level of illness severity might be best suited for SMS text messaging interventions. This finding further supports recommendations for standardized reporting of theories of change for behavioral interventions in academic research [79].

As this field emerges, high-quality studies will be required, despite rapidly evolving technical capabilities that can outdate services before they are able to be scaled up [80]. Preliminary usage data from primary studies in this review suggest adolescents will use and are satisfied with SMS text messaging services, but how satisfaction and engagement relate to direct health outcomes is unclear. In this review, over half of the included studies assessed some measure of acceptability or satisfaction. High scores on satisfaction outcomes across types of interventions, intensities, and countries of use point to possible ceiling effects of satisfaction measures [81]. Generating new knowledge about predictors of satisfaction and use (eg, treatment readiness, non–health-related texting use, degree of co-design [82]) could more usefully inform clinical practice guidelines for where these interventions are best matched to the needs of which adolescents, and under which circumstances. For example, adolescents across studies reported finding interventions helpful and acceptable and are satisfied with the experience even when they do not produce clinically significant changes (eg, [78]). The findings of this review have important implications for decision makers who are mandated to integrate services that show a return on investment. Mixed-effects studies that show interventions are “liked,” but have little impact on quality of life or symptoms, may present challenges for scale-up. Unpacking the relationship between perceived acceptability and clinical change will be an important direction for future research on SMS text messaging interventions. Better understanding of unidirectional versus bidirectional messaging design and their impact on engagement and exposure to the “dose” of SMS text messaging interventions could add needed insights. Preliminary research in other health fields with this population has established that bidirectional texting increases intervention effectiveness [83] and is a valuable line of inquiry for mental health and addiction–focused interventions.

Frequency and duration of interventions in this review, even for the same presenting clinical condition, ranged considerably from multiple times a day to several texts over the course of months. Future studies should consider the impact of habituation, response fatigue, and perceived “intrusiveness” of texts among adolescents using these interventions. Generating new knowledge about intervention intensity might also inform decision makers about where interventions “sit” within clinical workflow and provider/client relationships. For example, two-thirds of the studies in the review included some form of bidirectional interaction. It is possible that just messaging back and forth with clinic or research staff sufficiently evokes a relationship dynamic that bolsters perceived social connectedness and sense of well-being. Future work is needed to determine if it is the presence of ongoing communication at all or the specific therapeutic content of SMS text messaging interventions that results in improved outcomes. Studies in our review generally lacked details about how theory or therapeutic principles guided the content, frequency, and duration decisions around texts being sent to youth. It is promising that a number of studies reported some level of adolescent engagement as co-designers to potentially explicate some of these mechanisms of change and decisions around frequency and intensity. Richer descriptions of co-design models and theoretical frameworks used to shape intervention development would be beneficial.

This review points to methodological decisions and implementation considerations (eg, study remuneration practices, self-referral versus provider referral, and researcher-led communication) that might impact adherence and outcomes once interventions are rolled out beyond the research cycle into full-service delivery. These will be important for future researchers to consider and suggest that hybrid implementation/effectiveness studies could be a valuable study design for this field. For example, nearly half of the studies in the review reported providing remuneration for study participants. Would engagement be as high if youth did not receive remuneration for participating, but the service was offered as standard of care? Over a third of the studies reported that human resources required specialized training in order to support the SMS text messaging intervention or that technical troubleshooting for youth was required. If those tasks were to be integrated into a clinician’s existing workflow, how might that impact costs or sustainability? A significant gap identified in our review of implementation features was that none of the studies in the review provided costing information. The ability of this field to communicate cost-effectiveness data to policy makers that incorporate implementation costs is central to the likelihood of interventions being adopted by practitioners and would be in alignment with emerging standards for reporting.
A systematic review of the use of research evidence in public health decision-making processes recommended that more research target the needs of decision makers [86]. A strength of this review is that in adopting Levac’s steps for scoping reviews, the research questions, data extraction and analysis, and reporting approaches were codeveloped with an interdisciplinary team of people with lived experience in mental health services and clinicians as well as local health system decision makers, through the MSSU. Their involvement was critical in identifying the diverse data of interest to be explored in this review. This scoping review, therefore, is set apart from other recent reviews by systematically highlighting persistent reporting gaps in the costing, sustainability planning, privacy and security, and technical infrastructure elements of SMS text messaging behavioral interventions for adolescents that are vital for translating research into real-world services. Knowing what has not been reported in the published literature is equally important for decision makers who are tasked with evaluating risks and return on investment. Recommendations for reporting on behavioral interventions have been around for over a decade [87,88] as well as new guidelines focused on mobile phone–based intervention reporting [89]. Gaps outlined in this review highlight knowledge-sharing opportunities on these important elements. Researchers in this rapidly advancing field can support decision makers to more quickly mobilize efforts to integrate effective, evidence-based SMS text messaging interventions by following these guidelines and integrating locally relevant information needs.

**Limitations**

There are several limitations to this review. Although inclusion criteria were kept as broad as possible while maintaining focus on the research question, only 31 studies met inclusion criteria. While this is a significantly larger number than any of the other related recent reviews, a search of the grey literature was not completed. The search strategy was also limited as a result of search terminology related to mental health and addictions, which has an extremely diffused and nuanced lexicon across different fields. This issue has been raised by other researchers [90] and reported as a significant obstacle to integrating mental health services within broader health care systems [91]. This relatively small and highly heterogeneous sample raises an issue of publication bias and increases the importance of assessing for bias within individual studies. Within currently recommended scoping review methods, quality appraisals are not typically undertaken. While this review did not conduct risk of bias appraisals, systematically broad mapping of basic study methods did identify discrete areas of methodological weakness, which have been detailed and summarized to inform future work. In the context of this burgeoning intervention literature, it was important to include a range of designs to fully scope what is currently known. Finally, this review combined results from trials with early pilot studies, program evaluations, and qualitative investigations, resulting in a wide range of interventions, follow-up durations, and study aims. A review of only high-quality studies may produce different summary conclusions.

**Conclusion**

Text message interventions are increasingly explored in adolescent mental health and addictions. However, the research to date has important limitations in terms of the heterogeneity of interventions and implementation characteristics of studies. There are significant gaps regarding the hypothesized theoretical and therapeutic mechanisms driving observed outcomes, which contributes to the lag in translating research into policy and practice. Impact on direct health outcomes, cost considerations, predictors of participant engagement, and ongoing sustainability or return on investment are underreported aspects of interventions studied to date. While studies broadly reported high levels of satisfaction with SMS text messaging interventions among adolescents, rigorous study designs are needed to parse out success features and how satisfaction related to engagement. Ideally, future research should formally compare characteristics (eg, number of texts, frequency, duration) in order to measure and adjust each of these parameters to meet the needs of adolescents as they change over time. Multiple lines of innovative inquiry in the use of SMS text messaging interventions for adolescent mental health and addiction are possible and promising.

**Acknowledgments**

We would like to thank Swati Rathore for her contributions to screening, data extraction, and codebook development. We want to acknowledge the support of all first-voice advisors, clinicians, administrators, and external partners who participated in the Maritime SPOR SUPPORT Unit “UniCity Team Priority Project” that guided this work. This work was supported by the Maritime SPOR SUPPORT Unit, which receives financial support from the Canadian Institutes of Health Research, the Nova Scotia Department of Health and Wellness, the New Brunswick Department of Health, the Nova Scotia Health Research Foundation, and the New Brunswick Health Research Foundation. The opinions, results, and conclusions reported in this paper are those of the authors and are independent from the funding sources. No endorsement by the MSSU or the named funding partners is intended or should be inferred.
Authors' Contributions

SM made substantial contributions to the design of the paper, conducted data screening and extraction, interpreted the results, drafted components of the paper, and approved the final version.

SJ made substantial contributions to the design of the work, contributed to the methodological plan, interpreted the results, critically revised drafts of the paper, and approved the final version.

SC made substantial contributions to the design of the work, interpreted the results, critically revised drafts of the paper, and approved the final version.

LAC made contributions to the design of the work, contributed to the methodological plan, interpreted the results, critically revised drafts of the paper, and approved the final version.

AM made substantial contributions to the design of the work, interpreted the results, critically revised drafts of the paper, and approved the final version.

LW made substantial contributions to the design of the paper, contributed to the statistical analysis, interpreted of the results, drafted the paper, and approved the final version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File, 79 KB - mental_v8i1e16508_app1.docx ]

Multimedia Appendix 2
Characteristics of included studies.
[DOCX File, 23 KB - mental_v8i1e16508_app2.docx ]

Multimedia Appendix 3
Overview of intervention features and targeted population.
[DOCX File, 19 KB - mental_v8i1e16508_app3.docx ]

References


72. Pisani AR, Wymaan PA, Guridotta K, Schmeek-Cone K, Anderson CL, Judd E. Mobile Phone Intervention to Reduce Youth Suicide in Rural Communities: Field Test. JMIR Ment Health 2018 May 31;5(2):e10425 [FREE Full text] [doi: 10.2196/10425] [Medline: 29853439]
Abbreviations

MSSU: Maritime SPOR SUPPORT Unit
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
MacDougall et al

©Sarah MacDougall, Susan Jerrott, Sharon Clark, Leslie Anne Campbell, Andrea Murphy, Lori Wozney. Originally published in JMIR Mental Health (http://mental.jmir.org), 08.01.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Mental Health, is properly cited. The complete bibliographic information, a link to the original publication on http://mental.jmir.org/, as well as this copyright and license information must be included.
A Mindfulness-Based Intervention for Student Depression, Anxiety, and Stress: Randomized Controlled Trial

Paul Ritvo¹*, PhD; Farah Ahmad²*, MPH, PhD, MBBS; Christo El Morr³*, PhD; Meysam Pirbaglou¹, PhD; Rahim Moineddin⁴, PhD; MVC Team⁴*

¹School of Kinesiology and Health Science, Faculty of Health, York University, Toronto, ON, Canada
²School of Health Policy and Management, York University, Toronto, ON, Canada
³Dalla Lana School of Public Health, University of Toronto, Toronto, ON, Canada
⁴York University, Toronto, ON, Canada
*these authors contributed equally

Corresponding Author:
Christo El Morr, PhD
School of Health Policy and Management
York University
4700 Keele St,
Toronto, ON, M3J1P3
Canada
Phone: 1 4167362100
Email: elmorr@yorku.ca

Abstract

Background: University students are experiencing higher levels of distress and mental health disorders than before. In addressing mental health needs, web-based interventions have shown increasing promise in overcoming geographic distances and high student-to-counselor ratios, leading to the potential for wider implementation. The Mindfulness Virtual Community (MVC) program, a web-based program, guided by mindfulness and cognitive behavioral therapy principles, is among efforts aimed at effectively and efficiently reducing symptoms of depression, anxiety, and perceived stress in students.

Objective: This study’s aim was to evaluate the efficacy of an 8-week MVC program in reducing depression, anxiety, and perceived stress (primary outcomes), and improving mindfulness (secondary outcome) in undergraduate students at a large Canadian university. Guided by two prior randomized controlled trials (RCTs) that each demonstrated efficacy when conducted during regular university operations, this study coincided with a university-wide labor strike. Nonetheless, the students’ response to an online mental health program on a disrupted campus can provide useful information for anticipating the impact of other disruptions, including those related to the COVID-19 pandemic as well as future disruptions.

Methods: In this parallel-arm RCT, 154 students were randomly allocated to an 8-week MVC intervention (n=76) or a wait-list control (WLC) condition (n=78). The MVC intervention included the following: (1) educational and mindfulness video modules, (2) anonymous peer-to-peer discussions, and (3) anonymous, group-based, professionally guided, 20-minute videoconferences. Study outcomes were evaluated at baseline and at 8-week follow-up using the following: Patient Health Questionnaire-9 (PHQ-9), the Beck Anxiety Inventory (BAI), the Perceived Stress Scale (PSS), and the Five Facets Mindfulness Questionnaire Short Form (FFMQ-SF). Generalized estimation equations with an AR (1) covariance structure were used to evaluate the impact of the intervention, with outcome evaluations performed on both an intention-to-treat (ITT) and per-protocol (PP) basis.

Results: Participants (n=154) included 35 males and 117 females with a mean age of 23.1 years. There were no statistically significant differences at baseline between the MVC and WLC groups on demographics and psychological characteristics, indicating similar demographic and psychological characteristics across the two groups. Results under both ITT and PP approaches indicated that there were no statistically significant between-group differences in PHQ-9 (ITT: β=−0.44, P=.64; PP: β=−0.62, P=.053), BAI (ITT: β=−2.06, P=.31; PP: β=−2.32, P=.27), and FFMQ-SF (ITT: β=1.33, P=.43; PP: β=1.44, P=.41) compared to WLC. There was a significant difference for the PSS (ITT: β=−2.31, P=.03; PP: β=−2.38, P=.03).

Conclusions: During a university labor strike, the MVC program led to statistically significant reductions in PSS compared to the WLC group, but there were no other significant between-group differences. Comparisons with previous cycles of intervention testing, undertaken during nondisrupted university operations, when efficacy was demonstrated, are discussed.
online intervention; randomized controlled trial; university student; depression; anxiety; stress; mental health; efficacy; intervention

Introduction

The impact of COVID-19 and other disruptions on education, generally, and secondary education, specifically, can have major student mental health effects. COVID-19–related shutdowns of face-to-face high school education have been associated with significant student dropout rates in multiple cities in the United States, as high proportions of students, engaged face-to-face, failed to connect online (13%) [1]. With the escalating infection rates of both spring and fall 2020, many university courses that were initiated in face-to-face formats were completed online [2,3], with considerable uncertainties about the fall 2020 semester and the impact of COVID-19 outbreaks on campus life generally.

However, the response of students to other campus crises can provide useful data in anticipating the effects of other disruptions. In 2018, during a 4-year project assessing the effects of mindfulness-based cognitive behavioral therapy (M-CBT), York University was impacted by a faculty and teaching assistant strike. Most students respected picket lines and did not attend classes until the strike resolved, which occurred after the end of the usual calendar-defined semester. Despite the strike, study recruitment and intervention testing proceeded. Here, we report the results of that randomized controlled trial (RCT).

The RCT undertaken was identical to two prior studies [4,5] on the same campus site under noncrisis conditions. As the studies were undertaken months apart, there were minimal environmental differences, other than seasonal change. In the initial study [4], 113 students (mean age 24.8 years) participated in an 8-week online M-CBT program and significant between-group benefits were found on measures of depression (Patient Health Questionnaire-9 [PHQ-9]), anxiety (Beck Anxiety Inventory [BAI]), quality of life (Quality of Life Scale [QOLS]), and mindfulness (Five Facet Mindfulness Questionnaire - Short Form [FFMQ-SF]), favoring the intervention group (MVC) compared to wait-list controls (WLC). In the follow-up study [5], 159 students (mean age 22.5 years) participated in the same 8-week online M-CBT program and significant between-group differences were again found on identical measures of depression (PHQ-9), anxiety (BAI), mindfulness (FFMQ-SF), and quality of life in intervention participants, compared to wait-list controls. These two demonstrations of intervention efficacy prepared us to investigate how the campus crisis may have affected student responses.

All three studies were motivated by the rising prevalence of mental health disorders, including depression and anxiety, among college students worldwide, prior to the COVID-19 pandemic [6]. In the United States, for example, analyses of college data show that mental health disorders are among the top 5 diagnostic categories seen at college health services and are responsible for the highest number of visits (4.93) per student [7]. Multiple US studies have suggested there is an increasing prevalence of mental health disorders, especially depression and anxiety, among undergraduate college and university students [7-14].

Additionally, of concern for online researchers, there are possible links between decreased youth mental health and increased online activity (ie, “screen time”) [15-17]. Strong arguments and reliable data support the perspective that online activities, particularly social media engagement, have unhealthy impacts for some youth populations [15-17]. Disruptions of regular face-to-face classes at universities and high schools may have elevated levels of unhealthy “screen time” for select populations. Accordingly, it is worthwhile to ascertain what changes might have occurred when university students experiencing a disruption that likely increased time spent online used an online mental health intervention.

A student survey of 32 Canadian postsecondary institutions indicated a high prevalence of high anxiety (56.5%), hopelessness (54%), seriously depressed mood (37.5%), and overwhelming anger (42%) [18]. The mental health problems seen among North American students are also apparent worldwide, as the World Health Organization (2018) reported increasing mental disorders in college and university students [6].

Despite student distress, the face-to-face counseling offered in colleges and universities has not kept pace with demand. For example, from 2007 to 2012, full-time enrollment in the Ontario (Canada) college system increased by 26% while the number of counselors employed in the college system increased by only 4.6% [19]. This discrepancy has resulted in observations of underserved students and overwhelmed counselors amid the increasing distress among students.

Mindfulness-based interventions have been demonstrated to positively impact psychological and physical health [20-22], with several meta-analyses demonstrating impacts across clinical and nonclinical populations [22-27]. However, with large numbers of students (50,000-60,000 on some campuses), there may not be sufficient numbers of trained personnel to convey helpful mindfulness-based practices directly.

Accordingly, we developed a web-delivered M-CBT program (aimed at creating a Mindfulness Virtual Community, or MVC) to reduce depression, anxiety, and stress in university students and, as mentioned above, previously reported on two RCTs targeting Canadian university students that indicated efficacy [4,5].
Methods

Trial Design and Ethical Approval
This study was a two-arm parallel-design RCT comparing the web-based Mindfulness Virtual Community program to a wait-list control group. The Human Participant Research Committee at York University provided research ethics approval for the RCT (Certificate number: e2016 - 345).

Participants and Recruitment
Eligibility criteria were applied to recruit actively enrolled undergraduates aged ≥18 years, with English-language fluency and self-reported confidence in completing the study. Students were excluded if they reported substance abuse or episodes of psychosis during the month prior to the trial.

The study was advertised using study posters, class announcements, and email invitations via listservs of student associations in the Faculties of Health and Liberal Arts. Interested students contacted the research staff via email or phone and were screened for student registration, substance abuse, and indications of psychoses. If abuse or psychotic behaviors “interfered in routine life within the last month,” students were excluded and provided with a list of accessible mental health resources. In addition, a registered clinical psychologist could be contacted directly if there was a perceived need for interim mental health counseling.

Eligible and willing students received detailed in-person information about the study and provided informed written consent. Participants had the option to receive an honorarium of Can $50 (US $39) or 2% in course grade (for professors who gave permission for this option) or three credits (equivalent to 2% course grade) in the Undergraduate Research Participation Pool (URPP) of the Department of Psychology. Each participant also received a resource list that included information about health and social services on campus and in the community (eg, the 24/7 “Good to Talk” helpline for postsecondary students in Ontario). Our protocol included a safety mechanism whereby participants were asked verbally and on the consent form to contact the research staff if they felt distress during the trial period so that “limited counselling with a clinical psychologist could be arranged, if needed.” No such requests arose during the reported study period.

A sample of 480 students (160 students per group) was recruited over 3 semesters (Fall 2017, Winter 2018, and Fall 2018). The 3 samples were not combined due to the campus environment differences related to the 3-month strike in Winter 2018. Here, we report on a sample of n=154 students in a two-arm RCT.

Randomization
Participating students were randomized to the MVC intervention or the wait-list (control) using 1:1 block randomization. The randomized allocation sequence was computer-generated by an off-site research team member and allocations were concealed in sequentially numbered opaque envelopes [28]. The envelopes were only opened after written consent was obtained, ensuring participants and staff were blind prior to the allocation. Each participant in the MVC group received a unique ID and a temporary password; participants changed passwords after their first login, while IDs remained the same to reduce the potential of multiple accounts or identities. Participants in all groups completed online questionnaires at baseline (T1) and 8 weeks (T2).

Intervention
The MVC intervention was 8 weeks in duration and featured the following: (1) 12 student-specific mental health modules conveyed by online video, (2) 3 anonymous discussion boards dedicated to depression, anxiety, and stress, and (3) an anonymous 20-minute group-based live videoconference led by a moderator (with a master’s degree in psychology), during which students raised and discussed topics covered in the modules (Figure 1).

Figure 1. The Mindfulness Virtual Community design.
principles, with the topics informed by prior focus group study [35,36]. The choice of moving and still images used in the videos involved collaborative work (PR, CE, and FA). The module topics and video durations are presented in Table 1. The role played by the online mindfulness moderator was informed by a pilot study [4].

Table 1. Topics and duration of modules.

<table>
<thead>
<tr>
<th>Topics</th>
<th>Education videos (duration in minutes and seconds)</th>
<th>Mindfulness videos (duration in minutes and seconds)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overcoming stress, anxiety, and depression</td>
<td>7:09</td>
<td>9:00</td>
</tr>
<tr>
<td>Mindfulness and being a student</td>
<td>5:18</td>
<td>9:14</td>
</tr>
<tr>
<td>Mindfulness for better sleep</td>
<td>4:40</td>
<td>8:13</td>
</tr>
<tr>
<td>Thriving in a fast-changing world</td>
<td>7:23</td>
<td>8:23</td>
</tr>
<tr>
<td>Healthy intimacy</td>
<td>7:32</td>
<td>9:33</td>
</tr>
<tr>
<td>Destigmatization</td>
<td>6:13</td>
<td>9:12</td>
</tr>
<tr>
<td>No more procrastination</td>
<td>3:42</td>
<td>10:48</td>
</tr>
<tr>
<td>Pain reduction and mindfulness</td>
<td>3:48</td>
<td>9:48</td>
</tr>
<tr>
<td>Healthy body image</td>
<td>5:44</td>
<td>9:54</td>
</tr>
<tr>
<td>Healthier eating</td>
<td>10:10</td>
<td>9:26</td>
</tr>
<tr>
<td>Overcoming trauma</td>
<td>6:01</td>
<td>9:43</td>
</tr>
<tr>
<td>Relationships with family and friends</td>
<td>7:49</td>
<td>8:09</td>
</tr>
</tbody>
</table>

New modules were regularly released during the 8-week intervention period. Following release, they were accessible to students for the remaining intervention period. The videoconferences were offered biweekly in three 20-minute evening sessions. The students in the intervention group received email reminders from project staff prior to the release of each module and prior to the live videoconferences. Access to the internet was assumed to be a minor obstacle as the prior focus group study revealed that 94.4% of the students had access to a smartphone and 93.1% had access to laptops or personal home computers. All participants had free internet access on campus and nearly all reported internet access through smartphones and/or laptop computers [38].

The MVC was constructed in partnership with an industry partner (ForaHealthyme Inc) and designed to be a virtual environment supportive of personal mindfulness practice and related CBT self-help. It facilitated mutual help interactions between participants, and between participants and the moderator. The two categories of users were students and health professionals (who moderated the discussion board dialogues and led the live videoconferences). All users used a login and a password to gain access.

Once logged in, each student could do the following: (1) access the educational and mindfulness video modules, (2) access 3 peer-to-peer discussion boards, 1 for each of the 3 mental health conditions targeted by the RCT (anxiety, stress, and depression), (3) notify the moderator about any message posting that represented a problem to the student (eg, online bullying), (4) access a calendar to book an upcoming videoconference, (5) access a virtual “room” that allowed videoconferencing (camera and microphone being off as default) and private text-based chatting with the moderator, and (6) a resource page with contact information for various social and health services.

Once the moderator logged in, they had access to the same options as the students, as well as several additional features: the ability to delete any message on the discussion boards deemed potentially distressing to other users, the ability to populate the calendar with dates and times for upcoming videoconference sessions, the ability to start a videoconference session (camera turned on by default), and the ability to respond privately to incoming text messages in the videoconferencing virtual “room.”

The moderator had weekly supervision sessions with the team psychologist (PR) to optimize responses to the videoconferences and submitted weekly written reports (without individual names) about topics raised by students and responses to them. The content of the modules and the platform structure remained unchanged during the 8-week intervention; the name of the university that received the research grant and the name of the partnering information technology company appeared on the main page of the platform.

Study Outcomes and Measures Used

The primary outcomes were depression, anxiety, and perceived symptoms of stress. Depression symptoms were assessed with the PHQ-9 [38], where each item is rated on a 0-3 scale and total scores range from 0 to 27 (0-4 indicates minimal/subclinical depression, 5-9 indicates mild depression, 10-14 moderate, 15-19 moderately severe, and ≥20 indicates severe depression). Anxiety was measured using the 21-item BAI [39], where each item is rated on a 0-3 scale and the total score range is 0 to 63 (0-7 indicates a minimal anxiety level, 8-15 a mild anxiety level,
16-25 a moderate anxiety level, and 26-63 a severe anxiety level). For the measurement of stress, we used the 10-item PSS [40], where each item is rated on a 0-4 scale, and the total score range is 0 to 40 (scores of 0-13 indicate mild levels of stress, 14-26 moderate, and 27-40 high). The secondary outcomes were quality of life, life satisfaction, and mindfulness. We used the 16-item Quality of Life Scale (QOLS) [41]; each item is rated on a scale from 1 to 7 and the total score ranges from 16 to 112. Student life satisfaction was measured using the 6-item Brief Multidimensional Students’ Life Satisfaction Scale-Peabody Treatment Progress Battery (BMSLSS-PTPB) [42]; each item is rated on a scale from 1 to 5, and item scores are averaged together to give a total score that ranges from 1 to 5. The level of mindfulness was measured by the 24-item FFMQ-SF [43]; each item is rated on a scale from 1 to 5, and the total score range is 24 to 120. The subscales in the FFMQ-SF are nonreactivity to inner experience (5 items), observing (4 items), acting with awareness (5 items), describing (5 items), and nonjudging of inner experience (5 items). We assessed each of the scales for internal consistency within the T1 and T2 data sets, and Cronbach α ranged from .87 and .90 for PHQ-9, .94 and .90 for BAI, .88 and .90 for PSS; and .85 and .84 for FFMQ-SF at T1 and T2 respectively. Participants also completed a sociodemographic questionnaire at T1 that inquired about age, gender, birth country, years lived in Canada, first language, relationship status, and ethnic heritage.

All outcomes and other variables were measured by the self-report questionnaires at T1 and T2. The primary RCT outcomes were depression, anxiety, and perceived stress, while mindfulness was measured as a secondary outcome. It was hypothesized that symptom scores for depression, anxiety, and stress at T2 would be significantly lower in the MVC group compared with the WLC group, and that mindfulness scores at T2 would be significantly higher in the MVC group (compared with WLC). The outcomes were measured with the following validated scales: PHQ-9 [40], BAI [41], PSS [42], and FFMQ-SF [43]. Participants also completed a sociodemographic questionnaire at T1.

**Sample Size and Statistical Analysis**

The sample size was calculated for 80% power and 5% type I error to detect a standardized effect size of 0.5 or larger. The required sample size was 63 students per arm. Our recruitment goal was 80 participants per arm, assuming an attrition rate of ~20%.

Descriptive statistics for demographic and psychological characteristics were calculated at baseline, and potential between-group differences were assessed using chi-square tests of independence for categorical variables, and t tests for numeric variables. To evaluate the impact of the intervention on study outcomes (ie, depression, anxiety, perceived stress, and mindfulness), generalized estimating equations with an AR (1) covariance structure were employed. Outcome evaluations were performed on both an intention-to-treat (ITT) and per-protocol (PP) basis, with the group × time interaction indicating a between-group change in study outcomes. As the missing data were minimal (<10% overall) and considered to be missing at random, multiple imputations were used to estimate missing observations. For all outcomes, effect sizes are presented for between-group comparisons at follow-up, evaluated by Cohen d and calculated as difference between MVC and WLC means divided by their pooled standard deviations. Cohen d effect size for between-group and within-group (repeated measures) comparisons are calculated according to procedures outlined in Lakens [44].

**Results**

**Participants**

Study participants were 154 undergraduate university students randomized to MVC (n=76) or WLC (n=78) conditions. Of the 154 participants with complete baseline assessments, 7 participants (9.2%) within the MVC and 1 (1.3%) WLC participant dropped out of the study without completing follow-up assessments. As indicated in Table 2, there were no statistically significant between-group differences at baseline in age, gender, country of birth, first language, relationship status, and ethnicity, indicating similar proportions of demographic characteristics across the MVC and WLC groups. Similarly, there were no statistically significant between-group differences in health status and access to private mental health services, with 74.7% of participants across both groups indicating good/very good/excellent health, and 60.4% of participants (across both groups) indicating no access to private mental health services. In addition, there were no significant between-group differences in hours spent at work (paid and unpaid) and in physical activity. In relation to psychological characteristics, no statistically significant between-group differences were detected at baseline between MVC and WLC scores on PHQ-9 (P=.88), BAI (P=.86), PSS (P=.60), and FFMQ-SF (P=.44), and in the following subscales: nonreactivity (P=.87), observing (P=.63), acting with awareness (P=.56), describing (P=.40), and nonjudgment (P=.44).
Table 2. Participants’ demographic characteristics at baseline.

<table>
<thead>
<tr>
<th>Demographic characteristics</th>
<th>All (n=154)</th>
<th>Mindfulness Virtual Community (n=76)</th>
<th>Wait-list control (n=78)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)a</td>
<td>23.10 (8.09)</td>
<td>22.02 (5.52)</td>
<td>24.18 (9.95)</td>
<td>.10</td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Male</td>
<td>35 (22.7)</td>
<td>18 (23.7)</td>
<td>17 (21.8)</td>
<td>.96</td>
</tr>
<tr>
<td>- Female</td>
<td>117 (76.0)</td>
<td>57 (75.0)</td>
<td>60 (76.9)</td>
<td></td>
</tr>
<tr>
<td>- Other</td>
<td>2 (1.3)</td>
<td>1 (1.3)</td>
<td>1 (1.3)</td>
<td></td>
</tr>
<tr>
<td>Country of birth, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Canada</td>
<td>91 (59.1)</td>
<td>46 (60.5)</td>
<td>45 (57.7)</td>
<td>.72</td>
</tr>
<tr>
<td>- Other</td>
<td>63 (40.9)</td>
<td>30 (39.5)</td>
<td>33 (42.3)</td>
<td></td>
</tr>
<tr>
<td>Years in Canada, mean (SD)b</td>
<td>10.73 (7.05)</td>
<td>11.05 (6.21)</td>
<td>10.42 (7.82)</td>
<td>.73</td>
</tr>
<tr>
<td>First language, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- English</td>
<td>103 (66.9)</td>
<td>50 (65.8)</td>
<td>53 (67.9)</td>
<td>.78</td>
</tr>
<tr>
<td>- Other</td>
<td>51 (33.1)</td>
<td>26 (34.2)</td>
<td>25 (32.1)</td>
<td></td>
</tr>
<tr>
<td>Relationship status, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Single, not in a relationship</td>
<td>79 (51.3)</td>
<td>40 (52.6)</td>
<td>39 (50.0)</td>
<td>.70</td>
</tr>
<tr>
<td>- Single, in a relationship</td>
<td>56 (36.4)</td>
<td>27 (35.5)</td>
<td>29 (37.2)</td>
<td></td>
</tr>
<tr>
<td>- Married or common law</td>
<td>13 (8.4)</td>
<td>5 (6.6)</td>
<td>8 (10.3)</td>
<td></td>
</tr>
<tr>
<td>- Divorced, separated, widowed, or other</td>
<td>6 (3.9)</td>
<td>4 (5.3)</td>
<td>2 (2.6)</td>
<td></td>
</tr>
<tr>
<td>Ethnicity, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- White</td>
<td>40 (26.0)</td>
<td>17 (22.4)</td>
<td>23 (29.5)</td>
<td>.69</td>
</tr>
<tr>
<td>- Black</td>
<td>25 (16.2)</td>
<td>11 (14.5)</td>
<td>14 (17.9)</td>
<td></td>
</tr>
<tr>
<td>- South Asian</td>
<td>37 (24.0)</td>
<td>21 (27.6)</td>
<td>16 (20.5)</td>
<td></td>
</tr>
<tr>
<td>- Other</td>
<td>35 (22.7)</td>
<td>19 (25.0)</td>
<td>16 (20.5)</td>
<td></td>
</tr>
<tr>
<td>- Multiple ethnicities</td>
<td>17 (11.0)</td>
<td>8 (10.5)</td>
<td>9 (11.5)</td>
<td></td>
</tr>
<tr>
<td>Self-rated health, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Poor or fair</td>
<td>39 (25.3)</td>
<td>23 (30.3)</td>
<td>16 (20.5)</td>
<td>.38</td>
</tr>
<tr>
<td>- Good</td>
<td>60 (39.0)</td>
<td>28 (36.8)</td>
<td>32 (41.0)</td>
<td></td>
</tr>
<tr>
<td>- Very good or excellent</td>
<td>55 (35.7)</td>
<td>25 (32.9)</td>
<td>30 (38.5)</td>
<td></td>
</tr>
<tr>
<td>Access to private mental health care, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Yes</td>
<td>61 (39.6)</td>
<td>31 (40.8)</td>
<td>30 (38.5)</td>
<td>.77</td>
</tr>
<tr>
<td>- No</td>
<td>93 (60.4)</td>
<td>45 (59.2)</td>
<td>48 (61.5)</td>
<td></td>
</tr>
<tr>
<td>Weekly hours of activities, mean (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Paid work</td>
<td>7.78 (10.60)</td>
<td>7.76 (9.22)</td>
<td>7.81 (11.84)</td>
<td>.98</td>
</tr>
<tr>
<td>- Unpaid work (including volunteer work)</td>
<td>2.87 (5.42)</td>
<td>2.89 (5.81)</td>
<td>2.85 (5.04)</td>
<td>.96</td>
</tr>
<tr>
<td>- Vigorous physical activitiesc</td>
<td>1.68 (1.87)</td>
<td>1.75 (1.76)</td>
<td>1.62 (1.97)</td>
<td>.66</td>
</tr>
<tr>
<td>Patient Health Questionnaire score, mean (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- 0-9</td>
<td>9.60 (6.14)</td>
<td>9.53 (5.83)</td>
<td>9.67 (6.50)</td>
<td>.88</td>
</tr>
<tr>
<td>- ≥10</td>
<td>77 (50.0)</td>
<td>39 (51.3)</td>
<td>38 (48.7)</td>
<td>.74</td>
</tr>
<tr>
<td>Beck Anxiety Inventory score, mean (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- 0-21 (low)</td>
<td>16.88 (13.32)</td>
<td>16.68 (13.56)</td>
<td>17.06 (13.17)</td>
<td>.86</td>
</tr>
<tr>
<td>- 22-35 (moderate)</td>
<td>100 (65.0)</td>
<td>49 (64.5)</td>
<td>51 (65.4)</td>
<td>.69</td>
</tr>
<tr>
<td>- ≥36 (high)</td>
<td>35 (22.7)</td>
<td>16 (21.0)</td>
<td>19 (24.4)</td>
<td></td>
</tr>
<tr>
<td>- 16.88 (13.32)</td>
<td>19 (12.3)</td>
<td>11 (14.5)</td>
<td>8 (10.2)</td>
<td></td>
</tr>
<tr>
<td>Demographic characteristics</td>
<td>All (n=154)</td>
<td>Mindfulness Virtual Community (n=76)</td>
<td>Wait-list control (n=78)</td>
<td>P value</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>------------</td>
<td>-------------------------------------</td>
<td>-------------------------</td>
<td>---------</td>
</tr>
<tr>
<td><strong>Perceived Stress Scale score, mean (SD)</strong></td>
<td>21.30 (7.66)</td>
<td>21.63 (7.72)</td>
<td>20.99 (7.63)</td>
<td>.60</td>
</tr>
<tr>
<td>0-13 (low)</td>
<td>25 (16.2)</td>
<td>11 (14.5)</td>
<td>14 (17.9)</td>
<td>.74</td>
</tr>
<tr>
<td>14-26 (moderate)</td>
<td>90 (58.4)</td>
<td>44 (57.9)</td>
<td>46 (59.0)</td>
<td></td>
</tr>
<tr>
<td>27-40 (high)</td>
<td>39 (25.3)</td>
<td>21 (27.6)</td>
<td>18 (23.1)</td>
<td></td>
</tr>
<tr>
<td><strong>Five-Facet Mindfulness Questionnaire score, mean (SD)</strong></td>
<td>74.15 (12.42)</td>
<td>73.36 (11.22)</td>
<td>74.92 (13.52)</td>
<td>.44</td>
</tr>
</tbody>
</table>

*aBased on n=152 participants (missing data on n=2).
*bBased on n=63 participants born outside of Canada.
*cBased on n=151 participants (missing data on n=3).

**Outcome Evaluations**

Table 3 shows the means and standard deviations for study outcomes at baseline and at 8-week follow-up for the MVC and WLC groups. Table 3 further includes effect sizes for the mean difference between MVC and WLC at 8-week follow up. The PHQ-9 and the FFMQ-SF exhibited negligible between-group differences at 8 weeks, while the between-group effect sizes for the BAI, PSS, and FFMQ-SF nonjudgment subscale were in the small range (d=0.20-0.24).
Table 3. Descriptive statistics for depression, anxiety, stress, and mindfulness scores.

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Intention to treat</th>
<th></th>
<th></th>
<th>Per protocol</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mindfulness Virtual Community, mean (SD)</td>
<td></td>
<td></td>
<td>Mindfulness Virtual Community, mean (SD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>n=76</td>
<td>n=78</td>
<td></td>
<td>n=69</td>
<td>n=77</td>
<td></td>
</tr>
<tr>
<td>Patient Health Questionnaire-9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>9.67 (6.50)</td>
<td>9.53 (5.83)</td>
<td>N/A&lt;sup&gt;a&lt;/sup&gt;</td>
<td>9.80 (6.63)</td>
<td>9.42 (5.78)</td>
<td>N/A</td>
</tr>
<tr>
<td>8 weeks</td>
<td>7.76 (6.12)</td>
<td>8.05 (6.25)</td>
<td>0.05</td>
<td>7.81 (6.41)</td>
<td>8.05 (6.30)</td>
<td>0.04</td>
</tr>
<tr>
<td>Beck Anxiety Inventory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>16.68 (13.56)</td>
<td>17.06 (13.17)</td>
<td>N/A</td>
<td>16.91 (13.56)</td>
<td>16.91 (13.19)</td>
<td>N/A</td>
</tr>
<tr>
<td>8 weeks</td>
<td>12.15 (10.50)</td>
<td>14.58 (12.29)</td>
<td>0.21</td>
<td>12.29 (10.84)</td>
<td>14.61 (12.37)</td>
<td>0.20</td>
</tr>
<tr>
<td>Perceived Stress Scale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>21.63 (7.72)</td>
<td>20.99 (7.63)</td>
<td>N/A</td>
<td>21.52 (7.74)</td>
<td>20.99 (7.68)</td>
<td>N/A</td>
</tr>
<tr>
<td>8 weeks</td>
<td>18.44 (7.51)</td>
<td>20.11 (7.83)</td>
<td>0.22</td>
<td>18.28 (7.82)</td>
<td>20.12 (7.88)</td>
<td>0.24</td>
</tr>
<tr>
<td>Five Facets Mindfulness Questionnaire-Short Form</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>73.35 (11.22)</td>
<td>74.92 (13.52)</td>
<td>N/A</td>
<td>73.22 (11.10)</td>
<td>74.92 (13.61)</td>
<td>N/A</td>
</tr>
<tr>
<td>8 weeks</td>
<td>75.65 (10.08)</td>
<td>75.89 (12.49)</td>
<td>0.02</td>
<td>75.62 (10.58)</td>
<td>75.89 (12.57)</td>
<td>0.02</td>
</tr>
<tr>
<td>Five Facets Mindfulness Questionnaire-Short Form, Nonreactivity subscale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>14.33 (3.79)</td>
<td>14.22 (4.31)</td>
<td>N/A</td>
<td>14.20 (3.77)</td>
<td>14.20 (4.34)</td>
<td>N/A</td>
</tr>
<tr>
<td>8 weeks</td>
<td>14.65 (3.26)</td>
<td>14.69 (3.58)</td>
<td>0.01</td>
<td>14.67 (3.39)</td>
<td>14.68 (3.60)</td>
<td>0.003</td>
</tr>
<tr>
<td>Five Facets Mindfulness Questionnaire-Short Form, Observing subscale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>13.38 (3.31)</td>
<td>13.65 (3.64)</td>
<td>N/A</td>
<td>13.38 (3.33)</td>
<td>13.69 (3.65)</td>
<td>N/A</td>
</tr>
<tr>
<td>8 weeks</td>
<td>13.46 (3.18)</td>
<td>13.90 (3.43)</td>
<td>0.13</td>
<td>13.45 (3.33)</td>
<td>13.90 (3.45)</td>
<td>0.13</td>
</tr>
<tr>
<td>Five Facets Mindfulness Questionnaire-Short Form, Acting with Awareness subscale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>15.54 (4.12)</td>
<td>15.92 (4.08)</td>
<td>N/A</td>
<td>15.72 (4.10)</td>
<td>15.92 (4.11)</td>
<td>N/A</td>
</tr>
<tr>
<td>8 weeks</td>
<td>16.10 (3.87)</td>
<td>16.39 (4.08)</td>
<td>0.07</td>
<td>16.09 (4.05)</td>
<td>16.40 (4.11)</td>
<td>0.08</td>
</tr>
<tr>
<td>Five Facets Mindfulness Questionnaire-Short Form, Describing subscale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>16.32 (4.16)</td>
<td>16.89 (4.15)</td>
<td>N/A</td>
<td>16.22 (4.30)</td>
<td>16.90 (4.17)</td>
<td>N/A</td>
</tr>
<tr>
<td>8 weeks</td>
<td>16.92 (3.51)</td>
<td>17.19 (4.26)</td>
<td>0.07</td>
<td>16.90 (3.68)</td>
<td>17.18 (4.29)</td>
<td>0.07</td>
</tr>
<tr>
<td>Five Facets Mindfulness Questionnaire-Short Form, Nonjudgement subscale</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>13.79 (3.28)</td>
<td>14.24 (3.99)</td>
<td>N/A</td>
<td>13.69 (3.28)</td>
<td>14.22 (4.01)</td>
<td>N/A</td>
</tr>
<tr>
<td>8 weeks</td>
<td>14.52 (3.48)</td>
<td>13.72 (3.69)</td>
<td>0.22</td>
<td>14.52 (3.65)</td>
<td>13.73 (3.71)</td>
<td>0.22</td>
</tr>
</tbody>
</table>

<sup>a</sup>N/A: not applicable.

Table 4 presents the results for ITT and PP models on the impact of the intervention on PHQ-9, BAI, PSS, and FFMQ-SF scores, including FFMQ-SF subscale scores. Except for PSS, there were no statistically significant reductions compared to WLC. For PSS, compared to the WLC group, the MVC group had a statistically significant reduction with both ITT and PP analytic approaches (ITT: β=−2.31, P=.03; PP: β=−2.38, P=.03). When compared to the WLC group, the MVC group also exhibited a slight increase in the FFMQ-SF nonjudgment subscale (ITT: β=1.25, P=.06; PP: β=1.31, P=.06).
### Table 4. Generalized estimating equations for depression, anxiety, stress, and mindfulness outcomes from baseline to 8 weeks.

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Intention to treat</th>
<th>Per protocol</th>
<th>P value</th>
<th>Per protocol</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SE), 95% CI</td>
<td></td>
<td></td>
<td>Mean (SE), 95% CI</td>
<td></td>
</tr>
<tr>
<td><strong>Patient Health Questionnaire-9</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>9.53 (0.66), 8.24 to 10.81</td>
<td>&lt;.001</td>
<td>9.42 (0.65), 8.13 to 10.70</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>-1.47 (0.63), -2.70 to -0.25</td>
<td>.02</td>
<td>-1.37 (0.63), -2.59 to -0.14</td>
<td>.03</td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>0.14 (0.99), -1.79 to 2.08</td>
<td>.88</td>
<td>0.38 (1.03), -1.63 to 2.40</td>
<td>.71</td>
<td></td>
</tr>
<tr>
<td>Group × Time</td>
<td>-0.44 (0.95), -2.31 to 1.43</td>
<td>.64</td>
<td>-0.62 (0.99), -2.57 to 1.32</td>
<td>.53</td>
<td></td>
</tr>
<tr>
<td><strong>Beck Anxiety Inventory</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>17.06 (1.48), 14.16 to 19.97</td>
<td>&lt;.001</td>
<td>16.91 (1.49), 13.98 to 19.84</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>-2.48 (1.51), -5.44 to 0.49</td>
<td>.10</td>
<td>-2.30 (1.52), -5.28 to 0.68</td>
<td>.13</td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>0.64 (1.23), -1.77 to 3.05</td>
<td>.86</td>
<td>0.004 (2.20), -4.31 to 4.32</td>
<td>.99</td>
<td></td>
</tr>
<tr>
<td>Group × Time</td>
<td>-2.06 (2.02), -6.03 to 1.92</td>
<td>.31</td>
<td>-2.32 (2.09), -6.42 to 1.77</td>
<td>.27</td>
<td></td>
</tr>
<tr>
<td><strong>Perceived Stress Scale</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>20.99 (0.86), 19.31 to 22.67</td>
<td>&lt;.001</td>
<td>20.99 (0.87), 19.28 to 22.69</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>-0.88 (0.69), -2.24 to 0.48</td>
<td>.20</td>
<td>-0.87 (0.70), -2.24 to 0.50</td>
<td>.22</td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>0.64 (1.23), -1.77 to 3.05</td>
<td>.86</td>
<td>0.53 (1.27), -1.95 to 3.02</td>
<td>.67</td>
<td></td>
</tr>
<tr>
<td>Group × Time</td>
<td>-2.31 (1.06), -4.39 to -0.24</td>
<td>.03</td>
<td>-2.38 (1.10), -4.54 to -0.22</td>
<td>.03</td>
<td></td>
</tr>
<tr>
<td><strong>Five Facets Mindfulness Questionnaire-Short Form</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>74.92 (1.52), 71.94 to 77.90</td>
<td>&lt;.001</td>
<td>74.92 (1.54), 71.90 to 77.94</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>0.97 (0.99), -0.97 to 2.90</td>
<td>.33</td>
<td>0.97 (1.00), -0.99 to 2.93</td>
<td>.33</td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>-1.57 (1.99), -5.46 to 2.33</td>
<td>.43</td>
<td>-1.70 (2.03), -5.69 to 2.28</td>
<td>.40</td>
<td></td>
</tr>
<tr>
<td>Group × Time</td>
<td>1.33 (1.69), -1.99 to 4.64</td>
<td>.43</td>
<td>1.44 (1.75), -1.99 to 4.87</td>
<td>.41</td>
<td></td>
</tr>
<tr>
<td><strong>Five Facets Mindfulness Questionnaire-Short Form, Nonreactivity subscale</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>14.22 (0.49), 13.27 to 15.17</td>
<td>&lt;.001</td>
<td>14.20 (0.49), 13.23 to 15.16</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>0.47 (0.38), -0.28 to 1.22</td>
<td>.22</td>
<td>0.48 (0.39), -0.28 to 1.24</td>
<td>.21</td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>0.11 (0.65), -1.16 to 1.39</td>
<td>.86</td>
<td>0.008 (0.66), -1.30 to 1.32</td>
<td>.99</td>
<td></td>
</tr>
<tr>
<td>Group × Time</td>
<td>-0.15 (0.64), -1.40 to 1.10</td>
<td>.81</td>
<td>-0.02 (0.67), -1.32 to 1.29</td>
<td>.98</td>
<td></td>
</tr>
<tr>
<td><strong>Five Facets Mindfulness Questionnaire-Short Form, Observing subscale</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>13.65 (0.41), 12.85 to 14.46</td>
<td>&lt;.001</td>
<td>13.69 (0.41), 12.88 to 14.50</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>0.25 (0.32), -0.39 to 0.88</td>
<td>.45</td>
<td>0.21 (0.32), -0.43 to 0.85</td>
<td>.52</td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>-0.27 (0.56), -1.36 to 0.82</td>
<td>.63</td>
<td>-0.31 (0.57), -1.44 to 0.81</td>
<td>.59</td>
<td></td>
</tr>
<tr>
<td>Group × Time</td>
<td>-0.16 (0.53), -1.20 to 0.88</td>
<td>.76</td>
<td>-0.14 (0.56), -1.23 to 0.95</td>
<td>.80</td>
<td></td>
</tr>
<tr>
<td><strong>Five Facets Mindfulness Questionnaire-Short Form, Acting with Awareness subscale</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>15.92 (0.46), 15.02 to 16.82</td>
<td>&lt;.001</td>
<td>15.92 (0.46), 15.01 to 16.83</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>0.47 (0.43), -0.38 to 1.31</td>
<td>.28</td>
<td>0.48 (0.44), -0.38 to 1.33</td>
<td>.27</td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>-0.38 (0.66), -1.67 to 0.91</td>
<td>.56</td>
<td>-0.20 (0.68), -1.52 to 1.13</td>
<td>.77</td>
<td></td>
</tr>
<tr>
<td>Group × Time</td>
<td>0.09 (0.66), -1.21 to 1.39</td>
<td>.89</td>
<td>-0.11 (0.68), -1.46 to 1.23</td>
<td>.87</td>
<td></td>
</tr>
<tr>
<td><strong>Five Facets Mindfulness Questionnaire-Short Form, Describing subscale</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>16.89 (0.47), 15.97 to 17.80</td>
<td>&lt;.001</td>
<td>16.90 (0.47), 15.97 to 17.82</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Time</td>
<td>0.30 (0.34), -0.36 to 0.97</td>
<td>.37</td>
<td>0.28 (0.34), -0.39 to 0.95</td>
<td>.41</td>
<td></td>
</tr>
<tr>
<td>Group</td>
<td>-0.57 (0.66), -1.87 to 0.73</td>
<td>.39</td>
<td>-0.68 (0.70), -2.05 to 0.69</td>
<td>.33</td>
<td></td>
</tr>
<tr>
<td>Group × Time</td>
<td>0.31 (0.50), -0.68 to 1.29</td>
<td>.54</td>
<td>0.40 (0.52), -0.63 to 1.43</td>
<td>.45</td>
<td></td>
</tr>
</tbody>
</table>
### Tables 5 and 6 show additional analyses that focused on discerning which subgroups were affected more or less by the intervention (given varying baseline depression, anxiety, and perceived stress levels). Across 3 levels of depressive symptoms per PHQ-9 categories (ie, mild, moderate, and severe symptom levels), intervention group reductions were greater than those of the control group. An exception was the moderately severe category, where the greater effect size observed was in the controls. We have highlighted the Cohen d effect sizes to facilitate comparisons. The moderately severe category has modest effect size differences.

When similar analyses involved female-male comparisons, the effect sizes and mean reductions were greater in females across mild, moderate, moderately severe, and severe categories. Reductions in cell size are noted, particularly for the male data, as the female study sample far exceeded the male sample (Tables 7 and 8).

### Table 5. Analysis of Patient Health Questionnaire-9 subgroups: intervention.

<table>
<thead>
<tr>
<th>Patient Health Questionnaire-9, intervention</th>
<th>Minimal (0-4)</th>
<th>Mild (5-9)</th>
<th>Moderate (10-14)</th>
<th>Moderately severe (15-19)</th>
<th>Severe (20-27)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (T1/T2)</td>
<td>2.20/4.86</td>
<td>6.84/5.41</td>
<td>11.95/8.79</td>
<td>16.54/12.70</td>
<td>23.33/12.33</td>
</tr>
<tr>
<td>Mean difference, Cohen d</td>
<td>2.66, 0.64</td>
<td>−1.43, 0.50</td>
<td>−3.16, 0.67</td>
<td>−3.84, 0.55</td>
<td>−11.0, 2.18</td>
</tr>
</tbody>
</table>

### Table 6. Analysis of Patient Health Questionnaire-9 subgroups: control.

<table>
<thead>
<tr>
<th>Patient Health Questionnaire-9, wait-list control</th>
<th>Minimal (0-4)</th>
<th>Mild (5-9)</th>
<th>Moderate (10-14)</th>
<th>Moderately severe (15-19)</th>
<th>Severe (20-27)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (T1/T2)</td>
<td>1.82/2.17</td>
<td>7.19/6.52</td>
<td>11.61/10.53</td>
<td>16.90/12.41</td>
<td>22.50/14.00</td>
</tr>
<tr>
<td>Mean difference, Cohen d</td>
<td>0.35, 0.20</td>
<td>−0.67, 0.19</td>
<td>−1.08, 0.29</td>
<td>−4.49, 0.68</td>
<td>−8.5, 1.52</td>
</tr>
</tbody>
</table>

### Table 7. Analysis of Patient Health Questionnaire-9 subgroups: male.

<table>
<thead>
<tr>
<th>Patient Health Questionnaire-9, intervention, females</th>
<th>Minimal (0-4)</th>
<th>Mild (5-9)</th>
<th>Moderate (10-14)</th>
<th>Moderately severe (15-19)</th>
<th>Severe (20-27)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (T1/T2)</td>
<td>2.36/3.37</td>
<td>7.00/5.31</td>
<td>11.92/7.82</td>
<td>16.44/12.19</td>
<td>22.60/11.00</td>
</tr>
<tr>
<td>Mean difference, Cohen d</td>
<td>1.01, 0.44</td>
<td>−1.69, 0.56</td>
<td>−4.1, 1.24</td>
<td>−4.25, 0.58</td>
<td>−11.6, 3.07</td>
</tr>
</tbody>
</table>

### Table 8. Analysis of Patient Health Questionnaire-9 subgroups: female.

<table>
<thead>
<tr>
<th>Patient Health Questionnaire-9, intervention, males</th>
<th>Minimal (0-4)</th>
<th>Mild (5-9)</th>
<th>Moderate (10-14)</th>
<th>Moderately severe (15-19)</th>
<th>Severe (20-27)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (T1/T2)</td>
<td>2.00/5.00</td>
<td>6.00/5.93</td>
<td>12.00/10.57</td>
<td>17.00/15.00</td>
<td>27.00/19.00</td>
</tr>
<tr>
<td>Mean difference, Cohen d</td>
<td>3.00, 0.97</td>
<td>−0.07, 0.33</td>
<td>−1.43, 0.18</td>
<td>−2.00, 0.37</td>
<td>−8.00, N/A</td>
</tr>
</tbody>
</table>

In established BAI categories (ie, low, moderate, and severe symptoms) intervention group reductions were less than control reductions in both low and high symptom categories, while in the moderate symptom category, intervention reductions were greater (Tables 9 and 10). When analyses focused on comparing females with males, the effect sizes and mean reductions were greater in females across moderate and high categories, while the low category results can be characterized as equal. There were reductions in cell size due to the female-male comparisons (Tables 11 and 12).
In perceived stress, intervention and control comparisons demonstrated, at the moderate and high stress levels, greater reductions in the intervention group with approximately equal reductions observed at the low level (Tables 13 and 14). When analyses focused on female-male comparisons, in the moderate and high levels, females demonstrated more change than males, while in the low group females had a negative intervention effect (their stress increased) while males had no intervention effects (ie, neither positive or negative; Tables 15 and 16).
Once again, the within-group reductions were nearly equivalent (Time 2<Time 1) by a mean (raw) scale score of –2.38. Whereas in Study 1 and 2, WLC scores increased over time (Time 2) in WLC subjects. Cross-study differences were also observable in the BAI indices, though it should be noted that the significant between-group differences absent in Study 3 appeared related to anxiety reductions in the WLC group. In Study 1 and Study 3, the WLC group depression scores (PHQ-9) increased over time (Time 2>Time 1; P<.05). In Study 1, significant between-group differences were found in Study 1 and Study 3, but not in Study 2. In Study 1 and 3, the WLC group scores were nearly equivalent (Time 1 and Time 2) as were the reductions achieved in the intervention group (Study 1: –3.10, Study 2: –3.19, in raw scores).

Altogether, females gained more benefit than males and intervention subjects benefitted more than controls. Gains were most evident in depression and perceived stress reductions. Anxiety results (BAI outcomes) were more variable, likely reflecting the immediate anxiety reduction effects in controls with the suspension of expected exams and assignments.

### Discussion

#### Principal Findings

This study was undertaken at a labor strike–affected university, following identical procedures described in two prior studies at the same site, conducted when no crises existed [4,5]. Accordingly, participant responses to the previous RCTs are usefully compared to responses in this study (referred to as “Study 3” below).

In the previously reported RCTs [4,5], significant between-group differences were observed in depression (PHQ-9), anxiety (BAI), quality of life (QOLS), and mindfulness (FFMQ-SF). In the first RCT, significant between-group differences were additionally found in perceived stress (PSS). In this study, a significant between-group difference was only found in the PSS (P=.03), while a marginal (P=.06) difference was observed in the Nonjudgment subscale of FFMQ-SF.

Despite the limited between-group differences in this disrupted-campus study, significant group × time effects suggested that the MVC intervention had a positive impact on students. However, while in Study 1 and 2 the WLC depression scores (PHQ-9) [4,5] increased over time (Time 2>Time 1; perhaps due to anticipating increasingly demanding academic tasks), in Study 3, the WLC subjects had substantially reduced depression scores (Time 1<Time 2, by –1.4 in mean raw score). As the intervention within-group reductions were nearly equivalent across studies 1, 2, and 3 (4.2, 4.8, and 4.5, respectively) indicating that reductions in the between-group differences in Study 3 were related to anxiety reductions in the WLC group. Other interesting cross-study differences were observable in PSS indices where significant between-group differences were found in Study 1 and Study 3, but not in Study 2. In Study 1 and 3, the WLC group scores were nearly equivalent (Time 1 and Time 2) as were the reductions achieved in the intervention group (Study 1: –3.10, Study 2: –3.19, in raw scores).

One interpretation of such study comparisons is that students experienced the crisis as reducing academic pressure, with the result of them experiencing less distress (represented in lower mean WLC group depression and anxiety scores). While the lowered distress was likely temporary (outcomes were assessed 8 weeks after baseline, and prior to resumed course activity), the pattern might have also affected motivations regarding participation in the intervention. If so, this would have contributed to the lack of significant between-group differences in depression and anxiety in this study. It is notable that while there were no significant between-group differences previously observed in Study 2 for the PSS, there was a significant difference observed in this study (Study 3) and Study 1. It seems that while the burdens of depression and anxiety in the WLC group were reduced in Study 3, there was not an equivalent reduction of the immediate stress represented in PSS items. Several PSS items refer to “unexpected happenings beyond control, anger and irritation due to loss of control,” and positive control (a reversed item) over “important things.” Such items reflect the lost control that can occur during a campus strike. Thus, significant between-group differences favoring the intervention could have occurred because intervention participants felt positive intervention effects related to the disruption caused by the strike. Why PSS differences were found in Study 1 is an interesting question for speculation. Notably, tense, publicized negotiations preceded the strike, during the period when Study 1 was undertaken. Students might have been aware of the increasing strike probability, with that awareness represented on the PSS intervention-attributable differences.

If the academic disruption was experienced as a temporary reduction in distressing academic pressures, the “strike” reaction is congruent with focus group findings preceding the 3 RCTs.

<table>
<thead>
<tr>
<th>Table 15. Analysis of Perceived Stress Scale categories: female.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived Stress Scale, intervention, female</td>
</tr>
<tr>
<td>Mean (T1/T2)</td>
</tr>
<tr>
<td>Mean difference, Cohen d</td>
</tr>
<tr>
<td>Mean difference, Cohen d</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 16. Analysis of Perceived Stress Scale categories: male.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived Stress Scale, intervention, males</td>
</tr>
<tr>
<td>Mean (T1/T2)</td>
</tr>
<tr>
<td>Mean difference, Cohen d</td>
</tr>
<tr>
<td>Mean difference, Cohen d</td>
</tr>
</tbody>
</table>
In closing, student participants in an 8-week RCT demonstrated behavior changes, although the study site was a strike-affected and disrupted campus. In two previously published RCTs at the study site, students experienced the disruption that occurred during the 8-week trial as a reprieve or pressure easing. The RCT demonstrated the assistance received from the intervention following the more immediate loss of control due to life disruptions (ie, strikes).

Several study limitations must be taken into account. This was not a single- or double-blinded trial. The study duration of 8 weeks did not allow us to test for longer-term effects (eg, 6 or 12 months), which would have resulted in a more fully assessed crisis effect. A further limitation is the high female preponderance in the control and intervention groups, as more precise future research projects would address gender differences through stratification with larger participant samples from multiple research sites (involving multiple universities and colleges). Missing data was also a limitation, somewhat mitigated by the use of a multiple imputation method. Lastly and importantly, given the marginal between-group differences found in this study in mindfulness self-report (FFMQ-SF), there was no measure of the participants’ mindfulness practice external to platform participation.

Conclusion

Our results suggest that an 8-week online M-CBT video-based program is an effective intervention for reducing perceived stress among undergraduate university students on a disrupted campus. To some degree, the online M-CBT interventions offer an opportunity to address mental health conditions in postsecondary populations under disrupted campus conditions that might have similarities to COVID-19–related disruptions, which may have resulted in face-to-face academic interactions being suspended.

Acknowledgments

The authors acknowledge the contribution of all students who gave their valuable time to participate in the study. We thank ForaHealthyMe.com Inc. as an industry partner in this project who provided great support. The work reported in this paper was funded by the Canadian Institutes for Health Research (CIHR) eHealth Innovations Partnership Program Grant (eHIPP; Grant No. EH1-143553). The project’s principal investigators are CE (nominated), FA, and PR.

Authors’ Contributions

CE, FA, and PR designed the study and questionnaire, received the funds, and contributed equally. PR led module development, providing written content and voice. MP analyzed the data. PR verified the analysis and prepared the first draft, and all authors provided critical feedback and revised it. The MVC Team members are (alphabetically): Sahir Abbas, BSc; Yvonne Bohr, PhD;
Manuela Ferrari, PhD; Wai Lun Alan Fung MD, ScD, FRCPC; Louise Hartley, PhD; Amin Mawani, PhD; Kwame McKenzie, MD, FRCPC; and Jan E. Odai, BA. These team members made contributions to several aspects of the project and results development. They all approve the final version and agree to be accountable for all aspects of the submitted paper. The trial protocol could be accessed on reasonable request to corresponding authors.

Conflicts of Interest
It is the understanding of the university and researchers that the Project Intellectual Property belongs to CEM, FA, and PR. The industry partner ForaHealthyMe.com owns all rights and title to the copyrights of any computer source code software that was developed out of this research project.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1362 KB - mental_v8i1e23491_app1.pdf]

References
13. Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration Data - National Survey on Drug Use and Health (NSDUH) Results. Substance Use and Mental Health Services Administration. URL: https://www.samhsa.gov/data/ [accessed 2020-07-28]


Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAI</td>
<td>Beck Anxiety Inventory</td>
</tr>
<tr>
<td>BMSLSS-PTPB</td>
<td>Brief Multidimensional Students’ Life Satisfaction Scale-Peabody Treatment Progress Battery</td>
</tr>
<tr>
<td>FFMQ-SF</td>
<td>Five Facets Mindfulness Questionnaire Short Form</td>
</tr>
<tr>
<td>ITT</td>
<td>intention to treat</td>
</tr>
<tr>
<td>M-CBT</td>
<td>mindfulness-based cognitive behavioral therapy</td>
</tr>
<tr>
<td>MVC</td>
<td>Mindfulness Virtual Community</td>
</tr>
<tr>
<td>PHQ-9</td>
<td>Patient Health Questionnaire-9</td>
</tr>
<tr>
<td>PP</td>
<td>per protocol</td>
</tr>
<tr>
<td>PSS</td>
<td>Perceived Stress Scale</td>
</tr>
<tr>
<td>QOLS</td>
<td>Quality of Life Scale</td>
</tr>
<tr>
<td>RCT</td>
<td>randomized controlled trial</td>
</tr>
<tr>
<td>WLC</td>
<td>wait-list control</td>
</tr>
</tbody>
</table>

©Paul Ritvo, Farah Ahmad, Christo El Morr, Meysam Pirbaglou, Rahim Moineddin, MVC Team. Originally published in JMIR Mental Health (http://mental.jmir.org/), 11.01.2021. This is an open-access article distributed under the terms of the Creative
Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Mental Health, is properly cited. The complete bibliographic information, a link to the original publication on http://mental.jmir.org/, as well as this copyright and license information must be included.
Comparing Effectiveness Between a Mobile App Program and Traditional Cognitive Behavior Therapy in Obsessive-Compulsive Disorder: Evaluation Study

Hyunchan Hwang¹, MD; Sujin Bae², PhD; Ji Sun Hong¹, MD, PhD; Doug Hyun Han¹, MD, PhD

¹Department of Psychiatry, Chung-Ang University Hospital, Seoul, Republic of Korea
²Office of Research, Chung-Ang University, Seoul, Republic of Korea

Abstract

Background: This study proposes a digital program for the treatment of mental illness that could increase motivation and improve learning outcomes for patients. Several studies have already applied this method by using an exposure and response prevention–inspired serious game to treat patients with obsessive-compulsive disorder (OCD).

Objective: We hypothesized that a mobile cognitive behavior therapy (CBT) program would be as effective in treating OCD as traditional offline CBT. In addition, the treatment efficacy in response to mobile CBT for OCD might be associated with increased brain activity within the cortico-striato-thalamo-cortical (CSTC) tract.

Methods: The digital CBT treatment program for OCD, OCfree, consists of 6 education sessions, 10 quests, and 7 casual games. Information was gathered from 27 patients with OCD (15 offline CBT and 12 OCfree CBT). During the 6-week intervention period, changes in clinical symptoms and brain function activity were analyzed.

Results: There was no significant difference in the change in OCD symptoms and depressive symptoms between the two groups. However, the OCfree group showed greater improvement in anxiety symptoms compared to the offline CBT group. Both offline CBT and OCfree CBT increased the functional connectivity within the CSTC tract in all patients with OCD. However, CBT using OCfree showed greater changes in brain connectivity within the thalamus and insula, compared to offline CBT.

Conclusions: OCfree, an OCD treatment app program, was effective in the treatment of drug-naïve patients with OCD. The treatment effects of OCfree are associated with increased brain connectivity within the CSTC tract. Multisensory stimulation by education, quests, and games in OCfree increases the activity within the thalamus and insula in patients with OCD.

(JMIR Ment Health 2021;8(1):e23778) doi:10.2196/23778

KEYWORDS
obsessive-compulsive disorder; exposure and response prevention; cognitive behavior therapy; cortico-striato-thalamo-cortical tract; functional connectivity; prevention; cognitive; mental illness; behavior therapy

Introduction

Overview

Obsessive-compulsive disorder (OCD) is a debilitating mental disorder associated with significant social and occupational impairments [1], affecting 2%-3% of the population worldwide [2]. It is diagnosed by the presence of obsessions, compulsions, or both [2]. Obsessions are characterized as intrusive thoughts or images that are often unwanted and repeat constantly. Compulsions, on the other hand, are defined as repetitive behaviors or mental thinking that people feel they need to do, often to counter the obsessions.

Along with obsession and compulsion, functional brain changes in patients with OCD have also been noted. Many studies have
found the cortico-striato-thalamo-cortical (CSTC) tract to be one of the crucial brain circuits involved in OCD [3,4]. Within the CSTC tract, patients with OCD showed different brain activity in response to various stimuli [5,6]. In response to emotion-related tasks, patients with OCD showed overactivation within the anterior cingulate cortex, insula, caudate head, and putamen, which are thought to play a part in salience, arousal, and habit responding [5]. In addition, underactivation was shown in the medial prefrontal cortex and posterior caudate, which are associated with cognitive and behavioral control [5]. Regions outside the CSTC tract are also of interest, as Zhang et al showed altered functional connectivity (FC) in resting-state functional magnetic resonance imaging (Rs-fMRI) between the cerebellum and CSTC circuit in OCD patients [6].

Of several treatment options for OCD, cognitive behavior therapy (CBT) with exposure and response prevention (ERP) is regarded as one of the first choices in many clinical guidelines [7,8]. The National Institute for Health and Care Excellence recommends intensive CBT (over 10 therapist hours including ERP) or a selective serotonin reuptake inhibitor (SSRI) as initial treatment [7]. The American Psychiatric Association also recommends CBT with ERP, SSRIs, or both as first-line treatments [8]. Recent reviews on the literature have found that CBT with ERP has larger effect sizes than pharmacotherapy, although interpreting this as CBT with ERP being better than medication must be avoided as there are many factors to consider [9].

However, CBT with ERP is not without its flaws. One of the main disadvantages of CBT is that it is time consuming, and efficacy is tied to the participant’s engagement with the therapy [10]. Moreover, Barnes et al found that homework, one of the key elements of CBT, was one of the main reasons for low adherence because it is linked to negative school homework experiences [10].

Many methods have been devised to overcome this disadvantage. The delivery of mental health services through the internet is one method being considered [11,12]. This method has the potential to increase patient engagement and availability [11]. In a randomized controlled trial of family-based treatment for OCD, internet delivered cases showed higher response rates than clinic cases, and this difference persisted after treatment, although the difference was not statistically significant [12]. Another method used is serious games. Eichenberg and Schott argued that serious games could increase motivation and improve learning outcomes, thereby supplementing some of the weaknesses of the conventional internet-mediated health program [13].

Serious games are defined as games developed with a purpose other than entertainment [14]. Notable examples of these are games designed to improve aircrew training [15] or education [16]. Serious games are also used in the medical field to help patients. A meta-analysis and systematic review showed that serious games for mental health were effective in reducing symptoms related to depression, autism spectrum disorder, post-traumatic stress disorder, attention-deficit/hyperactivity disorder, and alcohol use disorder [14]. Serious games have also been shown to lower anxiety and related symptoms. Kim et al reported that a serious game helped lower depression and anxiety in breast cancer patients [17]. A pilot study also showed that a serious game helped reduce anxiety and pain in children before day-care surgery [18]. More recently, Hong et al used an ERP-inspired serious game to treat OCD patients; the patients showed improved OCD symptoms, which correlated with increased brain connectivity between the dorsal anterior cingulate cortex and the prefrontal cortex, after 3 weeks of game play [19].

Hypothesis
To our knowledge, a mobile app using serious games with both ERP and CBT for OCD has not yet been developed. Therefore, we designed a mobile app based on these theories and conducted a randomized controlled trial comparing the developed program directly with traditional CBT with ERP in OCD patients. We hypothesized that the mobile CBT program would be as effective in the treatment of OCD as traditional offline CBT. In addition, the treatment efficacy in response to mobile CBT for OCD might be associated with increased brain activity within the CSTC tract.

Methods
Participants
Through advertisements for treatment of OCD, 32 patients with OCD were recruited from the Department of Psychiatry at Chung-Ang University Hospital. All patients with obsession or compulsive symptoms were screened using the structured Clinical Interview for the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), and diagnosed by a psychiatrist (DHH). Inclusion criteria were as follows: (1) age > 18 years, (2) diagnosed with OCD based on DSM-5, (3) drug naïve, and (4) right-handedness. Exclusion criteria were as follows: (1) IQ < 80; (2) history of medical or other psychiatric disorders; (3) history of substance use disorders; (4) contraindications to MRI scanning, including claustrophobia or metal implant; and (5) current psychotherapy or medication treatment.

Of the 32 patients with OCD, 1 patient was excluded due to psychotic symptoms of hallucinations. The remaining 31 patients were randomly classified into two groups: an offline CBT group (n=16) and a web-based CBT using OCfree group (n=15). One patient in the offline CBT group and 1 patient in the OCfree group were excluded due to taking medication for anxiety reduction, and another patient in the OCfree group was excluded due to a brain infarction finding in the baseline fMRI. In the OCfree group, 1 patient did not complete the study protocol because they did not want to visit the hospital during the COVID-19 pandemic. Finally, the information of 27 patients with OCD (15 offline CBT and 12 OCfree CBT) was analyzed. Participants who entered the trial received a maximum of 100,000 won (around US $89) during the whole trial to compensate for travel fees. The institutional review board of Chung-Ang University Hospital approved this study, and all participants provided written informed consent.
Study Design
A randomized and treatment-as-usual controlled design was applied for this study. Individual in-person CBT was selected as treatment-as-usual with reference to the American Psychiatric Association practice guidelines [20]. The guideline recommends CBT or medication (SSRI) as the first-line treatment for OCD, states that individual and group CBT seem equally effective, and mentions that internet-delivered CBT is promising and deserving of further research [20]. After screening, all patients with OCD were randomly assigned to receive offline CBT once per week or web-based CBT using OCfree daily for 6 weeks, according to the randomization sequence generated using SPSS version 24.0 (IBM Corp, Armonk, NY, USA), with a 1:1 allocation (offline CBT:OCfree). At baseline and after intervention, all patients with OCD were assessed with the Korean version of the Yale-Brown Obsessive Compulsive Scale (Y-BOCS) for OCD symptoms [21,22], the Korean version of the Beck Depression Inventory-II (BDI) for depressive symptoms [23,24], and the Korean version of the Beck Anxiety Inventory (BAI) for anxiety symptoms [25,26].

After baseline psychological scales and fMRI data were acquired, the offline CBT group had an hour-long individual session per week with a psychiatrist for 6 weeks. The CBT sessions were designed similarly to the traditional 10-session CBTs but shortened to match the online CBT session numbers. Homework was given and checked for each session.

The OCfree CBT group also had an individual psychiatrist assigned to meet the participant each week and oversee the CBT for 6 weeks. The psychiatrist in the OCfree group used the OCfree program to conduct CBT, and each session lasted approximately 40 minutes. The program has a scheduler system that assigns participants different programs within OCfree for their daily use, and the participant can use different parts of the program as many times as they wish. The psychiatrist would check the program each week for compliance.

OCfree Program
The mobile app (OCfree) for obsessive-compulsive disorder treatment consists of three categories: education, quests, and serious games. The education category consisted of 6 sessions: (1) learning about OCD: learn about symptoms, causes, and treatments; (2) analyzing obsessions: trigger factors for obsession and compulsive behaviors, compulsive infiltration, fearful ending; (3) understanding strategies with OCD; (4) customized treatment plan; (5) factors for change: a firm resolution; and (6) explanation of cognitive therapy. The learning time of each education session was 20-30 minutes. The quest category consists of 10 subcategories. Those are paired with each education session and supplied to patients as homework: (1) assessment of obsession and compulsion using 65 questions for obsession and 65 questions for compulsion, (2) analyzing the symptoms of obsession and compulsion with 8 panels, (3) understanding false beliefs related to obsession, (4) creating a customized treatment plan for you, (5) preparing for change by yourself, (6) working book for cognitive therapy and 4 adjuvant categories, (7) practice postponing anxiety, (8) ERP using imagination via voice recording, (9) identifying emotions, and (10) practicing choosing. The game category consists of 7 casual games: (1) shooting game, (2) break block game, (3) germ-removing game, (4) doubting and checking game, (5) symmetry and ordering game, (6) numbering and counting game, and (7) mental ritual game (Figure 1).

The shooting game is similar to the fixed shooter arcade game, Galaga. By controlling a spaceship, the players can destroy aliens while avoiding enemies’ projectiles. The spaceship is displaced with the object that OCD patients dislike (want to avoid) such as a needle, knife, death, germ, or airplane. The break block game is a modified version of the classic Breakout block game. The players use the paddle to bounce the ball and destroy the bricks. Behind the brick, there is a “word” that is associated with the obsession. When the bricks are completely destroyed, the “word” disappears. Before starting the block game, the OCD patients type the “word” that is associated with the obsession. In the germ-removing game, players can remove pictorial germs represented on the palm of a hand until only one germ remains. The OCD patient should wait 5 seconds before removing the last one. The number of germs increases when the stage is cleared. In the doubting and checking game, players can touch a pictorial faucet on a screen until the number “1” is represented. Like the germ-removing game, OCD patients should wait 5 seconds before the last touch (Figure 1). The objects of the checking can change in accordance with the patients’ obsessions, including a doorknob and gas valve. In the symmetry and ordering game, players can place books on a bookshelf in accordance with size and color. OCD patients should wait 5 seconds before arranging the last one. More books of various sizes and colors are presented after each stage is cleared. In the numbering and counting game, players type a number that they are preoccupied with. This number is the same as the number of eggs on the screen. The players touch the eggs to hatch them until the last egg remains. The OCD patients should wait 5 seconds before touching the last egg to hatch. In the mental ritual game, players type a word that they are preoccupied with. The word then multiplies and spreads on the screen. Of the multiplied words, 10% are modified in a different or wrong spelling (one or two characteristics are different), compared to the original spelling. OCD patients should only touch words written in the original spelling.
Figure 1. OCfree program: (1) learning about OCD: learn about symptoms, causes, and treatments, (2) analyzing obsessions: trigger factors for obsession and compulsive behaviors, compulsive infiltration, fearful ending, (3) understanding strategies with OCD, (4) understanding false beliefs related to obsession, (5) analyzing the symptoms of obsessions and compulsions with 8 panels, (6) creating a customized treatment plan for you, (7) shooting game, (8) break block game, (9) germ-removing game.

Brain Imaging Data Acquisition and Processing

A 3.0 Tesla Philips Achieva scanner was used to acquire Rs-MRIs. Only right-handed participants entered the trial due to reports showing functional and anatomical differences in the brain between right- and left-handedness [27]. A total of 230 volumes for 720 seconds were gathered using the following parameters: repetition time/echo time=3000/40 milliseconds, 40 slices, 64×64 matrix, 90° flip angle, 230-mm field of view, and 3-mm section thickness, without a gap. Using the programs of the Data Processing Assistant for Rs-fMRI [28] and the Rs-fMRI Data Analysis Toolkit (REST) [29], all acquired imaging data were prepared for preprocessing and processing. Brain activity within regions of interest (ROIs) was derived...
from the fractional amplitude of low-frequency fluctuations (fALFF), extracted using REST software. Seed-based FC analysis was performed using the seed ROI extracted from the previous step of correlation comparison between Y-BOCS and fALFF. More details of the fMRI data preprocessing and processing were described in our previous study [30].

**Statistical Analysis**

Demographic and clinical characteristics of the offline and OCfree groups were analyzed using the Mann-Whitney U test. The differences in sex ratio and symptom improvement between the two groups were analyzed using a chi-square test. Symptom improvement was defined as a decrease of 2.85 in 25% of Y-BOCS scores based on a reliable change index of symptoms (standard error of measure=1.03, effect size=0.91) [31]. Statistical significance was set at $P<.05$.

**Results**

**Demographic Characteristics and Clinical Scales**

There were no significant differences in age, years of education, Y-BOCS scores, BDI scores, and BAI scores between the offline CBT and OCfree groups at baseline (Table 1).

**Table 1. Demographic data and clinical scales.**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>OCfree group (n=12)</th>
<th>Offline CBT$^a$ group (n=15)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), mean (SD)</td>
<td>25.7 (7.7)</td>
<td>24.7 (10.7)</td>
</tr>
<tr>
<td>Sex, n</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Female</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>Education (years), mean (SD)</td>
<td>13.3 (1.8)</td>
<td>13.8 (2.1)</td>
</tr>
<tr>
<td>Economic status (income)$^b$, n</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Middle</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>High</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Y-BOCS$^c$, mean (SD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretreatment</td>
<td>21.9 (5.7)</td>
<td>19.5 (4.1)</td>
</tr>
<tr>
<td>Posttreatment</td>
<td>16.7 (5.4)</td>
<td>15.9 (4.7)</td>
</tr>
<tr>
<td>BDI$^d$, mean (SD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretreatment</td>
<td>19.3 (4.9)</td>
<td>16.5 (11.2)</td>
</tr>
<tr>
<td>Posttreatment</td>
<td>9.3 (5.3)</td>
<td>10.3 (9.5)</td>
</tr>
<tr>
<td>BAI$^e$, mean (SD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pretreatment</td>
<td>19.8 (10.7)</td>
<td>16.2 (11.7)</td>
</tr>
<tr>
<td>Posttreatment</td>
<td>9.4 (8.6)</td>
<td>11.0 (9.4)</td>
</tr>
</tbody>
</table>

$^a$CBT: cognitive behavior therapy.

$^b$Economic status (income): low, <US $20,000/year; middle, US $20,000–40,000/year; high, >US $40,000/year.

$^c$Y-BOCS: Yale-Brown Obsessive Compulsive Scale.

$^d$BDI: Beck Depression Inventory.

$^e$BAI: Beck Anxiety Inventory.

The number of improved OCD patients in the OCfree group (improvement vs nonimprovement: 8/12, 67% vs 4/12, 33%) was greater than that observed in the offline group (8/15, 53% vs 7/15, 47%), but the difference was not statistically significant ($\chi^2=0.5; P=.69$). There were also no significant differences in the change of Y-BOCS scores ($F=0.50; P=.48$) and BDI scores ($F=2.16; P=.16$) between the two groups. Compared to the offline CBT group, the OCfree group showed greater improvement in BAI scores ($F=5.74; P=.02$) (Figure 2).
Results of the Program Usage and Satisfaction Survey
On average, program compliance was 91.4%. The 6 education modules were used 6.0 times each, and the 10 quest modules were used 10.8 times each during the 6 weeks of program usage. The 7 game modules were used, on average, 12.1 times each.

An anonymous survey was taken at the end of the program, and 10 out of 12 participants completed it. The participants were given, among other things, a choice of 1 to 5 stars to measure overall satisfaction, 1 being the lowest and 5 being the highest. The overall satisfaction was 3.4 stars out of 5. The majority of 2- or 3-star ratings were due to minor errors in the program, giving it a somewhat crude feeling. The 4- and 5-star reviews stated that the program helped them get to know their obsessions and compulsions better. Of the 10 completed surveys, 5 (50%) said that they would like to continue using the program even...
after the trial, and 7 (70%) wished to recommend it to other people with similar symptoms. There have been no reports of adverse effects of the program. Adverse effects were checked by the psychiatrist conducting the CBT sessions each week and not asked about in the anonymous survey.

Correlation Between the Y-BOCS Scale and Brain Activity (fALFF)

In all patients with OCD, Y-BOCS scale scores were negatively correlated with fALFF within the right insular (Talairach code x, y, z: 48, 9, 3; T=5.41; \( P_{\text{uncorrected}}<0.001; k_E=20\), Brodmann area [BA] 13), right parietal supramarginal gyrus (x, y, z: 63, −45, 30; T=4.42; \( P_{\text{uncorrected}}<0.001; k_E=26\), BA 40), and left thalamus (x, y, z: −9, −21, 9; T=4.23; \( P_{\text{uncorrected}}<0.001; k_E=29\) (Figure 3).

Figure 3. Correlation between the Yale-Brown Obsessive Compulsive Scale and brain activity.

Comparison of the Changes in fALFF Between the OCfree Group and Offline CBT Group

During the intervention period, the fALFF within the right middle temporal gyrus (x, y, z: 48, −72, 21; T=5.50; \( P_{\text{uncorrected}}<0.001; k_E=101\), BA 39), right middle temporal gyrus (x, y, z: 54, −39, −12; T=4.41; \( P_{\text{uncorrected}}<0.001; k_E=73\), BA 20), right inferior temporal gyrus (x, y, z: 33, −6, −39; T=4.30; \( P_{\text{uncorrected}}<0.001; k_E=216\), BA 20), right superior frontal gyrus (x, y, z: 15, 54, −9; T=4.24; \( P_{\text{uncorrected}}<0.001; k_E=34\), BA 10), left inferior temporal gyrus (x, y, z: −48, −3, −36; T=4.22; \( P_{\text{uncorrected}}<0.001; k_E=39\), BA 20), and left superior frontal gyrus (x, y, z: −36, 45, 33; T=4.02; \( P_{\text{uncorrected}}<0.001; k_E=31\), BA 9) had increased in all patients with OCD (Figure 4).

During the intervention period, the OCfree group showed increased fALFF within the right parahippocampal gyrus (x, y, z: 39, −42, −6; T=4.60; \( P_{\text{uncorrected}}<0.001; k_E=35\), BA 19), compared to the offline CBT group.
Comparison of the Changes in FC Between the OCfree Group and Offline CBT Group

During the intervention period, the FC from the left thalamus to the right cerebellar tonsil (x, y, z: 12, −51, −45; T=3.18; $P_{\text{uncorrected}}<0.001$; $k_E=24$), right cerebellar inferior semilunar lobule (x, y, z: 30, −75, −39; T=3.08; $P_{\text{uncorrected}}<0.001$; $k_E=26$), and right insular (x, y, z: 36, 18, 3; T=3.01; $P_{\text{uncorrected}}<0.001$; $k_E=21$) had increased in all patients with OCD (Figure 5).

During the intervention period, the FC from the left thalamus to the left frontal rectal gyrus (x, y, z: −3, 24, −33; T=3.51; $P_{\text{uncorrected}}<0.001$; $k_E=30$), left inferior frontal gyrus (x, y, z: −48, 39, −12; T=3.39; $P_{\text{uncorrected}}<0.001$; $k_E=53$, BA 47), and left occipital lobe (x, y, z: 0, −69, 33; T=3.25; $P_{\text{uncorrected}}<0.001$; $k_E=39$, BA 7) was increased in the offline CBT group (Figure 6).

Figure 4. Comparison of the changes in fractional amplitude of low-frequency fluctuations between the OCfree and offline cognitive behavior therapy groups.

Figure 5. The changes in functional connectivity in all the patient groups (OCfree and offline cognitive behavior therapy group).
Discussion

Principal Findings

OCfree, an OCD treatment app program, was as effective at improving OCD symptoms as offline CBT for OCD. The severity of OCD assessed with the Y-BOCS scale in all patients was negatively associated with brain activity within the emotion perception network, including the thalamus and insular. Both offline CBT and OCfree CBT improved OCD symptoms and increased FC within the CSTC tract in all patients with OCD.
However, CBT using OCfree showed greater changes in fALFF within the thalamus and insular, compared to offline CBT.

The Effectiveness of OCfree, an OCD Treatment Program, on the Improvement of OCD Symptoms

The OCfree program was as effective as offline CBT for the improvement of OCD symptoms in OCD patients. Overall compliance with OCfree was 91.4% and overall satisfaction was rated 3.4 out of 5 stars. The internet delivery CBT system for OCD has already been reported to be as effective as offline CBT [12]. In our previous study, an ERP-inspired serious game for OCD improved symptoms in OCD patients [19]. Moreover, OCfree greatly improved anxiety compared to offline CBT. We believe that the OCFree web-based delivery system, including education and quests, may enable easy patient access to the treatment system, and serious games in OCfree may increase interest in treatment, as immersion is thought to be one of the merits of serious games. Easy access and frequent contact with patients in OCD management decrease patient anxiety [32-34]. In mood disorders, internet-assisted cognitive behavioral therapy is becoming an evidence-based cognitive treatment [35]. Serious game-assisted clinical treatments already suggest that serious games can increase affinity and treatment compliance in many areas, including cancer [36], obesity [37], and autism spectrum disorders [38].

Comparison of the Changes in fALFF Between the OCfree and Offline CBT Group

Comparisons before and after the treatment period in this study showed increased brain activity within the frontal and temporal lobes in all OCD patient groups. Compared to the offline CBT group, the OCfree group showed increased brain activity within the right parahippocampal gyrus. A deficit in emotional perception was reported in patients with OCD [39]. Due to this deficit, repetitive meaningless thoughts occurred in patients with OCD [39]. In several fMRI studies of OCD patients, decreased brain activity within the frontal and temporal lobes has been reported [40,41]. Chen et al reported that patients with OCD showed decreased brain activity within the left medial prefrontal cortex, compared to healthy subjects [42]. In patients with obsessive and compulsive symptoms due to temporal lobe infarction, SSRI treatment would improve the obsessive and compulsive symptoms [43]. The parahippocampal gyrus is known to play a crucial role in the perception of emotion [44]. Within the retrosplenial and posterior cingulate gyri, the parahippocampal gyrus is thought to play a crucial role in facial expression recognition [44]. Taken together, we believe that the OCfree program would improve OCD symptoms in patients with OCD. Moreover, the improvement may be due to the increased brain activity within the brain regions associated with emotional perception.

Comparison of the Changes in FC Between the OCfree and Offline CBT Group

The brain’s FC within the CSTC tract was increased in all patient groups. In several studies, altered (disconnected) brain FC within the CSTC tract in OCD patients has already been reported [6,45]. Based on these results, we suggest that both OCfree and offline CBT may present similar treatment mechanisms of increased FC within the CSTC tract.

Interestingly, compared to the offline CBT group, the OCfree group showed increased brain FC from the thalamus and insular. This result may be associated with various and multiple sources of sensory stimulation via education, quests, and games during play. The thalamus is thought to act as a hub that receives sensory signals from every sensory system and sends them to the associated cortex [46]. The insular functions are associated with sensorimotor processing, socioemotional processing, and cognitive functions [47]. Decreased brain activity within the thalamus [48] and insular [49] have been reported in patients with OCD. Considered together, OCfree with various stimulation systems may increase the thalamus and insular activity, compared to offline CBT. Although there was no difference in the efficacy of treatment between OCfree and offline CBT, future studies with a larger number of subjects may show greater efficacy in symptom improvement with OCfree, compared to offline CBT.

Limitations

There were several limitations to this study. First, the small sample size and short-term intervention period were not sufficient for generalizing the results. Second, due to the exclusion criteria for medication use, OCD patients with severe anxiety symptoms were excluded from this study. These exceptions can affect the results of the anxiety comparison between the two groups. In addition, although we could not find any individual with co-occurrence of OCD and claustrophobia, lifetime comorbidity between OCD and specific phobias has been reported to be 22% [50]. Data from a Mexican mental health survey showed that 24.8% of adolescents with a specific phobia were afraid of closed spaces [51]. For the results of this OCD brain study, OCD patients who experience anxiety in closed spaces were excluded. For this reason, readers should be cautious about generalizing current results. Finally, this study did not recruit a true control group without any formal structured CBT due to ethical limitations. Future studies should recruit a larger number of participants and consider comorbid conditions, including anxiety and mood fluctuations, with a truer control group.

Conclusions

OCfree, an OCD treatment program, was effective in the treatment of drug-naive patients with OCD. The treatment effects of OCfree are associated with increased brain connectivity within the CSTC tract. Multisensory stimulation by education, quests, and games in OCfree increased activity within the thalamus and insular regions in patients with OCD.
Conflicts of Interest
None declared.

References


Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA</td>
<td>Brodmann area</td>
</tr>
<tr>
<td>BAI</td>
<td>Beck Anxiety Inventory</td>
</tr>
<tr>
<td>BDI</td>
<td>Beck Depression Inventory</td>
</tr>
<tr>
<td>CBT</td>
<td>cognitive behavior therapy</td>
</tr>
<tr>
<td>CSTC</td>
<td>cortico-striato-thalamo-cortical</td>
</tr>
<tr>
<td>DSM-5</td>
<td>Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition</td>
</tr>
<tr>
<td>ERP</td>
<td>exposure and response prevention</td>
</tr>
<tr>
<td>fALFF</td>
<td>fractional amplitude of low-frequency fluctuations</td>
</tr>
<tr>
<td>OCD</td>
<td>obsessive-compulsive disorder</td>
</tr>
<tr>
<td>REST</td>
<td>Resting-state functional magnetic resonance imaging Data Analysis Toolkit</td>
</tr>
<tr>
<td>ROI</td>
<td>region of interest</td>
</tr>
<tr>
<td>Rs-fMRI</td>
<td>resting-state functional magnetic resonance imaging</td>
</tr>
<tr>
<td>SPM12</td>
<td>Statistical Parametric Mapping 12</td>
</tr>
<tr>
<td>SSRI</td>
<td>selective serotonin reuptake inhibitor</td>
</tr>
<tr>
<td>Y-BOCS</td>
<td>Yale-Brown Obsessive Compulsive Scale</td>
</tr>
</tbody>
</table>
A Blended Electronic Illness Management and Recovery Program for People With Severe Mental Illness: Qualitative Process Evaluation Alongside a Randomized Controlled Trial

Titus A A Beentjes1,2,3*, PhD; Betsie G I van Gaal1,4*, PhD; Hester Vermeulen1*, Prof Dr; Maria W G Nijhuis-van der Sanden1*, Prof Dr; Peter J J Goossens2,5*, Prof Dr

1Radboud Institute for Health Sciences, Radboud University Medical Center, IQ healthcare, Nijmegen, Netherlands
2Dimence Group Mental Health Care Centre, Deventer, Netherlands
3Saxion University of Applied Science, Center for Nursing Research, Deventer, Netherlands
4Faculty of Health and Social Studies, HAN University of Applied Sciences, Nijmegen, Netherlands
5Department of Public Health and Primary Care, Faculty of Medicine and Health Sciences, University Centre for Nursing and Midwifery, Ghent University, Ghent, Belgium

* all authors contributed equally

Corresponding Author:
Titus A A Beentjes, PhD
Dimence Group Mental Health Care Centre
Pikeursbaan 3
Deventer, 7411 GT
Netherlands
Phone: 31 651284459
Email: titus.beentjes@radboudumc.nl

Abstract

Background: We conducted a trial to test the electronic Illness Management and Recovery (e-IMR) intervention to provide conclusions on the potential efficacy of eHealth for people with severe mental illness (SMI). In the e-IMR intervention, we used the standard IMR program content and methodology and combined face-to-face sessions with internet-based strategies on the constructed e-IMR internet platform. During the trial, the e-IMR platform was sparsely used.

Objective: This study aimed to evaluate the added value of the e-IMR intervention and the barriers and facilitators that can explain the low use of the e-IMR platform.

Methods: This process evaluation was designed alongside a multicenter, cluster randomized controlled trial. In this study, we included all available participants and trainers from the intervention arm of the trial. Baseline characteristics were used to compare users with nonusers. Qualitative data were gathered at the end of the semi-structured interviews. Using theoretical thematic analyses, the data were analyzed deductively using a pre-existing coding frame.

Results: Out of 41 eligible participants and 14 trainers, 27 participants and 11 trainers were interviewed. Of the 27 participants, 10 were identified as users. eHealth components that had added value were the persuasive nature of the goal-tracking sheets, monitoring, and the peer testimonials, which had the potential to enhance group discussions and disclosure by participants. The low use of the e-IMR platform was influenced by the inflexibility of the platform, the lack of information technology (IT) resources, the group context, participants’ low computer skills and disabilities, and the hesitant eHealth attitude of the trainers.

Conclusions: The extent of eHealth readiness and correlations with vulnerabilities in persons with SMI need further investigation. This study shows that flexible options were needed for the use of e-IMR components and that options should be provided only in response to a participant’s need. Use of the e-IMR intervention in the future is preconditioned by checking the available IT resources (such as tablets for participants) providing computer or internet guidance to participants outside the group sessions, evaluating the eHealth attitude and skills of trainers, and tailoring eHealth training to increase the skills of future e-IMR trainers.

Trial Registration: Netherlands Trial Register NTR4772; https://www.trialregister.nl/trial/4621

International Registered Report Identifier (IRRID): RR2-10.1186/s12913-016-1267-z

doi:10.2196/20860
Introduction

Background

In mental health care, eHealth is expected to have great potential to increase access to care while being economically and socially efficient [1]. eHealth can be defined as making use of information technology (IT). In meta-analyses, eHealth interventions for persons with depressive and anxiety disorders are accepted and proven to be effective [2]. eHealth is also used for persons with severe mental illness (SMI). Persons with SMI are diagnosed with a psychiatric disorder that causes, and is because of, serious impairments in social and occupational functioning that lasts longer than at least a couple of years and necessitates coordinated multidisciplinary care [3]. eHealth for persons with SMI is used in a wide range of interventions, such as self-management, relapse prevention, promoting adherence to medications and/or treatment, psychoeducation, supporting recovery, and promoting health and wellness and symptom monitoring [4]. eHealth interventions for people with SMI are accepted and feasible [4], and they have potential to deliver effective education [5]. Unfortunately, conclusions on their effectiveness cannot be drawn [4,6]. A number of difficulties and barriers have been addressed concerning eHealth for persons with SMI (eg, cognitive impairments, lower IT experience [7]), which may explain the high attrition rates [8]. Blending face-to-face contact with eHealth is supposed to increase the therapeutic relationship and prevent attrition [9].

To contribute to consumer-oriented development and delivery of self-management electronic support programs, we developed and tested a blended version of the standardized, curriculum-based Illness Management and Recovery (IMR) program for people with SMI [10,11]. The standard IMR program provides information and teaches the skills necessary for managing an SMI effectively and working toward achieving personal recovery goals [12]. In accordance with the intervention mapping (IM) protocol [13] and in collaboration with target group members, we developed the e-IMR intervention to evaluate whether persons with SMI could benefit more from the IMR when making use of eHealth strategies in combination with face-to-face sessions [11]. On the e-IMR internet platform, the IMR curriculum was integrated, and we blended the use of this platform with face-to-face, group-wise delivery of the standard IMR program [11]. To evaluate the effectiveness of the e-IMR intervention compared with the standard IMR program, we conducted a multicenter, cluster randomized controlled trial [10,11].

The most striking finding of the trial was the low use of the e-IMR platform [10]; therefore, we could not conclude the effectiveness of the e-IMR intervention. Sieverink et al [14] reported that many eHealth evaluations show no or limited positive effects, which is strongly related to not using technologies in the desired way. Ben-Zeev et al [6] advised that the development of eHealth interventions for people with SMI must be coupled with examining the barriers and possible solutions. In addition, the IM protocol advises testing the effectiveness of an intervention and conducting a process evaluation to understand why an intervention did or did not work [13]. Therefore, we conducted this process evaluation alongside a randomized controlled trial to gain insights that will ultimately help to make adjustments to facilitate proper use of the e-IMR intervention specifically or of eHealth for people with SMI in general.

Objectives

This study aimed to identify the added value of the e-IMR intervention and the barriers and facilitators that can explain the low use of the e-IMR platform.

Methods

Study Design

We conducted a theoretical thematic analysis [15] alongside the trial. This qualitative method makes use of a pre-existing coding frame and provides a detailed analysis of the data [15]. Data were derived from semistructured interviews with participants and trainers held at the end point of the trial. This trial was registered in the Netherlands Trial Register (NL4621). We used the framework of Grol and Wensing [16-19], which frames the factors that potentially influence the effect of an intervention (Multimedia Appendix 1) [17]. Therefore, we focused on the e-IMR intervention itself and its implementation, the trial participants and their social context, the IMR trainers who provided the intervention, and their organizational context.

Study Population

In this study, we included all available participants and IMR trainers from the intervention arm of the e-IMR trial [10] (Figure 1). Information about inclusion, exclusion, and eligibility criteria and the effect of the e-IMR trial can be found elsewhere [10]. Participants in the intervention arm of the trial who completed at least the first module on the e-IMR platform or had logged into the e-IMR platform at least five times were defined as users. Nonusers either did not use the e-IMR platform or used it less than 5 times. Users were regarded as having had the opportunity to benefit from the e-IMR intervention and to reflect on it. The trainers of the group-wise–delivered e-IMR intervention were psychiatric nurses and peer professionals. A peer professional is a person with a lived experience of a mental illness, educated, and trained to become a professional capable of transferring knowledge and counseling other persons with a mental illness.
The e-IMR Intervention

The e-IMR intervention started with a *welcome page* explaining the use of the e-IMR platform and leading participants to the 11 modules. On the e-IMR platform, participants could fill in e-versions of the forms in the standard IMR, such as goal-tracking sheets, problem-solving sheets, sheets for tracking successful coping strategies, and a symptom-monitoring page. In addition, the e-IMR platform contained illustrative videos showing peer testimonials to encourage participants to talk more freely about themselves and to take steps in their recovery process. Further detailed information about the e-IMR intervention is shown in Multimedia Appendix 2.

Implementation of the e-IMR Platform

The e-IMR platform was introduced to the trainers and participants by the first author at the second group session. Participants were invited to use the e-IMR platform but were not obliged to use it at home because of the possible lack of resources. The trainers were educated on how to support participants in the use of the e-IMR platform, how to install it on a computer in the session room, and how to use it during the sessions. The registration forms on successful coping strategies and the symptom-monitoring page were introduced after the second module on *practical facts about mental illnesses*. Weekly emails with a link to the e-IMR platform led the participants directly to the symptom-monitoring page. After finishing any module, one of the trainers provided feedback to the participants via the platform and guided the participants to the next module. Further detailed information about the implementation of the e-IMR intervention is presented in Multimedia Appendix 2.

Halfway through the trial, we discussed the low use of the platform with the trainers and asked them to reintroduce the platform in the sessions and to motivate and guide the participants to use it at home to get as much experience with it as possible. With regard to this request and in addition to the original implementation strategy, in 4 out of the 7 groups, extra e-IMR lessons were organized outside the current IMR sessions (Multimedia Appendix 2).

Data Collection

Data were collected between January 2015 and October 2016. Three types of data were gathered: participants’ characteristics, log data of the use of the e-IMR platform, and qualitative data from semistructured interviews at the end point of the trial.

At baseline, the following data on participants’ characteristics were gathered: age, gender, diagnostic classification according to the *Diagnostic and Statistical Manual of Mental Disorders* (4th edition) [20], physical comorbidities, treatment history, cultural background, socioeconomic status, highest education, computer and internet availability and use, computer literacy, perceived computer skills, and the need for guidance when using a computer or the internet. The last 2 items were scored on a 5-point Likert scale, with the answer options of *strongly disagree* (1) to *strongly agree* (5). In addition, the following data from trainers were collected at baseline and used for this study: age, gender, profession, highest education, years of experience in mental health, and eHealth experience.

Log data about the actual use of the e-IMR intervention were derived from the e-IMR platform. These data were used to identify *users* and *nonusers*.

We conducted semistructured interviews at the end point of the trial with all available participants and trainers. After the halfway discussions with the trainers about the low use of the e-IMR platform, we discussed the potential influential factors and adapted the framework of Grol and Wensing [17]. Within each factor, we formulated a number of relevant determinants and accordingly set up the interview questions (Multimedia Appendix 1). The framework and questions were used as the interview topic list in semistructured interviews at the end point. The first author (TB) and research assistants performed the interviews with participants at their preferred location. The first author conducted interviews with the trainers. All semistructured interviews were audio recorded and transcribed verbatim. The transcripts were uploaded in Microsoft Excel (R).

Data Analyses

Descriptive statistics were used to present the outcomes for the groups of users and nonusers. Chi-square and Student *t* tests were carried out to compare the baseline characteristics and IT attitudes of the groups of users and nonusers. Quantitative data from the structured interviews at baseline and end point were analyzed using the Statistical Package for the Social Sciences (R) version 23 [21].
Data from the transcripts of the semistructured interviews were analyzed deductively using theoretical thematic analysis [15]. We used the following 7 steps:

1. All 3 authors (TB, BG, and PG) independently read and reread the transcripts from the participants and from the trainers for one e-IMR group and identified meaningful statements.
2. All 3 authors grouped the statements into the categories of the modified coding frame of Grol and Wensing [17].
3. All 3 authors triangulated their analyses thoroughly until consensus was reached, which means that discussions lasted until all agreed without any doubts.
4. The first author completed the analyses of the subsequent e-IMR groups according to the first 2 steps.
5. The first author formulated a description of the findings within each determinant and added verbatim examples.
6. All 3 authors discussed the description of the findings thoroughly until consensus was reached.
7. Finally, a composite description of the experiences and the use of the e-IMR intervention was written and discussed with all authors.

Results

Characteristics of Participants and Trainers

From the 7 groups, baseline characteristics were collected from 41 participants and 15 trainers (Multimedia Appendix 3). The mean age of the participants at baseline was 46.9 years (SD 11.6; n=41) and the majority had minimal income. The mean age of the trainers at baseline was 46.7 years (SD 8.8; n=15). In total, 9 trainers were psychiatric nurses and 5 were peer professionals. Of the 41 participants, 14 (34%) were identified as e-IMR users. The groups of users and nonusers only differed significantly according to gender (P<.042), with more men being nonusers.

Process Evaluation

At the end point, 27 participants (10 of whom were users) and 11 trainers were available to be interviewed (Figure 1 and Multimedia Appendix 3). A total of 14 participants (4 of whom were users) were not interviewed because they were too burdened by being interviewed. From all the e-IMR groups, at least one trainer was interviewed; 4 trainers were unavailable because of busy work schedules. In the following sections, the findings are reported according to the framework (Multimedia Appendix 1). In our findings, we used the terms users or nonusers to make it clear that among the participants, only users or nonusers reported the mentioned statement. We used the term participants when both users and nonusers reported the statement. The following section details the findings for the e-IMR intervention and its implementation, the trial participants and their social context, and the IMR trainers who provided the intervention and their organizational context. The determinants for these factors are illustrated by using quotes from participants coded with a P followed by 4 digits and either U or N (standing for user or nonuser, respectively) and by quotes from trainers coded with a T followed by 5 digits and either Pe or Nu (standing for peer professional or psychiatric nurse, respectively).

The e-IMR Intervention and Its Implementation

Regarding the e-IMR intervention, the following determinants are described: added value, accessibility, implementation fidelity, and feasibility.

Added Value

Users and trainers reported that the components of the e-IMR intervention had added value. One user stated that because of the platform, the standard IMR curriculum was easier to understand. Explanations on relevant subjects in the different modules, for instance, about symptoms, were easy to find using the buttons on the platform. A trainer mentioned that the time-consuming search in the textbook was no longer necessary. In 4 out of the 7 groups, peer testimonial videos were shown during the group sessions. Watching these videos was of great value to trainers and participants, enhancing discussions and disclosure. Participants found the peer testimonials very interesting and experienced recognition:

Yes, those videos … I liked them. Watching them was the first we did, and it became easier to talk about the subject. [T31002Pe]

However, sometimes the participants felt fearful when reminded of their own psychotic experiences.

Trainers and users reported the added value of the repetitive character of the goal-tracking sheets on the platform. Users easily tracked and celebrated their achievements. When only the hard copy module was used, the paper goal-track sheets were often lost, which hindered the monitoring of achievement over time:

So, your goals appear; that’s not in the book. . . . it’s not possible to drop your focus. You’re reminded of them ... [P1202U]

One user reported that the results of the weekly reminders to monitor symptoms led to a more objective interpretation of varying emotions, which increased personal insight. Another user did not benefit from this. A different user thought that the focus on symptoms was too strong, and one peer professional trainer mentioned that he experienced aversion to this assignment because of this focus on symptoms:

In every chapter it appears: How much did symptoms burden you? ... it’s too negative. I know it is meant to be positive .... But, huu [shivering], these symptoms again; f[...] off! [T51003Pe]

The users and trainers reported that they did not use the coping strategies and problem-solving sheets.

Accessibility

Most of the participants reported that the eHealth components were not easy to find. Out of 14 users, 6 reported having problems with logging on to the e-IMR platform at home. In 5 of the 7 groups, participants and trainers reported that accessing the platform during the sessions was problematic because of bugs when using certain browsers, problems with accounts, problems with logging in, and not having the appropriate IT resources:

http://mental.jmir.org/2021/1/e20860/
Someone from technical services helped them, but the trainers couldn’t get it running. The enthusiasm in the group to work with it was very low. So they stopped trying, and we worked with the book the rest of the time. [P1106U]

Implementation Fidelity
Trainers stated that they gave enough attention to motivate participants to use the e-IMR platform. Both trainers and participants reported that because of problems with accessing the platform and the aversive reaction of nonusers, the actual use of the platform during the sessions was low, apart from the peer testimonial videos. Moreover, participants reported that using the e-IMR platform at home was not discussed in later sessions. Some users felt that the trainers did not stimulate them and that they linked this to the fact that the use of the e-IMR platform was not obligatory:

It was like: “It is no obligation, I can do it, but ....” I think that when it got more attention, you’ll be able to see what it’s bringing you. [P1202U]

Feasibility
The participants and trainers reported the nonfeasibility of working on a computer with a projector and screen during the sessions. They estimated that it would be too time consuming to switch from 1 participant’s account to another. Furthermore, participants could not read their own homework notes when they watched another participant’s account on the projection screen. They thought that the use of a personal laptop or tablet could overcome this problem:

I wondered how a session would go when we do everything in the e-IMR, and nothing on paper. What if someone else is active on the screen, and then I can’t see my own notes? What did I write down at home? I won’t remember, unless we all have a tablet or laptop. [P4202N]

Trainers and users reported inflexibility of the platform, such as not being able to amend notes or skip an uninteresting module or change the module order. As the platform was not used adequately during the sessions, participants easily lost synchronicity: doing the e-IMR intervention on the platform at home and during the group sessions became 2 separate things. Nonusers reported that they stopped or did not start using the platform and the aversive reaction of nonusers, the actual use of the platform during the sessions was low, apart from the peer testimonial videos. Moreover, participants reported that using the e-IMR platform at home was not discussed in later sessions. Some users felt that the trainers did not stimulate them and that they linked this to the fact that the use of the e-IMR platform was not obligatory:

There are two things ... I was afraid to mix them up ... So, you do double work. You choose either the book or the platform ... not both. [P3104N]

The Participants
Regarding the participants, the following influencing determinants can be described: attitude, compliance, skills and knowledge, and resources.

Attitude
With regard to computers, nonusers reported that they postponed the use of computers, were not interested, did not have an affinity, felt that working with computers was impersonal, were too easily overstimulated by the overload of content on a computer screen, experienced a lack of control over what was happening in the computer, and had a preference for tangible paper and face-to-face communication. Some nonusers experienced fear and mistrust in the privacy protection of the e-IMR platform, not wanting to take the risk of others being able to read their notes:

I don’t know where my information goes when I am on the world wide Internet. I need control, always and ever ... I will get over-stimulated, all those things in my site, they really distract me. [P4207N]

Compliance
Some users said that they got lost and confused when confronted with the platform’s inflexibility or when they wanted to get through a backlog after a short period of not using the platform. Users missed additional stimuli from trainers to deal with this backlog. Not using or stopping use was related to vulnerability, such as wanting to avoid burdens because of duplication of effort, not feeling well enough, having sensory overload, a lack of concentration, dyslexia, perfectionism, or fear of failure:

Yes, in the group you can talk it out right away; that’s easy ... I did not like doing it on the computer. I think because of the upcoming emotions ... and being alone here at home, no one to talk with ... It just was too much for me, and I decided to stop using it. [P1207N]

In terms of vulnerability, the trainers added that the participants recently experienced psychosis, lived a chaotic life, lacked inquisitiveness and initiative, had low intelligence, or had learning disabilities. Learning new skills was reported to be too difficult when not feeling well. The opposite was also reported—feeling better halfway through the trial and then being able to use the platform:

First I thought: This looks handy; I can do it. I really intended to do so. But I got those mood swings and thought: Let me do it on paper; it’s ... what I am used to do ... and I will do it later when I feel well enough — then I will. But that did not work. [P4103N]

Skills and Knowledge
At baseline, 15% (6/41) participants reported that they had never used a computer and most participants (27/41, 66%) scored neutrally or agreed that they had good computer skills. At the end point, participants reported not being familiar with computers, being afraid of computer viruses from the internet, not knowing how to log in, and not being able to imagine how computers process their input:

I cannot work on the computer .... I did try to learn, but ... no. Terrible, I might be able in a year or so. Now I really cannot. [P1103U]

At baseline, 34% (22/41) of the participants did not agree that they needed guidance in working with computers. Of these participants, 29% (14/41) became users. At the end point, participants with a need for guidance reported reluctance in asking for help. A total of 3 participants became a user halfway through the trial with considerable support from the trainer. One
trainer illustrated how a user was helped with working on the e-IMR platform:

Well, I (trainer) was at the computer. She (a user) was sitting next to me, and I asked: “Shall I click here or there?” I typed the text and repeatedly asked: “Is this correct?” [T41001Nu]

**IT Resources**

In total, 8 participants (8/41, 20%) reported having no computer but one of them did become a user. Moreover, not having the internet, an email account, or finances to afford these resources was reported. Most (31/41, 76%) of the participants had minimal income:

No, really, I was angry; at that time, I had lost my computer. I did it the old-fashioned way. I was fed up with that at [...] computer.... [P3101N]

**The Social Context of the Participants**

Within the social context of the participants, the following determinants can be described: social support and group effect.

**Social Support**

A female user reported that getting help to use a computer from her partner caused irritation. She preferred the help of someone outside her family. Other participants reported that they had a partner with no computer skills. In total, 3 participants became users after getting help from relatives, friends, or trainers outside the group session:

I’ll tell you, I just met him, and he fixed the necessary update. I did not dare to open it, and that’s over now .... [P1204N]

**Group Effect**

In 4 out of the 7 e-IMR groups, the participants decided not to use the e-IMR platform during the sessions. A nonuser decided not to use the e-IMR platform at home because another person in the group (a user) was struggling obsessively with using the e-IMR at home. The users and trainers experienced a negative group attitude toward the e-IMR platform, for instance, when nonusers expressed their irritation when the e-IMR platform was discussed during the sessions:

Yes, those participants who were not active on the e-IMR platform were irritated and said: “Why talk about the e-IMR again? ....” [T11003Nu]

**Trainers**

Regarding trainers, the following determinants can be described: attitude and skills and knowledge.

**Attitude**

Most trainers reported not being computer minded or having a preference for face-to-face contact and tangible paper:

I ’m not that Internet-minded; nor is my colleague. ... My colleague prefers working with these flipcharts. [T42002Pe]

The trainers estimated that helping participants with the use of the platform during the sessions would take too much time. They differed on whether offering individual guidance to the participants was part of their job as an IMR trainer. The trainers doubted, and some did not offer lessons on using the e-IMR platform outside the group:

Yes, ... a participant had intentions to start, but had troubles with the computer firewall .... I was wondering, ... what can I do to lower barriers? One option was to install things on her computer, but I considered this was going too far. [T12001Pe]

Some trainers reported that they observed vulnerabilities, disabilities, lack of concentration, easy loss of self-esteem, lack of discipline, and struggle with computers in participants. The trainers suggested that participants belonged to a generation with less computer experience and thought that this was influential. Thus, some trainers stated that combining eHealth and SMI is a complete misfit, and they blamed the policy makers for this:

This trend is politically grounded ... this e-mental health, blah, blah. Well, I think people from behind their desk invented this. They do not know what people with SMI go through. [T11003Nu]

The trainers reported that working with the e-IMR intervention and motivating participants was an extra, burdensome effort. They felt that working with the e-IMR intervention disturbed the group sessions and that doing the IMR regularly and working with a group were already difficult. The trainers reported cautiousness in opposing the resistance of nonusers to the e-IMR intervention. Their priority was to work with the IMR content and prevent participant attrition from the sessions, and the e-IMR platform became an afterthought:

I think most important in the group is that we go on and follow the book. In fact, working on the e-IMR platform was a sideshow. [T12003Nu]

**Skills and Knowledge**

At baseline, one trainer had eHealth experience. Some trainers reported having had enough tools; however, others reported not having heard enough about the e-IMR intervention and the trial. The trainers gave differing reports on whether they had enough skills; some said they did not:

My colleague explained to me how to start the e-IMR platform, but when I am alone, like today, I can’t manage. [T42001Nu]

**The Organizational Context of the Trainers**

Regarding the organizational context of the trainers, the following determinants can be described: policy, IT resources, and workflow.

**Policy**

The trainers had difficulty logging on to the platform because of a privacy policy in their organization. The internet system of organizations had firewalls to protect the organizations’ IT environment for internet viruses. Owing to this, some websites and email addresses were identified as unsafe and were blocked:
Here [via our intranet], I can’t enter LinkedIn or Dropbox ... you can’t enter hardly anything. ... They’re afraid of viruses. [T31001Nu]

**IT Resources**

At the start of the trial, one organization had an IT environment that was compatible with the e-IMR intervention but the other organizations did not. The session rooms often lacked a computer, soundcard in the computer, Wi-Fi, a projector, and a screen. The trainers sought help from IT help desks in their organizations but could not resolve these problems. Some trainers were creative and determined to find a bypass outside the local IT environment:

S[...], to get the video’s work, the sound card was blocked, but I thought: “I won’t quit trying, I want to show them.” ... In the end, we made it. [T42002Pe]

**Workflow**

In the search for another session room, trainers were confronted with overly strict schedules. Another issue was about starting IMR groups and assigning IMR trainers on time. In a number of organizations, IMR groups could only be organized shortly before closing the trial period. The trainers reported that such workflow problems are business as usual. To fulfill the participants’ need for guidance with the e-IMR intervention, a number of trainers reported not having enough time in their work schedule:

But this person needs guidance every day. I do not know how to manage that. I don’t have time to do so. [T52001Nu]

**Discussion**

**Principal Findings**

In this study, we evaluated the added value of the e-IMR intervention and the barriers and facilitators that can explain the low use of the platform. The users and trainers had negative and positive experiences with the e-IMR intervention. The added value of the e-IMR intervention consisted of the peer testimonial videos, the persuasive nature of the monitoring page, and the weekly confrontations with their personal recovery goals. There were barriers in the platform’s inflexibility, the inflexible group-wise provision of the intervention, the hesitant attitude toward eHealth of the participants and trainers, the participants’ lack of IT resources, their low skills and knowledge of using the internet, and their being too overwhelmed by symptoms or disabled cognitive functioning, causing problems with using the e-IMR platform.

**Strength and Limitations**

The strength of this study is that it included people with low computer use, which enabled us to obtain a broad picture of the added values, barriers, and facilitators. A limitation of this study is that we cannot draw conclusions about the potential feasibility of the e-IMR intervention in individual treatment settings. The e-IMR intervention might work better in individual sessions, as it can be better tuned and tailored to the personal needs of the person with SMI. We estimate that the influence of the group attitude and the e-IMR intervention’s infeasibility in group settings were considerable. Unfortunately, the institutes where IMR is provided individually declined to participate in the trial.

**Comparison With Previous Work**

The weekly monitoring page worked out well for some users; however, for others, including a trainer, these reminders were disliked because of a strong focus on symptoms. For users, the weekly confrontations with personal recovery goals and actions worked better than the paper version. The peer testimonial videos were highly appreciated because of their potential to enhance group discussions and the disclosure of the participants. Peer testimonials fulfill the need for peer information and acknowledgment [22]; thus, watching this kind of video can be a pivotal experience that enhances reflection and discussion [23].

Users and trainers were confronted with the platform’s inflexibility when they wanted to emend previous notes, skip uninteresting modules or the monitoring page, and change the order of the modules. Therefore, the next version of the e-IMR intervention should be flexible to fit individual needs. Addressing personalization seems to be a key issue for future eHealth interventions for people with SMI [24]. The group-wise provision of the e-IMR intervention was experienced as infeasible because it was too time consuming to switch between the accounts of participants and the fact that participants were not able to look at their own notes. In addition, because of their e-IMR–averse attitude, participants chose not to use the e-IMR platform during the sessions. Unintentionally, the e-IMR platform and the face-to-face IMR session became 2 separate things. To overcome this group barrier, providing a tablet to participants was a widely heard suggestion. Providing devices to persons with SMI is known to support engagement in e-interventions [23]. This may also overcome the lack of IT resources in persons with SMI, which in our study group was present in 20% (n=41) of participants, comparable with the percentages found by Thomas et al [25]. A lack of IT resources was also present in the participating institutes. Future e-IMR–providing institutes need an open IT environment, open soundcards, strong computers, Wi-Fi for multiple tablets, an available projector plus screen, and a help desk. Technological resources are necessary to facilitate eHealth interventions [24].

The lack of computer skills and the preference for tangible paper and face-to-face communication of the participants can also be seen as barriers. Belonging to a generation with less computer experience might be an influence because low computer literacy is associated with higher age [26]. Similar to Williams et al [23], we identified log-in problems and problems with finding e-components on the platform, which contributed to the low use of the e-IMR platform. The participants in our study thought that their problems with learning and using the e-IMR platform were because of being too overwhelmed by their symptoms or a disability in their cognitive functioning. Executive functions, working memory, and sustained attention play an important role in using websites [7], and these functions are also highly associated with psychiatric illnesses [27,28]. To gain a clear picture of the correlation between the psychiatric health status of persons with SMI and their eHealth readiness, more research is necessary. Taking our findings regarding participants’
attitudes, compliance, and skills and knowledge, we concluded that most participants in our study were not yet ready to engage with eHealth. Berry et al [29] drew a comparable conclusion that persons with SMI have relatively low interest in and willingness to engage with eHealth interventions. We also concluded that there is a need for guidance in persons with SMI. Implementation of the renewed e-IMR intervention must coincide with an eHealth support intervention for participants. Successful use of internet-based interventions for persons with SMI is facilitated by training, support, and encouragement [23].

The eHealth attitudes, skills, and knowledge of trainers toward the e-IMR intervention are also barriers. Their hesitant attitude toward eHealth is based on their preference for tangible paper and face-to-face contact and their own low computer skills. The trainers in this study stopped promoting the e-IMR intervention so they would not burden the participants and avoid causing the participants to withdraw from the sessions. Identifying with the participants’ struggles and vulnerabilities might also have influenced the process. To illustrate, some of the trainers questioned the appropriateness of eHealth for people with SMI, blaming the policy makers. Dutch mental health nurses indicate that eHealth is not in line with the educational level, cultural background, or digital skills of mental health patients [30].

Williams et al [23] suggested a paternalistic attitude when workers determine the suitability of using eHealth interventions for persons with SMI. In this study, in 4 out of the 7 groups, trainers strived creatively to find solutions for showing the peer testimonial videos and organizing e-IMR lessons outside the group sessions. Owing to this effort, 3 of the non–computer-minded participants became users. Worker engagement is essential to the successful implementation of eHealth for persons with SMI [23]. Before implementing the renewed e-IMR intervention, it might be necessary to teach trainers how to use eHealth, become experienced, and resolve their hesitancy.

Despite our findings, the development of internet-based interventions is ongoing in our increasingly digitalizing society and health care. Strand et al [24] stated that the internet can play a transitional role in recovery-oriented practices, and Williams et al [23] identified its potential to elicit the personal values of persons with SMI and their treatment preferences. These promising statements make further development of the e-IMR intervention worthwhile.

Conclusions

The eHealth components of the e-IMR intervention that have added value are the persuasive nature of using goal-tracking sheets and monitoring and the potential of the peer testimonial videos to enhance group discussions and the disclosure of the participants. The low use of the e-IMR platform was influenced by its inflexibility, lack of IT resources, group context, lack of computer skills of the participants and their disabilities, and the hesitant eHealth attitude of the trainers. The extent of eHealth readiness and the correlations with vulnerabilities in persons with SMI need to be investigated further. Providing the e-IMR intervention in the future is preconditioned by the flexible use of components in response to a participant’s needs, checking the available IT resources in institutions, providing tablets to participants in group settings, providing computer or internet guidance to participants outside the group sessions, evaluating the eHealth attitude of trainers, and providing the necessary eHealth training to increase the skills of future e-IMR trainers.

Acknowledgments

This study was funded by the ZonMW (the Netherlands Organisation for Health Care Research and Development) program Tussen Weten en Doen (grant 520001001). The funder had no influence on the study design; the collection, analysis, and interpretation of the data; the writing of the report; and the decision to submit the paper for publication.

Authors’ Contributions

All authors contributed to the conception and design of the study. TB contributed to data collection. All authors contributed to the analysis and interpretation and provided drafting of the paper. All authors contributed to the critical revision of the paper for important intellectual content and the final approval of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Conceptual framework based on the barriers and incentives for change of different levels of health care, interview items, and data source.
[DOCX File, 123 KB - mental_v8i1e20860_app1.docx ]

Multimedia Appendix 2

An overview of the eHealth components and implementation of the electronic Illness Management and Recovery (e-IMR) intervention.
[DOCX File, 19 KB - mental_v8i1e20860_app2.docx ]
Multimedia Appendix 3
Personal characteristics of participants and trainers.

References


3. Delespaul PH, de Consensusgroep EP. Consensus regarding the definition of persons with severe mental illness and the number of such persons in the Netherlands. Tijdschr Psychiatr 2013;55(6):427-438 [FREE Full text] [Medline: 23864410]


Abbreviations

- e-IMR: electronic Illness Management and Recovery
- IM: intervention mapping
- IMR: Illness Management and Recovery
- IT: information technology
- RCT: randomized controlled trial
- SMI: severe mental illness

©Titus A A Beentjes, Betsie G I van Gaal, Hester Vermeulen, Maria W G Nijhuis-van der Sanden, Peter J J Goossens. Originally published in JMIR Mental Health (http://mental.jmir.org), 20.01.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Mental Health, is properly cited. The complete bibliographic information, a link to the original publication on http://mental.jmir.org/, as well as this copyright and license information must be included.
Correction: Evaluation of a Mobile App to Enhance Relational Awareness and Change During Cognitive Analytic Therapy: Mixed Methods Case Series

Stephen Kellett¹, BSc, MSc, D Clin; Katherine Easton², BSc, MRes, PhD; Martin Cooper³, BA, MSc; Abigail Millings², BSc, PhD; Melanie Simmonds-Buckley², BSc, PhD; Glenys Parry⁴, BA, PhD

¹Sheffield Health & Social Care NHS Foundation Trust, University of Sheffield, Sheffield, United Kingdom
²University of Sheffield, Sheffield, United Kingdom
³Sheffield Hallam University, Sheffield, United Kingdom
⁴Catalyse, UK, Sheffield, United Kingdom

Corresponding Author:
Stephen Kellett, BSc, MSc, D Clin
Sheffield Health & Social Care NHS Foundation Trust
University of Sheffield
1 Vicar Lane
Sheffield, S1 2LT
United Kingdom
Phone: 44 114 222 6537
Email: s.kellett@sheffield.ac.uk

Related Article:
Correction of: https://mental.jmir.org/2020/12/e19888/

do:10.2196/27159

In “Evaluation of a Mobile App to Enhance Relational Awareness and Change During Cognitive Analytic Therapy: Mixed Methods Case Series” (JMIR Ment Health 2020;7(12):e19888) the authors noted one error. The paper was inadvertently published with an incorrect list of affiliations. The original paper listed the authors and affiliations as follows:

Stephen Kellett¹, BSc, MSc, D Clin; Katherine Easton², BSc, MRes, PhD; Martin Cooper³, BA, MSc; Abigail Millings², BSc, PhD; Melanie Simmonds-Buckley², BSc, PhD; Glenys Parry⁴, BA, PhD

¹Sheffield Health & Social Care NHS Foundation Trust, Clinical Psychology Unit, University of Sheffield, Sheffield, United Kingdom
²University of Sheffield, Sheffield, United Kingdom
³Sheffield Hallam University, Sheffield, United Kingdom
⁴Catalyse, UK, Sheffield, United Kingdom

The corrected affiliations are as follows:

Stephen Kellett¹, BSc, MSc, D Clin; Katherine Easton², BSc, MRes, PhD; Martin Cooper³, BA, MSc; Abigail Millings², BSc, PhD; Melanie Simmonds-Buckley², BSc, PhD; Glenys Parry⁴, BA, PhD

¹Sheffield Health & Social Care NHS Foundation Trust, University of Sheffield, Sheffield, United Kingdom
²University of Sheffield, Sheffield, United Kingdom
³Sheffield Hallam University, Sheffield, United Kingdom
⁴Catalyse, Sheffield, United Kingdom

The correction will appear in the online version of the paper on the JMIR Publications website on January 15, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.
The Effects of Downloading a Government-Issued COVID-19 Contact Tracing App on Psychological Distress During the Pandemic Among Employed Adults: Prospective Study

Norito Kawakami¹, MD, PhD; Natsu Sasaki¹, MD; Reiko Kuroda², MD; Kanami Tsuno³, PhD; Kotaro Imamura¹, PhD

¹Department of Mental Health, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan
²Division for Environment, Health and Safety, The University of Tokyo, Tokyo, Japan
³School of Health Innovation, Kanagawa University of Human Services, Kawasaki, Japan

Corresponding Author:
Norito Kawakami, MD, PhD
Department of Mental Health
Graduate School of Medicine
The University of Tokyo
7-3-1 Hongo
Bunkyo-ku
Tokyo, 1130033
Japan
Phone: 81 358413521
Fax: 81 358413392
Email: nkawakami@m.u-tokyo.ac.jp

Abstract

Background: Downloading a COVID-19 contact tracing app may be effective in reducing users’ worry about COVID-19 and psychological distress.

Objective: This 2.5-month prospective study aimed to investigate the association of downloading a COVID-19 contact tracing app, the COVID-19 Contact Confirming Application (COCOA), released by the Japanese government, with worry about COVID-19 and psychological distress in a sample of employed adults in Japan.

Methods: A total of 996 full-time employed respondents to an online survey conducted May 22-26, 2020 (baseline), were invited to participate in a follow-up survey August 7-12, 2020 (follow-up). A high level of worrying about COVID-19 and high psychological distress were defined by baseline and follow-up scores on a single-item scale and the Kessler 6 (K6) scale, respectively. The app was released between the two surveys, on June 17. Participants were asked at follow-up if they downloaded the app.

Results: A total of 902 (90.6%) of 996 baseline participants responded to the follow-up survey. Among them, 184 (20.4%) reported that they downloaded the app. Downloading of the contact tracing app was significantly negatively associated with psychological distress at follow-up after controlling for baseline variables, but not with worry about COVID-19.

Conclusions: This study provides the first evidence that using a government-issued COVID-19 contact tracing app may be beneficial for the mental health of employed adults during the COVID-19 pandemic.

(JMIR Ment Health 2021;8(1):e23699) doi:10.2196/23699

KEYWORDS

coronavirus disease; digital contact tracing; mental health; working population; longitudinal study; COVID-19; contact tracing; surveillance; tracking; anxiety; distress
Introduction

Contact tracing is one of the most effective methods of controlling infectious diseases. During the 2020 COVID-19 pandemic, many smartphone apps for digital contact tracing have been developed and widely used [1-3]. These contact tracing smartphone apps automatically record contacts with other people through smartphones during everyday life. A user will be notified if someone he/she contacted recently had a positive test result for COVID-19 infection, and be encouraged to take a test for COVID-19. The use of these apps is considered acceptable by a majority of the public [4]; despite ethical concerns about people’s privacy [5]. Such apps are expected to slow the transmission of COVID-19 [6], although more research is required to confirm their effectiveness. Psychological distress was reported to increase in the community during the nationwide spread of COVID-19 [7-10], partly due to fear of COVID-19 infection [7,9]. Poor mental health has been considered another public health problem related to the COVID-19 outbreak [9]. As a contact tracing app could help users identify their own risk of infection with COVID-19, users can be assured that they are at low risk when no notification comes from the system, indicating that they have not had close contact with people infected by COVID-19. Thus, the use of a COVID-19 contact tracing app may be effective at reducing fear of and worry about COVID-19 and reducing users’ psychological distress. However, to date, no study has reported on the mental health effects of using a COVID-19 contact tracing app.

This prospective study aimed to examine the effect of the use of a COVID-19 contact tracing app released by the Japanese government in mid-June 2020 on the perceived threat from COVID-19 and psychological distress in a sample of employed adults in Japan. We examined the association between reported use of the COVID-19 contact tracing app and worry about COVID-19 and psychological distress at follow-up, adjusting for these variables at baseline.

Methods

Study Design and Participants

This study was a 2.5-month prospective study with two consecutive online surveys. Participants responded to an online “closed” survey conducted May 22-26, 2020 (baseline), which was part of a larger longitudinal study of full-time employees aged 20-59 years that were recruited by an internet survey company (Macrimill Inc) from a large pool (>1,300,000) of self-selected preregistered community-dwelling residents across Japan [11,12]. After excluding 36 respondents who were unemployed at baseline, as their financial or health condition may confound study findings, a total of 966 respondents were invited via email to participate in a follow-up survey on August 7-12, 2020 (follow-up). Between the two surveys, Japan experienced the second wave of the outbreak; in 78 days (May 27-August 12), the number of new COVID-19 cases per day increased from 22 (on June 8) to a maximum of 1595 (August 7), with an average of 440 per day [13]. All surveys were made using online questionnaires on a website that was specifically designed for the surveys and required an ID and password to log in. The survey system allowed each respondent to submit a questionnaire only once. Each questionnaire consisted of 73 questions displayed on one page, with automated completeness checks before the questionnaire was submitted; no review step was provided. The participants were given a small incentive (a token equivalent to 30 JPY [US $0.29]) for completing each questionnaire. Study participants were informed about the purpose of the study, the investigators, the length of the questionnaires, and processes related to the data (anonymization, place and duration of storage), and participated in the surveys voluntarily.

Use of the COVID-19 Contact Tracking App

On June 19, 2020, the Japanese government released a free contact tracking app for COVID-19, called the COVID-19 Contact Confirming Application (COCOA), for iOS and Android [14]. The app does not collect personal information such as phone numbers or location. Rather, the app records encounters with other phones that were within one meter and lasted for more than 15 minutes as encrypted data. The users are notified when a user who they have come into close contact with for more than 15 minutes reports a positive test result for COVID-19; the notification only occurs after the contact enters the positive result in the app and allows the result to be shared to other users anonymously. In addition, the validity of the app measuring the information was not reported and thus is unclear. The app had been downloaded 13.2 million times by August 14, 2020 [14]. By then, 252 (0.002%) users were confirmed to have COVID-19. Participants were asked if they downloaded this app (yes or no) at follow-up [14].

Measures

Worry About COVID-19

We used a single-item scale to measure worry about COVID-19 at baseline and follow-up, by asking “Do you worry about COVID-19?” and using a 6-point Likert-type response scale [11,12]. The responses were dichotomized into high (“strongly” to “somewhat positive”) and low (“somewhat negative” to “not at all”).

Psychological Distress

Psychological distress (depression and anxiety) in the last 30 days was measured using the Kessler 6 (K6) scale [15] at baseline and follow-up. Acceptable levels of reliability and validity of the Japanese version have been reported [16,17]. Psychological distress was defined as having a K6 score of ≥5, corresponding to a mild level of distress, according to a previous study [17]. For a sensitivity analysis, we also used severe psychological distress (K6 score ≥13) as an alternative outcome [18].

Demographic Variables

Participants were asked their sex (male or female), age (<35 years, 35-49 years, or ≥50 years), marital status (currently married or not), educational attainment (high school graduate and lower or university graduate and higher), and occupation (managers, non–manual workers, manual workers in non–health care settings, or health care workers); whether they work from home (yes including partially or never), live with a child up to home (yes including partially or never), live with a child up to
high school age (yes or no), or live in high-risk areas designated during the COVID-19 emergency between April and May 2020 by the government (yes or no) [19]; and if they had any chronic physical condition (any of 10 predetermined conditions) at baseline.

**Statistical Analysis**

The adjusted prevalence of worry about COVID-19 and high psychological distress at follow-up (controlling for baseline levels) were compared between those who did or did not use the COVID-19 contact tracing app (Mantel-Haenszel chi-square test). Multiple logistic regression analyses were conducted to estimate the effect (odds ratio [OR] and 95% CI) of worry about COVID-19 or high psychological distress at follow-up, adjusting for the baseline demographic variables and baseline values of these variables. No weighting was made to adjust for the nonrepresentativeness of the sample. Statistical significance was set at $P<.05$. SPSS (Version 26.0; IBM Corp) was used for analyses.

**Ethical Considerations**

Online informed consent was obtained from all participants with full disclosure and explanation of the purpose and procedures of this study. We explained that their participation was voluntary, and that they could withdraw consent for any reason, simply by not completing the questionnaire. This study was approved by the Research Ethics Committee of the Graduate School of Medicine/Faculty of Medicine at The University of Tokyo [number 10856-(2)(3)(4)(5)].

**Results**

A total of 902 (90.6%) of 996 participants at baseline responded to the follow-up survey. Among them, 184 (20.4%) reported that they downloaded the app. The respondents who downloaded the app were significantly more likely to be male, older, living with a child, university graduates or higher, and working from home than respondents who did not download the app (Table 1). Prevalence of high psychological distress was slightly but not significantly greater among app users at baseline, but the pattern was reversed at follow-up. Compared to the national labor force statistics in Japan, the sample was less represented by manual workers, while sex and age distributions were similar.

Downloading of the app was significantly negatively associated with psychological distress at follow-up after adjusting for psychological distress at baseline (adjusted prevalence of 42.2% and 49.6% for users and nonusers, respectively; OR 0.61, 95% CI 0.39-0.93; $P=.02$). Downloading of the contact tracing app was significantly negatively associated with psychological distress at follow-up in the group without high psychological distress at baseline ($P=.04$; Multimedia Appendix 1). Downloading of the app was not significantly associated with worry about COVID-19 at follow-up after adjusting for worry about COVID-19 at baseline (adjusted prevalence of 61.9% and 61.9%, respectively; OR 1.01, 95% CI 0.68-1.51; $P=.96$). Downloading of the app was significantly negatively associated with psychological distress at follow-up after adjusting for all covariates (OR 0.59, 95% CI 0.38-0.91; $P=.02$; Table 2). Downloading of the app was not significantly associated with worry about COVID-19 (OR 1.25, 95% CI 0.65-2.40; $P=.50$).

We conducted these multivariate analyses using age with a different categorization (20-29 years old, 30-39 years old, 40-49 years old, and 50-59 years old) and as a continuous variable, and obtained similar findings. When using the alternative definition of severe psychological distress (ie, K6 score of ≥13), downloading of the app was not significantly associated with severe psychological distress at follow-up (adjusted prevalence of 13.0% and 13.2% for users and nonusers, respectively; OR 0.88, 95% CI 0.54-1.44; $P=.71$) or after adjusting for all covariates (OR 0.95, 95% CI 0.54-1.69; $P=.92$).
Table 1. Demographic characteristics at baseline and psychological variables at baseline and follow-up among respondents who used or did not use the COVID-19 COntact COntraining Application (COCOA) at follow-up.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Total sample (N=902), n (%)</th>
<th>Nonusers (n=718), n (%)</th>
<th>App users (n=184), n (%)</th>
<th>P value(^a)</th>
<th>National Labour Force Survey(^b), %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex (female)</td>
<td>434 (48.1)</td>
<td>361 (50.2)</td>
<td>73 (39.7)</td>
<td>.001</td>
<td>45.2</td>
</tr>
<tr>
<td>Age, years</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20-34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35-49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50-60</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marital status (married)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Living with a child (yes)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education (university or higher)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Occupation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Managers</td>
<td>93 (10.3)</td>
<td>66 (9.2)</td>
<td>27 (14.7)</td>
<td>.16</td>
<td>1.3</td>
</tr>
<tr>
<td>Non–manual workers</td>
<td>480 (53.2)</td>
<td>384 (53.5)</td>
<td>96 (52.2)</td>
<td>—</td>
<td>35.5</td>
</tr>
<tr>
<td>Manual workers</td>
<td>232 (25.7)</td>
<td>190 (26.5)</td>
<td>42 (22.8)</td>
<td>—</td>
<td>57.7</td>
</tr>
<tr>
<td>Health care workers</td>
<td>97 (10.8)</td>
<td>78 (10.9)</td>
<td>19 (10.3)</td>
<td>—</td>
<td>5.6</td>
</tr>
<tr>
<td>Working from home (yes)</td>
<td>298 (33.0)</td>
<td>213 (29.7)</td>
<td>85 (46.2)</td>
<td>&lt;.001</td>
<td>ND</td>
</tr>
<tr>
<td>Chronic condition (any)</td>
<td>298 (33.0)</td>
<td>213 (29.7)</td>
<td>85 (46.2)</td>
<td>.29</td>
<td>ND</td>
</tr>
<tr>
<td>Living in prior high-risk areas</td>
<td>635 (70.4)</td>
<td>508 (70.8)</td>
<td>127 (69.0)</td>
<td>.65</td>
<td>ND</td>
</tr>
<tr>
<td>Worry about COVID-19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At baseline</td>
<td>513 (56.9)</td>
<td>403 (56.1)</td>
<td>110 (59.8)</td>
<td>.40</td>
<td>ND</td>
</tr>
<tr>
<td>At follow-up</td>
<td>559 (62.0)</td>
<td>442 (61.6)</td>
<td>117 (63.6)</td>
<td>.67</td>
<td>ND</td>
</tr>
<tr>
<td>Psychological distress (K6 score of ≥5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At baseline</td>
<td>414 (45.9)</td>
<td>323 (45.0)</td>
<td>91 (49.5)</td>
<td>.28</td>
<td>ND</td>
</tr>
<tr>
<td>At follow-up</td>
<td>434 (48.1)</td>
<td>352 (49.0)</td>
<td>82 (44.6)</td>
<td>.28</td>
<td>ND</td>
</tr>
</tbody>
</table>

\(^a\)Based on the chi-square test.

\(^b\)Proportions among employed adults aged 20-59 years (N=51.8 million) based on the Japan National Labour Force Survey 2019.

\(^c\)—: not available.

\(^d\)ND: no data.
Discussion

Principal Findings

This study found that downloading of the COVID-19 contact tracing app released by the Japanese government was significantly negatively associated with psychological distress in a sample of employed adults in Japan. This is the first evidence indicating that a COVID-19 contact tracing app may be beneficial for people’s mental health during the COVID-19 pandemic. Unexpectedly, downloading of the COVID-19 contact tracing app did not significantly influence worry about COVID-19. This may be interpreted as the users of the app becoming better able to cope with worry about COVID-19 (a stressor) and reduce their psychological distress, even though they still worry.

A contact tracing app decreased the prevalence of mild psychological distress by 7.4 points on average (a 15% reduction from its prevalence among nonusers). The effect size was rather small, but it might have a large impact on mental health (which has been reported to have deteriorated during the pandemic [7-9]) because there are a large number of potential users. However, we could not replicate the finding when using severe psychological distress (K6 scores of ≥13) as an indicator of poor mental health. This is partly attributable to the small sample size, with a lower prevalence of severe psychological distress than mild physical distress in the sample. Another possibility is that the effect of the app may be less clear for those with a severe level of psychological distress or recovering from it. This corresponds to our observation that downloading the app was significantly negatively associated with high psychological distress at follow-up only in the group without high psychological distress at baseline. The present finding should be replicated with mental health indicators measuring different levels of distress severity in a larger sample. The mechanism underlying the negative association between downloading the app and psychological distress is unclear. Downloading the app itself may serve as an attempt at active coping against the threat of COVID-19 infection for the users, which could improve their sense of control or self-efficacy and reduce psychological distress [20]. Another possibility is that users may have a better personal relationship with family, friends, or colleagues as a result of downloading the app because the users could be seen as “trustworthy” by those around them due to them taking
preventive measures (ie, using the app). This may be quite important in circumstances where physical distancing is encouraged and discrimination and stigma related to COVID-19 are increased [20]. Improved social relationships may lead people to have better social support, and thus their psychological distress could be reduced. The psychological effects of downloading a COVID-19 contact tracing app as well as the underlying mechanisms behind those effects should be investigated using relevant psychological theories and scales in future research.

While the app certainly provides users with objective information about the risk of COVID-19 infection, participants’ worry about COVID-19 (ie, perception of the risk of infection) was not different at follow-up between users of the contact tracing app and nonusers. In fact, the number of notifications that users had been in contact with a COVID-19–positive person was very limited (only 0.002% of users registered as having COVID-19 as of August 14, 2020, in the case of the COCOA app) [14]. However, users who do not receive contact notifications may not be assured that they have a low risk of being infected because there are many other opportunities for infection that the app cannot inform them about. This finding is consistent with a previous report indicating that implementing measures to protect against COVID-19 in the workplace was not associated with a reduction in the perceived risk of COVID-19 infection [11]. The effect of using the app may not be strong enough to decrease uncertainty [21] and change the perceived risk of infection.

Although this study indicated a possible benefit to mental health related to using a contact tracing app, there are other issues to be considered before encouraging the use of the app [21]. There are still privacy concerns because the anonymization process may not be perfect; it is unclear how long the collected data will be stored; and the system may be vulnerable to a cyberattack [5,22]. There are also concerns about the transparency and accuracy of such apps; it is hard for people to judge the quality and trustworthiness of the apps, even if the source code is published. Therefore, the apps should be pretested to indicate how accurately they measure proximity to and meaningful contacts with positive cases [22]. It is necessary to investigate the extent to which privacy concerns and possible measurement errors are ethically and psychologically acceptable. There is also concern about unintended side effects. Related to our study finding, for instance, the apps may increase the psychophysiological arousal of users as they continuously await the arrival of a notification, which could lead to physical complaints and sleep problems. Such an adverse effect of downloading this type of app should be investigated and if adverse effects exist, warnings should be communicated to the users. As all these issues could influence the use of the app itself, as well as the possible mental health benefit of using the apps, future research should be conducted to fully explore the ethical and behavioral aspects of the use of a contact tracing app.

Limitations

First, although the study employed a prospective design, downloading of the contact tracing app was measured at follow-up. Thus, the direction of causality between use of the app and psychological distress is unclear; it is conceivable that participants whose mental health has improved are more likely to start using the app. In addition, users had to complete a few steps after downloading the app to approve the use of the app and allow it to access the Bluetooth function. Thus, downloading the app may not necessarily be equal to use of the app. Some may have uninstalled the app soon after they downloaded it. Research using a more accurate assessment of the use of the app is needed to confirm our findings. Second, we investigated the effect of a government-released contact tracing app in this study. The findings may not be applicable to similar contact tracing apps released by commercial entities, which may be less trusted by people than government-released ones. Third, because the study sample included only employees, and might be demographically biased in some other way, the findings may not be generalizable to the general population (eg, they might not apply to unemployed individuals who might be experiencing financial hardship or have severe health conditions). The prevalence of psychological distress in this sample (40%-50%) was slightly higher than that reported in previous studies (34%-38%) [10]. The sample may be more distressed than the general population. Furthermore, the finding may not be generalizable to other countries with different policies, cultural norms, and behaviors (such as wearing masks) in response to the COVID-19 pandemic. The risk of infection and mortality may also be different from other countries, as Japan had a less stringent policy on social distancing (eg, lockdown), a higher proportion of people wearing masks [23], and a lower proportion of confirmed cases per capita [13] during the study period.

Fourth, while the sampling ratio of the respondents to the total employed population of Japan was extremely small, it is unlikely that some respondents were from the same family, social group, or workplace. However, if that was the case, it could result in an overestimation of the effect of the app. Fifth, behavioral patterns underlying the use of the app [24,25] and social desirability may confound the findings. For instance, participants may respond favorably to both questions about downloading the app and psychological distress. A relationship between the motivation to use the app or the duration of app use and psychological distress was not investigated in this study. The living environment, family environment, and social environment of the respondents may confound the findings. We could not investigate differences in the effect of using the app among various demographic groups, such as sex, age, and educational attainment, due to the small sample size. Future research should address these issues.

Conclusion

This prospective study found that downloading of the COVID-19 contact tracing app released by the Japanese government was significantly negatively associated with psychological distress in a sample of employed adults in Japan. Downloading of a COVID-19 contact tracing app may reduce psychological distress among users during the COVID-19 pandemic. However, other ethical and psychological issues related to contact tracing apps need to be fully discussed and resolved before recommending the app to public health officials.
Acknowledgments

This work was supported by a JSPS KAKENHI grant (number JP18H04072) and internal funds of the Department of Mental Health, Graduate School of Medicine, The University of Tokyo. These sponsors had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication.

Authors’ Contributions

NK was in charge of this study, supervising the process and providing his expert opinion. NS and NK organized the study design and analyzed the data. RK, KT, and KI ensured that questions related to the accuracy or integrity of all parts of the work were appropriately investigated and resolved. All authors participated in conducting the survey. NK wrote the first draft of the manuscript, and all other authors critically revised it. All authors approved the final version of the manuscript.

Conflicts of Interest

NK reports grants from Fujitsu LTD and SBAtWork Corp, and personal fees from the Occupational Health Foundation, Japan Dental Association, Sekisui Chemicals, Junpukai Health Care Center, Osaka Chamber of Commerce and Industry, as well as nonfinancial support from the Japan Productivity Center, none of which are related to the submitted work. The other authors declare no conflicts of interest.

Multimedia Appendix 1
Supplementary table.

References


Abbreviations
K6: Kessler 6
OR: odds ratio

©Norito Kawakami, Natsu Sasaki, Reiko Kuroda, Kanami Tsuno, Kotaro Imamura. Originally published in JMIR Mental Health (http://mental.jmir.org), 12.01.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License CC BY-NC 4.0.
License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Mental Health, is properly cited. The complete bibliographic information, a link to the original publication on http://mental.jmir.org/, as well as this copyright and license information must be included.
Psychological Impact of the COVID-19 Pandemic on Chinese Health Care Workers: Cross-Sectional Survey Study

Jie Ni¹, MD; Fang Wang¹, MD; Yihai Liu¹, PhD; Mingyue Wu¹, MD; Yan Jiang¹, MD; Yujie Zhou¹, MD; Yujie Zhou¹, MD; Dujuan Sha¹, MD

Nanjing Drum Tower Hospital, Nanjing, China

Corresponding Author:
Yujie Zhou, MD
Nanjing Drum Tower Hospital
Zhongshan Road 321
Nanjing
China
Phone: 1 8206299821
Email: jiji218506@sina.com

Abstract

Background: The outbreak of COVID-19 has dominated headlines worldwide. The number of infections has continued to rise and had reached 30,000 worldwide at the time this paper was written. Because of the high risk of nosocomial transmission, medical health care workers may be experiencing substantial psychological stress. This descriptive study aimed to identify psychosocial effects on hospital staff associated with working in a hospital environment during the COVID-19 outbreak.

Objective: Our survey participants included 57 frontline clinicians working at Wuhan First Hospital and 157 medical students working at Jiangsu Provincial People’s Hospital during the COVID-19 outbreak. The questionnaire we adopted included questions regarding the participants’ personal well-being, sociodemographic characteristics, and psychological status.

Methods: 57 frontline clinicians working in Wuhan First Hospital and 157 medical training students working in Jiangsu Provincial Peoples Hospital during this outbreak participated in our survey. The questionnaire we adopted included questions regarding the participants’ personal well-being, sociodemographic characteristics and the psychological status.

Results: The COVID-19 outbreak had psychological impacts both on formal workers and medical students. The psychological effects included sleep disorders, anxiety, and depression. There was no significant difference between the group of formal workers and medical students (P=.85), and more than 50% (30/54, 56%, vs. 83/157, 52.9%) of the respondents reported pandemic-related mental disorders.

Conclusions: Our study indicates that the high risk of SARS-CoV-2 exposure caused substantial psychological stress among health care workers. This finding emphasizes the need to promote psychological crisis intervention for medical personnel during this epidemic.

(JMIR Ment Health 2021;8(1):e23125) doi:10.2196/23125

KEYWORDS
2019-nCoV; COVID-19; frontline clinician; medical students; psychology

Introduction

A novel pneumonia associated with a coronavirus broke out suddenly in Wuhan, China in December 2019 [1]. The Chinese government reported that approximately five million residents left Wuhan and traveled to other provinces within China, while thousands of people reached other countries before the lockdown. The number of infections continued to rise and had reached 30,000 at the time of the writing of this paper according to real-time data on Weibo.com [2]. Along with the rapid expansion of the number of patients, health care systems worldwide were faced with difficult predicaments, including severe shortages of health care workers and medical materials [3]. Furthermore, health care providers had a greater likelihood of being exposed to the virus [4]. Research on the impact of previous epidemic outbreaks on the psychological well-being of health care workers showed that many health care workers presented high levels of psychological distress [5]. Although the Chinese government endeavored to guarantee the security of frontline clinicians, including organizing strict training,
providing adequate medical facilities, and discouraging off-work contact, little is known about the psychological effects of the COVID-19 outbreak on hospital workers. The main objective of our study was to investigate whether medical personnel were experiencing significant psychological conflict between their duties and their concern for their own safety [6] and to evaluate whether psychological intervention was necessary.

Methods

Sample
This was a cross-sectional study including 57 frontline clinicians working at Wuhan First Hospital and 157 medical students working at Jiangsu Provincial People’s Hospital in Nanjing during the COVID-19 outbreak. First, we distributed the web-based questionnaire among the Wuhan First Hospital frontline staff, and we collected 57 valid questionnaires on March 4, 2020. Later, we conducted a survey among young medical students to investigate their psychological changes associated with the COVID-19 pandemic, and we collected 157 valid surveys. The questionnaire administered to the frontline clinicians consisted of 4 main sections: basic demographic data, the Athens Insomnia Scale, the Self-Rating Anxiety Scale (SAS), and the Self-Rating Depression Scale (SDS). The questionnaire administered to the medical students did not contain the Athens Insomnia Scale. Owing to resource constraints, the distribution of the questionnaire was limited to only those who received the questionnaire on day 2 of data collection. We coded the response categories using the scoring method recommended by Goldberg and Williams [7] and calculated a total score. We used a threshold score of greater than 50 to identify the presence of emotional anxiety and a threshold score of greater than 53 to identify the presence of emotional depression.

Procedures
The questionnaires were completed on a voluntary basis by all willing respondents (physicians, nurses, and medical students working in hospitals during the crisis). We retrieved the completed questionnaires at a computer terminal. Reminders to complete the survey were sent to the volunteers by email.

Ethics Approval and Consent
The studies involving human participants were reviewed and approved by the Ethics Committee of Nanjing Drum Tower Hospital. Written informed consent to participate in this study was provided by the participants.

Statistical Analysis
We analyzed the data using SPSS 25.0 (IBM Corporation). Descriptive statistics were employed to organize the data collected from the survey. Chi-square tests were performed to compare the differences between groups. P values <.05 were considered statistically significant.

Results

Demographics

Wuhan Frontline Clinicians
A total of 54 physicians and nurses completed our survey voluntarily. The 54 respondents comprised 45 women (83%) and 9 men (17%) (Table 1). Among these medical staff, 26/54 (48%) were single, 26/54 (48%) were married, and 2/54 (4%) were divorced. Most of the 54 respondents were nurses (n=45, 83%), and physicians represented 17% of the sample. Among the physicians, 2 (22%) held senior professional titles; 6 (67%) held middle titles, and 1 (11%) held a junior title. Among the nurses, 2 (4%) had senior professional titles; 3 (7%) held middle titles, and 40 (89%) held junior titles. In this sample, 12/54 respondents (22%) had a master’s degree or above, and 42/54 (78%) had a bachelor’s or associate degree.
### Table 1. Demographic characteristics of the frontline clinicians at Wuhan First Hospital (N=54).

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Respondents with mental disorders (n=30), n (%)</th>
<th>Respondents without mental disorders (n=24), n (%)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>2 (7)</td>
<td>7 (29)</td>
<td>.07</td>
</tr>
<tr>
<td>Female</td>
<td>28 (93)</td>
<td>17 (71)</td>
<td></td>
</tr>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
<td>.75</td>
</tr>
<tr>
<td>20-30</td>
<td>14 (47)</td>
<td>11 (46)</td>
<td></td>
</tr>
<tr>
<td>30-40</td>
<td>14 (47)</td>
<td>10 (42)</td>
<td></td>
</tr>
<tr>
<td>≥40</td>
<td>2 (7)</td>
<td>3 (13)</td>
<td></td>
</tr>
<tr>
<td><strong>Education</strong></td>
<td></td>
<td></td>
<td>.44</td>
</tr>
<tr>
<td>College or bachelor’s degree</td>
<td>25 (83)</td>
<td>17 (71)</td>
<td></td>
</tr>
<tr>
<td>Master’s degree or above</td>
<td>5 (17)</td>
<td>7 (29)</td>
<td></td>
</tr>
<tr>
<td><strong>Occupation</strong></td>
<td></td>
<td></td>
<td>.27</td>
</tr>
<tr>
<td>Nurse</td>
<td>27 (90)</td>
<td>18 (75)</td>
<td></td>
</tr>
<tr>
<td>Physician</td>
<td>3 (10)</td>
<td>6 (25)</td>
<td></td>
</tr>
<tr>
<td><strong>Tenure in current occupation (years)</strong></td>
<td>12 (40)</td>
<td>9 (38)</td>
<td>.76</td>
</tr>
<tr>
<td>≤5</td>
<td>16 (53)</td>
<td>12 (50)</td>
<td></td>
</tr>
<tr>
<td>5-15</td>
<td>2 (7)</td>
<td>3 (13)</td>
<td></td>
</tr>
<tr>
<td><strong>Employment title</strong></td>
<td>23 (77)</td>
<td>17 (71)</td>
<td>.44</td>
</tr>
<tr>
<td>Junior</td>
<td>6 (20)</td>
<td>4 (17)</td>
<td></td>
</tr>
<tr>
<td>Middle</td>
<td>1 (3)</td>
<td>3 (13)</td>
<td></td>
</tr>
<tr>
<td><strong>Marital status</strong></td>
<td>14 (47)</td>
<td>12 (50)</td>
<td>.95</td>
</tr>
<tr>
<td>Single</td>
<td>15 (50)</td>
<td>11 (46)</td>
<td></td>
</tr>
<tr>
<td>Married</td>
<td>1 (3)</td>
<td>1 (4)</td>
<td></td>
</tr>
<tr>
<td>Divorced</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Nanjing Clinical Medicine Students**

After excluding respondents who had a history of anxiety, depression, or sleep disorders, we included 157 clinical medicine students in our study. From the received data, we found that none of the participants in our study had any symptoms of COVID-19 or a history of direct contact with SARS-CoV-2.

**Sleep and Psychological Status**

In contrast with the Nanjing clinical medicine students, Wuhan frontline clinicians were asked to complete an additional examination item: the Athens Insomnia Scale. In the group of frontline clinicians, the anxiety-related analysis showed that 9% (5/54) of participants had mild anxiety. Regarding the depression-related results, 35% (19/54) of the medical staff had mild depression, and 7% (4/54) had moderate depression. Synthesizing the results overall, we found that working experience, which was represented by work life, job title, and age, was highly associated with the survey results. People with more work experience had lower anxiety and depression rates. The results also showed that mental disorders were more common in nurses (Table 1).

When the risk of psychological symptoms was analyzed in a subsample of Nanjing clinical medicine students, we calculated that 47.1% (74/157) had mild depression and 1.2% (2/157) had moderate depression. Regarding the anxiety aspect, 4.0% (6/157) generally had mild depression. The distributions of the scores on the SAS and SDS for the medical students who participated in the study are shown in Table 2 and Table 3, respectively.

The high risk of psychological problems among health care providers observed in this study should evoke concern among psychologists.
Table 2. The distribution of the Self-Rating Anxiety Scale scores of the surveyed medical students (N=157).

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Score, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Anxiousness</td>
<td>100 (63.7)</td>
</tr>
<tr>
<td>Fear</td>
<td>128 (81.5)</td>
</tr>
<tr>
<td>Panic</td>
<td>125 (79.6)</td>
</tr>
<tr>
<td>Mental disintegration</td>
<td>142 (90.4)</td>
</tr>
<tr>
<td>Apprehension</td>
<td>68 (43.3)</td>
</tr>
<tr>
<td>Tremors</td>
<td>151 (96.2)</td>
</tr>
<tr>
<td>Body aches and pains</td>
<td>124 (79.0)</td>
</tr>
<tr>
<td>Easy fatigability, weakness</td>
<td>108 (68.8)</td>
</tr>
<tr>
<td>Restlessness</td>
<td>50 (31.8)</td>
</tr>
<tr>
<td>Palpitation</td>
<td>120 (76.4)</td>
</tr>
<tr>
<td>Dizziness</td>
<td>141 (89.8)</td>
</tr>
<tr>
<td>Faintness</td>
<td>151 (96.2)</td>
</tr>
<tr>
<td>Dyspnea</td>
<td>100 (63.7)</td>
</tr>
<tr>
<td>Paresthesia</td>
<td>150 (95.5)</td>
</tr>
<tr>
<td>Nausea and vomiting</td>
<td>126 (80.3)</td>
</tr>
<tr>
<td>Urinary frequency</td>
<td>132 (84.1)</td>
</tr>
<tr>
<td>Sweating</td>
<td>83 (52.9)</td>
</tr>
<tr>
<td>Face flushing</td>
<td>136 (86.6)</td>
</tr>
<tr>
<td>Insomnia</td>
<td>38 (24.2)</td>
</tr>
<tr>
<td>Nightmares</td>
<td>111 (70.7)</td>
</tr>
</tbody>
</table>
Table 3. The distribution of the Self-Rating Depression Scale scores of the surveyed medical students (N=157).

<table>
<thead>
<tr>
<th>Symptom</th>
<th>Score 1</th>
<th>Score 2</th>
<th>Score 3</th>
<th>Score 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Feeling downhearted</td>
<td>123 (78.3)</td>
<td>25 (15.9)</td>
<td>6 (3.8)</td>
<td>3 (1.9)</td>
</tr>
<tr>
<td>Morning severity</td>
<td>69 (43.9)</td>
<td>32 (20.4)</td>
<td>35 (22.3)</td>
<td>21 (13.4)</td>
</tr>
<tr>
<td>Easily crying</td>
<td>138 (87.9)</td>
<td>16 (10.2)</td>
<td>2 (1.3)</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>Insomnia</td>
<td>113 (72.0)</td>
<td>31 (19.7)</td>
<td>11 (7.0)</td>
<td>2 (1.3)</td>
</tr>
<tr>
<td>Lack of appetite</td>
<td>53 (33.8)</td>
<td>10 (6.4)</td>
<td>19 (12.1)</td>
<td>75 (47.8)</td>
</tr>
<tr>
<td>Decreased interest in sex</td>
<td>54 (34.4)</td>
<td>13 (8.3)</td>
<td>24 (15.3)</td>
<td>66 (42.0)</td>
</tr>
<tr>
<td>Weight loss</td>
<td>140 (89.2)</td>
<td>16 (10.2)</td>
<td>1 (0.6)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Constipation</td>
<td>118 (75.2)</td>
<td>31 (19.7)</td>
<td>7 (4.5)</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>Palpitation</td>
<td>131 (83.4)</td>
<td>22 (14.0)</td>
<td>3 (1.9)</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>Exhaustion</td>
<td>113 (72.0)</td>
<td>38 (24.2)</td>
<td>5 (3.2)</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>Difficulty in thinking</td>
<td>45 (28.7)</td>
<td>14 (8.9)</td>
<td>28 (17.8)</td>
<td>70 (44.6)</td>
</tr>
<tr>
<td>Scare capacity</td>
<td>51 (32.5)</td>
<td>13 (8.3)</td>
<td>26 (16.6)</td>
<td>67 (42.7)</td>
</tr>
<tr>
<td>Uneasiness</td>
<td>123 (78.3)</td>
<td>29 (18.5)</td>
<td>5 (3.2)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Despair</td>
<td>41 (26.1)</td>
<td>16 (10.2)</td>
<td>32 (20.4)</td>
<td>68 (43.3)</td>
</tr>
<tr>
<td>Emotion evoked</td>
<td>119 (75.8)</td>
<td>29 (18.5)</td>
<td>7 (4.5)</td>
<td>2 (1.3)</td>
</tr>
<tr>
<td>Hesitation</td>
<td>48 (30.6)</td>
<td>27 (17.2)</td>
<td>37 (23.6)</td>
<td>45 (28.7)</td>
</tr>
<tr>
<td>Futility</td>
<td>39 (24.8)</td>
<td>18 (11.5)</td>
<td>41 (26.1)</td>
<td>59 (37.6)</td>
</tr>
<tr>
<td>Feeling of living in a void</td>
<td>34 (21.7)</td>
<td>81 (51.6)</td>
<td>42 (26.8)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>No value</td>
<td>145 (92.4)</td>
<td>10 (6.4)</td>
<td>1 (0.6)</td>
<td>1 (0.6)</td>
</tr>
<tr>
<td>Interest loss</td>
<td>40 (25.5)</td>
<td>10 (6.4)</td>
<td>28 (17.8)</td>
<td>79 (50.4)</td>
</tr>
</tbody>
</table>

Discussion

The COVID-19 epidemic is one of the most virulent events that has ever threatened health care systems worldwide [8]. The results of this study show that hospital staff were significantly anxious during the pandemic, and their degrees of concern were moderately high. The greatest concerns of health care workers during the COVID-19 pandemic have been found to be infection and the potential consequences of the disease on their health [9]. At the time of writing of this paper, there had been over 3000 confirmed cases among health care workers [2]. The infection rate of medical personnel in Hubei was eight times higher than that of the general population for many reasons. Firstly, hospitals were the focus of confirmed patients; therefore, health care workers were by necessity susceptible to be exposed to the virus as well as to high exposure doses [3]. It has been proved that SARS-CoV-2 is transmitted through respiratory droplets, contact, and fecal-oral contact; even the eye is a possible transmission channel [10]. The generation of aerosols which only existed in hospital wards greatly exacerbated health care workers’ risk of infection [10]. As a result, the perceived risk of being infected was moderately high, and more than 50% (30/54, 55.6%) of respondents had psychological disorders.

Our results showed that relatively high numbers of health care workers experienced moderately high levels of worry during the pandemic, with nurses being more worried than physicians. This phenomenon is reasonable and can be explained by the following causes: (1) the nurses had more and closer contact with patients, and operations such as suctioning and collecting throat swabs exposed them to the disease; and (2) the nurses were generally younger and less educated than the physicians, so they lacked rich clinical experience and had received inadequate mental health education to cope with the difficult situation of the pandemic.

The results of this survey results revealed that the health care workers that came from other cities had higher rates of sleep problems. It is reasonable that the medical staff from outside Hubei Province were not familiar with the particular operation mode in this province compared with local workers. Furthermore, it was inevitable that they would feel fear when they left family and friends, not to speak of going to a dangerous place on their own [11]. Psychological counseling for frontline clinicians should be given attention, especially for those who left their original working environment and assisted during the outbreak in Wuhan.

In general, our research showed that frontline clinicians and clinical medicine students all tended to have higher rates of depression than of anxiety, and more research must be performed to determine the reasons for this finding. There was no significant difference in the rate of psychological disorders between frontline clinicians and medical students (P=.85).
was evident that frontline clinicians were at greater risk of infection; however, they had a stronger mentality.

Fortunately, most designated hospitals admitting patients infected with SARS-CoV-2 had established a shift system to allow medical staff to obtain sufficient rest. The government and people in general are making their best efforts to provide adequate backup resources for hospitals. We are hopeful that the psychological problems of medical workers can be alleviated if sufficient attention is given to this issue.

Conflicts of Interest
None declared.

References

Abbreviations

SAS: Self-Rating Anxiety Scale
SDS: Self-Rating Depression Scale

Edited by G Eysenbach; submitted 02.08.20; peer-reviewed by D Hu, P Liu; comments to author 04.08.20; revised version received 09.11.20; accepted 14.12.20; published 20.01.21.

Please cite as:
Psychological Impact of the COVID-19 Pandemic on Chinese Health Care Workers: Cross-Sectional Survey Study
JMIR Ment Health 2021;8(1):e23125
URL: http://mental.jmir.org/2021/1/e23125/
doi:10.2196/23125
PMID:33417754

©Jie Ni, Fang Wang, Yihai Liu, Mingyue Wu, Yan Jiang, Yujie Zhou, Dujuan Sha. Originally published in JMIR Mental Health (http://mental.jmir.org), 20.01.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Mental Health, is properly cited. The complete bibliographic information can be found at: http://mental.jmir.org/2021/1/e23125/
information, a link to the original publication on http://mental.jmir.org/, as well as this copyright and license information must be included.