Contents

Original Papers

Characterizing Use of a Multicomponent Digital Intervention to Predict Treatment Outcomes in First-Episode Psychosis: Cluster Analysis (e29211)
Shaunagh O'Sullivan, Lianne Schmaal, Simon D'Alfonso, Yara Toenders, Lee Valentine, Carla McEnery, Sarah Bendall, Barnaby Nelson, John Gleeson, Mario Alvarez-Jimenez ................................................................. 2

Telehealth-Based Psychoeducation for Caregivers: The Family Intervention in Recent-Onset Schizophrenia Treatment Study (e32492)
Kim Mueser, Eric Achtyes, Jagadish Gogate, Branislav Mancevski, Edward Kim, H Starr ................................................................. 20

Web-Based Single Session Intervention for Perceived Control Over Anxiety During COVID-19: Randomized Controlled Trial (e33473)
Michael Mullarkey, Mallory Dobias, Jenna Sung, Isaac Ahuvia, Jason Shumake, Christopher Beevers, Jessica Schleider ................................................................. 50

Improving Web-Based Treatment Intake for Multiple Mental and Substance Use Disorders by Text Mining and Machine Learning: Algorithm Development and Validation (e21111)
Sytske Wiegersma, Maurice Hidajat, Bart Schrieken, Bernard Veldkamp, Miranda Offl ................................................................. 68

A Serious Game for Young People With First Episode Psychosis (OnTrack>The Game): Qualitative Findings of a Randomized Controlled Trial (e33526)
Samantha Jankowski, Kathleen Ferreira, Franco Mascayano, Efly Donovan, Reanne Rahim, Michael Birnbaum, Sabrina Yum-Chan, Deborah Medoff, Bethany Marcogliese, Lijuan Fang, Terriann Nicholson, Lisa Dixon ................................................................. 85

Brief Digital Interventions to Support the Psychological Well-being of NHS Staff During the COVID-19 Pandemic: 3-Arm Pilot Randomized Controlled Trial (e34002)
Johannes De Kock, Helen Latham, Richard Cowden, Breda Cullen, Katia Narzisi, Shaun Jerdan, Sarah-Anne Munoz, Stephen Leslie, Andreas Stamatis, Jude Eze ................................................................. 96

Predicting Uptake of the COVID Coach App Among US Military Veterans: Funnel Analysis Using a Probability-Based Panel (e36217)
Beth Jaworski, Katherine Taylor, Kelly Ramsey, Adrienne Heinz, Sarah Steinmetz, Jason Owen, Jack Tsai, Robert Pietrzak ................................................................. 111

Review

Problematic Social Media Use in Adolescents and Young Adults: Systematic Review and Meta-analysis (e33450)
Holly Shannon, Katie Bush, Paul Villeneuve, Kim Hellemans, Synthia Guimond ................................................................. 38
Characterizing Use of a Multicomponent Digital Intervention to Predict Treatment Outcomes in First-Episode Psychosis: Cluster Analysis

Shaunagh O’Sullivan1,2, BA, MPsychSc; Lianne Schmaal1,2, BA, MSc, PhD; Simon D’Alfonso1,3, BASc (Hons), PhD; Yara Jo Toenders1,2, MSc; Lee Valentine1,2, BA, MSoC.Wk; Carla McEnery1,2, BPsychSc (Hons), PhD; Sarah Bendall1,2, BA (Hons), PGClinPsych, PhD; Barnaby Nelson1,2, MPsych, PhD; John F Gleeson4,5, BA (Hons), MPsych, PhD; Mario Alvarez-Jimenez1,2, BSc (Hons), MAResearchMeth, PhD, DClinPsy

1Orygen, Parkville, Australia
2Centre for Youth Mental Health, University of Melbourne, Melbourne, Australia
3School of Computing and Information Systems, University of Melbourne, Melbourne, Australia
4Health Brain and Mind Research Centre, Australian Catholic University, Melbourne, Australia
5School of Behavioural and Health Sciences, Australian Catholic University, Melbourne, Australia

Corresponding Author:
Shaunagh O’Sullivan, BA, MPsychSc
Orygen
35 Poplar Road
Parkville, 3052
Australia
Phone: 61 428 282 470
Email: shaunagh.osullivan@orygen.org.au

Abstract

Background: Multicomponent digital interventions offer the potential for tailored and flexible interventions that aim to address high attrition rates and increase engagement, an area of concern in digital mental health. However, increased flexibility in use makes it difficult to determine which components lead to improved treatment outcomes.

Objective: This study aims to identify user profiles on Horyzons, an 18-month digital relapse prevention intervention for first-episode psychosis that incorporates therapeutic content and social networking, along with clinical, vocational, and peer support, and to examine the predictive value of these user profiles for treatment outcomes. A secondary objective is to compare each user profile with young people receiving treatment as usual (TAU).

Methods: Participants comprised 82 young people (aged 16-27 years) with access to Horyzons and 84 receiving TAU, recovering from first-episode psychosis. In addition, 6-month use data from the therapy and social networking components of Horyzons were used as features for K-means clustering for joint trajectories to identify user profiles. Social functioning, psychotic symptoms, depression, and anxiety were assessed at baseline and 6-month follow-up. General linear mixed models were used to examine the predictive value of user profiles for treatment outcomes and between each user profile with TAU.

Results: A total of 3 user profiles were identified based on the following system use metrics: low use, maintained use of social components, and maintained use of both therapy and social components. The maintained therapy and social group showed improvements in social functioning ($F_{2,51}=3.58; P=.04$), negative symptoms ($F_{2,51}=4.45; P=.02$), and overall psychiatric symptom severity ($F_{2,50}=3.23; P=.048$) compared with the other user profiles. This group also showed improvements in social functioning ($F_{1,62}=4.68; P=.03$), negative symptoms ($F_{1,62}=14.61; P<.001$), and overall psychiatric symptom severity ($F_{1,63}=5.66; P=.02$) compared with the TAU group. Conversely, the maintained social group showed increases in anxiety compared with the TAU group ($F_{1,57}=7.65; P=.008$). No differences were found between the low use group and the TAU group on treatment outcomes.

Conclusions: Continued engagement with both therapy and social components might be key in achieving long-term recovery. Maintained social use and low use outcomes were broadly comparable with TAU, emphasizing the importance of maintaining engagement for improved treatment outcomes. Although the social network may be a key ingredient to increase sustained
engagement, as users engaged with this more consistently, it should be leveraged as a tool to engage young people with therapeutic content to bring about social and clinical benefits.

*(JMIR Ment Health 2022;9(4):e29211) doi:10.2196/29211*

**KEYWORDS**
digital intervention; digital health; youth mental health; psychotic disorders; clustering; usage metrics; log data; social networking

**Introduction**

**Background**

Evidence indicates the efficacy of specialist early intervention (SEI) services for first-episode psychosis (FEP) in achieving symptomatic remission during the first 2 years of treatment [1,2]. Despite this, the critical period for relapse extends to 5 years from the onset of psychosis, with 55% to 70% of individuals relapsing after 2 years [3,4]. Research has indicated that some treatment effects may not be sustained at 5 years, after 2 years of SEI has ceased [5,6].

Recently, 2 clinical trials addressed these limitations by evaluating the effects of extending SEI up to 5 years. Malla et al [7] found that clinical gains, in terms of remission of positive and negative psychotic symptoms, may be sustained if lower-intensity SEI is offered for an additional 3 years (on top of the 2 years already provided). However, findings from the trial by Albert et al [8] failed to demonstrate any additional benefits from extending SEI by 3 years, but this may be attributable to the high level of treatment provided to control participants in that study.

Although SEI has reported success in improving symptoms during the first 2 years of treatment, many young people with FEP continue to experience poor social and functional outcomes [9,10]. Although social and functional recovery is regarded by young people as the most important aspect of recovery [11], few FEP interventions have made this a primary target [12]. Fowler et al [10] addressed this by evaluating the effectiveness of social therapy in combination with early intervention services, with findings showing increases in structured activity, indicative of improved social functioning after 9 months. Furthermore, a randomized controlled trial (RCT) by Chang et al [13] found improvements in functional outcomes when SEI was extended by 1 year, but this was not sustained at 1- and 2-year follow-ups. Therefore, further research is needed to establish the effectiveness of longer-term interventions focusing on social and functional outcomes.

Digital interventions for FEP provide a unique opportunity to overcome the current limitations of treatment by providing continuous, engaging, and sustainable support to maintain long-term treatment effects [14]. It has been proposed that digital technologies can enhance care in FEP specifically by increasing access, enhancing current treatment, offering better predictive models, and accounting for clinical heterogeneity [15]. Some studies on the effectiveness of digital interventions for treating FEP and those with more established or sustained psychotic disorders have reported improvements in treatment outcomes such as social functioning [16], positive psychotic symptoms [17], negative psychotic symptoms [18], general psychopathology [18,19], overall psychiatric symptom severity [18], vocational outcomes [20], hallucination severity [19,21], hospital admissions [20,22], subjective well-being [16], social support [17], social connectedness [21], medication adherence [21], depression [1], and stress [14,17].

Although digital interventions have been associated with improved outcomes, they have also been associated with high attrition rates [23], and most do not typically extend beyond a 3-month period to focus on long-term recovery [24-26]. To address these limitations, Alvarez-Jimenez et al [1,27] pioneered a model of multicomponent digital interventions entitled *moderated online social therapy* (MOST). The MOST model integrates the following: (1) interactive psychosocial interventions, (2) social networking, (3) expert clinical moderation, and (4) peer support. To address attrition rates, MOST also aims to enhance long-term engagement by offering a shared, secure, and private social network for young people with similar mental health experiences.

The social networking and therapeutic elements of the MOST model were first applied in Horyzons, a world-first digital intervention aimed at maintaining long-term treatment effects and engagement to improve social functioning in young people recovering from FEP after receiving 2 years of SEI treatment [1,12]. Strengths and mindfulness-based approaches to therapy were adopted, with the aim of increasing self-efficacy and positive emotions, which have been linked to improved social functioning in psychosis [28,29]. The principles of self-determination theory (SDT) were also used with the aim of improving social functioning through increased intrinsic motivation [30].

A 4-week pilot study investigating the acceptability, safety, and clinical benefits of Horyzons indicated that the intervention was feasible, safe, and engaging and may enhance social connectedness in young people recovering from FEP [1]. An 18-month RCT of Horyzons has recently been completed, which found that Horyzons was effective in improving vocational outcomes and reducing presentations to hospital emergency services and hospital admissions compared with a control group receiving treatment as usual (TAU) [20]. Conversely, there were no differences between groups in social functioning over time. However, as there is limited evidence regarding the effectiveness of multicomponent digital interventions based on system use, it is difficult to determine the core therapeutic components of Horyzons, what outcomes they are associated with, and whether a specific pattern of use leads to improved social functioning in this population.

In line with SDT, multicomponent digital interventions based on the MOST model offer young people a high degree of choice over how and when they engage with the system, which...
increases flexibility in use. This increased flexibility increases the possibility of variation in use patterns or user trajectories. Distinct user profiles may exist in such multicomponent digital interventions, with users who may differ in use and engagement levels over time. The introduction of additional components, such as a therapeutic social network, is needed to address high attrition rates, increase engagement and tailor interventions to cater to the clinical needs and preferences of young people. However, new methods are needed to understand the complexities associated with determining which aspects and patterns of use lead to improved outcomes. Statistical modeling techniques such as growth mixture modeling can be used to identify different groups of users with similar trajectories over time. These techniques have previously been used for detecting similar symptom trajectories in mental health interventions [31]. K-means clustering techniques have also been used to identify and characterize participants based on unidimensional and multidimensional trajectories [32-34].

Objectives

Horyzons provides a unique opportunity to examine the relationship between multidimensional patterns of use and treatment outcomes by categorizing the use of multiple intervention components, such as therapeutic and social networking components. By gaining a better understanding of system use and user trajectories, and how they relate to treatment outcomes, multicomponent digital interventions could be further optimized to improve long-term recovery. Therefore, this study aims to examine the association between user profiles and treatment outcomes on Horyzons by (1) identifying user profiles based on 2D patterns of system use on both therapeutic and social components of the intervention, (2) characterizing the user profiles based on baseline demographic and clinical characteristics, and (3) examining the predictive value of the user profiles for treatment outcomes.

Methods

Study Design

Horyzons was a single-blind 18-month RCT, where participants with remitted FEP were randomly allocated to either TAU following 2 years of specialized care or TAU along with access to a moderated web-based social therapy intervention (Horyzons) [12]. Horyzons was based on the MOST model, which integrates (1) web-based therapy (Pathways and Steps), (2) peer-to-peer web-based social networking (the Café), (3) peer moderation, and (4) expert support by mental health clinicians and vocational workers. This RCT was registered on the Australian New Zealand Clinical Trials Registry (ACTRN12614000009617).

Ethics Approval

Ethical approval for the Horyzons RCT was granted by the Melbourne Health Research Ethics Committee (2013.146).

Participants

Participants comprised 86 young people allocated to the Horyzons intervention and 84 young people allocated to TAU, recruited from the Early Psychosis Prevention and Intervention Centre (EPPIC) at Orygen Youth Health, Melbourne, between October 2013 and January 2017. EPPIC is a specialist FEP program that provides 1 ½ to 2 years of specialized care to young people aged 15 to 24 years with FEP [35,36]. A total of 4 intervention participants who did not use the Horyzons platform independently and subsequently had no valid system use data were excluded from analyses. The remaining 82 intervention participants were aged between 16 and 27 years at randomization (mean 21, SD 2.88 years), and the 84 TAU participants were also aged between 16 and 27 years at randomization (mean 21, SD 2.83 years). Participants met clinical diagnosis for a first-episode psychotic disorder or mood disorder with psychotic features according to the Diagnostic and Statistical Manual of Mental Disorders 4th Edition (DSM-IV) [37], had not been treated with antipsychotic medication for >6 months before attending EPPIC, and showed remission of positive symptoms of psychosis for ≥4 weeks at the time of enrollment in the Horyzons study, as measured using the Positive and Negative Syndrome Scale [38].

Measures

Demographic and Clinical Characteristics

Data from baseline and the 6-month follow-up were used for this study’s analysis. Demographic information collected at baseline included sex, age, and vocational status. Baseline clinical characteristics included psychotic symptoms, levels of social functioning, depression, and anxiety, and are described in the Social Functioning, Psychotic Symptoms, and Depression and Anxiety sections below.

Social Functioning

Social functioning was measured using the Personal and Social Performance Scale (PSP) [39]. Ratings are based on functioning in the following four domains: (1) socially useful activities, (2) personal and social relationships, (3) self-care, and (4) disturbing and aggressive behaviors. The following four subscales of the First Episode Social Functioning Scale (FESFS) were also included to capture the full construct of social functioning: (1) living skills, (2) friends and activities, (3) intimacy, and (4) interacting with people [40]. These subscales were chosen based on their strong psychometric properties, independence from psychotic symptoms, and sensitivity to treatment effects [12]. The FESFS was designed specifically for young people with FEP.

Psychotic Symptoms

The Positive and Negative Syndrome Scale was used to assess psychotic symptoms, which included three subscales measuring (1) positive symptoms, (2) negative symptoms, and (3) general psychopathology [38]. The total score comprised all items from the three subscales, which indicated overall psychiatric symptom severity.

Depression and Anxiety

Depression was assessed using the Calgary Depression Scale for Schizophrenia [41]. Anxiety was assessed using the Depression Anxiety and Stress Scale [42].

https://mental.jmir.org/2022/4/e29211
**System Use Metrics**

System use metrics were extracted from the Horyzons web-based platform for each user for each day of their trial involvement. **Textbox 1** shows an overview of metrics representing aspects of use of the intervention’s therapeutic and social components.

<table>
<thead>
<tr>
<th>Therapy-related variables</th>
<th>Social networking–related variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of steps started (steps refer to the intervention modules)</td>
<td>Number of newsfeed posts (newsfeed refers to the social network)</td>
</tr>
<tr>
<td>Number of actions done (actions refer to the activities that comprise a step)</td>
<td>Number of newsfeed comments</td>
</tr>
<tr>
<td>Visited suggested content (suggested content refers to the therapeutic content recommended by clinical moderators)</td>
<td>Number of Talk it Out posts (Talk it Out refers to a problem-solving forum run by peer moderators)</td>
</tr>
<tr>
<td>Visited therapy (visiting therapy refers to visiting the homepage of the therapy component of the intervention)</td>
<td>Number of likes made</td>
</tr>
<tr>
<td>Visited notifications</td>
<td>Number of reactions made (reactions refer to short support messages in response to a post, eg, “thinking of you”)</td>
</tr>
<tr>
<td>Visited messages (messages refer to a private message section where moderators could contact participants directly)</td>
<td>Visited messages</td>
</tr>
<tr>
<td>Visited notifications</td>
<td>Visited notifications</td>
</tr>
<tr>
<td>Visited newsfeed</td>
<td>Visited Talk it Out</td>
</tr>
</tbody>
</table>

Therapeutic *Pathways* were divided into themes including understanding psychosis, identifying early warning signs to prevent relapse, identifying and exercising personal strengths, promoting social connections and positive emotions, and managing stress, anxiety, and depression. To increase usability, *Pathways* were further divided into short interactive *Steps*, for example, illustrating how to respond empathically to others (to foster positive connections). See **Multimedia Appendix 1** for an example of a *Step* on Horyzons. Each *Step* was accompanied by *Actions* or *Do its*, aiming to translate learning into behavior change, for example, suggestions on how to exercise empathy in specific contexts. Expert clinical moderators could also recommend *Pathways*, *Steps*, *Actions*, and *Talk it Outs* they felt would be relevant to different users via a private message, which would appear as a notification on the user’s dashboard.

Furthermore, users could visit the therapeutic component of Horyzons without completing any therapeutic content, for example, viewing what *Pathway* and *Step* was currently allocated to them.

The social network or the *Café* was led and moderated by peer-workers, who were trained young people who had a lived experience of mental illness. Participants were encouraged to communicate with one another to foster social support. Participants could post comments on the *Newsfeed* or like, respond, or react to comments that were already posted. Predeveloped *reactions* were designed to facilitate social support, for example, “I get you” and “thinking of you.”

**Textbox 1.** System use variables extracted from the Horyzons platform.

For this study, daily activity for both the therapeutic and social networking components was categorized. Daily activity was defined as the hierarchical level of system use per user per day based on the system use variables outlined in **Textbox 1**. A user was *inactive* when they did not display any activity on either the therapeutic or social component of Horyzons. Use was deemed *passive* when a user visited pages but did not actively engage with any content on either intervention component. A user was *engaged with therapy* when they started 1 step or completed 1 action. A user was *highly engaged with therapy* when they started >1 step, completed >1 action, or started at least one step and completed at least one action. Social use was deemed *moderate* when a user did not actively contribute to the social network but liked or reacted to at least one item. A user was *active* on the social network when they actively contributed via a post or comment. **Figure 1** shows the hierarchical categorization of daily activity into 2 dimensions.

<table>
<thead>
<tr>
<th>Daily Activity Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>For this study, daily activity for both the therapeutic and social networking components was categorized. Daily activity was defined as the hierarchical level of system use per user per day based on the system use variables outlined in <strong>Textbox 1</strong>. A user was <em>inactive</em> when they did not display any activity on either the therapeutic or social component of Horyzons. Use was deemed <em>passive</em> when a user visited pages but did not actively engage with any content on either intervention component. A user was <em>engaged with therapy</em> when they started 1 step or completed 1 action. A user was <em>highly engaged with therapy</em> when they started &gt;1 step, completed &gt;1 action, or started at least one step and completed at least one action. Social use was deemed <em>moderate</em> when a user did not actively contribute to the social network but liked or reacted to at least one item. A user was <em>active</em> on the social network when they actively contributed via a post or comment. <strong>Figure 1</strong> shows the hierarchical categorization of daily activity into 2 dimensions.</td>
</tr>
</tbody>
</table>
Statistical Analyses

Identifying User Profiles

K-means clustering for joint trajectories was implemented using the R package `km3d` (R Studio) to identify data-driven user profiles using Euclidean distance [44-46]. This is an unsupervised nonparametric technique that simultaneously partitions user trajectories from both the social and therapy dimensions into distinct cluster groups. This technique uses a hill-climbing expectation-maximization algorithm, alternating through various initialization methods until convergence is reached [45,47].

To run this analysis, participant trajectories were required to be of the same length. As such, we focused on the maximum number of days that all participants used Horyzons, which was 154 days. Each user’s first day on Horyzons consisted of an induction to the platform, so it could not be viewed as an independent system use. Therefore, daily activity (as per the hierarchical categories on both the social and therapeutic components of the intervention) from days 2 to 155 was used.

In terms of adherence, it was expected that participants would use Horyzons fortnightly to benefit from the intervention. On this basis, daily activity was transformed into 22 meaningful weekly scores, and these 22 weekly use scores were used as input features for the K-means clustering. Weekly scores comprised the maximum level of use per week, for example, if a user used Horyzons twice during week 1, and this consisted of passive use of the therapy dimension for 1 day (level of activity=2) and engaging with therapy on the second day (level of activity=3); they would obtain a score of 3 for week 1 on the therapy dimension (ie, the highest level of activity in that week).

No a priori hypothesis existed to substantiate the optimal number of clusters for analysis. Therefore, 2- to 4-cluster solutions were examined to account for complex patterns of system use found outside of a dichotomous high versus low use range. Furthermore, the sample was relatively small (N=82), suggesting that cluster solutions exceeding 4 would comprise too few participants per cluster. Cluster solutions with <15 participants in any cluster were excluded. K-means was rerun 100 times, each with different initial configurations, to ensure a global maximum was reached. A number of nonparametric fit indexes were used to compare cluster solutions, including the criteria developed by Calinski and Harabasz [48], Ray and Turi [49], and Davies and Bouldin [50]. A higher Calinski and Harabasz score indicates better fit, whereas lower Davies and Bouldin and Ray and Turi scores indicate better fit. In addition, cluster solutions were internally validated by calculating a Rand index, with scores closer to 1 indicating a higher likelihood of being assigned to the same cluster upon running 100 resamples [51]. Theoretical justifications and interpretability were also considered to select the optimal cluster solution.

Characterizing User Profiles

Differences between user profiles on demographic and baseline clinical characteristics were investigated using the 1-way analyses of variance and chi-square ($\chi^2$) tests for categorical variables.

Examining the Predictive Value of User Profiles for Treatment Outcomes

General linear mixed models were used to assess the associations between user profiles and treatment outcomes using the R package `lme4` [46,52]. Cluster group (user profile), time (baseline, 6-month follow-up), and group-by-time interaction were added as predictors. The predictive value of user profiles was assessed for social functioning, psychotic symptoms, depression, and anxiety. Sex, age, and days of untreated psychosis were added to the models as a priori determined covariates, as shorter days of untreated psychosis has been associated with improved outcomes and remission in FEP [53-58], male sex has been associated with poorer social and functioning outcomes in FEP [53], and age and sex may influence the use of the system. The models also controlled for baseline differences in the outcomes of interest. The effects of interest included (1) the main effect of group, (2) the main effect of time, and (3) the interaction between group and time. User IDs were added to the models as a random intercept effect, as they resolve the nonindependence associated with having multiple responses per user. As a secondary analysis, general linear mixed models were used to assess the associations between each individual user profile and TAU.

Results

Clusters Based on Joint Trajectories of System Use

The fit indexes for 2-, 3-, and 4-cluster solutions are reported in Multimedia Appendix 3. The 2-cluster solution was optimal based on all criteria, except for the Bayesian Information Criterion, where the 3-cluster solution showed the best solution. Cluster A (high-decreasing use) and B (low-decreasing use) trajectories remained consistent in the 2-, 3-, and 4-cluster solutions. The 2-cluster solution represented high-decreasing versus low-decreasing use, whereas the 3- and 4-cluster solutions represented more complex intermittent use, which existed based on visual inspections of the data (individual plots available upon request). Cluster C represented more intermittent and consistent use and added valuable information beyond high versus low use in terms of alternative user trajectories. On the basis of these
observations, the 3-cluster solution was selected as it was superior to the 4-cluster solution on all fit indexes, and the fit indexes were still relatively high compared with the 2-cluster solution. The 3-cluster solution also showed good internal validity based on the Rand index.

The trajectories of the user profiles based on the 3-cluster solution are shown in Figure 2. User profile A showed a rapid decrease in use on both the social and therapy dimensions after baseline and remained inactive for the following months; hence, this user profile was termed low use. User profile B showed initial high use on both dimensions, which decreased over time, with users remaining more active on the social dimension than on the therapy dimension (where use was mainly passive or inactive); therefore, user profile B was called maintained social. User profile C showed more variable but sustained use over time, remaining active on both dimensions, except during the final few weeks. User profile C also remained more engaged with the system’s therapy components than the other 2 user profiles and hence was called maintained therapy and social. The low use profile comprised 60% (49/82) of the users, the maintained social profile comprised 23% (19/82) of the users, and the maintained therapy and social profile comprised 17% (14/82) of the users.

Figure 2. User profile trajectories identified based on weekly hierarchical daily activity scores.

Characteristics of User Profiles
A 1-way between-groups analysis of variance indicated a statistically significant difference between user profiles on negative psychotic symptoms at baseline ($F_{2, 79}=6.375; P=.003$). Post hoc comparisons using the Tukey honest significant difference test indicated that the maintained therapy and social profile (mean 14.36, SD 4.99) had significantly higher symptoms than the maintained social profile (mean 10.48, SD 3.14) and the low use profile (mean 11.05, SD 3.55). No significant differences were observed between user profiles on any other clinical characteristics or on any demographic variables at baseline. A full overview of the results can be found in Multimedia Appendix 4.

Associations Between User Profiles and Treatment Outcomes
Significant group-by-time interaction effects were found for the primary outcome of social functioning as measured using the PSP, overall psychiatric symptom severity, and negative psychotic symptoms, with baseline effects accounted for (Table 1). Post hoc tests revealed that for social functioning, this interaction effect was accounted for by significant improvements for the maintained therapy and social profile from 0 to 6 months ($F_{1,11}=8.81; P=.01$), compared with the maintained social profile ($F_{1,15}=1.35; P=.26$) and the low use profile ($F_{1,28}=0.17; P=.68$; Figure 3). For overall psychiatric symptom severity, post hoc tests revealed that this interaction effect was accounted for by a significant decrease in symptoms for the maintained therapy and social profile ($F_{1,11}=5.99; P=.03$), compared with the maintained social profile ($F_{1,15}=1.71; P=.21$) and the low use profile ($F_{1,27}=0.004; P=.95$; Figure 4). In terms of negative symptoms, post hoc tests revealed that this interaction effect was accounted for by a significant reduction in symptoms for the maintained therapy and social profile ($F_{1,11}=10.94; P=.007$), compared with the maintained social profile ($F_{1,15}=0.66; P=.43$) and the low use profile ($F_{1,27}=0.98; P=.33$; Figure 5).
Table 1. Changes on outcomes from baseline to 6 months for user profiles.

<table>
<thead>
<tr>
<th></th>
<th>User profiles</th>
<th>Maintained social (n=19), mean (SD)</th>
<th>Maintained therapy and social (n=14), mean (SD)</th>
<th>Group x time interaction, F test (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low use (n=49), mean (SD)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSP(^a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>6.70 (1.97)</td>
<td>70.17 (3.15)</td>
<td>65.31 (3.64)</td>
<td>N/A(^b)</td>
<td>N/A</td>
</tr>
<tr>
<td>6 months</td>
<td>68.64 (2.62)</td>
<td>67.11 (3.15)</td>
<td>76.31 (3.64)</td>
<td>3.58 (2, 51)</td>
<td>.04</td>
</tr>
<tr>
<td>FESFS(^c) independent living skills</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>13.66 (0.33)</td>
<td>14.07 (0.52)</td>
<td>13.11 (0.60)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>6 months</td>
<td>13.59 (0.41)</td>
<td>14.04 (0.54)</td>
<td>13.03 (0.62)</td>
<td>0.003 (2, 45)</td>
<td>.99</td>
</tr>
<tr>
<td>FESFS interacting with people</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>12.96 (0.35)</td>
<td>12.52 (0.56)</td>
<td>11.90 (0.64)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>6 months</td>
<td>13.26 (0.44)</td>
<td>12.09 (0.58)</td>
<td>12.27 (0.66)</td>
<td>0.87 (2, 45)</td>
<td>.42</td>
</tr>
<tr>
<td>FESFS friends and activities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>18.96 (0.47)</td>
<td>17.50 (0.75)</td>
<td>17.61 (0.86)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>6 months</td>
<td>19.53 (0.62)</td>
<td>17.70 (0.78)</td>
<td>18.06 (0.89)</td>
<td>0.08 (2, 46)</td>
<td>.92</td>
</tr>
<tr>
<td>FESFS intimacy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>15.37 (0.45)</td>
<td>15.22 (0.75)</td>
<td>.52 (0.85)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>6 months</td>
<td>15.24 (0.58)</td>
<td>14.77 (0.75)</td>
<td>13.52 (0.90)</td>
<td>0.43 (2, 40)</td>
<td>.65</td>
</tr>
<tr>
<td>PANSS(^d) total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>45.15 (1.77)</td>
<td>43.69 (2.84)</td>
<td>48.91 (3.27)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>6 months</td>
<td>44.96 (2.22)</td>
<td>46.80 (2.84)</td>
<td>41.73 (3.27)</td>
<td>3.23 (2, 50)</td>
<td>.048</td>
</tr>
<tr>
<td>PANSS positive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>10.49 (0.54)</td>
<td>10.38 (0.86)</td>
<td>10.31 (1.00)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>6 months</td>
<td>10.29 (0.70)</td>
<td>11.31 (0.86)</td>
<td>9.89 (1.00)</td>
<td>0.60 (2, 52)</td>
<td>.55</td>
</tr>
<tr>
<td>PANSS negative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>10.73 (0.55)</td>
<td>11.24 (0.89)</td>
<td>13.62 (1.03)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>6 months</td>
<td>10.02 (0.72)</td>
<td>10.55 (0.89)</td>
<td>9.28 (1.03)</td>
<td>4.45 (2, 51)</td>
<td>.02</td>
</tr>
<tr>
<td>PANSS general psychopathology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>23.93 (1.06)</td>
<td>22.07 (1.70)</td>
<td>24.99 (1.96)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>6 months</td>
<td>24.65 (1.33)</td>
<td>24.92 (1.70)</td>
<td>22.56 (1.96)</td>
<td>2.34 (2, 50)</td>
<td>.11</td>
</tr>
<tr>
<td>CDSS(^e)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>3.91 (0.67)</td>
<td>3.31 (1.08)</td>
<td>2.78 (1.25)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>6 months</td>
<td>4.41 (0.84)</td>
<td>3.84 (1.08)</td>
<td>2.20 (1.25)</td>
<td>0.33 (2, 50)</td>
<td>.72</td>
</tr>
<tr>
<td>DASS(^f) anxiety</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>12.14 (1.35)</td>
<td>6.79 (2.02)</td>
<td>10.15 (2.45)</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>6 months</td>
<td>12.02 (1.77)</td>
<td>10.58 (2.18)</td>
<td>9.42 (2.53)</td>
<td>1.35 (2, 42)</td>
<td>.27</td>
</tr>
</tbody>
</table>

\(^a\)PSP: Personal and Social Performance Scale.
\(^b\)N/A: not applicable.
\(^c\)FESFS: First Episode Social Functioning Scale.
\(^d\)PANSS: Positive and Negative Syndrome Scale.
\(^e\)CDSS: Calgary Depression Scale for Schizophrenia.
\(^f\)DASS: Depression, Anxiety, and Stress Scale.
No significant group (user profile) by time associations were found for aspects of social functioning as measured by the FESFS, positive psychotic symptoms, general psychopathology, depression, or anxiety (Table 1). Furthermore, no main effects were found for differences between the profiles at each time point, and no main effects were found for changes over time for each profile on the outcomes.
Associations Between Individual User Profiles and TAU With Treatment Outcomes

Significant group-by-time interaction effects were found for social functioning as measured using the PSP, overall psychiatric symptom severity, and negative psychotic symptoms for the maintained therapy and social group versus the TAU group (Table 2). For social functioning, post hoc tests revealed that this interaction effect was accounted for by improvements for the maintained therapy and social group ($F_{1,11}=8.81; P=.01$) compared with the TAU group ($F_{1,56}=0.87; P=.35$), and a significant difference between groups at 6 month follow-up, with the maintained therapy and social group having higher social functioning scores than the TAU group ($F_{1,57}=5.82; P=.02$; Figure 6). In terms of overall psychiatric symptom severity, post hoc tests revealed that this interaction effect was accounted for by decreases in symptoms for the maintained therapy and social group ($F_{1,11}=5.99; P=.03$), compared with the TAU group ($F_{1,57}=0.78; P=.38$; Figure 7). In terms of negative symptoms, post hoc tests revealed that this interaction effect was accounted for by decreases in symptoms for the maintained therapy and social group ($F_{1,11}=10.94; P=.006$) compared with the TAU group ($F_{1,54}=0.71; P=.40$), and a significant difference between groups at baseline, with the maintained therapy and social group having higher symptoms than the TAU group ($F_{1,76}=4.35; P=.04$; Figure 8).

No significant group-by-time associations were found for treatment outcomes for the low use group versus the TAU group (Multimedia Appendix 5). Similarly, with the exception of anxiety, no significant group-by-time associations were found for treatment outcomes for the maintained social group versus the TAU group (Multimedia Appendix 6). Post hoc tests revealed that the interaction effect found for anxiety was accounted for by a significant decrease in symptoms for the TAU group ($F_{1,47}=6.50; P=.01$), a significant increase in symptoms for the maintained social group ($F_{1,12}=6.11; P=.03$), and a significant difference between groups at baseline, with the maintained social group having lower anxiety scores than the TAU group ($F_{1,72}=4.07; P=.047$; Figure 9).
Table 2. Changes on outcomes from baseline to 6 months for the maintained therapy and social and TAU<sup>a</sup> groups.

<table>
<thead>
<tr>
<th></th>
<th>TAU (n=84), mean (SD)</th>
<th>Maintained therapy and social (n=19), mean (SD)</th>
<th>Group × time interaction, F test (df)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PSP&lt;sup&gt;b&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>65.19 (1.57)</td>
<td>65.79 (3.82)</td>
<td>N/A&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
</tr>
<tr>
<td>6 months</td>
<td>66.89 (1.80)</td>
<td>76.79 (3.82)</td>
<td></td>
<td>.03</td>
</tr>
<tr>
<td><strong>FESFS&lt;sup&gt;d&lt;/sup&gt; independent living skills</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>13.71 (0.21)</td>
<td>13.12 (0.50)</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>6 months</td>
<td>13.71 (0.24)</td>
<td>13.04 (0.51)</td>
<td></td>
<td>.88</td>
</tr>
<tr>
<td><strong>FESFS interacting with people</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>12.78 (0.24)</td>
<td>11.94 (0.58)</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>6 months</td>
<td>12.62 (0.27)</td>
<td>12.31 (0.59)</td>
<td></td>
<td>.32</td>
</tr>
<tr>
<td><strong>FESFS friends and activities</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>18.47 (0.43)</td>
<td>17.74 (1.00)</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>6 months</td>
<td>18.25 (0.47)</td>
<td>18.20 (1.02)</td>
<td></td>
<td>.44</td>
</tr>
<tr>
<td><strong>FESFS intimacy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>14.82 (0.39)</td>
<td>14.56 (0.92)</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>6 months</td>
<td>14.77 (0.42)</td>
<td>13.47 (0.96)</td>
<td></td>
<td>.20</td>
</tr>
<tr>
<td><strong>PANSS&lt;sup&gt;e&lt;/sup&gt; total</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>44.26 (1.37)</td>
<td>48.57 (3.32)</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>6 months</td>
<td>45.72 (1.54)</td>
<td>41.40 (3.32)</td>
<td></td>
<td>.02</td>
</tr>
<tr>
<td><strong>PANSS positive</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>9.47 (0.43)</td>
<td>10.24 (1.05)</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>6 months</td>
<td>9.72 (0.50)</td>
<td>9.82 (1.05)</td>
<td></td>
<td>.62</td>
</tr>
<tr>
<td><strong>PANSS negative</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>11.00 (0.43)</td>
<td>13.42 (1.05)</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>6 months</td>
<td>10.64 (0.48)</td>
<td>9.09 (1.05)</td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>PANSS general psychopathology</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>23.78 (0.85)</td>
<td>24.92 (2.01)</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>6 months</td>
<td>25.39 (0.95)</td>
<td>22.50 (2.01)</td>
<td></td>
<td>.10</td>
</tr>
<tr>
<td><strong>CDSS&lt;sup&gt;f&lt;/sup&gt;</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>2.72 (0.41)</td>
<td>2.86 (0.99)</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>6 months</td>
<td>3.46 (0.46)</td>
<td>2.29 (0.99)</td>
<td></td>
<td>.22</td>
</tr>
<tr>
<td><strong>DASS&lt;sup&gt;g&lt;/sup&gt; anxiety</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>12.35 (1.16)</td>
<td>8.85 (2.80)</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>6 months</td>
<td>9.00 (1.32)</td>
<td>8.21 (2.89)</td>
<td></td>
<td>.39</td>
</tr>
</tbody>
</table>

<sup>a</sup>TAU: treatment as usual.

<sup>b</sup>PSP: Personal and Social Performance Scale.

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

<sup>d</sup>FESFS: First Episode Social Functioning Scale.

<sup>e</sup>PANSS: Positive and Negative Syndrome Scale.

<sup>f</sup>CDSS: Calgary Depression Scale for Schizophrenia.

<sup>g</sup>DASS: Depression, Anxiety, and Stress Scale.
Figure 6. Mean trends in PSP social functioning scores for the maintained therapy and social and TAU groups (95% CIs). PSP: Personal and Social Performance Scale; TAU: treatment as usual.

Figure 7. Mean trends in PANSS overall psychiatric symptom severity scores for the maintained therapy and social and TAU groups (95% CIs). PANSS: Positive and Negative Syndrome Scale; TAU: treatment as usual.

Figure 8. Mean trends in PANSS negative symptom scores for the maintained therapy and social and TAU groups (95% CIs). PANSS: Positive and Negative Syndrome Scale; TAU: treatment as usual.
Discussion

Principal Findings

This is the first study that characterized participants’ patterns of use of a multicomponent digital intervention (Horyzons), which included interactive therapy content and social networking, to predict treatment outcomes for young people recovering from FEP. Using a clustering procedure for joint trajectories, we identified three distinct user profiles: (1) low use, (2) maintained use of social components (maintained social), and (3) maintained use of both therapy and social components (maintained therapy and social). The maintained therapy and social profile had higher negative symptoms at baseline compared with the maintained social and low use profiles. The maintained therapy and social profile showed statistically significant improvements in social functioning and decreases in negative symptoms and overall psychiatric symptom severity compared with both the low use and maintained social profiles and the TAU group.

We used K-means clustering for joint trajectories to identify user profiles beyond that of a high versus low use dichotomy. Our approach accounted for the level of activity over time across the intervention components, going beyond categorizing the number of log-ins, which is limited in terms of meaningful engagement. In doing so, we found that the user profiles of individuals who demonstrated more variable but sustained use of both the therapy and social components over time were significantly associated with improved social functioning and clinical outcomes. In contrast, user profiles consisting of individuals with decreasing use on both social and therapy dimensions (ie, low use) demonstrated clinical outcomes comparable with that of TAU. The use of K-means clustering to identify use patterns in digital mental health interventions in the literature is a novel approach. A recent study by Sanatkar et al [59] used it to examine the association between engagement profiles (based on 2-month system use metrics) and depression and anxiety outcomes. They reported overall reductions in depressive and anxiety symptoms, but no differences were observed between the clusters. However, all users were somewhat engaged during this 2-month period, making it difficult to determine the optimal levels of use for improved treatment outcomes and how this was compared with nonuse. These findings differ from other research indicating that dropout rates in mental health apps are very rapid during the first month, with a retention rate of only 3.3% in the general population [60] and 0.5% to 28.6% completion rates or use beyond 6 weeks in interventions targeting depression and anxiety [61]. However, little is known about use patterns beyond a 2-month period, which our study examined.

Our findings indicated that the maintained use of both the therapy and social components was significantly associated with improvements in social functioning (the primary outcome of the Horyzons trial) compared with the other 2 user profiles. This is consistent with a recent pilot study investigating the effectiveness of a strengths- and mindfulness-based web-based social therapy for young people at ultrahigh risk of psychosis, which found increases in social functioning at 2-month follow-up [16]. It is worth noting that the MOST platform design is informed by SDT [62], which emphasizes meeting three key psychological needs to support motivation and behavioral change: (1) autonomy (feeling a sense of choice about one’s behavior), (2) competence (being able to bring about positive changes in desired outcomes), and (3) relatedness (feeling accepted by one’s social milieu). It may be the case that the combined system use (eg, the maintained therapy and social profile) aligned with both competence (therapy) and relatedness (social network), providing support for the SDT framework as a potentially mediating means in which to improve social functioning outcomes. Young people could engage in therapy on their own terms, which may also have promoted competence and autonomy. For example, choice in treatment (such as the choice young people had to complete therapy they felt was relevant to their needs on Horyzons) has been tied to the notion of individual autonomy [63]. Furthermore, there is evidence to suggest that moderated therapy, such as that offered on Horyzons, can promote self-competence in young people in particular [64].
Other improved outcomes, in terms of negative symptoms and overall psychiatric symptom severity, were observed for the maintained therapy and social group compared with the other 2 user profiles on Horyzons. These findings are consistent with those of the Horyzons RCT, which reported lower levels of negative symptoms compared with TAU from baseline to 12 months (which corresponded with a period of higher use of the Horyzons platform) [20], and lends support to the notion that Horyzons may improve negative symptoms for those young people with a certain level of engagement with the digital platform.

To be able to improve outcomes, our study suggests that sustained engagement with both the therapy and social networking components of Horyzons is required. Although these users comprised only 17% (14/82) of our sample, social functioning and negative symptoms are typically treatment resistant in FEP, which highlights the clinical significance of this finding [65,66]. Overall, 40% (33/82) of the users showed sustained use either on the social network alone (19/82, 23%) or on both the therapy and social networking aspects of the intervention (14/82, 17%), whereas 60% (49/82) of the participants were in the low use profile. This is an important observation, as a recent systematic search indicated a 15-day retention rate of 3.9% and a 30-day retention rate of 3.3% for mental health apps [60]; in contrast, 40% (33/82) of the users in our study showed more sustained use over 155 days. Furthermore, it is important to note that the low use profile did not mean nonuse. This cluster had a mean number of 12 log-ins to Horyzons over 6 months, indicating that these young people did engage with Horyzons but to a lesser extent than the maintained social (mean log-ins 142) and maintained therapy and social (mean log-ins 75) profiles. This also indicates that log-ins are not a good indicator of intervention effectiveness, as the maintained social profile had a higher number of log-ins than the maintained therapy and social profile but lower consistent engagement with therapy content. Therefore, rather than designing platforms to maximize log-ins, we need to design platforms to promote sustained engagement with the therapy and social components.

Exploratory analyses comparing each user profile to TAU further supported our main findings by demonstrating statistically significant improvements in social functioning, negative symptoms, and overall psychiatric symptom severity for the maintained therapy and social group compared with the TAU group. Conversely, increases in anxiety were observed for those in the maintained social group compared with those in the TAU group. An explanation for this may be that worsening of outcomes may lead to increased motivation to engage with social aspects and low motivation or perceived competence to engage with therapeutic content. Use of the social network was mostly moderate, with users mostly liking and reacting to posts (indicating passive use) rather than actively contributing via a post or comment. These findings are consistent with a study that found passive social media use to be associated with increased anxiety among adolescents [67]. Therefore, although the social network may increase engagement, as users engaged with this more consistently on Horyzons, it is important that it is designed to reduce anxiety and is leveraged to increase young people’s motivation to engage with therapeutic content for continued support, to bring about improved social and clinical outcomes. This is especially important given the critical 5-year period for the risk of relapse in FEP [4].

Our study confirmed that use is complex and that, although the maintained therapy and social profile showed improvements in outcomes, they also had higher negative symptoms at baseline. This raises the question of whether higher negative symptoms led to higher engagement, higher engagement led to improvements in outcomes, or both. For example, higher baseline symptoms may relate to perceived need to engage with therapeutic content. This is in line with previous research, which found that certain users only engaged with therapy until they completed what was relevant for them. Pung et al [68] offered participants self-help management strategies, similar to what was offered on Horyzons, for example, mindfulness steps and opportunities for social connection. Participants discontinued use after a skill was acquired but still had access to the intervention in case symptoms reemerged. These various reasons for use and disengagement may mask associations and contribute to mixed findings on the effectiveness of digital interventions. Although we controlled for baseline differences and our findings were consistent across social and clinical outcomes, we cannot make causal inferences about the change in outcomes in our study. Future research could address this by using multilevel models with an autoregressive lag [69] and examining the relationship between patterns of engagement and outcomes in real time [70].

Limitations

It should also be noted that this study had a number of methodological limitations. Although we controlled for baseline differences and key potential confounders, the findings of these analyses need to be interpreted with caution. First, the analyses comparing each user profile with TAU were exploratory and nonrandomized. Second, owing to the small sample size, we could not correct for multiple comparisons. That said, this is the first study to explore patterns of use in a multicomponent digital intervention for FEP and is arguably an informative starting point, as our results have significant clinical implications. Future research could build upon this contribution to the literature by replicating these analyses with larger sample sizes.

Conclusions

In conclusion, our findings indicate that sustained engagement with both the therapeutic and social networking components of Horyzons was key in improving social functioning, negative symptoms, and overall psychiatric symptom severity in young people with psychosis. This supports the therapeutic value of Horyzons and points to the need to capture complex patterns of use over time to determine key therapeutic targets and optimal use for improved outcomes. Going forward, this can be done in real time, with ongoing optimization of intervention features and management against key outcomes. This is a development in progress as digital interventions based on the MOST model are currently being implemented into clinical services as part of routine care, and novel methodologies including fast iterative
A/B testing and artificial intelligence optimization methods will be used to fast-track innovation and research translation [70]. These findings have real-world implications for the development of multicomponent digital interventions, as well as for the treatment of young people with psychosis through digital platforms. Future research will need to determine how to distill the contribution of specific aspects of the intervention and how components may work together to sustain user engagement and improve clinical outcomes. We are currently investigating this by determining which aspects of Horyzons system use lead to subsequent use by means of multiple convergent cross mapping.

Acknowledgments

The authors would like to thank the young people who participated in the Horyzons study and agreed to share their data for research purposes. The study was funded by the Mental Illness Research Fund (Victorian Government, Australia), the Australian National Health and Medical Research Council (NHMRC), and the Telstra Foundation. MAJ was supported by an Investigator Grant (APP1177235) from the NHMRC. LS was supported by an NHMRC Career Development Fellowship (1140764). SB was supported by the Ronald Philip Griffiths Fellowship. BN was supported by an NHMRC Senior Research Fellowship (1137687).

Conflicts of Interest

None declared.

Multimedia Appendix 1

Horyzons step: how to flourish.

[ PNG File, 249 KB - mental_v9i4e29211_app1.png ]

Multimedia Appendix 2

Horyzons social network.

[ PNG File, 269 KB - mental_v9i4e29211_app2.png ]

Multimedia Appendix 3

Fit indexes for each cluster solution.

[ DOC File, 32 KB - mental_v9i4e29211_app3.doc ]

Multimedia Appendix 4

Baseline demographic and clinical characteristics for user profiles.

[ DOC File, 57 KB - mental_v9i4e29211_app4.doc ]

Multimedia Appendix 5

Changes in outcomes from baseline to 6-months for the low use and treatment as usual groups.

[ DOC File, 41 KB - mental_v9i4e29211_app5.doc ]

Multimedia Appendix 6

Changes in outcomes from baseline to 6-months for the maintained social and treatment as usual groups.

[ DOC File, 41 KB - mental_v9i4e29211_app6.doc ]

References


Abbreviations

EPPIC: Early Psychosis Prevention and Intervention Centre
FEP: first-episode psychosis
FESFS: First Episode Social Functioning Scale
MOST: Moderated Online Social Therapy
PSP: Personal and Social Performance Scale
RCT: randomized controlled trial
SDT: self-determination theory
SEI: specialist early intervention
TAU: treatment as usual

©Shaunagh O’Sullivan, Lianne Schmaal, Simon D’Alfonso, Yara Jo Toenders, Lee Valentine, Carla McEnery, Sarah Bendall, Barnaby Nelson, John F Gleeson, Mario Alvarez-Jimenez. Originally published in JMIR Mental Health (https://mental.jmir.org), 07.04.2022. 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 https://mental.jmir.org/, as well as this copyright and license information must be included.
Telehealth-Based Psychoeducation for Caregivers: The Family Intervention in Recent-Onset Schizophrenia Treatment Study

Kim T Mueser¹, PhD; Eric D Achtyes²,³, MD, MS; Jagadish Gogate⁴, PhD; Branislav Mancevski⁵, MD, NCPsyA; Edward Kim⁵, MD, MBA; H Lynn Starr⁵, MD

¹Center for Psychiatric Rehabilitation, Boston University, Boston, MA, United States
²Michigan State University College of Human Medicine, East Lansing, MI, United States
³Cherry Health, Grand Rapids, MI, United States
⁴Janssen Research & Development, LLC, Titusville, NJ, United States
⁵Janssen Scientific Affairs, LLC, Titusville, NJ, United States

Corresponding Author: Branislav Mancevski, MD, NCPsyA
Janssen Scientific Affairs, LLC
1125 Trenton-Harbourton Road
Titusville, NJ, 08560
United States
Phone: 1 609 730 3115
Email: bmancevs@its.jnj.com

Abstract

Background: Schizophrenia is a lifelong illness that requires long-term treatment and caregiving. Family psychoeducation (FP) has been shown to lessen caregiver burden, improve caregiver functioning, and improve outcomes in patients. However, the impact of FP delivered specifically to caregivers on patient outcomes has not been well explored, particularly for early schizophrenia. Furthermore, there is a lack of research examining the benefits of telehealth-based psychoeducation for caregivers on either patient or caregiver outcomes.

Objective: The Family Intervention in Recent-Onset Schizophrenia Treatment (FIRST) study is a randomized controlled trial of patients with schizophrenia spectrum disorders and their caregivers, which is designed to evaluate the effect of telehealth-based, caregiver-focused, study-provided psychoeducation versus usual care (UC) on patient treatment failure (TF). The impact of study-provided psychoeducation on caregiver burden is also investigated.

Methods: Eligible patients and their designated caregivers were randomly assigned to either the study-provided psychoeducation (≤16 sessions of telehealth-based psychoeducation over 6 months) or UC group, stratified by antipsychotic treatment (paliperidone palmitate or oral antipsychotic). The major TF events (ie, psychiatric hospitalization or intervention, arrest or incarceration, and suicide attempts) were assessed at 3, 6, and 12 months after baseline. A proportional means model using mean cumulative function was used to assess between-group differences in the mean cumulative number of TF events over 12 months. Caregiver burden was assessed using the Involvement Evaluation Questionnaire and 12-item Short Form Health Survey.

Results: A total of 148 pairs of participants were enrolled in the study, of whom 96 (64.9%) patients and 94 (63.5%) caregivers completed the 12-month follow-up. The mean number of sessions in the study-provided psychoeducation group was 7.7 (SD 5.9). No differences were observed between the study-provided psychoeducation and UC groups in patient outcomes (rates of TF: 70% vs 67%; \( P = .90 \)) or measures of caregiver burden (assessment of caregiver distress and physical and mental health). However, post hoc analyses revealed lower relapse rates in patients who received paliperidone palmitate than in those who received oral antipsychotics at all time points. Although the FIRST study did not meet the primary end point, several key lessons were identified to inform future caregiver-focused, telehealth-based FP interventions. Lack of study-provided psychoeducation, focus on caregiver-only intervention, difficulties with enrollment, and caregiver–treatment team coordination may have affected the outcomes of the FIRST study.

Conclusions: Key insights from the FIRST study suggest the potential importance of supporting sufficient caregiver engagement; communication between clinicians, patients, and family members regarding treatment plans; and solidifying the relationship between clinicians providing psychoeducation to the caregiver and patient treatment team.

Trial Registration: ClinicalTrials.gov NCT02600741; http://clinicaltrials.gov/ct2/show/NCT02600741
schizophrenia; family psychoeducation; caregiver burden; recent-onset schizophrenia; telehealth

Introduction

Schizophrenia is a complex, lifelong illness that typically develops in young adults [1] and requires long-term treatment and caregiving, which are frequently provided by family members [2,3]. Caregivers often find that caring for a loved one with schizophrenia is difficult and struggle with social isolation, financial burden, and physical and emotional exhaustion [4,5]. Family psychoeducation (FP), a guideline-recommended complement to pharmacological treatment for schizophrenia, has been shown to lower burden and improve functioning in caregivers and can also lead to improved patient outcomes, including lower rates of relapse and hospitalization [6-11]. However, FP is often unavailable or underused, partially because of implementation barriers such as scheduling difficulties and lack of access to care from specialists [12-15].

To address this unmet need, web-based or telehealth-based models of psychoeducation that offer private at-home sessions have been developed [16-18]. Compared with usual care (UC), web-based FP interventions involving caregiver support, patient psychoeducation, and mutual patient–caregiver support have been found to be successful in lowering stress, reducing symptoms, increasing perceived social support for patients with schizophrenia, and improving the illness knowledge of caregivers [19,20]. Family interventions during the early phase of illness have been studied; however, the efficacy of FP interventions delivered exclusively to caregivers is still being explored.

The Family Intervention in Recent-Onset Schizophrenia Treatment (FIRST) study was designed to evaluate the impact of FP given specifically to caregivers on the outcomes of patients with schizophrenia spectrum disorder under their care and family burden. In the FIRST study, FP was delivered using Healios Inc, doing business as MyHealios, a telehealth-based study-provided psychoeducation (SPPE) and skills training intervention. MyHealios was developed to incorporate common components of efficacious caregiver-oriented FP interventions during the patients’ early phase of illness; the FP program was individualized to each caregiver to include education about schizophrenia and its treatment and skills training to improve communication, problem solving, and coping [21-23]. The MyHealios live web-based sessions were clinician led, enabling caregivers to access professional services from home. This paper reports the primary findings of the FIRST study and outlines other key learnings of the study.

Methods

Study Design and Patients

The FIRST study (NCT02600741) was a randomized controlled trial of patients with schizophrenia spectrum disorders and their caregivers, that was conducted to evaluate the overall effect of caregiver-focused study-provided psychoeducation and skills training compared with UC on the number of treatment failure (TF) events in patients (Multimedia Appendix 1). The study design was informed by a meta-analysis of caregiver-directed psychosocial interventions [24]. The FIRST study was initiated on July 24, 2015, and completed on July 5, 2018. The study sites were 31 community mental health centers in the United States, which provide routine clinical care to patients with schizophrenia. The study investigators received formal training through an investigator meeting and other training provided by the sponsor. Study participants were patients with diagnoses of schizophrenia, schizoaffective disorder, or schizophreniform disorder, aged 18 to 35 years, who were receiving paliperidone palmitate or oral antipsychotics, as prescribed by their clinician. Participants must have had ≥1 TF within 6 months of screening, defined as psychiatric hospitalization, intensive outpatient psychiatric treatment or partial hospitalization, psychiatric emergency department visit, crisis center visit, mobile crisis unit intervention, arrest or incarceration, or suicide attempt. Caregivers were individuals who provided the patient with assistance and care. They could be members of the immediate or extended family, friends, neighbors, or significant others. Caregivers were included if they were aged ≥18 years, had verbal interaction with the patient ≥2 times a week, had internet access, and had not received formal psychoeducation in the past 12 months. After screening, caregivers were randomly assigned in a 1:1 ratio to the study-provided psychoeducation or UC, stratified by patient antipsychotic treatment (paliperidone palmitate or oral antipsychotics; Figure 1). If the caregiver was unable or unwilling to continue participation in the study, the caregiver was not replaced; however, the patient was followed up. If a patient withdrew from the study, both the patient and caregiver were discontinued from the study.
Figure 1. FIRST study design. In the FIRST study, caregivers randomized to the study-provided psychoeducation received up to 16 study-provided psychoeducation and skills training sessions within a 6-month period. UC consisted of caregiver support that was customarily provided by the study site (if any). FIRST: Family Intervention in Recent-Onset Schizophrenia Treatment; UC: usual care.

Ethical Considerations

The study protocol was approved by an institutional review board (ID #5146C) and conducted in accordance with the Declaration of Helsinki and was consistent with Good Clinical Practices and applicable regulatory requirements. Patients and their legally acceptable representatives provided written informed consent. Further details of the study design can be accessed on the ClinicalTrials.gov page for the FIRST study [25].

Interventions

Caregivers randomly assigned to the study-provided psychoeducation group were invited to attend up to 16 live web-based sessions of MyHealios, a telehealth-based FP and skills training program for caregivers of patients with schizophrenia over a 6-month period. Each caregiver was assigned a trained and certified masters-level clinician who was independent of, and had no communication with, the patient’s UC team. All MyHealios clinicians received formal training and a training manual. They also underwent a certification process before conducting FP sessions with the caregivers enrolled in the study. The clinician who developed the FP curriculum and supervised the caregiver sessions was a PhD-level clinical psychologist with expertise in FP interventions for schizophrenia. Regular (eg, weekly) supervision was provided to the certified clinicians throughout the study. The adherence of clinicians to the FP and skills training program and manual was routinely evaluated using a 10-point fidelity scale based on observations of recorded caregiver sessions (including items such as agenda setting, collaboration, efficient use of time, interpersonal effectiveness, and following the structure of skills training), with fidelity ratings provided as feedback to clinicians and incorporated into supervision.

The MyHealios clinicians worked with the caregivers through live web-based sessions on a one-on-one basis throughout the program. Each session was 40 minutes in length and was conducted on the web at a time convenient for the caregiver. The web interface included live videos of both the caregiver and clinician, as well as a chat window to facilitate communication and caregiver participation in interactive activities. The number of delivered sessions and topics were determined jointly by the caregiver and clinician, with the teaching information and skills individually tailored to the caregiver. During each session, the caregiver presented problems that arose from caring for the patient and elaborated with specific examples. The clinician offered training and guidance on the appropriate methods to manage the identified problems.

Sessions were planned to occur weekly at the beginning of the program and decrease in frequency over the next 6 months as participants learned how to apply the skills in their day to day lives. A total of 3 modules were identified for initial completion by all caregivers (engagement and goal setting, communications, problem solving and goal achievement). Caregivers could then elect to complete any of the other modules in any order (coping, relapse prevention, delusions, low levels of activity, schizophrenia, anxiety, bipolar disorder, hallucinations, crisis identification and management, alcohol and drugs, depression, engaging the treatment team, and treatment adherence).

The UC group received support routinely provided by caregivers at the study sites. In both groups, patients and their caregivers were followed up for ≤12 months after the baseline assessment.
Assessments

Assessments, including those of TF events, were evaluated at baseline and at 3, 6, and 12 months. Patient illness self-management was evaluated using the self-reported Illness Management and Recovery (IMR) scale [26]. This self-reported scale contains 15 questions, each of which is answered on a 5-point Likert scale, with higher scores indicating better recovery status. The IMR total score (range 15-75) was derived as the sum of the 15 item scores. The severity of psychotic symptoms was rated using the Clinical Global Impression of Severity (CGI-S) scale [27] by a member of the patient’s treatment team (not a family clinician) who was not masked to the treatment assignment. The CGI-S rating scale rates the severity of a participant’s psychotic condition based on a 7-point global assessment of symptom severity from 1 (normal, not ill) to 7 (most extremely ill).

Caregiver-reported assessments were conducted at the same time as patient assessments. The Involvement Evaluation Questionnaire (IEQ) [28] was used to measure caregiver distress, and the 12-item Short Form Health Survey (SF-12) [29] was used to measure overall perceived physical health (physical component score [PCS]) and mental health (mental health component score [MCS]). The IEQ is designed to measure the consequences of caregiving on family members and friends of patients with schizophrenia. All items are scored on a scale of 0 (never) to 4 (always), and the total score ranges from 0 to 108. Higher IEQ scores indicate higher levels of caregiver burden. The SF-12 is a self-administered 12-item questionnaire designed to cover 8 domains of functional health status and well-being: physical functioning, role limitations because of physical health problems, bodily pain, general health perceptions, vitality, social functioning, role limitations because of emotional problems, and mental health. These scales are scored from 0 to 100, with higher scores indicating better health. A 1-week recall period was used for PCS and MCS.

Safety was assessed based on reported adverse events (AEs) and serious AEs (SAEs). AEs and SAEs were reported for patients, and only SAEs were reported for caregivers.

Medical resource utilization, including hospitalizations, emergency department visits, and outpatient services for patients and caregivers, was recorded on a patient health resource utilization form by chart abstraction and an interview or questionnaire if data were missing.

Statistical Analyses

The primary efficacy end point was the mean cumulative number of TF events experienced by patients over the 12-month study period. A proportional means model using the mean cumulative function was used to assess between-group differences in the mean cumulative number of TF events over 12 months. The mean cumulative function, as a function of time, was defined as the expected (mean) number of TF events in a given time interval since study day 1. The mean cumulative function for recurrent events and Kaplan–Meier (for time to the first event) analyses were performed for overall TF because of any event and for TF because of each of the events specified in the definition of TF. For secondary outcomes, changes from baseline to 3, 6, and 12 months in IEQ, IMR, SF-12, and CGI-S scores were analyzed using a mixed model repeated measures methodology with terms for study group, time, study group by time interaction, and baseline score. In addition, treatment-emergent AEs (TEAEs) were presented according to the treatment group (defined by the antipsychotic medication at baseline: paliperidone palmitate or oral antipsychotics).

The TF rate in the control group was assumed to be 0.50 based on a previous study with a similar end point [30]. The effect size in terms of a risk ratio of 0.60 was obtained from a meta-analysis of 18 randomized controlled studies examining the effect of face-to-face psychoeducation for caregivers on similar end points [24].

Results

Disposition

Owing to difficulties in study enrollment, recruitment was discontinued before the target enrollment of 300 pairs was met, resulting in underpowered statistical analyses. A total of 170 patient–caregiver pairs were screened in the study; 19 pairs had screening failures (some with more than 1 reason for a total of 21 screening failures [Figure 2]). As a result, 151 (88.8%) were randomly assigned to study-provided psychoeducation or UC; of these 151 pairs, 148 (98%) patient–caregiver pairs were included in the all-randomized analysis set (study-provided psychoeducation, n=73, 49.3%; UC, n=75, 50.7%). Of the 148 participants, 96 (64.9%) patients and 94 (63.5%) caregivers completed 12 months of follow-up; 52 (35.1%) patients and 54 (36.5%) caregivers discontinued participation before 12 months (Figure 2).
Figure 2. Disposition of study pairs in the FIRST study. Study patient pairs comprised individuals with schizophrenia and their designated caregivers. Patients could have ≥1 reason for screen failure. FIRST: Family Intervention in Recent-Onset Schizophrenia Treatment; UC: usual care.

Demographics and Baseline Characteristics

Patients’ and caregivers’ demographics and baseline characteristics were generally balanced across the study-provided psychoeducation and UC groups (Tables 1 and 2). The median patient age was 25.0 (range 18 to 35) years, suggesting that patients were early in their disease course; most patients were male (111/148, 75%), White (84/148, 56.8%), and living with family or friends (131/148, 88.5%). Of the 148 participants, at baseline, 49 (33.1%) patients were receiving paliperidone palmitate, and 99 (66.9%) were receiving oral antipsychotics. The mean CGI-S score was 4.2, indicating moderate severity of illness (Table 1). The median caregiver age was 52.5 (range 21-76) years, with most being female (116/148, 78.4%), White (87/148, 58.8%), and a parent of the patient (112/148, 75.7%). Baseline IEQ total scores, SF-12 PCS scores, and SF-12 MCS scores were similar in the study-provided psychoeducation and UC groups (Table 1).

Caregivers who discontinued participation early tended to be nonparent relatives with lower self-reported health on the SF-12 and high burden scores on the IEQ at baseline (Table 2). The baseline demographic and clinical characteristics of the patients whose caregivers discontinued participation by month 12 were similar.
Table 1. Demographics and baseline characteristics of patients with schizophrenia spectrum disorders and caregivers by study group (N=148).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Intervention&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Usual care (n=75)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Study-provided psychoeducation (n=73)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Patients</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>25.3 (4.7)</td>
<td>25.1 (5.0)</td>
<td>25.2 (4.8)</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>54 (74)</td>
<td>57 (76)</td>
<td>111 (75)</td>
</tr>
<tr>
<td>Female</td>
<td>19 (26)</td>
<td>18 (24)</td>
<td>37 (25)</td>
</tr>
<tr>
<td><strong>Race, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>45 (61.6)</td>
<td>39 (52)</td>
<td>84 (56.8)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>21 (28.8)</td>
<td>31 (41.3)</td>
<td>52 (35.1)</td>
</tr>
<tr>
<td>Multiple or other</td>
<td>7 (9.6)</td>
<td>4 (5.3)</td>
<td>11 (7.4)</td>
</tr>
<tr>
<td>Unknown or not reported</td>
<td>0 (0)</td>
<td>1 (1.3)</td>
<td>1 (0.7)</td>
</tr>
<tr>
<td><strong>Ethnicity, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>13 (17.8)</td>
<td>17 (22.7)</td>
<td>30 (20.3)</td>
</tr>
<tr>
<td>Not Hispanic or Latino</td>
<td>60 (82.2)</td>
<td>57 (76)</td>
<td>117 (79.1)</td>
</tr>
<tr>
<td>Unknown or not reported</td>
<td>0 (0)</td>
<td>1 (1.3)</td>
<td>1 (0.7)</td>
</tr>
<tr>
<td><strong>Living status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At home with family or friends</td>
<td>63 (86.3)</td>
<td>68 (90.7)</td>
<td>131 (88.5)</td>
</tr>
<tr>
<td>At home alone</td>
<td>5 (6.8)</td>
<td>5 (6.7)</td>
<td>10 (6.8)</td>
</tr>
<tr>
<td>Sheltered living</td>
<td>2 (2.7)</td>
<td>0 (0)</td>
<td>2 (1.4)</td>
</tr>
<tr>
<td>Other</td>
<td>3 (4.1)</td>
<td>1 (1.3)</td>
<td>4 (2.7)</td>
</tr>
<tr>
<td><strong>Diagnosis, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>44 (60.3)</td>
<td>38 (50.7)</td>
<td>82 (55.4)</td>
</tr>
<tr>
<td>Schizoaffective disorder</td>
<td>31 (42.5)</td>
<td>34 (45.3)</td>
<td>65 (43.9)</td>
</tr>
<tr>
<td>Schizophreniform disorder</td>
<td>0 (0)</td>
<td>5 (6.7)</td>
<td>5 (3.4)</td>
</tr>
<tr>
<td><strong>Functioning, mean (SD)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMR&lt;sup&gt;b&lt;/sup&gt; total score</td>
<td>48.3 (6.8)</td>
<td>49.2 (7.1)</td>
<td>48.8 (7.0)</td>
</tr>
<tr>
<td>CGI-S&lt;sup&gt;c&lt;/sup&gt; score</td>
<td>4.1 (1.1)</td>
<td>4.3 (1.1)</td>
<td>4.2 (1.1)</td>
</tr>
<tr>
<td><strong>Caregivers</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>52.1 (11.2)</td>
<td>49.0 (12.5)</td>
<td>50.5 (11.9)</td>
</tr>
<tr>
<td><strong>Sex, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>14 (19.2)</td>
<td>18 (24)</td>
<td>32 (21.6)</td>
</tr>
<tr>
<td>Female</td>
<td>59 (80.8)</td>
<td>57 (76)</td>
<td>116 (78.4)</td>
</tr>
<tr>
<td><strong>Race, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>47 (64.4)</td>
<td>40 (53.3)</td>
<td>87 (58.8)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>18 (24.7)</td>
<td>31 (41.3)</td>
<td>49 (33.1)</td>
</tr>
<tr>
<td>Multiple or other</td>
<td>7 (9.6)</td>
<td>2 (2.7)</td>
<td>9 (6.1)</td>
</tr>
<tr>
<td>Not reported or unknown</td>
<td>1 (1.4)</td>
<td>2 (2.7)</td>
<td>3 (2.0)</td>
</tr>
<tr>
<td><strong>Ethnicity, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>9 (12.3)</td>
<td>16 (21.3)</td>
<td>25 (16.9)</td>
</tr>
<tr>
<td>Not Hispanic or Latino</td>
<td>64 (87.7)</td>
<td>58 (77.3)</td>
<td>122 (82.4)</td>
</tr>
<tr>
<td>Not reported</td>
<td>0 (0)</td>
<td>1 (1.3)</td>
<td>1 (0.7)</td>
</tr>
<tr>
<td>Parameter</td>
<td>Intervention(^a)</td>
<td>Usual care (n=75)</td>
<td>Total (n=73)</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>------------------</td>
<td>--------------</td>
</tr>
<tr>
<td><strong>Relationship with patient, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent(^d)</td>
<td>56 (76.7)</td>
<td>56 (74.7)</td>
<td>112 (75.7)</td>
</tr>
<tr>
<td>Sibling</td>
<td>2 (2.7)</td>
<td>6 (8.0)</td>
<td>8 (5.4)</td>
</tr>
<tr>
<td>Other relative</td>
<td>5 (6.8)</td>
<td>4 (5.3)</td>
<td>9 (6.1)</td>
</tr>
<tr>
<td>Spouse or partner</td>
<td>6 (8.2)</td>
<td>5 (6.7)</td>
<td>11 (7.4)</td>
</tr>
<tr>
<td>Friend</td>
<td>3 (4.1)</td>
<td>2 (2.7)</td>
<td>5 (3.4)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (1.4)</td>
<td>2 (2.7)</td>
<td>3 (2.0)</td>
</tr>
<tr>
<td><strong>IEQ(^e) score, mean (SD)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>30.8 (16.4)</td>
<td>29.7 (17.3)</td>
<td>30.2 (16.8)</td>
</tr>
<tr>
<td>Tension</td>
<td>8.1 (5.4)</td>
<td>6.8 (6.0)</td>
<td>7.5 (5.7)</td>
</tr>
<tr>
<td>Supervision</td>
<td>3.1 (3.3)</td>
<td>3.8 (3.8)</td>
<td>3.5 (3.6)</td>
</tr>
<tr>
<td>Worrying</td>
<td>11.3 (6.2)</td>
<td>9.7 (5.8)</td>
<td>10.5 (6.0)</td>
</tr>
<tr>
<td>Urging</td>
<td>10.7 (5.9)</td>
<td>11.9 (7.1)</td>
<td>11.3 (6.6)</td>
</tr>
<tr>
<td>SF-12(^f) PCS(^g) score, mean (SD)(^h)</td>
<td>50.3 (10.3)</td>
<td>51.7 (9.4)</td>
<td>51.0 (9.8)</td>
</tr>
<tr>
<td>SF-12 MCS(^i) score, mean (SD)(^h)</td>
<td>45.3 (10.4)</td>
<td>48.7 (9.5)</td>
<td>47.0 (10.1)</td>
</tr>
</tbody>
</table>

\(^a\) All-randomized analysis set (all caregivers or patients who were randomly assigned and entered the study).

\(^b\) IMR: Illness Management and Recovery.

\(^c\) CGI-S: Clinical Global Impression of Severity.

\(^d\) Includes stepparents, foster parents, and adoptive parents.

\(^e\) IEQ: Involvement Evaluation Questionnaire.

\(^f\) SF-12: 12-item Short Form Health Survey.

\(^g\) PCS: physical component summary.

\(^h\) For SF-12 (PCS and MCS), there were 43 caregivers in the discontinued early group, 102 in the completed study group, and 145 in the total group.

\(^i\) MCS: mental component summary.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Caregiver discontinuation status(a)</th>
<th>Completed study (n=103)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patients</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td></td>
<td>25.2 (5.1)</td>
<td>25.2 (4.8)</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>31 (68.9)</td>
<td>80 (77.7)</td>
<td>111 (75)</td>
</tr>
<tr>
<td>Female</td>
<td>14 (31.1)</td>
<td>23 (22.3)</td>
<td>37 (25)</td>
</tr>
<tr>
<td><strong>Race, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>26 (57.8)</td>
<td>58 (56.3)</td>
<td>84 (56.8)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>17 (37.8)</td>
<td>35 (34.0)</td>
<td>52 (35.1)</td>
</tr>
<tr>
<td>Multiple or other</td>
<td>1 (2.2)</td>
<td>10 (9.7)</td>
<td>11 (7.4)</td>
</tr>
<tr>
<td>Unknown or not reported</td>
<td>1 (2.2)</td>
<td>0 (0)</td>
<td>1 (0.7)</td>
</tr>
<tr>
<td><strong>Ethnicity, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>8 (17.8)</td>
<td>22 (21.4)</td>
<td>30 (20.3)</td>
</tr>
<tr>
<td>Not Hispanic or Latino</td>
<td>37 (82.2)</td>
<td>80 (77.7)</td>
<td>117 (79.1)</td>
</tr>
<tr>
<td>Unknown or not reported</td>
<td>0 (0)</td>
<td>1 (1.0)</td>
<td>1 (0.7)</td>
</tr>
<tr>
<td><strong>Living status, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>At home with family or friends</td>
<td>41 (91.1)</td>
<td>90 (87.4)</td>
<td>131 (88.5)</td>
</tr>
<tr>
<td>At home alone</td>
<td>3 (6.7)</td>
<td>7 (6.8)</td>
<td>10 (6.8)</td>
</tr>
<tr>
<td>Sheltered living</td>
<td>0 (0)</td>
<td>2 (1.9)</td>
<td>2 (1.4)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (2.2)</td>
<td>3 (2.9)</td>
<td>4 (2.7)</td>
</tr>
<tr>
<td><strong>Diagnosis, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>29 (64.4)</td>
<td>53 (51.5)</td>
<td>82 (55.4)</td>
</tr>
<tr>
<td>Schizoaffective disorder</td>
<td>19 (42.2)</td>
<td>46 (44.7)</td>
<td>65 (43.9)</td>
</tr>
<tr>
<td>Schizophreniform disorder</td>
<td>0 (0)</td>
<td>5 (4.9)</td>
<td>5 (3.4)</td>
</tr>
<tr>
<td><strong>Functioning, mean (SD)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMR(^b) total score, mean (SD)</td>
<td>46.0 (6.2)</td>
<td>50.0 (7.0)</td>
<td>48.8 (7.0)</td>
</tr>
<tr>
<td>CGI-S(^c) score, mean (SD)</td>
<td>4.2 (1.0)</td>
<td>4.2 (1.2)</td>
<td>4.2 (1.1)</td>
</tr>
<tr>
<td><strong>Caregivers</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>47.8 (13.3)</td>
<td>51.7 (11.2)</td>
<td>50.5 (11.9)</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>9 (20)</td>
<td>23 (22.3)</td>
<td>32 (21.6)</td>
</tr>
<tr>
<td>Female</td>
<td>36 (80)</td>
<td>80 (77.7)</td>
<td>116 (78.4)</td>
</tr>
<tr>
<td><strong>Race, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>28 (62.2)</td>
<td>59 (57.3)</td>
<td>87 (58.8)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>16 (35.6)</td>
<td>33 (32)</td>
<td>49 (33.1)</td>
</tr>
<tr>
<td>Multiple or other</td>
<td>0 (0)</td>
<td>9 (8.7)</td>
<td>9 (6.1)</td>
</tr>
<tr>
<td>Not reported or unknown</td>
<td>1 (2.2)</td>
<td>2 (1.9)</td>
<td>3 (2)</td>
</tr>
<tr>
<td><strong>Ethnicity, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>7 (15.6)</td>
<td>18 (17.5)</td>
<td>25 (16.9)</td>
</tr>
<tr>
<td>Not Hispanic or Latino</td>
<td>38 (84.4)</td>
<td>84 (81.6)</td>
<td>122 (82.4)</td>
</tr>
<tr>
<td>Not reported</td>
<td>0 (0)</td>
<td>1 (1)</td>
<td>1 (0.7)</td>
</tr>
</tbody>
</table>

\(^a\) Caregiver discontinuation status; \(^b\) IMR: IMPACT Measure for Recovery; \(^c\) CGI-S: Clinician Global Impression of Severity.
<table>
<thead>
<tr>
<th>Relationship with patient, n (%)</th>
<th>Caregiver discontinuation status(a)</th>
<th>Completed study (n=103)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Discontinued early (n=45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parent(d)</td>
<td>29 (64.4)</td>
<td>83 (80.6)</td>
<td>112 (75.7)</td>
</tr>
<tr>
<td>Sibling</td>
<td>4 (8.9)</td>
<td>4 (3.9)</td>
<td>8 (5.4)</td>
</tr>
<tr>
<td>Other relative</td>
<td>5 (11.1)</td>
<td>4 (3.9)</td>
<td>9 (6.1)</td>
</tr>
<tr>
<td>Spouse or partner</td>
<td>4 (8.9)</td>
<td>7 (6.8)</td>
<td>11 (7.4)</td>
</tr>
<tr>
<td>Friend</td>
<td>2 (4.4)</td>
<td>3 (2.9)</td>
<td>5 (3.4)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (2.2)</td>
<td>2 (1.9)</td>
<td>3 (2)</td>
</tr>
<tr>
<td><strong>IEQ(e) score, mean (SD)</strong></td>
<td><strong>Total</strong></td>
<td><strong>Discontinued early (n=45)</strong></td>
<td><strong>Completed study (n=103)</strong></td>
</tr>
<tr>
<td></td>
<td>33.2 (18.2)</td>
<td>29.0 (16.1)</td>
<td>30.2 (16.8)</td>
</tr>
<tr>
<td>Tension</td>
<td>8.6 (6.1)</td>
<td>7.0 (5.6)</td>
<td>7.5 (5.7)</td>
</tr>
<tr>
<td>Supervision</td>
<td>4.2 (4.6)</td>
<td>3.1 (3.0)</td>
<td>3.5 (3.6)</td>
</tr>
<tr>
<td>Worrying</td>
<td>10.7 (6.2)</td>
<td>10.5 (6.0)</td>
<td>10.5 (6.0)</td>
</tr>
<tr>
<td>Urging</td>
<td>12.1 (7.2)</td>
<td>10.9 (6.2)</td>
<td>11.3 (6.6)</td>
</tr>
<tr>
<td>SF-12(f) PCS(g) score, mean (SD)(h)</td>
<td>48.5 (11.4)</td>
<td>52.0 (8.9)</td>
<td>51.0 (9.8)</td>
</tr>
<tr>
<td>SF-12 MCS(i) score, mean (SD)(h)</td>
<td>46.4 (10.2)</td>
<td>47.2 (10.0)</td>
<td>47.0 (10.1)</td>
</tr>
</tbody>
</table>

\(a\) Safety analysis set (all caregivers or patients who entered the study).

\(b\) IMR: Illness Management and Recovery.

\(c\) CGI-S: Clinical Global Impression of Severity.

\(d\) Includes stepparents, foster parents, and adoptive parents.

\(e\) IEQ: Involvement Evaluation Questionnaire.

\(f\) SF-12: 12-item Short Form Health Survey.

\(g\) PCS: physical component summary.

\(h\) For SF-12 (PCS and MCS), there were 43 caregivers in the discontinued early group, 102 in the completed study group, and 145 in the total group.

\(i\) MCS: mental component summary.

**Extent of Exposure to Caregiver Support and Education Program**

In the study-provided psychoeducation group, the mean number of caregiver sessions received was 7.7 (SD 5.88), and the median was 8 (range 0-16). Of the 73 participants, 40 (55%) caregivers who were randomly assigned to the study-provided psychoeducation intervention group received at least half of the modules (ie, ≥8 sessions); 12 (16%) caregivers did not receive any sessions, and 7 (10%) caregivers received only 1 session; 9 (12%) caregivers received 15 training sessions, and 3 (4%) received a maximum of 16 sessions (Multimedia Appendix 2). Of the 73 participants, 61 (84%) caregivers received at least one session, of whom all (n=73, 100%) received the *engagement and goal setting* module, 52 (85%) received the *communications* module, 40 (66%) received the *problem solving and goal achievement* module, and 35 (57%) received the *coping* module. The other modules were assigned to <50% of the caregivers (Tables 3 and 4). Caregivers who received fewer sessions were younger and more likely to be spouses or partners than those who received more sessions (Multimedia Appendix 3). Of the 75 caregivers in the UC group, 59 (79%) received no support services, and 7 (9%) were provided with case management or individual counseling or therapy.
Table 3. Summary of different modules administered to caregivers during study-provided psychoeducation (N=73)a.

<table>
<thead>
<tr>
<th>Caregiver-focused study-provided psychoeducation module description</th>
<th>Paliperidone palmitate (n=25)b, n (%)</th>
<th>Oral antipsychotics (n=48)b, n (%)</th>
<th>Total, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engagement and goal setting</td>
<td>22 (100)</td>
<td>39 (100)</td>
<td>61 (100)</td>
</tr>
<tr>
<td>Communications</td>
<td>18 (82)</td>
<td>34 (87)</td>
<td>52 (85)</td>
</tr>
<tr>
<td>Problem solving and goal achievement</td>
<td>15 (68)</td>
<td>25 (64)</td>
<td>40 (66)</td>
</tr>
<tr>
<td>Coping</td>
<td>13 (59)</td>
<td>22 (56)</td>
<td>35 (57)</td>
</tr>
<tr>
<td>Release prevention</td>
<td>8 (36)</td>
<td>10 (26)</td>
<td>18 (30)</td>
</tr>
<tr>
<td>Delusions</td>
<td>7 (32)</td>
<td>10 (26)</td>
<td>17 (28)</td>
</tr>
<tr>
<td>Low levels of activity</td>
<td>3 (14)</td>
<td>9 (23)</td>
<td>12 (20)</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>3 (14)</td>
<td>7 (18)</td>
<td>10 (16)</td>
</tr>
<tr>
<td>Anxiety</td>
<td>4 (18)</td>
<td>5 (13)</td>
<td>9 (15)</td>
</tr>
<tr>
<td>Bipolar</td>
<td>3 (14)</td>
<td>2 (5)</td>
<td>5 (8)</td>
</tr>
<tr>
<td>Hallucinations</td>
<td>3 (14)</td>
<td>2 (5)</td>
<td>5 (8)</td>
</tr>
<tr>
<td>Crisis identification and management</td>
<td>2 (9)</td>
<td>2 (5)</td>
<td>4 (7)</td>
</tr>
<tr>
<td>Alcohol and drugs</td>
<td>0 (0)</td>
<td>3 (8)</td>
<td>3 (5)</td>
</tr>
<tr>
<td>Depression</td>
<td>0 (0)</td>
<td>2 (5)</td>
<td>2 (3)</td>
</tr>
<tr>
<td>Engaging the treatment team</td>
<td>1 (5)</td>
<td>0 (0)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Treatment adherence</td>
<td>0 (0)</td>
<td>1 (3)</td>
<td>1 (2)</td>
</tr>
</tbody>
</table>

aAll-randomized analysis set (all caregivers who were randomly assigned and entered the study).
bA total of 12 caregivers (paliperidone palmitate, n=3; oral antipsychotics, n=9) did not receive any modules; percentages are given as a proportion of the caregivers receiving modules.

Table 4. Summary of different modules administered to caregivers during caregiver support in usual care (N=75)a.

<table>
<thead>
<tr>
<th>Usual care provided</th>
<th>Paliperidone palmitate (n=24), n (%)</th>
<th>Oral antipsychotics (n=51), n (%)</th>
<th>Total, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>21 (88)</td>
<td>38 (75)</td>
<td>59 (79)</td>
</tr>
<tr>
<td>Case management</td>
<td>1 (4)</td>
<td>6 (12)</td>
<td>7 (9)</td>
</tr>
<tr>
<td>Individual counseling or therapy</td>
<td>1 (4)</td>
<td>6 (12)</td>
<td>7 (9)</td>
</tr>
<tr>
<td>NAMIb</td>
<td>2 (8)</td>
<td>3 (6)</td>
<td>5 (7)</td>
</tr>
<tr>
<td>Group counseling or therapy</td>
<td>0 (0)</td>
<td>2 (4)</td>
<td>2 (3)</td>
</tr>
<tr>
<td>Option to join NAMI family-to-family education program</td>
<td>0 (0)</td>
<td>1 (2)</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Live interaction</td>
<td>0 (0)</td>
<td>1 (2)</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Supportive therapy</td>
<td>0 (0)</td>
<td>1 (2)</td>
<td>1 (1)</td>
</tr>
<tr>
<td>Website link</td>
<td>0 (0)</td>
<td>1 (2)</td>
<td>1 (1)</td>
</tr>
</tbody>
</table>

aAll-randomized analysis set (all caregivers who were randomly assigned and entered the study).
bNAMI: National Alliance on Mental Illness.

**Efficacy**

**TF Events**

A total of 89 TF events occurred during the study. Approximately 37% (23/63) of participants in the study-provided psychoeducation group and 37% (25/67) of participants in the UC group had at least 1 TF due to any event. TF rates were not associated with baseline CGI-S scores and did not differ between the study-provided psychoeducation and UC groups (P=0.90; Figure 3). Most TF events were because of psychiatric hospitalization (61/89, 69%) or psychiatric emergency department visits (13/89, 15%). Post hoc analyses also showed lower relapse rates in patients who received paliperidone palmitate than in those who received oral antipsychotics at all time points (Figure 4). Exploratory post hoc analyses were performed to investigate whether higher levels of caregiver participation in the study-provided psychoeducation intervention were associated with improved patient TF outcomes. There was no significant difference in the mean number of TFs because of any event.
between caregivers who received >8 sessions versus the overall UC group (36% vs 37%; \( P=.76 \)). In the study-provided psychoeducation group, TF rates were notably higher in patients whose caregivers received at least one session than in patients of caregivers who received 15 to 16 sessions (10/13, 77% vs 4/12, 33%; Table 5).

**Figure 3.** Cumulative mean functions of treatment failure because of any event in the study-provided psychoeducation and UC groups UC: usual care.

**Figure 4.** Treatment failure rates by antipsychotic treatment strata (post hoc analysis). Efficacy analysis set (n=130, all patients who entered the study and had at least one postbaseline efficacy assessment). UC: usual care.

**Table 5.** Summary of secondary end points: caregiver and patient secondary outcomes.

<table>
<thead>
<tr>
<th>Training sessions</th>
<th>Number of participants, N</th>
<th>Patients with ≥1 treatment failure, n (%)</th>
<th>Total treatment failures, n</th>
<th>Treatment failure rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>63</td>
<td>23 (36.5)</td>
<td>44</td>
<td>0.70</td>
</tr>
<tr>
<td>0-1</td>
<td>13</td>
<td>5 (38.5)</td>
<td>10</td>
<td>0.77</td>
</tr>
<tr>
<td>2-14</td>
<td>38</td>
<td>15 (39.5)</td>
<td>30</td>
<td>0.79</td>
</tr>
<tr>
<td>15-16</td>
<td>12</td>
<td>3 (25)</td>
<td>4</td>
<td>0.33</td>
</tr>
</tbody>
</table>

**Secondary Outcomes**

Caregiver IEQ total scores, SF-12 PCS and MCS scores, patient IMR total scores, and CGI-S scores improved from baseline to the follow-up assessments for both the study-provided psychoeducation and UC groups (Tables 6-8). However, there were no statistically significant differences in change from baseline between groups at any time point (\( P>.05 \) for all comparisons). Similar decreases from baseline in health resource use at months 6 and 12 were observed in the study-provided psychoeducation and UC groups.
Table 6. Summary of secondary end points: health resource utilization outcomes.

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>3 months</th>
<th>6 months</th>
<th>12 months</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Study-provided psychoeducation</td>
<td>UC&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Study-provided psychoeducation</td>
</tr>
<tr>
<td><strong>Caregiver outcomes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IEQ&lt;sup&gt;b,c&lt;/sup&gt; change from baseline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS&lt;sup&gt;d&lt;/sup&gt; mean (SE)</td>
<td>−2.95 (1.59)</td>
<td>−5.08 (1.54)</td>
<td>−6.96 (1.59)</td>
</tr>
<tr>
<td>Difference (study-provided psychoeducation vs UC)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS mean (95% CI)</td>
<td>2.13 (−2.25 to 6.52)</td>
<td>2.13 (−2.25 to 6.52)</td>
<td>0.66 (−3.79 to 5.10)</td>
</tr>
<tr>
<td>P value</td>
<td>.34</td>
<td>.34</td>
<td>.77</td>
</tr>
<tr>
<td>SF-12&lt;sup&gt;e&lt;/sup&gt; MCS&lt;sup&gt;f,g&lt;/sup&gt; change from baseline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS mean (SE)</td>
<td>0.63 (1.16)</td>
<td>1.27 (1.12)</td>
<td>2.86 (1.07)</td>
</tr>
<tr>
<td>Difference (study-provided psychoeducation vs UC)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS mean (95% CI)</td>
<td>−0.64 (−3.84 to 2.56)</td>
<td>−0.64 (−3.84 to 2.56)</td>
<td>−0.05 (−3.07 to 2.96)</td>
</tr>
<tr>
<td>P value</td>
<td>.69</td>
<td>.69</td>
<td>.97</td>
</tr>
<tr>
<td>SF-12 PCS&lt;sup&gt;h&lt;/sup&gt; change from baseline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS mean (SE)</td>
<td>−0.59 (0.89)</td>
<td>−0.75 (0.86)</td>
<td>−2.38 (0.87)</td>
</tr>
<tr>
<td>Difference (study-provided psychoeducation vs UC)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS mean (95% CI)</td>
<td>0.16 (−2.29 to 2.60)</td>
<td>0.16 (−2.29 to 2.60)</td>
<td>−0.79 (−3.23 to 1.64)</td>
</tr>
<tr>
<td>P value</td>
<td>.90</td>
<td>.90</td>
<td>.52</td>
</tr>
<tr>
<td><strong>Patient outcomes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IMR&lt;sup&gt;i,j&lt;/sup&gt; change from baseline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS mean (SE)</td>
<td>1.12 (0.75)</td>
<td>1.04 (0.72)</td>
<td>4.34 (0.89)</td>
</tr>
<tr>
<td>Difference (study-provided psychoeducation vs UC)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS mean (95% CI)</td>
<td>0.09 (−1.97 to 2.14)</td>
<td>0.09 (−1.97 to 2.14)</td>
<td>0.94 (−1.50 to 3.38)</td>
</tr>
<tr>
<td>P value</td>
<td>.93</td>
<td>.93</td>
<td>.45</td>
</tr>
<tr>
<td>CGI-S&lt;sup&gt;k,l&lt;/sup&gt; change from baseline</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS mean (SE)</td>
<td>−0.18 (0.10)</td>
<td>−0.12 (0.10)</td>
<td>−0.24 (0.11)</td>
</tr>
<tr>
<td>Difference (study-provided psychoeducation vs UC)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LS mean (95% CI)</td>
<td>−0.06 (−0.34 to 0.22)</td>
<td>−0.06 (−0.34 to 0.22)</td>
<td>0.06 (−0.25 to 0.37)</td>
</tr>
<tr>
<td>P value</td>
<td>.66</td>
<td>.66</td>
<td>.69</td>
</tr>
</tbody>
</table>

<sup>a</sup>UC: usual care.
<sup>b</sup>IEQ: Involvement Evaluation Questionnaire.
<sup>c</sup>Higher scores on the IEQ indicate a higher caregiver burden.
<sup>d</sup>LS: least squares.
<sup>e</sup>SF-12: 12-item Short Form Health Survey.
<sup>f</sup>MCS: mental component summary.
<sup>g</sup>Higher scores on the SF-12 indicate better health.
<sup>h</sup>PCS: physical component summary.
<sup>i</sup>IMR: Illness Management and Recovery.
<sup>j</sup>Higher scores on the IMR indicate better recovery status.
Table 7. Summary of secondary end points: decreases in health resource utilization (N=148).

<table>
<thead>
<tr>
<th>Health resource use</th>
<th>Baseline, n (%)</th>
<th>6 months, n (%)</th>
<th>12 months, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospitalizations</td>
<td>46 (31)</td>
<td>10 (7)</td>
<td>6 (4)</td>
</tr>
<tr>
<td>Emergency department visits</td>
<td>72 (49)</td>
<td>17 (12)</td>
<td>17 (12)</td>
</tr>
<tr>
<td>Intensive outpatient treatment</td>
<td>11 (7)</td>
<td>3 (2)</td>
<td>0 (0)</td>
</tr>
</tbody>
</table>

Table 8. Decreases in health resource utilization.

<table>
<thead>
<tr>
<th>Health resource utilization decrease from baseline to month 12</th>
<th>Study-provided psychoeducation group</th>
<th>Usual care group</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reductions in hospitalizations</td>
<td>62% to 21%</td>
<td>61% to 18%</td>
</tr>
<tr>
<td>Emergency department visits</td>
<td>41% to 12%</td>
<td>56% to 11%</td>
</tr>
<tr>
<td>Intensive outpatient treatment</td>
<td>6 to 0 patients</td>
<td>5 to 0 patients</td>
</tr>
</tbody>
</table>

Safety

Of the 148 patients, 84 (56.8%) reported at least 1 TEAE during the study (Multimedia Appendix 4). No TEAEs were considered to be related to study-specific procedures. In total, 3 deaths were reported (n=1, 33% suicide; n=1, 33% drug overdose; and n=1, 33% cerebral hemorrhage), all in the UC group; none were considered related to trial-specific procedures. Safety in the paliperidone palmitate group was consistent with the known safety profile of paliperidone palmitate in adults, with no new events identified [31-33].

Discussion

Principal Findings and Key Learnings

No differences were observed over the 12-month study period between the study-provided psychoeducation and UC groups in either patient outcomes (TFs such as relapse, illness management, and change in clinical functioning) or caregiver outcomes (burden and physical and mental health functioning), with both groups showing significant improvement. This study aimed to fill the gap in the evidence base for FP by providing information on the effects of FP delivered specifically to caregivers using a telehealth-based platform. FP programs share several common characteristics but can vary considerably in length, setting, and content [34]. Although the results of this study did not show a benefit of the FP intervention at the level of exposure reached, consideration of the study limitations and additional key insights is important for the continued development of efficacious telehealth FP interventions.

Studies of caregiver-directed psychosocial interventions with positive outcomes have typically been longer (mean 57 weeks) and have provided more overall sessions (mean 28 sessions) than this study [24]. The duration of the study-provided psychoeducation program was also shorter than the minimum duration of 9 months recommended for FP by some experts [8,34]. However, other factors may have also played a role in the null results. Among the caregivers assigned to the study–provided psychoeducation group, there was a moderate amount of module completion, with 55% (40/73) of caregivers receiving >8 sessions. Although 16% (12/73) of caregivers received either 15 or 16 sessions of the intervention, 26% (19/73) of caregivers received either 0 or 1 session. The findings of exploratory analyses suggest that the wide range of participation in study-provided psychoeducation may have limited our ability to detect group differences. Furthermore, for caregivers who were engaged in study-provided psychoeducation, the psychoeducational modules that focused on relapse prevention, schizophrenia, and treatment adherence were received by <50% of caregivers despite the relevance of these topics to coping with a recent TF experienced by a family member. Therefore, limited participation in the study-provided psychoeducation and limited attention to psychoeducation about relapse prevention might have resulted in caregivers receiving insufficient information to avert events such as relapses and hospitalizations.

Most published studies on FP have evaluated models that include patients in the intervention. Since the inception of FP in the 1970s, several models have evolved to meet the needs of families, including FP and support [35,36], behavioral family therapy [37], and multi-family group therapy [38]. Studies of in-person family- and caregiver-focused psychoeducation programs have shown significant benefits over UC [6,7,24]. A meta-analysis of 18 randomized controlled trials of caregiver-directed psychosocial interventions for schizophrenia demonstrated significant improvements compared with UC in hospitalizations, relapse, and other patient outcomes, including visits to emergency departments, suicide attempts, and deaths [24]. A meta-analysis of 21 randomized controlled trials of interventions for informal caregivers found improved experiences of caring, increased quality of life, and reduced psychological distress among caregivers [7]. In the FIRST study, patients were not directly involved in the study-provided psychoeducation program, and caregivers were the primary focus. It is possible that the inclusion of both caregivers and patients in sessions has greater potential to improve outcomes over treatment with UC [19,20]. Furthermore, caregivers enrolled in the study-provided psychoeducation intervention were expected to identify their own educational needs and guide treatment by selecting most of the educational modules taught.
in the program. Research has shown that individuals often misjudge their knowledge or competence [39].

An unexpectedly large percentage of caregivers (54/148, 36.5%) discontinued participation in the study. The most common reasons for discontinuation were withdrawal of consent (17/148, 11.5%), others (17/148, 11.5%); which included administrative reasons [eg, lost to follow-up and nonadherence with study procedures] and personal reasons [eg, moved out of town and no longer serving as a caregiver], lost to follow-up (13/148, 8.8%), and physician’s decision (5/148, 3.4%). Although caregiver demographic factors were similar between those who discontinued the study and those who completed the study, 80.6% (83/103) of caregivers who completed the study were parents of the patient compared with only 64% (29/45) of caregivers who dropped out. It is possible that the parents of patients may have been more committed and motivated to continue the study than caregivers who were not parents of the patient. In addition, per protocol, when a patient discontinued participation in the study, their caregivers were also discontinued. This may have also contributed to the high discontinuation rate among the caregivers.

The baseline characteristics of caregivers in the FIRST study may help to identify caregivers who are likely to sufficiently engage with a telehealth-based study-provided psychoeducation intervention and those who may need additional support to fully engage. In a post hoc analysis of the study-provided psychoeducation group comparing baseline characteristics of caregivers receiving ≤8 sessions with those receiving >8 sessions (Multimedia Appendix 3), caregivers who received >8 sessions were more likely to be older and parents of individuals with schizophrenia. Furthermore, except for the IEQ subscale score of worrying, the baseline IEQ total and subscale scores were lower among those who received >8 sessions, indicating lower caregiver burden. It is possible that caregivers with a higher burden may have been too distressed to engage in the program, regardless of the convenience of internet-based access to interventions, and dropped out early. As noted earlier, caregivers who discontinued participation within the first 12 months of the FIRST study were also more likely to be nonparent relatives with poorer health (Table 1). This finding may help future researchers develop strategies for adherence to treatment that may improve attendance, engagement, and continuous caregiver involvement.

Another limitation of this study was that the sample size was smaller than intended, which may have affected the ability to draw specific conclusions. In addition, patients were eligible for enrollment only if they had experienced at least one TF within 6 months of screening, indicating a high degree of clinical severity, and the observed TF rate in the FIRST study was higher than expected for comparable studies with similar sample sizes. The recovery period following a TF event (eg, psychiatric hospitalization) may be a particularly vulnerable period that requires an additional level of support not examined in this study to facilitate better outcomes. Furthermore, the median age of the patients in the FIRST study was 25.0 years, indicating that they were also early in the course of their illness. Typically, many patients have difficulty accepting their diagnosis [40] and experience high levels of stress, mood symptoms, and suicidal ideation during early illness [6]. The risk of relapse is very high during this period and can predict disease progression [6]. Implementing effective interventions early to prevent repeated relapses may reduce the associated decline in cognition and functioning [6].

The study-provided psychoeducation intervention was delivered across many study sites [31], which differed in the standard services provided for both the study-provided psychoeducation and UC groups. Another limitation of the implementation of the study-provided psychoeducation intervention is that the clinician provided by MyHealios was not a member of the treatment team; therefore, progress in the program was not integrated with patient care. This also precluded the ability of the clinician to relay potentially important clinical information learned from the caregiver to the treatment team about changes in the patient’s condition (eg, emergence of early signs of relapse and treatment nonadherence).

The results of this study coincide with a critical moment for telehealth interventions. Although telehealth interventions were only used by 8% of Americans in 2019, engagement with telehealth has grown dramatically in acceptance during the COVID-19 pandemic [41,42]. For example, in a community mental health authority in Michigan (Network180), the rates of telehealth services increased from 5% before the pandemic to 84% during the peak of the pandemic in 2020 [43]. In addition, many mental health professionals have recommended the ethical use of telehealth interventions to provide continued support and care to patients and caregivers throughout the pandemic rather than in-person interventions, noting that telehealth support can be just as effective and may result in fewer missed visits [44-46]. Insights on best practices for web-based delivery of mental health interventions are critically needed, and new models are under development [47]. Further research using FP methods, taking the lessons learned from the FIRST study into account, is warranted.

Conclusions

The findings from this study provide valuable insights into a supplemental telehealth-based FP provided in the treatment of patients with early-phase schizophrenia spectrum disorders receiving paliperidone palmitate or oral antipsychotic medication. Key insights include the potential importance of supporting sufficient caregiver engagement; communication between clinicians, patients, and family members regarding treatment plans; and ensuring a link between clinicians providing psychoeducation to patients and the rest of their treatment team. Future studies in which telehealth interventions include caregiver–patient sessions and multicarere group sessions are warranted [19]. Meanwhile, traditional methods of delivering FP to caregivers and patients with schizophrenia spectrum disorders continue to have great potential for reducing caregiver burden and improving patient outcomes. As more telehealth psychoeducation platforms become available, we anticipate a continued exploration of how to adapt these important support programs to telehealth, with the goal of increasing benefits to patients and families.
Acknowledgments

The authors thank ApotheCom for editorial and writing services, which were funded by Janssen Scientific Affairs, LLC. The authors also thank the patients and caregivers who participated in this study. The authors would also like to thank the investigators at the following sites who participated in this research study: Southern Illinois University School of Medicine; Cutting Edge Research Group; North County Clinical Research; Oklahoma Clinical Research Center; Murdoch Children’s Research Institute; Neuropsychiatry and Mohave Mental Health Clinic; Frontier Institute and Spokane Mental Health; UConn Health Center; Centerstone Mental Health and Research Institute (Bloomington, Indiana); Centerstone Research Institute (Nashville, Tennessee); Unison Behavioral Health Group; Zain Research; Nathan Kline Institute for Psychiatric Research; Atlanta Behavioral Research, LLC; KCRCMD Inc; Psychiatric and Behavioral Solutions, Fountain House; Mental Health Centers of Denver, Medical Research Group of Central Florida; Louisiana State University Health Sciences Center Psychiatry; LifeStream Behavioral Center, Inc; University of California Davis Health System; Richmond Behavioral Associates; BioBehavioral Research of Austin; Mid-America Clinical Research, LLC; University of Illinois at Chicago Medical Center Comprehensive Clinical Services, PC; Lincoln Research; Focus and Balance Research, LLC; University of Pittsburgh Medical Center; Galen Research. The authors also thank Katie Ashcroft and Healios, Inc. for their considerable contributions to this study.

Authors’ Contributions

This study was funded by Janssen Scientific Affairs, LLC. The sponsor was involved in the design and conduct of the study and the collection, analysis, and interpretation of data. All the authors provided direction and comments on the manuscript, reviewed and approved the final version before submission, made the final decision about where to publish these data, and approved submission to this journal. Authors had full access to the study data and take responsibility for data integrity and the accuracy of the analyses.

Conflicts of Interest

EDA has received consulting fees from F Hoffmann-La Roche; served on advisory boards for Indivior, Janssen, Neurocrine Biosciences, Sunovion, and Otsuka or Lundbeck; received research support from Alkermes, Astellas, Avanir, Biogen, Boehringer Ingelheim, InnateVR, Janssen, National Network of Depression Centers, Neurocrine Biosciences, Novartis, Pear Therapeutics, Pine Rest Foundation, Otsuka, Takeda, and Vanguard Research Group; has owned stock in AstraZeneca, Johnson & Johnson, Pfizer, Inc, and Moderna; and served as an investigator in this study but was not paid to be an author of this manuscript. JG, BM, and HLS are employees of Janssen Scientific Affairs, LLC, and stockholders of Johnson & Johnson, Inc. EK is a former employee of Janssen Scientific Affairs, LLC, and a stockholder of Johnson & Johnson, Inc.

Multimedia Appendix 1
CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) checklist (version 1.6.1).

Multimedia Appendix 2
Number of psychoeducation sessions provided to caregivers in the study-provided psychoeducation group (N=148).

Multimedia Appendix 3
Caregiver demographics and baseline characteristics by number of study-provided psychoeducation sessions (≤8 vs >8; safety analysis set).

Multimedia Appendix 4
Summary of safety data.

References


25. Family Intervention in Recent Onset Schizophrenia Treatment (FIRST). ClinicalTrials. URL: https://clinicaltrials.gov/ct2/show/NCT02600741 [accessed 2022-03-29]


Abbreviations

AE: adverse event  
CGI-S: Clinical Global Impression of Severity  
FIRST: Family Intervention in Recent-Onset Schizophrenia Treatment  
FP: family psychoeducation  
IEQ: Involvement Evaluation Questionnaire  
IMR: Illness Management and Recovery  
MCS: mental health component score  
PCS: physical component score  
SAE: serious adverse event  
SF-12: 12-item Short Form Health Survey  
TEAE: treatment-emergent adverse event  
TF: treatment failure  
UC: usual care
Review

Problematic Social Media Use in Adolescents and Young Adults: Systematic Review and Meta-analysis

Holly Shannon¹,², BSc; Katie Bush¹, BA; Paul J Villeneuve²,³*, PhD; Kim GC Hellemans²*, PhD; Synthia Guimond¹,²,⁴*, PhD

¹Department of Psychiatry, The Royal's Institute of Mental Health Research, University of Ottawa, Ottawa, ON, Canada
²Department of Neuroscience, Carleton University, Ottawa, ON, Canada
³School of Mathematics and Statistics, Carleton University, Ottawa, ON, Canada
⁴Department of Psychoeducation and Psychology, Université du Québec en Outaouais, Gatineau, QC, Canada
*these authors contributed equally

Corresponding Author:
Synthia Guimond, PhD
Department of Psychiatry
The Royal's Institute of Mental Health Research
University of Ottawa
1145 Carling Ave
Ottawa, ON, K1Z 7K4
Canada
Phone: 1 6135708402
Email: synthia.guimond@theroyal.ca

Abstract

Background: Technology is ever evolving, with more and more diverse activities becoming possible on screen-based devices. However, participating in a heavy screen-based lifestyle may come at a cost. Our hypothesis was that problematic social media use increased the prevalence of mental health outcomes.

Objective: This study seeks to systematically examine problematic social media use in youth and its association with symptoms of depression, anxiety, and stress.

Methods: A systematic search was conducted to identify studies in adolescents and young adults, using the databases Engineering Village, Psycinfo, Pubmed, and Web of Science. A total of 18 studies were identified, with a total of 9269 participants in our review and included in the meta-analysis.

Results: Our metaregression shows moderate but statistically significant correlations between problematic social media use and depression ($r=0.273$, $P<.001$), anxiety ($r=0.348$, $P<.001$), and stress ($r=0.313$, $P<.001$). We did not find evidence of heterogeneity of these summary correlations by age, gender, or year of publication.

Conclusions: This study provides further evidence of the association between problematic social media use and negative mental health among adolescents and young adults and supports future research to focus on the underlying mechanisms of problematic use of social media.

Trial Registration: PROSPERO CRD42021222309; https://tinyurl.com/2p9y4bjx

(JMIR Ment Health 2022;9(4):e33450) doi:10.2196/33450

KEYWORDS
problematic social media use; depression; anxiety; stress

Introduction

Technology is ever evolving, with more and more diverse activities becoming possible on screen-based devices. With this increasing engagement in the digital world, social networking sites have become an increasingly popular activity, especially among younger populations [1]. Adolescents and young adults represent a unique population in terms of social media users, as they are the first generations to grow up in a highly digitized society. Social media use is highly normative among young individuals: In 2016, 97.5% of young adults in the United States reported using at least one social media site regularly [2].
However, participating in a heavy screen-based lifestyle may come at a cost. A wealth of evidence suggests higher levels of social media use are associated with symptoms of anxiety [3-5], symptoms of depression [3,6-8], decreased psychological well-being [9], lower self-esteem [3], psychological distress [10-12], and loneliness [5]. A meta-analysis in young adults reports a small correlation between depressive symptoms and adolescent social media use, defined by frequency of use [13]. However, along with the evidence supporting the negative impacts of social media use, some reports suggest there may exist positive outcomes following use. For example, social media use has also been linked to higher quality of life, social support, well-being, and reduced stress [14,15].

Aside from excessive use of social media, typically defined on the basis of hours of use, the term of problematic use characterizes individuals who experience addiction-like symptoms as a result of their social media use [5]. Problematic social media use reflects a non–substance related disorder by which detrimental effects occur as a result of preoccupation and compulsion to excessively engage in social media platforms despite negative consequences [16]. While there exists no official diagnostic term or measurement, Andreasen et al [17] developed the Facebook Addiction Scale, which measures features of substance use disorder such as salience, tolerance, preoccupation, impaired role performance, loss of control, and withdrawal, to systematically score problematic Facebook use. This scale has been widely used to conceptualize problematic use as a behavioral addiction and has therefore also been modified to measure overall problematic social media use, instead of focusing on Facebook specifically [18]. Similar to high frequencies of social media use, problematic social media use has also been associated with poor mental health outcomes such as depression, anxiety, decreased well-being, and lower self-esteem [1,17,19-22]. A recent meta-analysis by Cunningham et al [23] found that problematic social media use was a stronger predictor of depressive symptoms when compared to the measure of time spent on social networking sites. Therefore, based on previous evidence, problematic social media use may be more imperative to examine than hours spent on social media platforms.

Researchers recognize youth and students as a vulnerable group compared to adults because their increased use of social media is occurring during a time of identity formation, where they are free to explore various life possibilities and develop new values [2]. Furthermore, their use occurs when critical brain circuits involved in emotion regulation and motivation are continuing to undergo development [24]. As social media plays a large role in their day-to-day lives, patterns and frequency of use have the potential to become problematic. On this level, youth are more at risk for facing cyberbullying [25], finding it difficult to disengage from the media and allowing it to interfere with their social relationships [26]; this in turn puts them at risk for experiencing negative emotional and psychosocial outcomes [27]. Therefore, younger individuals are a vulnerable group of social media users, and it is important to better understand the outcomes for well-being that are associated with this type of problematic social media use. Yet, the magnitude of impact social media has on adolescents and emerging adults, especially when considering problematic use, remains unclear.

With this background, we systematically examined and summarized, with the most current evidence, the strength of association between problematic social media use and multiple mental health outcomes. Specifically, we considered depressive symptoms, anxiety symptoms, and stress. Our a priori hypothesis was that problematic social media use adversely impacts all mental health outcomes measured. In addition, age, gender, and year of publication were investigated as covariates in the relationship between problematic social media use and all mental health outcome variables.

Methods

This meta-analysis was registered with the International Prospective Register of Systematic Reviews (PROSPERO; CRD42021222309). The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines were followed [28].

Inclusion and Exclusion Criteria

This systematic review included measures of problematic social media use, with depressive symptoms, anxiety symptoms, and stress as outcome measures, assessed by validated instruments. The studies included were cross-sectional and provided a measure of association between problematic use and at least one of the mental health outcomes. Studies must have included a measure of problematic use from the participants; simply indicating if the participant was a user of social media was not acceptable (eg, grouping users vs nonusers of social media). Social media use was also examined in general, without focusing on specific activities (eg, studies looking at specific screen content or comparisons on social media platforms, etc) or a specific platform (eg, Facebook). Problematic social media use scales must have been validated to specifically measure social media use in terms of addictive use, comprising criteria used when measuring substance use disorders. Studies included were restricted to English language, and ages 12 to 30 years. Studies were excluded if they only measured frequency or problematic use of the internet in general, as social media use specifically must have been measured. Studies were also excluded if social media was being used as a treatment/intervention or in a focus-group setting. Finally, studies were excluded if they only measured social media use in clinical populations.

Literature Search

A systematic literature search was conducted in April 2021 using the databases Engineering Village, PsycInfo, Pubmed, and Web of Science using the terms “social media,” “social networking,” “mental health,” “depression,” “depressive symptoms,” “anxiety,” and “stress.” These search terms were used to quantify social media use in terms of problematic use.

Assessment of Quality

All eligible studies were assessed for quality using an adapted version of the Newcastle-Ottawa quality assessment scale for cross-sectional studies, which was used to score the risk of bias for each study [29]. All studies were independently rated by HS.
and KB and given a score out of 10. Conflicts in scoring were resolved by discussion (Multimedia Appendix 1).

**Data Extraction**

For each study identified as eligible, the following information was extracted: study identification (authors, year of publication, and country conducted), study design (sample size, age range, mean age, gender, and questionnaire used to measure problematic social media use), outcome variables (questionnaire used to measure each outcome and measure of association). See Multimedia Appendix 1 for questionnaires used to measure problematic use and outcome variables for each study included in the meta-analysis.

**Statistical Analysis**

To quantify the association between problematic social media use and depressive symptoms, anxiety symptoms, and stress, we used the Pearson correlation coefficient ($r$). Problematic use was considered on a continuum, based on the score obtained from the questionnaire used, which measures problematic use as addiction-like tendencies. All data analysis was performed using the statistical software Stata (Stat Corp) [30]. A random effects model was used, as it does not assume a common effect size across studies. The variance of $r$ was calculated in order to obtain the standard error for each correlation coefficient. The effect size in all groups of analysis had a 95% confidence interval. Publication bias was evaluated by producing a funnel plot, and by performing the Egger test. Age, gender, and year of publication were investigated as covariates by adding mean age, the percentage of male participants reported, and publication year for each study into separate metaregression analyses.

**Ethical Considerations**

Since meta-analyses do not need Institutional Review Board approval, the authors did not seek ethics approval.

**Results**

**Literature Search**

The literature search yielded 2846 articles, with 2410 articles remaining after duplicates were removed (Figure 1). Articles were screened based on titles and abstract to remove any records that were not quantitative, did not assess one of the outcomes, or were longitudinal. After the first screening, 417 (17.30%) articles were considered to be eligible and were then screened based on full text to exclude any remaining records that did not meet the inclusion criteria. Of the remaining articles, 4 were excluded as they were reporting results using dichotomized continuous variables. These studies separated participants into groups based on the scores of their respective scales, and therefore could not be used in our meta-analysis. Additionally, any unpublished data were obtained by contacting the corresponding author. One study included reported statistics distinct to two separate samples; therefore, the two samples were coded independently [31]. The results from Kim et al [32] were excluded from the metaregression, as mean age was not reported or received when contacted. The correlation from Giordano et al [33] with problematic social media use was reported as a combined score of depressive and anxiety symptoms, which therefore could not be included in the meta-analysis. However, all variables were pooled together for the metaregression analyses, so they were included when examining age, gender, and publication year as covariates. Details on the final 18 studies and 9269 total participants included are summarized in Table 1.

Figure 1. Flow chart of the search process and studies included.
Table 1. Summary of included studies on the relationship between social media use and outcome variables (note that not all studies measured all three outcomes. Giordano et al [33] assessed anxiety and depression combined and was therefore only included in the meta-regression analyses).

<table>
<thead>
<tr>
<th>First author (year)</th>
<th>Sample size</th>
<th>Female, n (%)</th>
<th>Male, n (%)</th>
<th>Age (years) range (mean)</th>
<th>Country</th>
<th>Problematic use and depression (r)</th>
<th>Problematic use and anxiety (r)</th>
<th>Problematic use and stress (r)</th>
<th>Problematic use and depression and anxiety combined (r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Holm gren (2017) [34]</td>
<td>442</td>
<td>228 (51.6)</td>
<td>214 (48.4)</td>
<td>18-21 (18.86)</td>
<td>United States</td>
<td>0.29</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Wang (2018) [35]</td>
<td>365</td>
<td>190 (52)</td>
<td>175 (48)</td>
<td>14-18 (16.29)</td>
<td>China</td>
<td>0.18</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Apaolaza (2019) [36]</td>
<td>346</td>
<td>179 (51.7)</td>
<td>167 (48.3)</td>
<td>17-26 (18.73)</td>
<td>Spain</td>
<td>N/A</td>
<td>N/A</td>
<td>0.49</td>
<td>N/A</td>
</tr>
<tr>
<td>Hou (2019) [37]</td>
<td>641</td>
<td>477 (74.4)</td>
<td>164 (25.6)</td>
<td>17-25 (19.9)</td>
<td>China</td>
<td>0.22</td>
<td>0.22</td>
<td>0.11</td>
<td>N/A</td>
</tr>
<tr>
<td>Kir caburun (2019)</td>
<td>470</td>
<td>280 (59.6)</td>
<td>190 (40.4)</td>
<td>14-18 (16.29)</td>
<td>Turkey</td>
<td>0.03</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Mitra (2019) [38]</td>
<td>264</td>
<td>164 (62.2)</td>
<td>100 (37.8)</td>
<td>18-25 (21.56)</td>
<td>India</td>
<td>0.39</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Chen (2020) [39]</td>
<td>437</td>
<td>308 (70.5)</td>
<td>129 (29.5)</td>
<td>16-30 (24.21)</td>
<td>China</td>
<td>N/A</td>
<td>0.29</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Kim (2020) [32]</td>
<td>209</td>
<td>31 (14.8)</td>
<td>178 (85.2)</td>
<td>15-18 (N/A)</td>
<td>China</td>
<td>N/A</td>
<td>0.20</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Kir caburun, Demetro- vics (2020) [40]</td>
<td>344</td>
<td>282 (82)</td>
<td>62 (18)</td>
<td>18-25 (20.87)</td>
<td>Turkey</td>
<td>0.22</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Kir caburun, Griffiths (2020) [41]</td>
<td>460</td>
<td>281 (61)</td>
<td>179 (39)</td>
<td>18-26 (19.74)</td>
<td>Turkey</td>
<td>0.34</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Stockdale (2020) [26]</td>
<td>385</td>
<td>204 (53)</td>
<td>181 (47)</td>
<td>17-19 (18.01)</td>
<td>United States</td>
<td>0.28</td>
<td>0.24</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Wong (2020) [42]</td>
<td>300</td>
<td>178 (59.3)</td>
<td>122 (40.7)</td>
<td>18-24 (20.89)</td>
<td>Hong Kong</td>
<td>0.336</td>
<td>0.344</td>
<td>0.384</td>
<td>N/A</td>
</tr>
<tr>
<td>Yildiz (2020) [43]</td>
<td>451</td>
<td>214 (47.5)</td>
<td>237 (52.5)</td>
<td>13-17 (15.5)</td>
<td>Turkey</td>
<td>N/A</td>
<td>0.58</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Brailovskaia; Lithuani an sample (2021) [31]</td>
<td>1640</td>
<td>1123 (68.5)</td>
<td>517 (31.5)</td>
<td>18-29 (19.09)</td>
<td>Lithuania</td>
<td>0.305</td>
<td>0.329</td>
<td>0.246</td>
<td>N/A</td>
</tr>
<tr>
<td>Brailovskaia; German sample (2021) [31]</td>
<td>727</td>
<td>548 (75.4)</td>
<td>179 (24.6)</td>
<td>18-29 (21.47)</td>
<td>Germany</td>
<td>0.396</td>
<td>0.461</td>
<td>0.411</td>
<td>N/A</td>
</tr>
<tr>
<td>Giordano (2021) [33]</td>
<td>428</td>
<td>218 (50.9)</td>
<td>210 (49.1)</td>
<td>13-19 (17.38)</td>
<td>United States</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>0.314</td>
</tr>
<tr>
<td>He (2021) [44]</td>
<td>218</td>
<td>218 (100)</td>
<td>0 (0)</td>
<td>19-23 (19.6)</td>
<td>China</td>
<td>N/A</td>
<td>N/A</td>
<td>0.23</td>
<td>N/A</td>
</tr>
<tr>
<td>Kilincel (2021) [45]</td>
<td>1142</td>
<td>722 (63.2)</td>
<td>420 (36.8)</td>
<td>12-18 (15.6)</td>
<td>Turkey</td>
<td>N/A</td>
<td>0.417</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

aN/A: not applicable.

Problematic Social Media Use and Depressive Symptoms

When examining depression as an outcome, 11 studies presented associations between problematic social media use in adolescents and young adults. The Center of Epidemiologic Studies Depression Scale was most commonly used to measure depressive symptoms. The summary metaregression correlation between problematic social media use and depressive symptoms was 0.273 (95% CI 0.215-0.332, P<.001). There was heterogeneity in the measures of association across the studies (Figure 2) with an $\Gamma^2=83.2\%$, $Q^2=59.69$, and $P<.001$. The funnel plot (Multimedia Appendix 1) shows slight asymmetry, suggesting slight publication bias, however Egger’s test for small-study effects was not significant ($P=.35$).
Problematic Social Media Use and Anxiety Symptoms

When examining anxiety symptoms as an outcome, 9 studies were identified measuring an association with problematic social media use in adolescents and young adults. The Depression Anxiety Stress Scale was most commonly used to measure anxiety symptoms. The summary metaregression correlation between problematic social media use and anxiety symptoms was 0.348 (95% CI 0.270-0.426, *P* < .001). There was substantial heterogeneity in the measures of association across the studies (Figure 3) with an I² = 91.6%, Q² = 94.75, *P* < .001. The funnel plot (Multimedia Appendix 1) shows asymmetry, suggesting some publication bias being present; however, the Egger test for small-study effects was not significant (*P* = .30).
Finally, when examining stress as an outcome, only 6 studies were identified measuring an association with problematic social media use in adolescents and young adults. The summary metaregression correlation between problematic social media use and stress was 0.313 (95% CI 0.203-0.423, \( P < .001 \)). There was heterogeneity in the measures of association across the studies (Figure 4) with an \( I^2 = 92.6\% \), \( Q^2 = 67.59 \), \( P < .001 \). The funnel plot (Multimedia Appendix 1) shows symmetry, suggesting no publication bias, with no significant bias from the Egger test as well (\( P = .79 \)).
**Moderators of Problematic Social Media Use**

The metaregression assessing the impact of age as a covariate on the relationship between problematic social media use and all mental health outcomes combined showed that age was not significantly moderating the relationships ($P=.75$). When examining gender as a covariate in the relationship between problematic social media use and all mental health outcomes, gender did not significantly moderate the relationship ($P=.25$). Finally, year of publication also did not significantly moderate the relationship between problematic social media use and all mental health outcomes when added as a covariate ($P=.09$). See Multimedia Appendix 1 for metaregression plots.

**Discussion**

**Principal Findings**

This meta-analysis reports outcome measures of depression, anxiety, and stress in association with problematic social media use, specifically among adolescents and young adults. There is evidence for a significant relationship between problematic social media use in youth and negative mental health outcomes, particularly higher depression and anxiety symptoms, and greater stress. The strongest correlation was observed with anxiety; however, this also presented the most heterogeneity, likely due to the variety of assessments used to quantify symptoms of anxiety in the individual studies.

Although the correlations are moderate, this meta-analysis provides further evidence for the possible harms of problematic social media use. Previous meta-analyses examining time spent on social media and mental health show very small effect sizes, with most correlations being reported below $r=0.20$ [46-48]. One explanation for previously small correlations observed is the variability of social media content itself and the ways individuals are using or viewing their social media accounts. There has been evidence of multiple variables that can influence the severity of mental health outcomes such as night time-specific use, passive use, the number of social media platforms, motives for using social media, and so on [3,49-52]. Problematic social media use is a distinct pattern of use characterized by “addiction-like” symptoms based on behavioral and psychological attributes. It is characterized not only by time spent on social media, but also by measuring the extent of symptoms similar to a substance-related disorder, such as withdrawal, tolerance, and dependence [22]. Therefore, problematic social media use could represent a more clinically meaningful behavior to direct research, as a stronger relationship is seen with adverse mental health symptoms compared to previous studies investigating time spent on social networking sites or screen time in general [23,53,54].

The influence of age is still highly debated with evidence pointing toward younger social media users being more likely to have worse mental health symptoms compared to older users [55], whereas others have found no significant age effect with time spent on social media [56]. Cunningham et al [23] found...
age did not moderate the relationship between problematic social media use and depression; however, this study was performed in a mainly adult sample. Likewise, in our meta-analysis, age did not significantly moderate the relationship between the outcome variables combined and problematic use. This is likely due to the restricted age range, as the mean age between individual studies were analogous. Higher social media usage, along with developmental vulnerabilities, in adolescents and young adults has been proposed to explain the higher association with worse mental health outcomes compared to adults [57,58]. However, when looking specifically at mental health associated with problematic social media use as a behavior, the severity of reported problematic use symptoms may be more imperative to consider rather than age. Future research could directly compare adolescents to adults to examine if a difference in correlational strength is present, specifically when measuring problematic use.

Gender was examined as a moderator by including the percentage of male participants from each study into a metaregression analysis. Gender did not significantly moderate the relationship between problematic social media use and mental health, suggesting the association between mental health symptoms and problematic use of social media is not different between genders. Studies included in this meta-analysis did not specify if they assessed biological sex. Future research should provide more specific results for each group for both sex and genders to allow future meta-analyses to summarize this information and provide insight into gaps in the current literature on problematic use of social media [23,51,59].

Year of publication did not significantly moderate the relationship between problematic social media use and mental health outcomes. Although there are increased rates of social media use in adolescents and young adults over time, this may not be directly pertinent in the strength of the association between mental health and problematic use [23,60]. Year of publication may be more indicative of the prevalence of social media use as it increases with the growing use of technology [60]. Along with previous data, it is suggested that mental health symptoms associated with problematic social media use do not appear to be worsening over time; however, longitudinal studies exploring this specific aspect are needed.

**Strengths and Limitations**

This study is not without limitations. The number of studies included in each meta-analysis was limited; therefore, the results are somewhat limited in power. Secondly, the results are based on cross-sectional correlational data. Therefore, a causal relationship cannot be inferred from the direct impact of social media on mental health outcomes of depressive symptoms, anxiety symptoms, or stress. It is possible that there are likely bidirectional effects between poor mental health and social media use [61]. In addition, the research studies included in the meta-analysis used did not report assessing the presence of a clinical diagnosis; therefore, it is unknown how many participants already had a known or possible clinical psychiatric diagnosis. This could influence the results of the outcome variables being measured, as it is unknown if individuals are more likely to have negative social media experiences or consequences as a result of using social media compared to individuals without a mental health diagnosis. Although the included questionnaires were previously validated, the majority of the research relies on self-report measures, also presenting as a limitation to the results reported.

**Future Directions**

Overall, there is a lack of research providing evidence on the mental health outcomes of social media use, particularly patterns of problematic use in younger populations. In order to thoroughly understand the direct implications of problematic social media use, longitudinal studies could aid in providing more causational conclusions, as cross-sectional methodology is limited in its ability to draw conclusions beyond correlation [62]. In addition to a longitudinal design, understanding the biological basis of problematic use could contribute to understanding vulnerability to negative mental health outcomes.

Future studies exploring the relationship between problematic social media and mental health outcomes would also benefit from including more detailed information on how participants are using various platforms. Indeed, there are several other scales exploring social media use that explore motivations for and mood associated with use (eg, social media use integration scale), which may provide greater depth of understanding around these associations. Finally, in traditional clinical practices for substance use disorders, treatment is often based on abstinence. For problematic social media use, total abstinence may not be a realistic option in today’s technology-based culture. Therefore, there should also be an increasing focus on identifying healthy ways to use social media in order to avoid the development of problematic use.

**Conclusions**

The findings from this study provide further evidence of the association between problematic social media and negative mental health outcomes of depression, anxiety, and stress among adolescents and young adults. Although there is a large amount of evidence pointing toward the negative impacts of social media on mental health, there is still a need for further research to provide conclusive results on the causal relationship and how social media can be used without taking a toll on the mental health of users. Considering the omnipresence of social media among youth, more resources should be allocated to better understand the relationship between use and mental health symptoms and to prevent such negative outcomes.

**Conflicts of Interest**

None declared.

Multimedia Appendix 1
Quality assessment rating for each study included in the meta-analysis. Questionnaires used for each study to measure problematic use and the outcome variables. Funnel plots and metaregressions.

References


Abbreviations

PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROSPERO: Prospective Register of Systematic Reviews
Original Paper

Web-Based Single Session Intervention for Perceived Control Over Anxiety During COVID-19: Randomized Controlled Trial

Michael Mullarkey, PhD; Mallory Dobias, MA; Jenna Sung, MA; Isaac Ahuvia, BA; Jason Shumake, PhD; Christopher Beevers, PhD; Jessica Schleider, PhD

Department of Psychology, Stony Brook University, Stony Brook, NY, United States

Corresponding Author:
Michael Mullarkey, PhD
Department of Psychology
Stony Brook University
100 Nicolls Rd
Stony Brook, NY, 11794
United States
Phone: 1 6316326000
Email: Michael.Mullarkey@stonybrook.edu

Abstract

Background: Anxiety is rising across the United States during the COVID-19 pandemic, and social distancing mandates preclude in-person mental health care. Greater perceived control over anxiety has predicted decreased anxiety pathology, including adaptive responses to uncontrollable stressors. Evidence suggests that no-therapist, single-session interventions can strengthen perceived control over emotions like anxiety; similar programs, if designed for the COVID-19 context, could hold substantial public health value.

Objective: Our registered report evaluated a no-therapist, single-session, online intervention targeting perceived control over anxiety in the COVID-19 context against a placebo intervention encouraging handwashing. We tested whether the intervention could (1) decrease generalized anxiety and increase perceived control over anxiety and (2) achieve this without decreasing social-distancing intentions.

Methods: We tested these questions using a between-subjects design in a weighted-probability sample of US adults recruited via a closed online platform (ie, Prolific). All outcomes were indexed via online self-report questionnaires.

Results: Of 522 randomized individuals, 500 (95.8%) completed the baseline survey and intervention. Intent-to-treat analyses using all randomized participants (N=522) found no support for therapeutic or iatrogenic effects; effects on generalized anxiety were \( d = -0.06 \) (95% CI –0.27 to 0.15; \( P = .48 \)), effects on perceived control were \( d = 0.04 \) (95% CI –0.08 to 0.16; \( P = .48 \)), and effects on social-distancing intentions were \( d = -0.02 \) (95% CI –0.23 to 0.19; \( P = .83 \)).

Conclusions: Strengths of this study included a large, nationally representative sample and adherence to open science practices. Implications for scalable interventions, including the challenge of targeting perceived control over anxiety, are discussed.

Trial Registration: ClinicalTrials.gov NCT04459455; https://clinicaltrials.gov/show/NCT04459455

(Keywords)

anxiety; COVID-19; single-session intervention; SSI; perceived control; intervention; mental health; control; online intervention; telemedicine; telehealth; scalable

Introduction

Background

Social isolation, looming threats of infection, and declining confidence in our abilities to cope have been spurred by the COVID-19 pandemic [1]. For many, this cocktail of stressors has led to increased anxiety. Based on large surveys of health care workers [2], their family members [3], and the general population [4] in China during the COVID-19 pandemic’s escalation, 24%-33% of people met criteria for an anxiety disorder. These rates are roughly double the point prevalence rate of anxiety disorders from a previous representative sample [5].
Likewise, levels of anxiety symptoms appear to be rising among US citizens. The COVID-19 pandemic has had wide and enduring negative effects on the mental health of individuals across the life span [6]. In a nationally representative survey conducted March 11-15, 2020 (n=1216), 32% of American adults reported worry due to COVID-19 had negatively impacted their mental health, and this rate climbed to 45% when the same question was asked in another nationally representative sample conducted March 25-30, 2020 (n=1226) [7].

Certainly, not all of these shifts reflect increases in pathological anxiety (versus situation-appropriate worry), but reasons remain for serious clinical concern. Increased time spent worrying about COVID-19 relates to more severe anxiety pathology—both in health care workers [8] and the general population in China [9]. Increased anxiety symptoms could also have negative public health effects during a pandemic. For example, 28% of people with anxiety disorders seek treatment in emergency rooms each year [10], frequently due to somatic symptoms with no medical cause (eg, panic symptoms like unspecified chest pain) [11]. As anxiety rates increase, so too could these often-unnecessary hospital visits, further exacerbating patient burden in already overwhelmed emergency medicine departments. Identifying an intervention to facilitate independent coping with anxiety—ideally, one that is brief and easily scalable—could help mitigate negative effects of increasing anxiety nationwide.

Perceived control, or one’s subjectively felt ability to control one’s environment and inner experiences, prospectively predicts lower distress during and following numerous uncontrollable stressors, from experiencing sexual assault [12] to recovering from breast cancer surgery [13] and a heart attack [14]. If one perceives control over their ability to reduce anxious responses (eg, racing thoughts, pounding heart), theory suggests that one is likely to experience less distress, regardless of actual control [15,16]. Empirical evidence consistently supports this idea. Individuals reporting lower perceived control of their internal experiences exhibit higher levels of anxiety (-ranging from nonclinical to clinical levels), regardless of objective levels of control [17,18]. Adults in community and nonclinical samples reporting lower perceived internal control have shown higher prevalence rates of anxiety disorders and more severe anxiety symptomatology versus those reporting higher levels of perceived internal control [19-21]. With respect to prospective associations [22], lower levels of perceived internal control have predicted higher future anxiety symptom severity in adults (including both social and generalized anxiety severity). Likewise, a meta-analysis exploring low perceived control as a transdiagnostic risk factor for anxiety disorders [23] found, across studies and diagnoses, perceived control was negatively linked with both trait and pathological anxiety severity. Experimental evidence suggests that one can reliably decrease anxiety related to low perceived internal control by increasing one’s capacity to alter their own thoughts, emotions, and experiences—for instance, by teaching specific skills or strategies to manage inner experiences [24-27].

During the present pandemic, one’s perceived control over the circumstances may be (and remain) understandably low: No individual can slow its course single-handedly. In fact, perceived control of one’s environment is largely unrelated to anxiety following circumstantial stressors (eg, undergoing basic military training) [28]. However perceived control over one’s own anxiety may remain high—and can be augmented—even amid uncontrollable circumstances like a psychiatric hospitalization [29]. In contrast with perceived control over one’s environment, perceived control over one’s own anxiety is negatively associated with generalized anxiety symptoms, even after controlling for the Big Five personality traits [22], and prospectively predicts fewer daily internalizing symptoms over the course of a month [30]. Psychosocial treatments such as intensive psychiatric hospitalization [29] and acceptance and commitment therapy [31] appear to increase perceived control of anxiety, and these increases are associated with declines in anxiety symptoms during treatment. Although we cannot be certain perceived control will demonstrate the same links to anxiety symptoms during a population-level uncontrollable event (eg, a global pandemic), it is a promising preliminary target for intervention. Importantly, preliminary evidence suggests that a self-administered, online, and single session intervention (SSI) can increase perceived control over emotions like anxiety from pre- to postintervention [32], in turn predicting decreases in anxiety severity 9 months later [33].

SSIs consistently demonstrate similar effect sizes to multisession therapies on mental health outcomes [34-36]. Meta-analytic evidence also indicates self-administered, online SSIs yield similar effect on mental health outcomes as therapist-directed SSIs [34,37]. Brief interventions that teach simple, repeatable skills, such as goal setting, may exert larger effects on psychopathology compared with “information-only” and norm-referencing interventions [37]. In fact, early research identifies a combination of (1) normalizing experiences via neuroscientific explanations of mental health difficulties (with care to not communicate these difficulties are thereby inherent and unchangeable [38]), (2) providing testimonials from others to reinforce this norm and introduce examples of repeatable skill use, and (3) empowering participants as helpers by asking them to practice the repeatable skill during the session and share their advice for how to implement the repeatable skill with others [39], as potentially helpful components of self-administered SSIs. This format allows for “minimal” interventions that retain efficacy [40-42]. Indeed, interventions as short as 5 minutes can improve mental health-related outcomes [37], consistent with findings that interventions of similar length can improve academic performance [43] and increase later egalitarian actions [44]. This intervention format can also more effectively scale up to meet the mental health needs of large numbers of people than traditional face-to-face therapy or longer online treatments [45,46].

Online, self-administered SSIs are also more easily, rapidly testable in representative samples than interventions requiring clinician contact (either in-person or remotely). Clinical trials of mental health treatments are generally underpowered [47] and nonrepresentative of the general population [48]. In larger, more representative clinical trials of clinician-dependent treatments, the recruitment process often requires several years [49]. This timeline is wholly incompatible with testing interventions to mitigate harms of immediate crises, including the COVID-19 pandemic. Rather, such tests require
interventions that may be evaluated rapidly, iterated if necessary, and disseminated appropriately while the pandemic is still ongoing. Online, self-administered SSIs fulfill these criteria, as there is evidence weighted-probability samples can be collected in 2 to 3 days for survey research [50], and SSIs found to be efficacious could be disseminated immediately, and broadly, online [51]. Even if participants are half as willing to complete an SSI program embedded within a survey, compared with a survey on its own, baseline data collection could still be completed in less than 1 week.

We therefore evaluated whether an online, self-administered SSI designed to strengthen perceived control over anxiety in the context of the COVID-19 pandemic (Contain COVID Anxiety) increased perceived control over anxiety immediate post-SSI and decreases general anxiety 2 weeks later more than a placebo, handwashing-plan SSI (Remain COVID Free) in a weighted-probability sample of the United States (n=500, See Multimedia Appendix 1 for the full text of both SSIs). See Table 1 for all confirmatory hypotheses and guidelines for interpretations of results.

Our primary hypothesis concerned whether the Contain COVID Anxiety SSI decreased generalized anxiety symptoms 2 weeks later more than the placebo SSI.
### Table 1. Design table.

<table>
<thead>
<tr>
<th>Question and hypotheses</th>
<th>Sampling plan (eg, power analysis)</th>
<th>Analysis plan&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Interpretation given to different outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Does the Contain COVID Anxiety SSI&lt;sup&gt;b&lt;/sup&gt; decrease generalized anxiety symptoms from baseline to 2 weeks later more than the placebo SSI Remain COVID Free?</td>
<td>H1: Generalized anxiety decreases more when participants are randomized to the Contain COVID Anxiety SSI than when participants are randomized to the Remain COVID Free placebo SSI.</td>
<td>H1: Test for assumptions and apply transformations as necessary. Take the 2-week follow-up generalized anxiety mean and enter it as the dependent variable in a linear model with the baseline generalized anxiety mean and treatment condition as predictors.</td>
<td>Support for H1: If the ( P ) value for condition is &lt;0.0167 in the H1 linear model and the 95% CI for the difference in generalized anxiety is negative and does not include 0 when participants are randomized to the Contain COVID Anxiety SSI, we will reject H0 and interpret the study as supporting Contain COVID Anxiety decreasing generalized anxiety more than the placebo SSI Remain COVID Free. Lack of support for H1: If the ( P ) value for the equivalence test described in the “Support for H0” section is &gt;0.0167, we will interpret the study as producing evidence that the Contain COVID Anxiety SSI is neither superior nor equivalent or inferior to the Remain COVID Free SSI at decreasing generalized anxiety.</td>
</tr>
<tr>
<td>H0: Generalized anxiety either does not increase more when participants are randomized to the Contain COVID Anxiety SSI than when participants are randomized to the Remain COVID Free placebo SSI, or generalized anxiety increases more when participants are randomized to the Remain COVID Free placebo SSI than when participants are randomized to the Contain COVID Anxiety placebo SSI.</td>
<td>H0: n=150 for 95% power</td>
<td>H0: Test for assumptions and apply transformations as necessary. Take the 2 weeks later generalized anxiety mean and predict it with the baseline generalized anxiety mean. Enter the mean and SD of the standardized residuals from that model for when the condition is Remain COVID Free and the mean and SD of the standardized residuals from that model when the condition is Contain COVID Anxiety into a between-groups equivalence test with equivalence bounds of ( d=-0.66 ) to ( d=0.33 ).</td>
<td>Support for H0: If the ( P ) value for condition is &lt;0.0167 in the H1 linear model and the 95% CI for the difference in generalized anxiety is positive and does not include 0 when participants are randomized to the Contain COVID Anxiety SSI or the ( P ) value or intervention order is &gt;0.0167 in the H1 linear model, we will run the between-groups equivalence test described in the analytic plan for H0. If the equivalence test has a ( P ) value &lt;0.0167, we will interpret the results as indicating the Contain COVID Anxiety SSI was equivalent or inferior to the Remain COVID Free SSI at improving generalized anxiety. Lack of support for H0: If the ( P ) value for the equivalence test described in the “Support for H0” section is &gt;0.0167, we will interpret the study as producing evidence that the Contain COVID Anxiety SSI is neither superior nor equivalent or inferior to the Remain COVID Free SSI at decreasing generalized anxiety.</td>
</tr>
</tbody>
</table>

**Does the Contain COVID Anxiety SSI increase perceived control over anxiety from baseline to immediately post-SSI more than the placebo SSI Remain COVID Free?**

| H1: Perceived control over anxiety increases more when participants are randomized to the Contain COVID Anxiety SSI than when participants are randomized to the Remain COVID Free placebo SSI. | H1: n=350 for 95% power | H1: Test for assumptions and apply transformations as necessary. Take the post-SSI perceived control over anxiety mean and enter it as the dependent variable in a linear model with baseline perceived control over anxiety mean and treatment condition as predictors. | Support for H1: If the \( P \) value for condition is <0.0167 in the H1 linear model and the 95% CI for the difference in perceived control over anxiety is positive and does not include 0 when participants are randomized to the Contain COVID Anxiety SSI, we will reject H0 and interpret the study as supporting Contain COVID Anxiety increasing perceived control over anxiety more than the placebo SSI Remain COVID Free. Lack of support for H1: If the \( P \) value for the equivalence test described in the “Support for H0” section is >0.0167, we will interpret the study as producing evidence that the Contain COVID Anxiety SSI is neither superior nor equivalent or inferior to the Remain COVID Free SSI at improving perceived control over anxiety. |
The placebo SSI. This pattern of results would indicate a larger decrease in generalized anxiety symptoms from immediately pre-SSI to 2 weeks later when participants were randomized to the Contain COVID Anxiety SSI instead of the placebo SSI. This pattern of results would indicate a larger decrease in generalized anxiety symptoms occurs pre-SSI to 2 weeks later in the Contain COVID Anxiety SSI compared with the placebo SSI.

### Hypothesis 1

We expected a larger decrease in generalized anxiety symptoms from immediately pre-SSI to 2 weeks later when participants were randomized to the Contain COVID Anxiety SSI compared with the placebo SSI.

#### Sampling plan

<table>
<thead>
<tr>
<th>Question and hypotheses</th>
<th>Sampling plan (eg, power analysis)</th>
<th>Analysis plan</th>
<th>Interpretation given to different outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>H0: Perceived control over anxiety either does not increase more when participants are randomized to the Contain COVID Anxiety SSI than when participants are randomized to the Remain COVID Free placebo SSI or perceived control over anxiety increases more when participants are randomized to the Remain COVID Free placebo SSI than when participants are randomized to the Contain COVID Anxiety placebo SSI.</td>
<td>H0: n=150 for 95% power</td>
<td>H0: Test for assumptions and apply transformations as necessary. Take the post-SSI perceived control over anxiety mean and predict it with the baseline perceived control over anxiety mean. Enter the mean and SD of the standardized residuals from that model for when the condition is Remain COVID Free and the mean and SD of the standardized residuals from that model when the intervention order is Contain COVID Anxiety into a between groups equivalence test with equivalence bounds of ( d=0.21 ).</td>
<td>Support for H0: If the ( P ) value for the equivalence test described in the “Support for H0” section is &gt;.0167, we will interpret the results as indicating the Contain COVID Anxiety SSI was equivalent or inferior to the Remain COVID Free SSI at improving perceived control over anxiety. Lack of support for H0: If the ( P ) value for the equivalence test in the analysis plan column is &lt;.0167, we will interpret the result as support for the hypothesis that social distancing intentions are not increased or decreased pre to post both Contain COVID Anxiety (even if the paired ( t ) test for H0 in the analysis plan column also has a ( P ) value &lt;.0167).</td>
</tr>
<tr>
<td>H0: Social distancing intentions do not increase or decrease pre-Contain COVID Anxiety SSIs to immediate post-Contain COVID Anxiety SSI.</td>
<td>H1: n=154 for 95% power</td>
<td>H1: Test for assumptions and apply transformations as necessary. Enter the mean and SD of the social distancing intention composites at baseline and post-Contain COVID Anxiety SSI (only among people who were randomized to the Contain COVID Anxiety) into a paired equivalence test with equivalence bounds of ( d=-0.33 ) to ( d=0.33 ).</td>
<td>Support for H1: The ( P ) value for the paired equivalence test described in the analysis plan column is &lt;.0167. We will interpret this result as support for the hypothesis that social distancing intentions are not increased or decreased pre to post both Contain COVID Anxiety (even if the paired ( t ) test for H0 in the analysis plan column also has a ( P ) value &lt;.0167).</td>
</tr>
<tr>
<td>H0: Social distancing intentions either increase or decrease pre- to immediate post-Contain COVID Anxiety SSI</td>
<td>H0: n=156 for 95% power</td>
<td>H0: Test for assumptions and apply transformations as necessary. Enter the mean and SD of the social distancing intention composites at baseline and post-Contain COVID Anxiety SSI (only among people who were randomized to the Contain COVID Anxiety) into a paired ( t ) test.</td>
<td>Support for H0: The ( P ) value for the paired equivalence test in the analysis plan column is &gt;.0167, and the ( P ) value for the paired ( t ) test in the analysis plan column is &lt;.0167. We will interpret this result as supporting the hypothesis that social distancing intentions either increased or decreased as the result of completing the Contain COVID Anxiety SSI. We will examine the direction of the effect by looking at the direction of the effect size (positive effect size = increase in social distancing intentions; negative effect size = decrease in social distancing intentions).</td>
</tr>
</tbody>
</table>

\( a \)Refer to the R code on the open science framework page for the power analysis and analysis plan [52].

\( b \)SSI: single-session intervention.

---

### Does completing the Contain COVID Anxiety SSI have an association with social distancing intentions statistically equivalent to 0?

#### Hypothesis 1

We also tested whether the active SSI Contain COVID Anxiety can be deployed at scale without reducing intentions to engage in social distancing. Social distancing remains the most potent known nonpharmacological intervention to reduce the spread of the SARS-Cov-2 virus [53]. Directly testing whether completing Contain COVID Anxiety has the negative side effect of reduced social distancing intentions, which could in turn predict reduced social distancing behaviors [54], is crucial to determining whether the intervention can be responsibly tested and disseminated at scale. Thus, we developed this intervention.
with an eye toward not undermining social distancing intentions. However, social distancing intentions were not a direct target of the intervention, so we also did not expect to see an increase in these intentions as a result of completing the intervention.

**Hypothesis 2**

We hypothesized that engaging in the Contain COVID Anxiety SSI would have an association with pre- to post-SSI change in social distancing intentions that is statistically equivalent to 0. This pattern of results would indicate completing the intervention is not meaningfully associated with intentions to socially distance.

We were also interested in whether the SSI designed to increase perceived control over anxiety did, in fact, increase perceived control over anxiety immediate post-SSI more than the placebo SSI.

**Hypothesis 3**

We expected a larger increase immediate pre- to immediate post-SSI in perceived control over anxiety to occur for participants randomized to the Contain COVID Anxiety SSI, relative to those randomized to the placebo SSI. This pattern of results would indicate a larger increase in perceived control over anxiety occurs pre- to post-Contain COVID Anxiety SSI compared with the placebo SSI.

**Present Study**

This study tested the efficacy of decreasing generalized anxiety symptoms (Hypothesis 1) and increasing perceived control of anxiety with the Contain COVID Anxiety SSI in a nationally representative sample (Hypotheses 1 and 3) and the safety of testing and making the SSI available at scale (Hypothesis 2). This is a crucial step toward providing open-access, evidence-based resources to help the US population more effectively cope with their anxiety during the COVID-19 pandemic.

**Methods**

**Ethics Information**

The study was approved by the Institutional Review Board (IRB) of Stony Brook University under protocol #2020-00204. As required by US law, a description of this study is available at Clinicaltrials.gov. We have also included a CONSORT checklist in Multimedia Appendix 2.

**Participants and Design**

We recruited a sample of 500 participants representing a weighted-probability sample of the United States through Prolific, an online platform connecting researchers and participants. Prolific is designed specifically for use in the scientific research context, unlike older online crowdsourcing platforms (eg, Mechanical Turk) that were designed for broader use (eg, by marketing and advertising companies to outsource labor) [55,56]. To address particular needs of the academic research community, Prolific provides estimates of the available population for a given study, enables confidential collection of human subjects data, allows for prescreening and exclusion of participants based on individual study criteria, prevents duplicate responses by limiting users to one Prolific account that is verified by built-in data quality checks using cookies and IP address, and directly facilitates longitudinal data collection [56].

Upon signing up via Prolific to volunteer for scientific research studies (ie, click a button on the Prolific website that reads, “Participate: Take part in engaging research, earn cash, and help improve human knowledge!”), Prolific users provide sociodemographic and personal background information; they then receive invitations via email to take part in studies for which they qualify, whenever such studies are made available by research teams around the world. To date, Prolific has been used in hundreds of psychological scientific studies, including many focused on adults experiencing mental health problems [57,58]. Prolific allows for informed consent to be provided digitally. In this study, individuals volunteered to participate by providing digital informed consent within an online Qualtrics survey developed by the study team, after being presented with an IRB-approved study information and consent form. This consent form included all relevant information about the benefits, risks, and compensation related to study participation.

University affiliation (Stony Brook University) was visible to participants on consent forms and on the first page of both SSI programs (Contain COVID Anxiety SSI and the placebo Remain COVID Free SSI). Once a user completed the study, the project was no longer viewable on their Prolific account.

This study’s weighted-probility sample was stratified on age, sex, and ethnicity. To help further maximize the generalizability of our findings, there were no prescreening inclusion or exclusion criteria other than having a Prolific ID, being at least 18 years old (able to provide consent), and residing in the United States. However, study participation required access to, and comfort using, a device connected to the Internet. We also recruited 8 pilot participants from the Prolific platform before recruiting this weighted probability sample to ensure data were being collected properly, and these pilot participants’ data were not used in confirmatory analyses of this study. All pilot data are available on the open science framework page [52].

We conducted intent-to-treat analyses including all participants who were randomized to a study condition (n=522, see [52]). We sought to prevent missing data by requesting responses to each question (with a reminder at the end of each page if participants had not answered a question) and imputed missing data using the expectation-maximization and bootstrapping algorithm implemented with Amelia II in R [59].

We used a between-subjects design; participants were randomized to receive either the active Contain COVID Anxiety SSI (50% allocation) or the placebo Remain COVID Free SSI (50% allocation). The sequence determining randomization of condition was automatically generated using the randomizer within Qualtrics Survey Software (no blocking was used for this randomization), making the randomization process double-blind. To triple-blind our analysis process, the last author (JS) downloaded the data from Qualtrics and recoded the variable indicating to which SSI the participants were randomized before sending the data to the first author (MM) who performed the primary analyses. Therefore, the primary analyses were conducted without the knowledge of which condition is which.
Power analyses were conducted using a “smallest effect size of interest” approach, where we aimed to be powered to detect the smallest effect size corresponding with a subjectively meaningful difference in participants’ experiences [60]. For hypothesis 1, using simulations conducted in R, we determined we would need a sample size of 400 for 95% power to detect the smallest effect size of interest for this hypothesis (d=0.33, as determined by a conservative estimate from a previous SSI for general anxiety). For hypothesis 2, using the TOSTER package in R, we determined we would have >95% power to detect whether the beginning to end effect of the Contain COVID Anxiety SSI on social distancing intentions falls within the equivalence bounds of d=–0.33 to d=0.33 with an n of 250. We chose these equivalence bounds based on not wanting any negative side effects on social distancing intentions to be greater than the smallest effect size of interest for our primary outcome (general anxiety). For hypothesis 3, using simulations conducted in R, we determined we would have >95% power to detect the smallest effect size of interest, d=0.21, the smallest change in perceived control due to an intervention to predict later decreases in anxiety observed in previous SSI work at n=500 [33]. We retained greater than 95% power by both recruiting enough participants to account for 20% attrition at the 2-week follow-up and using multiple imputation techniques to carry out an intent-to-treat approach. Further, we also conducted sensitivity tests for each hypothesis to examine the range of estimates of the effects observed if all missing data are assumed to be in either the 75th or 25th percentile of change for each key variable—thereby quantifying what our estimates would look like if our data were not missing at random due to unobserved confounders. See the publicly available code for the power analysis [52].

**Procedure**

The entire procedure was conducted online via the Qualtrics Survey Platform, which participants were linked to directly from Prolific. After providing informed consent, participants spent approximately 8 minutes filling out pre-intervention questionnaires including demographics, depression symptoms, generalized anxiety symptoms, self-hatred, access to mental health treatment, and COVID-19–related stressors.

Immediately following answering these questions and immediately prior to the SSI intervention, participants completed the Anxiety Control Questionnaire—Emotion Control (ACQ-EC) scale, the Hand Washing Intentions scale, and several questions about social distancing asked in national surveys to measure how long do you believe others would be willing to engage in the following behaviors?” and then “Avoid going out to a bank provided by the CDC [63]: All start with “Starting today, you will ___.” As validated measures for social distancing intentions do not yet exist, we propose to use these scales directly related to confirmatory hypotheses and quality checks. See Multimedia Appendix 1 for a list of questionnaires included.

**Measures**

**Anxiety Symptoms**

The Generalized Anxiety Disorder-7 (GAD-7; [61]) measures clinical anxiety symptom severity, based on diagnostic criteria for generalized anxiety disorder. The GAD-7 includes 7 items asking respondents how often, during the last 2 weeks, they were bothered by each of 7 anxiety symptoms. Response options are “not at all,” “several days,” “more than half the days,” and “nearly every day;” scored as 0, 1, 2, and 3, respectively; thus, total sum-scores may range from 0 to 21, and average scores range from 0 to 3. The GAD-7 has shown adequate reliability and strong convergent validity with other anxiety scales [61]. The GAD-7 is frequently used in large-scale treatment and dissemination studies as a generic measure of change in anxiety symptoms [62].

**Perceived Control Over Anxiety**

The ACQ-EC [19] measures how much perceived control participants have over their anxiety, the primary outcome of the study. It is one of the 3 validated subscales of the Anxiety Control questionnaire and contains 4 items (eg, “I am able to control my level of anxiety.”) rated on a 0 ("Strongly Disagree") to 5 ("Strongly Agree") scale. The potential mean scores of the scale (the score of interest for testing hypothesis 1 at all 3 times points) therefore range from 0 to 5. The scale has a well-validated factor structure in a nonclinically selected sample, is strongly associated with anxiety and depression symptoms, and has demonstrated good internal consistency in previous investigations [19].

**Social Distancing Intentions**

The following Social Distancing Intentions questions, the secondary outcome of the study, are part of a standardized item bank provided by the CDC [63]: All start with “Starting today, for how long do you believe others would be willing to engage in the following behaviors?” and then “Avoid going out to a restaurant, bar, or club;” “Avoid going to a family gathering like a birthday party or wedding or funeral;” or “Avoid going to a social gathering with friends, peers, or coworkers (not including relatives)” on a scale from 1 (“Less than a month”) to 4 (“4 months or more”). As validated measures for social distancing intentions do not yet exist, we propose to use these scales...
questions given these items are drawn from a standardized item bank provided by the CDC to better facilitate cumulative science (as other researchers will also utilize these items). At the suggestion of a reviewer, we changed the wording of these questions to ask about participants’ beliefs about others’ willingness to engage in these behaviors to reduce potential social desirability bias in responding. The potential mean scores of the scale (the score of interest for testing hypothesis 2 at both time points) therefore range from 1 to 4.

**Comprehension Questions**

We used comprehension questions as an initial quality check to ensure participants comprehended the core messages of both SSIs. These questions go beyond traditional attention check items, which can be answered incorrectly even by attentive participants [64]. Following each intervention, we asked 2 multiple choice questions with 4 potential response options—1 correct answer, 1 incorrect answer that contains material from the intervention not relevant to answering the current question, and 2 incorrect responses referencing material not contained in the intervention. The exact questions can be found in Multimedia Appendix 1. We initially required at least 75% of participants to answer both comprehension check questions correctly following the Contain COVID Anxiety SSI to pass the quality check, though see the “Comprehension Check Questions” section for further discussion as these questions did not appear to index intervention fidelity in this context.

**Single-Session Interventions**

**Contain COVID Anxiety SSI**

This active SSI was developed following current recommendations for evidence-based SSI design to target mental health–related outcomes [39]. Participants first received normalizing scientific information (including neuroscience findings) that help explain why increased anxiety during the COVID-19 pandemic is a typical response. They then read testimonials from 3 other people in the United States who have applied a 3-step action plan for coping more effectively with their anxiety. These 3 steps were (1) reminding themselves increased anxiety is a typical response during a pandemic before writing down some anxiety-provoking events they can’t control and some anxiety-provoking events they can control, (2) picking one of the anxiety-provoking events they can control, and (3) deciding on 1 small step to cope more effectively with the anxiety-provoking event they can control. Participants were then empowered as helpers by us asking for their permission to share their action plan with others to help them more effectively cope with pandemic-related anxiety. The entire intervention took approximately 8 minutes and was completed entirely within the Qualtrics survey platform.

**Remain COVID Free SSI**

This placebo SSI was developed to mirror the structure of the Contain COVID Anxiety SSI, discuss COVID-19–related content, and do so without as many of the potential active ingredients of effective SSIs. Participants received scientific information about how soap kills the SARS-CoV-2 virus but no neuroscience information related to behaviors or behavior change. Participants were told didactically that there is only one way to wash their hands effectively, by following this 3-step plan: (1) deciding on 10 times a day to wash their hands, (2) putting reminders in their calendar or setting alarms on their phone to remind them to wash their hands, and (3) singing happy birthday to their favorite celebrity twice while washing their hands. They then read 3 testimonials from other people who had implemented this plan, but they did not make a plan themselves. They therefore also did not have the opportunity to share their plan to prosocially help others. The entire intervention took approximately 8 minutes and was completed entirely within the Qualtrics survey platform.

**Analysis Plan**

**Testing Participant Dropout**

We first tested for dropout from the study due to intervention assignment. For example, people could differentially drop out when receiving the active Contain COVID Anxiety. Thus, we tested for differential dropout using a z test of differential proportions, in which we compared the proportion of people who dropped out before completing the study (0 = No, 1 = Yes) as a function of treatment condition (0 = Remain COVID Free, 1 = Contain COVID Anxiety). If the P value for this test was <.05, we planned to interpret that dropout as dependent on condition assignment and preregistered that we would not be able to interpret the effects of intervention assignment on outcome (ie, we would not be able to test Hypotheses 1 and 3). If the P value was >.05 for this test, we preregistered that we would assume dropout was not dependent on condition assignment.

**Data Aggregation for Hypothesis Testing**

We then created 2 separate scores for the GAD-7 to reflect baseline and 2 weeks post-SSI scores by taking the mean of the 7 items at each time point (score range at each time point: 0-3). We then created 2 separate scores for the ACQ-EC to reflect baseline and immediate post-SSI by taking the mean of the 4 items at each time point (score range at each time point: 0-5). We also calculated the mean of the 3 Social Distancing Intentions questions (score range at each time point: 1-4) at baseline and immediate post-SSI to calculate composite social distancing intentions scores. Following the creation of these composites, we imputed any missing data using the expectation-maximization and bootstrapping algorithm implemented with Amelia II in R [59]. These imputed data sets allowed for more conservative intent-to-treat analyses than listwise deletion or last-observation carried forward [65]. We imputed as many data sets as there were percentages of missing data for an outcome—rounding up to the next highest percentage (eg, If 2.4% of data were missing on an outcome, we created 3 imputed data sets). This process allowed us to retain high power even in the presence of missing data [66].

Consistent with best practices, we included all predictors from the statistical model (baseline value of imputed outcome, either perceived control over anxiety or social distancing intentions, and intervention order) and all baseline variables expected to be associated with the outcome variable (for generalized anxiety and perceived control over anxiety: Inventory of Depression and Anxiety Symptoms Dysphoria mean score, having received
mental health treatment in the past 12 months or not, and self-hate scale mean score; for social distancing intentions: age, gender [male, female, nonbinary], education level, and income level). Imputed data were analyzed using the tidyverse package in R [67]. Cohen d effect sizes and 95% CIs for analyses were calculated using t values for the treatment effect obtained from the analyses with the MOTE package in R [68]. We also conducted sensitivity analyses for all 3 hypotheses, in which all missing data for confirmatory outcomes were assumed to be in the 25th or 75th percentile of change in those outcomes observed in the sample. These analyses allowed us to examine the potential range of estimates for our hypotheses if we assumed the data were not missing at random but were instead impacted by unobserved confounders. See Multimedia Appendix 3 for the full imputation code and analytic strategy.

Testing Hypothesis 1

We tested whether the Contain COVID Anxiety SSI decreased scores on the GAD-7 immediately pre-SSI to 2 weeks later more than the Remain COVID Free SSI using a linear regression approach. We entered baseline GAD-7 score and condition as predictors of the follow-up GAD-7 mean score. We expected to see a larger decrease in GAD-7 score when the participants were randomized to Contain COVID Anxiety SSI compared with when they were randomized to the placebo Remain COVID Free SSI. This pattern of differences would indicate a decrease in generalized anxiety disorder symptoms to a greater extent due to the Contain COVID Anxiety SSI compared with the Remain COVID Free SSI. We preregistered that a P value <.0167 (to Bonferroni correct for multiple comparisons) for condition in a linear model with a larger decrease in GAD-7 occurring when randomized to the Contain COVID Anxiety SSI would allow us to reject the null hypothesis that the difference between conditions was 0. We planned to confirm the pattern of differences by examining the sign of the condition coefficient and descriptive pattern of means based on condition. See Table 1 for all alternative interpretations of results.

Testing Hypothesis 2

We tested whether completing both interventions had an effect on social distancing intentions statistically equivalent to 0 using a paired-equivalence test. We entered baseline and postintervention social distancing intentions mean scores and SDs into a paired-equivalence test with equivalence bounds of $d = -0.33$ to $d = 0.33$. We preregistered that a P value <.05 would allow us to reject the null hypothesis that the effect of completing both interventions was statistically different from 0. If the $P$ value for this paired-equivalence test was nonsignificant, we preregistered that we would run a paired $t$ test with baseline and postintervention social distancing intentions scores to determine if the association of the Contain COVID Anxiety intervention on social distancing intentions was significantly different from 0. We preregistered that if the $P$ value was <.0167 (to Bonferroni correct for multiple comparisons), we would reject the null hypothesis that there was no difference pre to post within the active SSI intervention. See Table 1 for all alternative interpretations of results.

Testing Hypothesis 3

We tested whether the Contain COVID Anxiety SSI increased scores on the ACQ-EC more than the Remain COVID Free SSI using a linear regression approach. We entered baseline ACQ-EC scores and condition as predictors of immediate post-SSI ACQ-EC score. We expected to see a larger increase in ACQ-EC score when the participants were randomized to the Contain COVID Anxiety SSI compared with when they were randomized to the placebo Remain COVID Free SSI. This pattern of differences would indicate an increase in perceived control over anxiety to a greater extent due to the Contain COVID Anxiety SSI compared with the Remain COVID Free SSI. We preregistered that a $P$ value <.0167 (to Bonferroni correct for multiple comparisons) for condition in a linear model with a larger increase in ACQ-EC occurring when randomized to the Contain COVID Anxiety SSI would allow us to reject the null hypothesis that the difference between conditions was 0. We planned to confirm the pattern of differences by examining the sign of the condition coefficient and descriptive pattern of means based on condition. See Table 1 for all alternative interpretations of results.

Results

Participant Demographics

Of the 529 participants who began the survey, 522 participants were randomized to achieve the weighted-probability sample of 500 (7 participants exited the survey prior to randomization, and 22 participants exited the survey prior to completion of the baseline survey; ie, 94.5% and 95.8% completion rates among individuals who started the baseline survey and among those who were randomized, respectively). All demographics for all randomized participants are reported by treatment condition in Table 2. Participants in both groups were experiencing, on average, mild anxiety (GAD-7 sum scores of 5.25-5.39), which were similar to the GAD-7 values assumed in our a priori power analysis (5.73). The sample appeared to be representative of the United States in terms of gender, age, and race/ethnicity. All responses were collected between September 13, 2020, and September 29, 2020.
### Table 2. Demographics by treatment condition.

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Treatment received</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Active Contain COVID Anxiety (n=261)</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>46.02 (15.65)</td>
</tr>
<tr>
<td><strong>Race/ethnicity, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>American Indian and/or Alaska Native</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Asian</td>
<td>19 (7.3)</td>
</tr>
<tr>
<td>African American</td>
<td>31 (12.0)</td>
</tr>
<tr>
<td>Hispanic or Latino/a</td>
<td>16 (6.1)</td>
</tr>
<tr>
<td>Native Hawaiian or Pacific Islander</td>
<td>1 (0.4)</td>
</tr>
<tr>
<td>White, non-Hispanic</td>
<td>187 (72.0)</td>
</tr>
<tr>
<td>More than one race</td>
<td>4 (1.5)</td>
</tr>
<tr>
<td>Other</td>
<td>3 (1.1)</td>
</tr>
<tr>
<td><strong>Gender, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Agender</td>
<td>2 (0.8)</td>
</tr>
<tr>
<td>Genderqueer or gender fluid</td>
<td>3 (1.1)</td>
</tr>
<tr>
<td>Man</td>
<td>127 (49.0)</td>
</tr>
<tr>
<td>Trans man</td>
<td>3 (1.1)</td>
</tr>
<tr>
<td>Woman</td>
<td>125 (48.0)</td>
</tr>
<tr>
<td>Other</td>
<td>1 (0.4)</td>
</tr>
<tr>
<td><strong>Sexual orientation, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Asexual</td>
<td>6 (2.3)</td>
</tr>
<tr>
<td>Bisexual</td>
<td>20 (7.7)</td>
</tr>
<tr>
<td>Gay</td>
<td>3 (1.1)</td>
</tr>
<tr>
<td>Heterosexual</td>
<td>218 (84.0)</td>
</tr>
<tr>
<td>Lesbian</td>
<td>2 (0.8)</td>
</tr>
<tr>
<td>Pansexual</td>
<td>3 (1.1)</td>
</tr>
<tr>
<td>Queer</td>
<td>5 (1.9)</td>
</tr>
<tr>
<td>Questioning or unsure</td>
<td>2 (0.8)</td>
</tr>
<tr>
<td>Other</td>
<td>2 (0.8)</td>
</tr>
<tr>
<td>Unknown</td>
<td>0 (0)</td>
</tr>
<tr>
<td><strong>Education, n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Less than high school degree</td>
<td>1 (0.4)</td>
</tr>
<tr>
<td>High school degree</td>
<td>27 (10.0)</td>
</tr>
<tr>
<td>Some college, no degree</td>
<td>74 (28.0)</td>
</tr>
<tr>
<td>Associate degree</td>
<td>26 (10.0)</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>77 (30.0)</td>
</tr>
<tr>
<td>Master’s degree</td>
<td>46 (18.0)</td>
</tr>
<tr>
<td>Professional degree</td>
<td>4 (1.5)</td>
</tr>
<tr>
<td>Doctorate</td>
<td>6 (2.3)</td>
</tr>
<tr>
<td><strong>Annual income (US $), n (%)</strong></td>
<td></td>
</tr>
<tr>
<td>Less than 10,000</td>
<td>17 (6.5)</td>
</tr>
<tr>
<td>10,000-19,999</td>
<td>24 (9.2)</td>
</tr>
<tr>
<td>20,000-29,999</td>
<td>29 (11.0)</td>
</tr>
</tbody>
</table>
### Demographics

<table>
<thead>
<tr>
<th>Treatment received</th>
<th>Placebo Remain COVID Free (n=261)</th>
<th>Active Contain COVID Anxiety (n=261)</th>
</tr>
</thead>
<tbody>
<tr>
<td>30,000-39,999</td>
<td>24 (9.2)</td>
<td>23 (8.8)</td>
</tr>
<tr>
<td>40,000-49,999</td>
<td>21 (8.0)</td>
<td>21 (8.0)</td>
</tr>
<tr>
<td>50,000-59,999</td>
<td>26 (10.0)</td>
<td>22 (8.4)</td>
</tr>
<tr>
<td>60,000-69,999</td>
<td>15 (5.7)</td>
<td>10 (3.8)</td>
</tr>
<tr>
<td>70,000-79,999</td>
<td>28 (11.0)</td>
<td>20 (7.7)</td>
</tr>
<tr>
<td>80,000-89,999</td>
<td>15 (5.7)</td>
<td>12 (4.6)</td>
</tr>
<tr>
<td>90,000-99,999</td>
<td>15 (5.7)</td>
<td>12 (4.6)</td>
</tr>
<tr>
<td>100,000-149,999</td>
<td>27 (10.0)</td>
<td>30 (11.0)</td>
</tr>
<tr>
<td>150,000 or more</td>
<td>21 (8.0)</td>
<td>19 (7.3)</td>
</tr>
</tbody>
</table>

### Relationship status, n (%)

- No current relationship: 118 (45.0) vs. 101 (39.0)
- Relationship, not living together: 20 (7.7) vs. 25 (9.6)
- Relationship, living together: 22 (8.4) vs. 24 (9.2)
- Engaged: 3 (1.1) vs. 3 (1.1)
- Married: 108 (41.0) vs. 98 (38.0)

### Has children, n (%)

- 117 (45.0) vs. 127 (49.0)

### Health insurance covers mental health, n (%)

- 195 (75.0) vs. 196 (75.0)

### Received psychotherapy in the past year, n (%)

- 56 (21.0) vs. 54 (21.0)

### Received medication for mental health in the past year, n (%)

- 56 (21.0) vs. 54 (21.0)

### Perceived need for mental health treatment in the past year, n (%)

- 87 (33.0) vs. 89 (34.0)

### Baseline IDAS<sup>a</sup>-Dysphoria (1-5), mean (SD)

- 2.00 (0.90) vs. 2.09 (0.92)

### Baseline GAD-7<sup>b</sup> (0-3), mean (SD)

- 0.75 (0.74) vs. 0.77 (0.75)

### Baseline self-hate (1-7), mean (SD)

- 2.07 (1.65) vs. 2.20 (1.59)

### Baseline perceived control over anxiety (0-5), mean (SD)

- 2.88 (1.34) vs. 2.85 (1.29)

### Baseline Hand Washing Intentions (1-7), mean (SD)

- 5.08 (1.59) vs. 5.01 (1.61)

### Baseline social distancing intentions of others (1-4), mean (SD)

- 2.27 (1.12) vs. 2.32 (1.12)

---

<sup>a</sup>IDAS: Inventory of Depression and Anxiety Symptoms.  
<sup>b</sup>GAD-7: Generalized Anxiety Disorder-7.

### Testing Participant Dropout

There was no evidence participants were significantly more likely to drop out of either condition at the 2-week follow-up (25/261, 9.6% dropped out from the Remain COVID Free SSI, and 18/261, 6.9% dropped out from the Contain COVID Anxiety SSI; *P*=.34). However, there was some evidence participants dropped out during the baseline survey significantly more often if they were randomized to the Contain COVID Anxiety SSI (20/261, 7.7% dropped out) versus the Remain COVID Free SSI (1/261, 0.4%; *P*<.001). Therefore, we interpreted the results for hypotheses 2 and 3 (which involve immediate postintervention outcomes) under conditions in which dropout is not presumed to be random (ie, a sensitivity test in which those who dropped out are assumed to change far more or less than average; see the preregistered sensitivity test in the publicly available code [52]). We also conducted this sensitivity test for Hypothesis 1, as unmeasured confounding can occur even if dropout does not significantly differ between conditions. All participants who were randomized were included in the intent-to-treat analyses (n=522).

### Testing Comprehension Questions

During our piloting of the Prolific platform (as outlined in our preregistered message), we noticed a substantial portion of participants were not answering the comprehension check questions correctly despite providing face-valid qualitative and quantitative data. We updated our comprehension check questions to attempt to align them more with completing the intervention with fidelity. However, among all participants who were randomized to the Contain COVID Anxiety SSI and answered a comprehension check question, 52.3% (126/241) answered both comprehension questions correctly. To examine whether this phenomenon was a function of the questions or...
lack of fidelity to the intervention, we developed a systematic qualitative coding system focused on fidelity for each qualitative response in the Contain COVID Anxiety SSI group. To be coded as having a high-fidelity qualitative response, the participant had to respond not only to the prompt with related content (a more general comprehension check) but also to the prompt as instructed (eg, a response enumerating concrete coping strategies to a prompt instructing participants to validate their own anxiety would be marked as a low fidelity response; see the publicly available code for the full qualitative coding system for fidelity check [52]).

We double-coded a random 20% of intervention responses (48 participants with 6 responses each, effective n=288) and found 87.13% average fidelity across these participants’ responses. Further, answering both comprehension check questions correctly shared only 0.01% of the variance with each participant’s fidelity score across their qualitative responses. We therefore determined that the comprehension check questions were poor indicators of completing the intervention with fidelity and chose to proceed with our planned analyses.

**Testing Hypothesis 1**

In full intent-to-treat analyses with all participants who were randomized (n=522), we did not find support for the alternative hypothesis (t_{260}=–0.71, P=.48; d=–0.06, 95% CI –0.27 to 0.15) and did find support for the null (noninferiority to placebo) hypothesis (t_{260}=3.76, P<.001). These results were unchanged when we conducted a sensitivity test to determine whether results differed when participants who dropped out were assumed to have (1) experienced GAD-7 changes in the 25th percentile of the sample or (2) experienced GAD-7 changes in the 75th percentile of the sample (see publicly available code for the sensitivity tests for all hypotheses [52]). Therefore, we found evidence in favor of the placebo (Remain COVID Free SSI) being equally strong or stronger than the active condition (Contain COVID Anxiety SSI) in reducing generalized anxiety 2 weeks later. These results held when these tests were conducted in only the weighted-probability sample (n=500) and in only participants who answered both comprehension questions correctly (n=126). See the publicly available code for the sensitivity analysis [52].

**Testing Hypothesis 2**

To make it possible to generate fully invertible matrices necessary to produce imputations, participant gender was dropped from the imputation model. In this case, the alternative hypothesis was operationalized as a change in social distancing intentions of others pre- to immediate post-Contain COVID Anxiety being statistically equivalent within a range of d = –0.33 to 0.33, while the null hypothesis was operationalized as a change in social distancing intentions in the same circumstance falling outside the effect range of d from –0.33 to 0.33. In full intent-to-treat analyses with all participants who were randomized to the Contain COVID Anxiety SSI (n=261), we found support for the alternative hypothesis (t_{260}=4.63, P<.001) and did not find support for the null hypothesis (t_{260}=0.70, P=.48; d=0.04, 95% CI –0.08 to 0.16). However, these results changed to unclear support for either the null or alternative hypothesis when we conducted a sensitivity test to determine whether results differed when participants who dropped out were assumed to have experienced (1) social distancing intentions of others changes in the 25th percentile of the sample or (2) social distancing intentions of others changes in the 75th percentile of the sample. Therefore, we found evidence that the participants in the Contain COVID Anxiety condition were statistically equivalent to participants in the Remain COVID Free condition in experiencing changes in social distancing intentions, though this result could be influenced by unmeasured confounding in participant dropout. These results held when these tests were conducted in only the weighted-probability sample (n=250) and in only participants who answered both comprehension questions correctly (n=126). See the publicly available code for the sensitivity analysis [52].

**Testing Hypothesis 3**

In full intent-to-treat analyses with all participants who were randomized (n=522), we did not find support for the alternative hypothesis (t_{260}=–0.21, P=.83; d=–0.02, 95% CI –0.23 to 0.19) and did find support for the null (noninferior to placebo) hypothesis (t_{260}=2.40, P=.001). However, these results changed to unclear support for either the null or alternative hypothesis when we conducted a sensitivity test to determine whether results differed participants who dropped out were assumed to have experienced (1) ACQ-EC changes in the 25th percentile of the sample or (2) ACQ-EC changes in the 75th percentile of the sample. Therefore, we found evidence in favor of the placebo (Remain COVID Free) being equally strong or stronger than the active condition (Contain COVID Anxiety) in increasing perceived control over anxiety immediately postintervention, though this result could be influenced by unmeasured confounding in participant dropout. These results held when these tests were conducted in only the weighted-probability sample (n=500) and in only participants who answered both comprehension questions correctly (n=387). Within-group effect sizes were negligible in both the Contain COVID Anxiety (t_{260}=1.03; d=0.06, 95% CI –0.06 to 0.19) and Remain COVID Free (t_{260}=1.63; d=0.10, 95% CI –0.02 to 0.22). See the publicly available code for the sensitivity analysis [52].

**Discussion**

**Principal Findings**

Compared with a placebo control, a self-guided SSI for US adults did not improve short-term generalized anxiety or perceived control over anxiety during the COVID-19 pandemic. This high-powered randomized controlled trial (RCT), which used a nationally representative US sample, also demonstrated that this intervention did not worsen short-term generalized anxiety or perceived control. There was also statistically equivalent to zero iatrogenic movement within the intervention condition of beliefs in others’ willingness to social distance.
Comparison With Prior Work

Interest in the use of brief, e-mental health interventions has increased substantially during the COVID-19 pandemic across the general adult population [69], and a large majority of these tools have minimal or no empirical support [70]. Even face-valid interventions containing evidence-based components may not necessarily improve mental health outcomes, and many mental health applications are used only once [71]. In this sample, a SSI for a community sample of adults, containing components associated with both proximal and longer-term mental health improvements in adolescents, did not lead to anxiety-related improvements above and beyond a placebo control. These differences could be due, at least in part, to sample characteristics: This study’s sample was older and more age-diverse than those for whom other self-guided SSIs have improved perceived control, anxiety, and depression [32,42,51], and participants were recruited from the broader US community rather than a clinically high-risk subgroup. Further, prior well-powered trials of SSIs targeting adults have significantly improved non-anxiety outcomes—such as positive psychotherapy expectancies [72] and positive parenting behaviors and distress tolerance in high-symptom individuals [73]—but self-guided SSI effects on clinical outcomes in adults outside of substance and alcohol use problems [37] have not been previously explored. It is also possible that the intervention tested by this study was simply not therapeutically effective, but that other interventions targeting similar outcomes in a similar sample may still be.

Accordingly, these results are the first to suggest that perceived control over anxiety and generalized anxiety symptoms may in fact be difficult to move in general adult samples via self-guided SSIs, at least in this nationally representative sample. Within-group effect sizes for perceived control over anxiety was negligible in both the active and placebo conditions, in contrast to within-group SSI effects seen in trials targeting adolescents. Further, nonzero increases were observed in generalized anxiety symptoms in both the active and placebo conditions over 2 weeks. Therefore, it is not the case that participants benefited from either condition (a placebo effect) but rather that they benefited from neither condition on targeted outcomes.

This design did not contain a wait list control condition, and we cannot explicitly rule out that receiving either light-touch intervention would have resulted in a smaller increase in generalized anxiety disorder symptoms compared with receiving nothing. This pattern of within-group effect sizes (ie, increasing generalized anxiety symptoms over time in both conditions) is consistent only with potentially preventative, as opposed to therapeutic, on average effects compared with “no treatment” control. Although we found no evidence of iatrogenic movement on social distancing intentions of others within our SSI, the lack of iatrogenic effects in other e-mental health interventions cannot be guaranteed without testing those outcomes directly. E-mental health applications hold promise in increasing mental health treatment access, [74] and well-powered tests of effectiveness must accompany (or ideally precede) dissemination if we wish to reduce overall mental health burden (eg, reducing subclinical anxiety symptoms) across general adult populations rather than solely the number of people without mental health support. Further, especially in the context of a pandemic, direct tests of iatrogenic outcomes should be included as primary outcomes in tests of single-session and light-touch mental health interventions.

We would like to propose 2 complementary paths toward building and understanding the impacts of effective SSIs for anxiety in adults, based on the results of this trial, which may generalize to evaluations of other light-touch interventions as well. First, given substantial heterogeneity in individual-level responses to any mental health intervention (including the SSI tested here), we recommend that researchers and program developers collect data necessary to build predictive models of individual-level response to SSIs. Predictive models require much larger sample sizes than typical clinical trials collect to identify subgroups of best responders. For example, recent simulation studies demonstrated that clinical trials may need as many as 500 participants per treatment arm to recover reliable predictions about who would benefit most from which treatment (ie, questions of moderation effects; [75])—far larger than typical mental health treatment RCTs (average n=52) [76]. Trials of self-guided SSIs create opportunities to quickly recruit large samples while retaining a rigorous experimental design. These larger sample sizes, combined with advances in feature engineering, could facilitate nuanced and definitive analyses regarding which individuals will (or will not) benefit from an extremely light-touch intervention. Such analyses could help situate self-guided SSIs within a stratified care system [77], where (for example) adults more and less likely to benefit from low-intensity support for anxiety are referred directly to the best-fit level of care.

Second, we recommend the systematic incorporation of qualitative and user-experience data into trials of self-guided SSIs. It has been posited that SSIs targeting adolescent mental health problems may show acceptability and efficacy, at least in part, because they do not “feel” like interventions to youth—that is, they are designed to be nonstigmatizing to users [39]. However, systematic qualitative data around participants’ experiences with SSIs for mental health are scarce, and this hypothesis has not been systematically tested. Collecting and analyzing qualitative and user experience data could clarify how people view self-guided SSIs as similar or different to longer-term and face-to-face interventions and whether these perceptions differ across distinct populations (eg, youth versus adults, given that many elements of youth-directed SSI design were developed through a developmentally specific lens). User experience data may be analyzed using both qualitative (eg, grounded theory) and quantitative (eg, topic modeling) methods to leverage this important information as much as possible during iterative intervention development.

Limitations

There are certainly limitations to what this study can conclude. First, this study was conducted during the COVID-19 pandemic, and it is unclear whether the nonzero increases in generalized anxiety within both groups reflected the many structural challenges of pandemic conditions (which a self-guided SSI cannot change) or would have occurred regardless. Examination
of within-group effect sizes in self-guided SSI trials conducted after the COVID-19 pandemic ends should examine whether negligible to slightly increasing within-group effect sizes persist for clinical anxiety in unselected adult samples. Other work suggests that certain outcomes, such as parenting behaviors and distress tolerance, may be modifiable via self-guided SSIs in high-symptom adults even during the pandemic [73]. Second, our original quality check measure—3 multiple-choice comprehension check items, created specifically for this trial—proved invalid as a gauge of intervention fidelity, sharing only 0.01% of variance with a subsequently developed, more rigorous, qualitatively coded intervention fidelity metric. This improved qualitatively coded fidelity measure showed that participants were highly successful in completing the interventions as intended, per their written responses to within-program prompts. Thus, it is unlikely the null results are due to the lack of participants engaging with and understanding the intervention content. It is also possible that the study’s sample, while representative of the US public across demographic variables, was subject to selection bias owing to their participation in Prolific. Additionally, should these SSIs be disseminated outside of an RCT context (ie, not posted as a paid research opportunity on Prolific), it is possible that a different pattern of results may emerge. Finally, although we did not observe differential dropout for our primary outcome 2 weeks later, there was higher dropout in the intervention group than in the placebo group during the baseline session containing the interventions. This pattern fits with sensitivity tests indicating that, if dropout did not occur at random, our statistical conclusions about perceived control over anxiety and the social distancing intentions of others become unclear. However, across all other sensitivity analyses, we found support for the null hypotheses, and within-group effect sizes would remain negligible regardless of dropout across conditions. Finally, this study was conducted in a US context, and its results cannot be assumed to generalize to other countries.

Conclusions
Compared with a placebo control, an 8-minute, self-guided SSI for US adults did not improve short-term generalized anxiety nor perceived control over anxiety during the COVID-19 pandemic. Additionally, neither condition yielded any iatrogenic movement in a key public health behavior (assumed social distancing intentions of others). Our rigorous methods and well-powered sample bolster confidence in these results, which carry direct implications for future research on self-guided SSIs for mental health problems—both for anxiety in adults and more broadly. As with so many interventions targeting complex, individual-level problems, key questions for SSI research remain: “which intervention, for whom, and under what circumstances?” [8].

Acknowledgments
This work was supported by research funding from the Department of Psychology at Stony Brook University to JS. We thank Ms. Sharon Chen, Mr. Akash Shroff, and Ms. Julia Fassler for piloting and providing key feedback on survey and intervention materials evaluated in this study.

Data Availability
Data from this study are publicly available [52].

Authors’ Contributions
MM, MD, JS, and JS contributed to study conceptualization. MM, JS, MD, and JS contributed to the intervention design. MM wrote the original draft of the manuscript. JS, MD, JS, IA, CB, and JS contributed to manuscript editing and revisions. MD wrote the original draft of the primary analysis code. JS wrote the original draft of the power analysis code. MM edited the analysis code/power analysis code. MM, MD, IA, and JS will be responsible for data curation, investigation, and formal analysis. JS contributed to funding acquisition and supervision.

Conflicts of Interest
The funders have/had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript. Unrelated to the present study, JS presently receives grant and research support from the National Institutes of Health (DP5OD28123), the Klingenstein Third Generation Foundation, the American Psychological Foundation, and Limbix Inc; all grants support research evaluating brief or single-session mental health interventions. MD receives grant support from a Stony Brook University Graduate Research Fellowship, and MD and JS receive research funding from the Psi Chi Honor Society for research on single-session interventions. JS, MD, and MM are under contract with New Harbinger Publications to co-author a therapeutic workbook for adolescents. JS is under contract with Oxford University Press to co-edit a book on low-intensity mental health interventions for youth. CB reports receiving grant support from the National Institute of Health and compensation for his work as journal editor for Sage Publications Inc. The authors report no other financial or nonfinancial conflicts.

Multimedia Appendix 1
Full survey and interventions.
References


35. Schleider JL, Weisz J. Can less be more? The promise (and perils) of single-session youth mental health interventions. The Behavior Therapist 2017;40(7):256-261 [FREE Full text]


Abbreviations

ACQ-EC: Anxiety Control Questionnaire-Emotion Control
CDC: Centers for Disease Control and Prevention
GAD-7: Generalized Anxiety Disorder-7
IRB: institutional review board
RCT: randomized controlled trial
SSI: single-session intervention

©Michael Mullarkey, Mallory Dobias, Jenna Sung, Isaac Ahuvia, Jason Shumake, Christopher Beevers, Jessica Schleider. Originally published in JMIR Mental Health (https://mental.jmir.org), 12.04.2022. 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 https://mental.jmir.org/, as well as this copyright and license information must be included.
Improving Web-Based Treatment Intake for Multiple Mental and Substance Use Disorders by Text Mining and Machine Learning: Algorithm Development and Validation

Sytske Wiegersma¹, MSc; Maurice Hidajat², BSc; Bart Schrieken², MSc; Bernard Veldkamp¹, Prof Dr; Miranda Olff³,4, Prof Dr

¹Department of Research Methodology, Measurement and Data Analysis, University of Twente, Enschede, Netherlands
²Interapy PLC, Amsterdam, Netherlands
³Department of Psychiatry, Amsterdam University Medical Centres, location Academic Medical Centre, Amsterdam, Netherlands
⁴ARQ National Psychotrauma Centre, Diemen, Netherlands

Abstract

Background: Text mining and machine learning are increasingly used in mental health care practice and research, potentially saving time and effort in the diagnosis and monitoring of patients. Previous studies showed that mental disorders can be detected based on text, but they focused on screening for a single predefined disorder instead of multiple disorders simultaneously.

Objective: The aim of this study is to develop a Dutch multi-class text-classification model to screen for a range of mental disorders to refer new patients to the most suitable treatment.

Methods: On the basis of textual responses of patients (N=5863) to a questionnaire currently used for intake and referral, a 7-class classifier was developed to distinguish among anxiety, panic, posttraumatic stress, mood, eating, substance use, and somatic symptom disorders. A linear support vector machine was fitted using nested cross-validation grid search.

Results: The highest classification rate was found for eating disorders (82%). The scores for panic (55%), posttraumatic stress (52%), mood (50%), somatic symptom (50%), anxiety (35%), and substance use disorders (33%) were lower, likely because of overlapping symptoms. The overall classification accuracy (49%) was reasonable for a 7-class classifier.

Conclusions: A classification model was developed that could screen text for multiple mental health disorders. The screener resulted in an additional outcome score that may serve as input for a formal diagnostic interview and referral. This may lead to a more efficient and standardized intake process.

(JMIR Ment Health 2022;9(4):e21111) doi:10.2196/21111

KEYWORDS

supervised text classification; multi-class classification; screening; mental health disorders; computerized CBT; automated intake and referral

Introduction

Background

Mental and substance use disorders such as anxiety, mood, alcohol and drug use, eating, and depressive disorders have been listed among the leading causes of global disability over the past years [1]. Annual studies show that between 2010 and 2016, these disorders accounted for approximately 18%-19% of the global burden of disease, measured in years lived with disability [2]. The proportion of people living with a mental disorder has remained practically unchanged in recent years (approximately 15.6%, 17.6%, and 19% for the global, European, and Dutch...
populations, respectively). However, because of population growth, absolute numbers of people diagnosed with a mental disorder have increased by 72 million globally and by 2 million in Europe between 2010 and 2016. For the Netherlands, despite an initial decrease in numbers by 15,000 from 2010 to 2014, there was an increase by 4000 between 2014 and 2016.

This growing number of people requiring mental health care each year makes preventing and detecting mental disorders, implementing early interventions, and improving treatments and mental health care access to public health and research priorities [3,4]. Mental health disorders are usually treated through medication or psychotherapy such as cognitive behavioral therapy (CBT), of which psychotherapy is generally seen as the first-line treatment [5]. However, mental health treatments are often underused [6] or delayed for many years [7]. Especially in low- and middle-income countries, there is a huge treatment gap in mental health care; 75% of the people experiencing anxiety, mood, impulse control, or substance use disorders remain untreated [8]. The reasons for this could be individual patient factors (eg, embarrassment, lack of time, and geographic influences); provider factors (eg, undertreatment and lack of skill in treating mental health problems); or systemic factors such as limited access to, or limited availability of, mental health providers, resulting in waiting lists [6].

This calls for more efficient, accurate, and accessible screening and treatment methods [9,10]. Modern technologies are increasingly recognized as a means of improving the accessibility of care and advancing the assessment, treatment, and prevention of mental health disorders. Creative, low-cost approaches should be used to increase access to (trauma-focused) CBT and other treatments [11]. An example of such an approach is web-based self-help, which is an increasingly available alternative for a range of disorders. Web-based self-help can be therapist-guided or not, and although some studies reported equal effects for guided and unguided web-based treatment (eg, for social anxiety disorders [12] and depression [13]), most research endorses the importance of at least minimal, regular therapist guidance in psychological interventions [14,15]. Web-based therapist-guided treatment such as computerized CBT is found to be approximately as effective as face-to-face treatment for several mental health disorders (eg, depression, anxiety, and burnout) [16-18].

In the Netherlands, 1 party offering web-based, therapist-assisted CBT is Interapy, a web-based mental health clinic approved by the Dutch health regulatory body. Interapy conducts screening, treatment, and outcome measurement on the web. Patient intake and diagnosis is performed using validated self-report instruments, followed by a diagnostic interview by telephone, after which patients are referred to a protocolled disorder-specific treatment. The treatment consists of a fixed set of evidence-based homework assignments provided through the Interapy platform and uses standardized instructions that are tailored to the patient by a therapist. After submitting the homework assignments, the patient receives asynchronous personal feedback and new instructions [14].

This form of web-based therapy generates large quantities of digital text data to be processed manually by the treating therapist. Textual data contain a lot of information that could be used more efficiently in the screening and treatment process through the application of text mining techniques. Text mining is generally used to automatically explore patterns and extract information from unstructured text data [19]. There is a large body of literature on text mining applications in the field of psychiatry and mental health; 2 recent systematic literature reviews provide a useful overview of the scope and limits, general analytic approaches, and performance of text mining in this context [20,21]. Abbe et al [20] concluded that text mining should be seen as a key methodological tool in psychiatric research and practice because of its ability to deal with the ever-growing amount of (textual) mental health data derived from, for example, medical files, web-based communities, and social media pages. However, despite the amount of data that are generated, assembling large, high-quality mental health text data sets has been found to be difficult [21]. With regard to the analytic approach, in most studies, predictive models are developed using supervised learning algorithms such as support vector machines (SVMs) and verified using k-fold cross-validation [21].

A way in which text mining can be put to use in mental health care practice concerns the detection of mental disorders. Previous studies showed that text mining can be used successfully in screening for posttraumatic stress disorder (PTSD) and depression [22,23]. He et al [22] developed an automatic screening model for PTSD using textual features from self-narratives posted on a forum for trauma survivors. On the basis of a set of highly discriminative keywords and word combinations extracted from the narratives using text mining techniques, they developed a text classifier that could accurately distinguish between trauma survivors with and those without PTSD. They concluded that automatic classification based on textual features is a promising addition to the current screening and diagnostic process for PTSD that can be easily implemented in web-based diagnosis and treatment platforms for PTSD and other psychiatric disorders. Neuman et al [23] developed an automatic screening system for depression using a depression lexicon based on metaphorical relations and relevant conceptual domains related to depression harvested from the internet. This lexicon was used to screen texts from open questions on a mental health website and a set of general blog texts for signs of depression and was found to classify texts that included signs of depression very accurately.

Although both studies showed the technical potential of automatic text classification in screening for mental disorders, they applied a proxy or a self-reported diagnosis instead of a direct, formal diagnosis by a psychiatrist as the classification criterion. In addition, both studies developed a binary classifier that focused on recognizing only a single specific disorder (PTSD or depression) at a time, which is the case in most studies that apply text mining to detect mental disorders [20,21]. However, in practice, for many patients who register with mental health complaints or sign up for web-based treatment, it is not clear beforehand which disorder they should be screened for. In this case, a multi-class classifier, screening for multiple different mental disorders at once, would be more useful than a binary classifier screening for only a single prespecified
disorder. Finally, it is pointed out that most natural language processing tools are currently designed for exploring English texts [20]. Although, indeed, text mining and language processing tools are mainly developed for the English language, the methods and techniques underlying the text analysis process are not necessarily language dependent. The development of models for different languages depends mainly on the availability of training and testing corpora and not so much on the methods and techniques used, as will be demonstrated in this study.

Objectives
This study investigates if and to what extent automatic text classification can improve the current web-based intake procedure of a Dutch web-based mental health clinic. The current intake questionnaire (see Methods section) consists of open and multiple-choice questions. The multiple-choice answers are converted to scores on four scales (somatization, depression, distress, and anxiety) as well as estimates of symptom severity, required level of care, suicide and psychosis risk, and drug dependence. These scores lead to an automatically generated indicative referral advice. This advice and the answers to the open questions are used by the therapist as input for the subsequent diagnostic telephone interview to arrive at a formal diagnosis and referral advice. However, the current questionnaire does not cover all disorders for which treatment is offered by Interapy, and the textual answers to the open questions remain to be processed and interpreted by the therapist. An automatic text screener may provide the therapist with more specific additional information, making the intake process more efficient and standardized.

Therefore, a multi-class text-classification model has been developed to screen for a range of different mental disorders with the aim of referring newly registered patients to the most fitting treatment. The focus is on a selection of treatments currently offered by Interapy for anxiety and panic disorders, PTSD, mood disorder (including depressive disorders), eating disorder, substance use disorders, and somatic symptom disorders. These will be referred to, respectively, as Anxiety, Panic, PTSD, Mood, Eating, Addiction, and Somatic throughout the rest of this paper. The treatment choice was made based on the amount of text data that was readily available from the Interapy database at the time of this research. This study adds to existing research in that (1) the patients in our sample have an official clinical diagnosis made by a therapist; (2) our data set consists of patients with a variety of mental health disorders, enabling us to develop a multi-class text classifier; and (3) the derived texts and the resulting classifier are in Dutch and as such provide an example of non-English text mining efforts applied in mental health care research and practice.

Methods

Methods Overview
The multi-disorder screening model was developed based on text and questionnaire data collected through the web-based intake environment of Interapy. This section describes the methods and techniques used to develop the supervised text-classification model and evaluate its performance.

Data Set
We used pretreatment scores on a self-reported questionnaire and text data derived from 3 open questions collected within the web-based intake environment. The patients are Dutch adults and adolescents who were referred to one of Interapy’s web-based treatments by their general practitioner and diagnosed by a therapist. All participants have given permission for their treatment data to be used for anonymized research by Interapy to improve and evaluate their treatments through informed consent. The electronic patient database was queried in July 2017. For each treatment, all available data were retrieved, excluding incomplete or double entries. For treatments for which large quantities of data were available, a random sample of 1100 patients was drawn to distribute the available data across the classes more evenly.

Web-Based Questionnaire
After signing up, new patients were asked to fill in the Digitale Indicatiehulp Psychische Problemen (DIPP; Digital Indication Aid for Mental Health Problems) questionnaire, an approved and validated decision support tool developed by Interapy and the HSK group, a national organization for psychological care in the Netherlands [24,25]. The DIPP questionnaire consists of the Dutch version [26] of the Four-Dimensional Symptom Questionnaire [27,28], complemented with several multiple-choice and open questions. The 4D Symptom Questionnaire contains 50 multiple-choice questions measuring distress, depression, anxiety, and somatization, which are dimensions of common psychopathology [27]. The complementary questions relate to current symptoms, treatment goals, anamnesis, psychosis risk, substance use, and medication. The DIPP questionnaire was originally developed, validated, and published in Dutch. A translated version of the questionnaire is provided in Multimedia Appendix 1. The answers to the following three open questions were used to develop the text-classification model:

1. Can you briefly describe your main symptom or symptoms?
2. What would you like to achieve with a treatment?
3. Have there been any events (such as a divorce, loss of job, or accident) that, in your opinion, affect your current symptoms, and if so, what are they?

The information collected through the DIPP questionnaire results in scores on four scales: somatization, depression, distress, and anxiety. Each patient is then assigned a weight to indicate symptom severity and level of care (no care, general practice mental health care, basic mental health care: short, basic mental health care: moderate, basic mental health care: intensive, and specialist mental health care). The outcome is verified by a semistructured diagnostic interview over the telephone, which results in a formal referral advice and diagnosis. Intake, diagnosis, referral, and treatment are all conducted by a CBT-certified health psychologist.

Automated Text-Screening Model

Supervised Classification
To screen future textual answers on the 3 open questions of the DIPP questionnaire for the presence of anxiety and panic...
disorders, PTSD, mood disorders, eating disorders, substance addiction, or somatic symptom disorders, a supervised multi-class text classifier was developed. It is called a supervised classifier because it was developed based on an existing set of text fragments provided with the correct diagnostic labels. The answers to all 3 questions were combined into 1 text document per patient. The formal referral advice based on the DIPP questionnaire scores and the diagnostic interview was used as the diagnostic label to be predicted by the model. The classifier is multi-class because the model refers each input text to 1 of multiple classes: the 7 disorders present in the input corpus.

The development of a supervised classification model follows a 2-phase strategy: a model-training phase and a label-prediction phase. This section explains the steps taken in each phase. The complete classification procedure is shown graphically in Figure 1.

Figure 1. Supervised text classification model procedure. In the training phase, the model is trained on labeled feature sets extracted from the input texts. In the prediction phase, the trained model is used to predict labels for new, unlabeled feature sets extracted from the input texts.

Training
During training, text features (words or word combinations) are extracted from each input text, converting the texts to labeled feature sets. These labeled feature sets are used as input for the machine learning algorithm, which generates a multi-class model by selecting the most informative features for each class.

Preprocessing
Standard preprocessing steps such as tokenization (splitting texts into separate tokens such as words, numerical expressions, and punctuation) and normalization (removing punctuation, converting capital letters to lowercase letters, and stripping off accents) were applied to process all texts at the word level [29]. All words were brought back to their core, meaning-baring stem using the Snowball Stemmer, a standard stemming algorithm available for many languages, including Dutch [30]. The resulting set of words for each input text is termed the vocabulary and consists of tokens, all used words or word combinations, and types, all unique words or word combinations used [31].

Feature Extraction
To convert the resulting vocabularies to feature sets suitable as input for the machine learning algorithm, the dimensionality of the feature space was reduced by feature extraction and feature selection techniques. For feature extraction, different document representation and vectorization schemes were compared. The document representations considered were unigrams, N-grams, and N-multigrams, which are single words, sequences of N words, and variable-length sequences of maximum N words, respectively [32]. The vectorization schemes refer to the specified term weights, for which we used normalized term frequency [33] or term frequency–inverse document frequency [34].

Feature Selection
Stop word removal, minimal document frequency, and the Pearson chi-square test were used to select the most informative features. Stop word removal was considered because stop words are generally not expected to contribute to the meaning of the text [29], although other studies contradict this [35]. In addition, words that only occur sparsely throughout the complete corpus (document frequency) may also be removed [36]. The most informative features (features with the highest chi-square values) are found by ranking features based on their Pearson chi-square value, a common and highly efficient method that measures the independence among corpora by comparing the observed and expected feature occurrences in each class [33]. The optimal number of features to select is determined by an exhaustive parameter grid search, which will be further explained in the section Analytical Strategy.

Machine Learning Algorithm
The selected features and their corresponding labels from the training set form the labeled feature sets that were used as input for the machine learning algorithm. The SVM [37] was used because this is a high-performing and robust classification algorithm that deals well with high-dimensional data such as text [36]. As SVMs were originally intended for binary classification tasks, multi-class (K-class) classification tasks
were split into $K$ binary classification tasks following the one-against-all (O-a-A, also known as one-versus-rest) or the one-against-one (also known as one-versus-one) decomposition strategy.

The one-against-one strategy, which compares each pair of classes separately [38,39], is generally considered a better approach when dealing with class imbalance, as was present in our data set. However, this strategy requires substantially more computational resources because many pairwise SVMs need to be trained. We therefore applied the widely used O-a-A strategy, which compares each single class with the remaining classes [38,39]. This strategy is the most commonly used, thanks to its computational efficiency and interpretability. To compensate for the class imbalance, a class-weighting scheme was used where classes were weighted to be inversely proportional to the class frequencies in the complete data set (as proposed by King and Zeng [40]). This puts more emphasis on the information extracted from the smaller classes and prevents the highly present classes from overshadowing the classification model.

The SVM with O-a-A strategy was implemented in the linear support vector classifier within the LIBLINEAR library developed by Fan et al [41]. Finally, 2 hyperparameters could be optimized for the SVM model: the kernel parameter $\gamma$ [42], which controls training and testing error [42], and the regularization parameter $C$, which controls training and testing error [42]. We used a linear kernel as is common in text classification [36] and optimized the regularization parameter in the grid search (see Analytical Strategy).

**Prediction**

During prediction, text features of new, unlabeled input texts were extracted and converted to feature sets following the same strategy used during training. Following the O-a-A approach, we fitted 7 SVMs, 1 for each disorder, alternately comparing 1 strategy used during training. Following the O-a-A approach, we fitted 7 SVMs, 1 for each disorder, alternately comparing 1 class against all the remaining classes (the negative class).

Confusion Matrix

The performance of the classifier was measured by comparing the predicted labels with the known labels for each class using a confusion matrix. A confusion matrix displays the instances in the predicted classes per column and the true classes per row, directly visualizing the number of correctly labeled documents on the diagonal and the errors (mislabeled documents) in the surrounding cells [31]. Table 1 shows the confusion matrix for a 7-class classifier with classes A-G.

![Confusion Matrix](https://mental.jmir.org/2022/4/e21111)

Table 1. Confusion matrix for the 7-class classifier: comparison of true and predicted class labels for classes A-G.

<table>
<thead>
<tr>
<th>True label</th>
<th>Predicted label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class A</td>
<td>$TP_A$ $^{ab}$</td>
</tr>
<tr>
<td>Class B</td>
<td>$E_{B,A}$</td>
</tr>
<tr>
<td>Class C</td>
<td>$E_{C,A}$</td>
</tr>
<tr>
<td>Class D</td>
<td>$E_{D,A}$</td>
</tr>
<tr>
<td>Class E</td>
<td>$E_{E,A}$</td>
</tr>
<tr>
<td>Class F</td>
<td>$E_{F,A}$</td>
</tr>
<tr>
<td>Class G</td>
<td>$E_{G,A}$</td>
</tr>
</tbody>
</table>

$^a$TP: true positive.

The values on the diagonal (in italics) show the correctly predicted class labels. The off-diagonal values show the prediction errors.

**Performance Metrics**

The correct predictions (TPs and TNs) and errors (FPs and FNs) were then used to calculate performance metrics for each class. Bird et al [31] define several metrics, the simplest of which is accuracy, a measure for the proportion of correctly labeled input texts in the test set. The recall, also called sensitivity or TP rate, indicates how many of the text documents with a true (known) positive label were identified as such by the classifier and is

https://mental.jmir.org/2022/4/e21111
calculated for each class by using the following formula: TP/(TP+FN). The precision (also known as positive predictive value) is calculated for each class by using the formula TP/(TP+FP) and concerns the proportion of positively predicted text documents where the true (known) label was indeed positive. The harmonic mean of the precision and recall, $2 \times (\text{Precision} \times \text{Recall})/ (\text{Precision} + \text{Recall})$, is the $F_1$ score. The overall performance scores for the classifier were calculated by averaging the performance scores of all classes (ie, all 7 binary SVMs that were fitted following the O-a-A approach). We used weighted macroaveraged scores because this accounts for class imbalance; as this method gives equal weight to each class, it prevents the most occurring classes from dominating the model [45].

**Analytical Strategy**

To prevent model evaluation bias, different subsets of the data were used to train, validate, and test the model. A nested $k$-fold cross-validation strategy was adopted, using a 5-fold cross-validated grid search in the inner loop for model selection and 5-fold cross-validation in the outer loop for model evaluation (see Figure 2 for a schematic representation). To make sure all classes were represented in each fold in approximately the same proportions as in the complete data set, stratified sampling [46] was used in both cross-validation loops.

**Figure 2.** Nested 5-fold cross-validation scheme. The validation strategy consists of an inner and an outer 5-fold cross-validation loop. In the inner loop an exhaustive parameter grid search is conducted using data from the development set to select the best combination of parameter settings. The selected model is then tested on the held-out test set from the outer loop to evaluate final model performance. Both loops are being iterated 5 times, alternately using each fold as test set (outer loop) or validation set (inner loop) once.

For the outer loop, the data set was first split into 5 folds, alternately defining 4 folds as the development set for model selection and setting aside 1 fold as a test set for assessing final model performance and generalization. To optimize the different model parameters, an exhaustive parameter grid search was conducted in the inner loop. In this grid search, all possible combinations of parameter values were fitted on the data set in search of the combination resulting in the highest performance score. The following model parameters and parameter values were compared:

- Choice of representation scheme: unigrams, bigrams, trigrams, or 3-multigrams
- Term weights: term frequency or term frequency–inverse document frequency
- Stop words: included or excluded
- Minimal document frequency: 1, 2, 3
- Optimal number of features: ranging from 1 to 500, increasing with steps of 20
- Regularization parameter $C$: 1, 2, 3, 10, 100, 1000

The search can be guided by any performance metric. We used the $F_1$ score because this is the preferred metric when working with imbalanced data sets. The grid search also uses a 5-fold cross-validation approach, splitting the development set into 5 folds, alternately using 4 folds for training and the remaining fold for validation. This is repeated until every fold has been used as the validation set once. The parameter combination that resulted in the highest mean weighted $F_1$ score over all validation sets was selected as the final model. The generalization performance of the selected model was estimated by again calculating the mean weighted $F_1$ score, but this time over all test sets from the outer cross-validation loop.

**Text-Classification Tool**

The process of model development by means of nested stratified $k$-fold cross-validated grid search is fully automated in a blind text-classification tool developed by the authors. This tool can be used to develop and test a text-classification model on any available text data set without human insight into the data set (hence blind). It can be installed and used locally. After
installation, no external packages are required; therefore, there is no need to send sensitive information over the internet for external text processing or analysis. An extensive description of the tool, the model development process, and the results on different test data sets will be published in a forthcoming paper by the authors. The tool was applied and described previously in a master’s thesis [47].

Ethics Approval
This study was approved by the Behavioral, Management, and Social Sciences Ethics Committee of the University of Twente (approval number 220089).

Results

Data Set
Table 2 shows the demographic characteristics and DIPP questionnaire results of the patients and the lexical characteristics of their documents for each class. The class labels are Addiction (substance use disorders), Panic (anxiety disorders with panic attacks), Anxiety (anxiety disorders without panic attacks), PTSD, Mood (mood disorders, including depressive disorders), Eating (eating disorders), and Somatic (undifferentiated somatoform and other somatic symptom disorders).

Table 2. Patient and lexical characteristics (N=5863).

<table>
<thead>
<tr>
<th>Variable</th>
<th>Addiction (n=197)</th>
<th>Anxiety (n=1100)</th>
<th>Panic (n=1100)</th>
<th>PTSD(^a) (n=1016)</th>
<th>Somatic (n=1100)</th>
<th>Mood (n=1100)</th>
<th>Eating (n=250)</th>
<th>Total (N=5863)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographic characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>18 (9.14)</td>
<td>362 (32.91)</td>
<td>394 (35.82)</td>
<td>498 (49.02)</td>
<td>500 (45.45)</td>
<td>265 (24.09)</td>
<td>180 (72)</td>
<td>2217 (37.81)</td>
</tr>
<tr>
<td>Male</td>
<td>34 (17.26)</td>
<td>176 (16)</td>
<td>174 (15.82)</td>
<td>119 (11.71)</td>
<td>197 (17.91)</td>
<td>166 (15.09)</td>
<td>8 (3.20)</td>
<td>874 (14.91)</td>
</tr>
<tr>
<td>Unknown(^b)</td>
<td>145 (73.60)</td>
<td>562 (51.09)</td>
<td>532 (48.36)</td>
<td>399 (39.27)</td>
<td>403 (36.64)</td>
<td>669 (60.82)</td>
<td>62 (24.80)</td>
<td>2772 (47.28)</td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>37.9 (15.0)</td>
<td>36.5 (14.2)</td>
<td>36.3 (13.8)</td>
<td>36.5 (13.1)</td>
<td>41.2 (11.7)</td>
<td>39.2 (14.4)</td>
<td>30.8 (10.0)</td>
<td>37.7 (13.6)</td>
</tr>
<tr>
<td><strong>DIPP(^c) questionnaire results: 4DSQ(^d) scales, mean (SD)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety</td>
<td>5.8 (5.3)</td>
<td>8.0 (5.0)</td>
<td>11.9 (5.5)</td>
<td>9.3 (6.3)</td>
<td>5.8 (4.9)</td>
<td>6.6 (5.3)</td>
<td>5.8 (5.6)</td>
<td>8.1 (5.8)</td>
</tr>
<tr>
<td>Depression</td>
<td>3.9 (3.8)</td>
<td>3.3 (3.1)</td>
<td>4.1 (3.7)</td>
<td>4.8 (3.8)</td>
<td>3.5 (3.1)</td>
<td>6.3 (3.7)</td>
<td>4.4 (3.8)</td>
<td>4.4 (3.7)</td>
</tr>
<tr>
<td>Distress</td>
<td>19.0 (8.4)</td>
<td>19.2 (7.5)</td>
<td>20.5 (7.6)</td>
<td>23.6 (6.9)</td>
<td>21.5 (6.9)</td>
<td>23.7 (6.8)</td>
<td>19.1 (8.2)</td>
<td>21.5 (7.5)</td>
</tr>
<tr>
<td>Somatization</td>
<td>10.5 (6.8)</td>
<td>11.1 (6.6)</td>
<td>15.3 (6.9)</td>
<td>14.7 (7.4)</td>
<td>13.6 (6.7)</td>
<td>12.6 (6.9)</td>
<td>12.4 (7.1)</td>
<td>13.3 (7.1)</td>
</tr>
<tr>
<td><strong>Level of care, n (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No care</td>
<td>15 (7.61)</td>
<td>62 (5.64)</td>
<td>28 (2.55)</td>
<td>31 (3.05)</td>
<td>61 (5.55)</td>
<td>55 (5)</td>
<td>13 (5.20)</td>
<td>265 (4.52)</td>
</tr>
<tr>
<td>General practice</td>
<td>46 (23.35)</td>
<td>198 (18)</td>
<td>165 (15)</td>
<td>90 (8.86)</td>
<td>171 (15.55)</td>
<td>183 (16.64)</td>
<td>19 (7.60)</td>
<td>872 (14.87)</td>
</tr>
<tr>
<td>Basic: short</td>
<td>11 (5.58)</td>
<td>127 (11.55)</td>
<td>92 (8.36)</td>
<td>93 (9.15)</td>
<td>110 (10)</td>
<td>102 (9.27)</td>
<td>8 (3.20)</td>
<td>543 (9.26)</td>
</tr>
<tr>
<td>Basic: moderate</td>
<td>4 (2.03)</td>
<td>90 (8.18)</td>
<td>69 (6.27)</td>
<td>41 (4.04)</td>
<td>84 (7.64)</td>
<td>34 (3.09)</td>
<td>7 (2.80)</td>
<td>329 (5.61)</td>
</tr>
<tr>
<td>Basic: intensive</td>
<td>23 (11.68)</td>
<td>340 (30.91)</td>
<td>340 (30.91)</td>
<td>244 (24.02)</td>
<td>457 (41.55)</td>
<td>283 (25.73)</td>
<td>29 (11.60)</td>
<td>1716 (29.27)</td>
</tr>
<tr>
<td>Specialist</td>
<td>98 (49.75)</td>
<td>283 (25.73)</td>
<td>406 (36.91)</td>
<td>517 (50.89)</td>
<td>217 (19.72)</td>
<td>443 (40.27)</td>
<td>174 (69.60)</td>
<td>2138 (36.47)</td>
</tr>
<tr>
<td>Lexical characteristics: words (N), mean (SD)</td>
<td>55.1 (55.0)</td>
<td>71.7 (69.5)</td>
<td>68.0 (103.5)</td>
<td>75.1 (157.0)</td>
<td>70.9 (74.9)</td>
<td>65.5 (75.2)</td>
<td>76.4 (72.4)</td>
<td>69.9 (98.2)</td>
</tr>
</tbody>
</table>

\(^a\)PTSD: posttraumatic stress disorder.
\(^b\)For patients who entered the study through their general practitioner, the gender is not registered; as such, gender is unknown for a large group of patients.
\(^c\)DIPP: Digitale Indicatiehulp Psychische Problemen (Digital Indication Aid for Mental Health Problems).
\(^d\)4DSQ: Dutch 4D Symptom Questionnaire. For the 4DSQ, trichotomized 5-point scale responses on each subscale are reported (see the study by Terluin et al [27] for the exact scoring method). Scores are considered moderately elevated (>10, >2, >8, >10) or strongly elevated (>20, >5, >12, >20) for distress, depression, anxiety, and somatization, respectively.

The demographic information (Table 2) shows that for those patients whose gender is known, more women than men had registered for all treatments except for Addiction. The mean age of the sample was 37.7 (SD 13.6) years, where patients treated for eating disorders were considerably younger (mean 30.8, SD 10.0 years) and patients treated for somatic disorders slightly older (mean 41.2, SD 11.7 years). The DIPP questionnaire results show that patients in treatment for panic attacks had the highest anxiety and somatization scores compared with those in other treatments. Patients treated for mood disorders scored higher on the depression and distress scale than those treated for other disorders. From the lexical characteristics, it can be concluded that the texts written by patients treated for addiction were considerably shorter: the
mean number of words was 55.1 (SD 55.0) compared with an overall mean number of words of 69.9 (SD 98.2) for the complete sample. Patients with PTSD and eating disorders wrote relatively longer answers (mean 75.1, SD 157.0, and mean 76.4, SD 72.4, respectively).

**Screening Model**

**Overview**

In the exhaustive grid search in the inner 5-fold cross-validation loop, all possible combinations of parameter values listed in the Analytical Strategy section were compared to find the model with the highest performance score. This resulted in a linear support vector classifier with a weighted $F_1$ score of 0.471. The selected model consisted of 470 unigrams (single words) weighted by term frequency. For this model, stop words were excluded and the selected keywords had to occur in at least one of the documents in the training set. The optimal value found for the regularization parameter $C$ was 1. An overview of the selected model parameters is presented in Table 3.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Best value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remove stop words</td>
<td>Yes</td>
</tr>
<tr>
<td>Minimal $a$ documents</td>
<td>1</td>
</tr>
<tr>
<td>Representation scheme</td>
<td>Unigrams</td>
</tr>
<tr>
<td>Term weight</td>
<td>Term frequency</td>
</tr>
<tr>
<td>Select $k^b$ best features</td>
<td>470</td>
</tr>
<tr>
<td>Regularization parameter $C$</td>
<td>1</td>
</tr>
</tbody>
</table>

$a$: number of documents a feature should be present in.

$b$: number of most informative features selected.

**Most Informative Features**

The 50 most informative unigrams (from hereon referred to as “keywords”) are listed in Table 4. The keyword column contains the translated English keywords, followed by the Dutch stemmed keywords in parentheses. The large chi-square values and highly significant $P$ values (when applying the O-a-A strategy, chi-square value $>3.84$ is required to indicate significant differences [$P<.05$]) show that there are significant differences between the observed and expected frequencies with which the keywords occur in texts written by patients with different disorders. These keywords are considered informative and were therefore included in the model. The remaining columns show the frequency with which each keyword occurs in each class (classes being the disorders for which the patients are being treated). For each keyword, the class in which it occurs most is presented in italics. This shows that especially for the eating disorder, many highly distinctive keywords were found: 22 of the 50 keywords have the highest frequency of occurrence in Eating. Some keywords have a high occurrence in several of the classes; for example, the word *fear* occurs often in the classes Panic (N=574), Anxiety (N=411), and PTSD (N=205). Of the top 50, none of the keywords occurs the most in Anxiety, and only a few have the highest occurrence in Mood and Addiction.
### Table 4. The 50 most informative features (keywords) of the multi-class classifier with the highest chi-square values and significant ($P<.05$) $P$ values.

<table>
<thead>
<tr>
<th>English keyword (Dutch stem)</th>
<th>Chi-square (df)</th>
<th>$P$ value</th>
<th>Addiction$^a$</th>
<th>Anxiety$^a$</th>
<th>Eating$^a$</th>
<th>Mood$^a$</th>
<th>PTSD$^{ab}$</th>
<th>Panic$^a$</th>
<th>Somatic$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>food (eten)</td>
<td>437.0 (1)</td>
<td>&lt;.001</td>
<td>1</td>
<td>18</td>
<td>218$^c$</td>
<td>19</td>
<td>20</td>
<td>32</td>
<td>22</td>
</tr>
<tr>
<td>binge (eetbui)</td>
<td>407.3 (1)</td>
<td>&lt;.001</td>
<td>0</td>
<td>3</td>
<td>121$^c$</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>fear (angst)</td>
<td>126.6 (1)</td>
<td>&lt;.001</td>
<td>17</td>
<td>411</td>
<td>25</td>
<td>98</td>
<td>205</td>
<td>574</td>
<td>82</td>
</tr>
<tr>
<td>eating disorder (eetstoornis)</td>
<td>100.9 (1)</td>
<td>&lt;.001</td>
<td>0</td>
<td>1</td>
<td>33</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>panic attacks (paniekaanval)</td>
<td>96.6 (1)</td>
<td>&lt;.001</td>
<td>0</td>
<td>13</td>
<td>2</td>
<td>12</td>
<td>21</td>
<td>196</td>
<td>11</td>
</tr>
<tr>
<td>to vomit (brak)</td>
<td>93.1 (1)</td>
<td>&lt;.001</td>
<td>0</td>
<td>6</td>
<td>28</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>bulimia (boulimia)</td>
<td>78.4 (1)</td>
<td>&lt;.001</td>
<td>0</td>
<td>1</td>
<td>26</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>eating pattern (eetpatron)</td>
<td>75.8 (1)</td>
<td>&lt;.001</td>
<td>0</td>
<td>0</td>
<td>24</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>weight (gewicht)</td>
<td>69.9 (1)</td>
<td>&lt;.001</td>
<td>0</td>
<td>0</td>
<td>26</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>to throw up (overgev)</td>
<td>62.2 (1)</td>
<td>&lt;.001</td>
<td>2</td>
<td>16</td>
<td>39</td>
<td>0</td>
<td>1</td>
<td>19</td>
<td>4</td>
</tr>
<tr>
<td>panic (paniek)</td>
<td>57.7 (1)</td>
<td>&lt;.001</td>
<td>8</td>
<td>42</td>
<td>4</td>
<td>22</td>
<td>49</td>
<td>185</td>
<td>23</td>
</tr>
<tr>
<td>eat (eet)</td>
<td>53.4 (1)</td>
<td>&lt;.001</td>
<td>2</td>
<td>6</td>
<td>33</td>
<td>2</td>
<td>4</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>drink (drink)</td>
<td>48.0 (1)</td>
<td>&lt;.001</td>
<td>20</td>
<td>5</td>
<td>2</td>
<td>8</td>
<td>2</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>eating behavior (eetgedrag)</td>
<td>44.4 (1)</td>
<td>&lt;.001</td>
<td>0</td>
<td>0</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>nightmares (nachtmerries)</td>
<td>42.3 (1)</td>
<td>&lt;.001</td>
<td>0</td>
<td>7</td>
<td>6</td>
<td>78</td>
<td>8</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>binge (vreetbui)</td>
<td>40.9 (1)</td>
<td>&lt;.001</td>
<td>0</td>
<td>0</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>work (werk)</td>
<td>39.5 (1)</td>
<td>&lt;.001</td>
<td>30</td>
<td>214</td>
<td>26</td>
<td>238</td>
<td>172</td>
<td>232</td>
<td>531</td>
</tr>
<tr>
<td>past (verled)</td>
<td>37.4 (1)</td>
<td>&lt;.001</td>
<td>5</td>
<td>74</td>
<td>11</td>
<td>65</td>
<td>188</td>
<td>73</td>
<td>47</td>
</tr>
<tr>
<td>healthy (gezond)</td>
<td>36.8 (1)</td>
<td>&lt;.001</td>
<td>4</td>
<td>21</td>
<td>50</td>
<td>30</td>
<td>17</td>
<td>37</td>
<td>20</td>
</tr>
<tr>
<td>overeating (overet)</td>
<td>35.6 (1)</td>
<td>&lt;.001</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>sense (zin)</td>
<td>34.5 (1)</td>
<td>&lt;.001</td>
<td>20</td>
<td>41</td>
<td>17</td>
<td>198</td>
<td>78</td>
<td>56</td>
<td>103</td>
</tr>
<tr>
<td>to lose weight (afvall)</td>
<td>30.6 (1)</td>
<td>&lt;.001</td>
<td>2</td>
<td>1</td>
<td>21</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>eating problems (eetproblem)</td>
<td>30.3 (1)</td>
<td>&lt;.001</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>scared (bang)</td>
<td>30.1 (1)</td>
<td>&lt;.001</td>
<td>13</td>
<td>205</td>
<td>22</td>
<td>65</td>
<td>131</td>
<td>206</td>
<td>54</td>
</tr>
<tr>
<td>to attack (aannvall)</td>
<td>29.5 (1)</td>
<td>&lt;.001</td>
<td>2</td>
<td>6</td>
<td>1</td>
<td>8</td>
<td>22</td>
<td>74</td>
<td>7</td>
</tr>
<tr>
<td>to compensate (compensier)</td>
<td>28.3 (1)</td>
<td>&lt;.001</td>
<td>0</td>
<td>3</td>
<td>11</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>fat (dik)</td>
<td>28.2 (1)</td>
<td>&lt;.001</td>
<td>0</td>
<td>3</td>
<td>12</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>anxious (angstig)</td>
<td>27.6 (1)</td>
<td>&lt;.001</td>
<td>6</td>
<td>152</td>
<td>8</td>
<td>62</td>
<td>102</td>
<td>168</td>
<td>43</td>
</tr>
<tr>
<td>tired (moes)</td>
<td>27.2 (1)</td>
<td>&lt;.001</td>
<td>12</td>
<td>66</td>
<td>10</td>
<td>145</td>
<td>88</td>
<td>66</td>
<td>214</td>
</tr>
<tr>
<td>panic attack (paniekaanval)</td>
<td>27.1 (1)</td>
<td>&lt;.001</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>55</td>
<td>3</td>
</tr>
<tr>
<td>drug (drug)</td>
<td>26.3 (1)</td>
<td>&lt;.001</td>
<td>14</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>raped (verkracht)</td>
<td>23.6 (1)</td>
<td>.001</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>44</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>accident (ongeluk)</td>
<td>23.0 (1)</td>
<td>.001</td>
<td>7</td>
<td>26</td>
<td>1</td>
<td>20</td>
<td>87</td>
<td>30</td>
<td>24</td>
</tr>
<tr>
<td>overweight (overgewicht)</td>
<td>22.9 (1)</td>
<td>.001</td>
<td>0</td>
<td>1</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>to smoke (blow)</td>
<td>22.6 (1)</td>
<td>.001</td>
<td>10</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>hyperventilation (hyperventilatie)</td>
<td>22.5 (1)</td>
<td>.001</td>
<td>2</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>4</td>
<td>51</td>
<td>7</td>
</tr>
<tr>
<td>tired (vermoeid)</td>
<td>22.5 (1)</td>
<td>.001</td>
<td>7</td>
<td>33</td>
<td>4</td>
<td>60</td>
<td>35</td>
<td>38</td>
<td>134</td>
</tr>
<tr>
<td>alcohol (alcohol)</td>
<td>22.5 (1)</td>
<td>.001</td>
<td>15</td>
<td>9</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>abuse (misbruik)</td>
<td>21.1 (1)</td>
<td>.002</td>
<td>5</td>
<td>9</td>
<td>0</td>
<td>6</td>
<td>53</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>obsession (obsessie)</td>
<td>21.1 (1)</td>
<td>.002</td>
<td>0</td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>flashback ( flashback)</td>
<td>20.7 (1)</td>
<td>.002</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>27</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
Performance Metrics

Table 5 reports the performance scores of the final model for each class. The model performs especially well in screening for eating disorders. The high precision (0.75) for this class means that 75% (41/55) of the patients whom the model classified as having an eating disorder were indeed referred to a treatment for eating disorders by the therapist. The high recall (0.82) shows that 82% (41/50) of the patients who were referred to a treatment for eating disorders by the therapist were also identified as such by the model. The model screens the least effective for addiction and anxiety. Only 25% (13/52) of the patients who were classified by the model as having an addiction and 44% (77/175) of the patients with anxiety were also identified as such by the therapist. Of the patients referred to treatments for addiction and anxiety by the therapist, respectively, only 33% (13/40) and 35% (77/220) were also found by the model. The overall accuracy of the classifier is 0.49, meaning that 49.28% (578/1173) of the predictions made by the model were correct. For a 7-class classifier this exceeds random guessing, which would be 14% (1/7).

Table 5. Performance metrics final model: per class and average performance scores for the final model (N=1173).

<table>
<thead>
<tr>
<th>Disorder</th>
<th>Patients in test set, n (%)</th>
<th>Precision</th>
<th>Recall</th>
<th>F₁ score</th>
<th>Overall accuracyänder</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addiction</td>
<td>40 (3.41)</td>
<td>0.25</td>
<td>0.33</td>
<td>0.28</td>
<td>—</td>
</tr>
<tr>
<td>Anxiety</td>
<td>220 (18.76)</td>
<td>0.44</td>
<td>0.35</td>
<td>0.39</td>
<td>—</td>
</tr>
<tr>
<td>Eating</td>
<td>50 (4.26)</td>
<td>0.75</td>
<td>0.82</td>
<td>0.78</td>
<td>—</td>
</tr>
<tr>
<td>Mood</td>
<td>220 (18.76)</td>
<td>0.44</td>
<td>0.50</td>
<td>0.47</td>
<td>—</td>
</tr>
<tr>
<td>PTSD</td>
<td>203 (17.31)</td>
<td>0.57</td>
<td>0.52</td>
<td>0.54</td>
<td>—</td>
</tr>
<tr>
<td>Panic</td>
<td>220 (18.76)</td>
<td>0.57</td>
<td>0.55</td>
<td>0.56</td>
<td>—</td>
</tr>
<tr>
<td>Somatic</td>
<td>220 (18.76)</td>
<td>0.46</td>
<td>0.50</td>
<td>0.48</td>
<td>—</td>
</tr>
<tr>
<td>Weighted average</td>
<td>N/Aänder</td>
<td>0.50</td>
<td>0.49</td>
<td>0.49</td>
<td>0.49</td>
</tr>
</tbody>
</table>

aAccuracy is the overall accuracy of the classifier averaged over all classes.
bData not available for separate classes.
cPTSD: posttraumatic stress disorder.
dN/A: not applicable.

Confusion Matrix

The confusion matrix in Table 6 contains the absolute counts and normalized values (counts corrected by the number of documents present in each class, in %) for the true and predicted labels. The normalized values are the most useful because these indicate the proportion of correctly predicted labels for each class, independent of the class sizes. The normalized values on the diagonal show that the classifier screens the best for Eating (41/50, 82% correct), followed by Panic (121/220, 55%), PTSD (105/203, 51.7%), Somatic (111/220, 50.5%), Mood (110/220, 50%), Anxiety (77/220, 35%), and Addiction (13/40, 32.5%). Of the 1173 patients in the test set, this screener referred 578 (49.28%) to the correct treatment.
Table 6. Confusion matrix for the 7-class classifier: absolute and normalized values (%) for the true versus predicted class labels.

<table>
<thead>
<tr>
<th>True disorder</th>
<th>Predicted disorder</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Addiction</td>
</tr>
<tr>
<td>Addiction (N=40), n (%)</td>
<td>13 (32.5)(^b)</td>
</tr>
<tr>
<td>Anxiety (N=220), n (%)</td>
<td>11 (5)</td>
</tr>
<tr>
<td>Eating (N=50), n (%)</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Mood (N=220), n (%)</td>
<td>11 (5)</td>
</tr>
<tr>
<td>PTSD (N=203), n (%)</td>
<td>2 (1)</td>
</tr>
<tr>
<td>Panic (N=220), n (%)</td>
<td>4 (1.8)</td>
</tr>
<tr>
<td>Somatic (N=220), n (%)</td>
<td>10 (4.5)</td>
</tr>
</tbody>
</table>

\(^a\)PTSD: posttraumatic stress disorder.

\(^b\)The diagonal cells show the correctly predicted labels (in italics). The off-diagonal cells show the prediction errors for each class.

The normalized confusion matrix is plotted in Figure 3 to give a more direct visual presentation of which classes are being misclassified. The darker the blue tones, the higher the proportions in that cell. The perfect classifier would have a dark blue diagonal line, surrounded by white cells. The plot confirms that Eating is rarely misclassified. Most confusion occurs for Addiction, which is often mislabeled as a mood or somatic disorder. In addition, mood and somatic disorders are often confused with each other, as are panic and anxiety disorders.

Figure 3. Normalized confusion plot. Visual presentation of the true versus predicted class labels. The darker the tone, the higher the proportion in the corresponding cell. PTSD: posttraumatic stress disorder.
Final Model Evaluation
The 5-fold cross-validation grid search was conducted 5 times in the inner loop, iteratively using 4 of the 5 folds from the outer loop as the development set once. This resulted in 5 weighted $F_1$ scores: one for each final model selected in the inner cross-validation loop that was tested on the test set in the outer cross-validation loop. The weighted $F_1$ scores for the 5 outer test folds were 0.49, 0.49, 0.47, 0.46, and 0.47. The scores are relatively close to each other, meaning that the classifier generates stable results. The mean weighted $F_1$ score over the 5 iterations was 0.48 (SD 0.01). This is the estimated generalization performance, the performance that can be expected when the final model is applied to new data sets in the future.

Discussion
Principal Findings
This study aims to improve the intake procedure of a web-based mental health therapy provider by using multi-class text classification to automatically screen textual answers on open questions from an intake questionnaire for a range of different mental health disorders. The resulting classification model turned out to be especially effective in screening for Eating, correctly identifying 82% (41/50) of the patients with an eating disorder. This is comparable to binary classifiers in previous studies; for example, for PTSD (80% correct; performance score for the SVM model based on unigrams) [22] or depression (84% correct) [23]. The correct classification rates for the other disorders were substantially lower: Panic, 55% (121/220); PTSD, 51.7% (105/203); Mood, 50% (110/220); Somatic, 50.5% (111/220); Anxiety, 35% (77/220); and Addiction, 32.5% (13/40), resulting in an overall accuracy of 49.28% (578/1173). This is a reasonable score for a 7-class classification model, although not high enough to make strong and accurate referrals for all treatments.

The difference in performance is also reflected in the selected keywords, of which many are highly discriminative for Eating. For example, simple words such as food, binge, weight, or bulimia are clearly related to eating disorders while sparsely being used in texts written by patients with other disorders. For the remaining disorders, the keywords found are more generally related to fears and feelings and occur more in all classes except for Eating and thus are less discriminative. For example, fear and scared are selected as keywords for Panic, but they also have high occurrences in Anxiety and PTSD. Sense is a keyword for Mood, but it is also highly used in texts written by patients with somatic disorders, whereas the somatic keyword tired is also used often in texts written by patients with a mood disorder.

As a result, the model could not accurately differentiate between mood and somatic disorders as well as between panic and anxiety disorders. None of the 50 most informative keywords was related mostly to Anxiety, for which one of the lowest classification performances was reported.

The reasons for the overlap in keywords for different disorders may be symptom overlap (in case symptoms are part of the defining symptom set of multiple disorders) and nonspecificity of defining symptoms (in case symptoms also occur regularly in persons without the disorder), both issues resulting from definitional choices made in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition [48]. For example, PTSD has overlapping symptom criteria with depression, generalized anxiety disorder, and panic disorder [49]. When (future) patients are asked to describe their most important symptoms (1 of the 3 open intake questions, the answers to which were used to develop our model; Multimedia Appendix 1), because symptoms for several disorders overlap, it is not surprising that descriptions and thus keywords for these disorders will also overlap.

The low screening performance for Addiction could be because only a very small number of patients with addiction were present in the data set (n=197), and as such the machine learning algorithm was provided with inadequate training data for this class. However, for Eating, not many more patients were included (n=250), and for this class the classifier performed very well. Another reason could be that patients in Addiction were found to write shorter texts; on average, the mean number of words used by patients in the Addiction class is 55.1 (SD 55.0) versus an average of 69.9 (SD 98.2) over all classes and even 76.4 (SD 72.4) for the Eating class (Table 2). This shows that patients with an eating disorder provide a more extensive description of their symptoms, treatment goals, and anamnesis than patients with addiction. Because of this, less information is available for Addiction than for Eating, which makes it hard for the machine learning algorithm to learn key features for this class.

The results further show that the classifier has difficulty differentiating mood from somatic disorders and panic from anxiety disorders. For mood and somatic disorders this can be explained by the fact that most patients with somatic disorders are commonly found to have an underlying mood disorder [50]. The difficulty in distinguishing between panic and anxiety disorders could be because panic disorder is actually classified as a type of anxiety disorder in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition [51]. Despite the underlying similarity, we expected that panic disorders could be easily distinguished from anxiety disorders because of their distinctive characteristics. Although the classifier found quite a few significant keywords for Panic (eg, fear, panic attack, and panic), these words also occurred often in texts written by patients with Anxiety and PTSD and thus were not discriminative enough. In contrast, none of the top 50 keywords had the highest frequency of occurrence in the Anxiety class, meaning no highly discriminative keywords were found for Anxiety. As Panic and Anxiety are closely related, merging the 2 classes into one would probably improve the performance of the screener. However, this would reduce the practical applicability of the screener because the goal is to refer patients to the most suitable treatment offered by the health care provider, which offers separate treatments for Panic and Anxiety.

Theoretical and Practical Contributions
First, this study extends the findings of previous research on text-classification applications in mental health care in that it
investigates the use of a multi-class classifier instead of a binary classifier, which is predominantly used [20,21]. This way it is possible to screen for multiple disorders at once, without the need to make prior assumptions regarding the type of disorder a new patient signs up with. Second, this study shows an application of text mining and natural language processing applications originally developed for English text to non-English, in this case Dutch, mental health data. Although most of the scientific publications in this area focus on English data and tools [20,21], most underlying processes and techniques are not language dependent and as such can be easily applied to non-English texts. Finally, our data set contained high-quality class labels, consisting of official clinical diagnoses made by a therapist, enabling us to compare the labels predicted by the classifier to an official gold standard instead of a proxy. The quality of the labels is highly important for the performance, validity, and clinical applicability of the developed model, and acquiring large, high-quality mental health text data sets is found to be challenging [21].

For the web-based mental health provider, the developed text screener provides an additional outcome score that can be used as input for the automatically generated indicative diagnosis and for the formal diagnostic interview by the therapist. Although the overall performance of the classifier still needs to be improved, the classifier was able to distinguish eating disorders very well. As an eating disorder is currently not reported as a separate scale in the DIPP questionnaire (which reports on anxiety, depression, distress, and somatization), the text screener provides additional information that was not available from the multiple-choice questions.

This study further shows how text mining, specifically text classification, can add value to current (web-based) mental health care practice because it can be used for more efficient screening, intake, or treatment referral. As described previously, mental health problems often remain undiagnosed and untreated. This can partly be attributed to the fact that most people are only seen by primary care providers who do not always recognize mental health conditions because of comorbidity between physical and psychological diseases. Magruder et al [8] therefore propose that primary care clinicians should receive more training on the recognition of these conditions. However, even after being diagnosed, patients often remain untreated because of the scarcity of health care resources. To scale up the mental health workforce, the World Health Organization [52] has proposed to shift caregiving to mental health workers with lower qualifications or even lay helpers under the supervision of highly qualified health workers [8]. An alternative way of reducing the workload for mental health workers is to increase the use of modern technologies in screening, providing treatment, and monitoring treatment outcomes. Instead of (or in addition to) extra training for primary care providers, an automatic screening tool could also aid in the recognition of mental health problems, and instead of shifting care to lower-qualified or lay helpers, mental health providers could be supported by modern technology. The automatic screener described in this paper should be seen as an example of this.

Limitations

An important limitation of our classifier is that it is not capable of dealing with comorbidity. Comorbidity is an important issue; 45% of the patients with psychiatric disorders are reported to meet the criteria for ≥2 disorders within the same year [48]. As stated earlier, it is not unusual for patients with somatic disorders to have an underlying mood disorder [50], whereas mood disorders are commonly found to co-occur with anxiety disorders [48]. Substance use disorders are also often found to co-occur with other mental health disorders; for drug use disorders in particular, high associations with anxiety (especially panic disorder) and affective (mood) disorders have been reported [53-55]. The main limitation of this study is that although the multi-class classifier can screen for multiple disorders at once, it does not take into account the possibility that a patient can have a combination of multiple disorders simultaneously (comorbidity). This may explain why the screener did not prove to be very capable when it came to distinguishing between some disorders, which indicates the need for a multi-label classifier that can screen for combinations of disorders instead of only a single disorder.

Another limitation may be the fact that we used a blinded tool to develop the automatic screening model. Some might state that to develop a model, at least some insight into the input data is required to actively monitor the development process. However, the tool was tested and applied in a previous study by the authors and in a master’s thesis [47] in which the process and outcomes were confirmed. This tool enabled us to work on sensitive information without any insight into the textual content, on a local computer, and without the need to send the information over the internet for processing and analysis, thereby reducing not only the risk of privacy issues, but also the risk of possible confirmation bias because of prior knowledge. However, by using a tool, one is limited by the choice of models and parameters made beforehand during the development of the tool. Adding to, or changing, the tool’s settings based on new insights is quite laborious because this requires developing, updating, and installing a new version. Therefore, we chose to use a common and proven classifier and analytic approach [21].

Yet another limitation could be the definition of the classes and class imbalance. The classes used in this study are defined by the specific diagnoses for which treatment is offered by the mental health clinic Interapy, instead of symptomatology. The performance of the classifier might be improved by grouping together comorbid disorders or disorders with overlapping symptoms (e.g., combine somatic and mood disorders or panic and anxiety disorders). However, because this would decrease the practical usability of the screener, we chose to keep these classes separate. Model performance may also be influenced by class balance (or imbalance), that is, the extent to which the texts are evenly distributed across the classes. The classes Addiction and Eating were strongly underrepresented in our data set, and despite the use of class weights and stratified samples, performance for the Addiction class especially was poor. In contrast, the highest performance was reported for the Eating class; therefore, it seems that as long as the text content is discriminative enough, even small samples may provide enough information to make strong predictions.
Future Research

Future research should focus first of all on improving the overall performance of the classifier. The current screener does not show a high enough performance for all classes, which might be solved by trying alternative classification algorithms or machine learning strategies such as a multi-label strategy to deal with comorbidity. In addition to adopting a multi-label approach, exploring a multistage learning system also seems a useful next step. Multistage models (eg, cascade classifiers) use a staged decision process in which the output of a model (the first stage) is used as the input for a successive model (the second stage), and so on. Multistage models are widely used in medical practice, and physicians use this approach for the stepwise exclusion of possible diagnoses [56]. Several studies show that multistage classifiers outperform the single-stage classifiers generally used in supervised multi-class classification tasks; for example, in the prediction of liver fibrosis degree [57] and in distinguishing among levels of dementia [56]. For our screener it could be useful to first classify the disorders into more general groups of (possibly) overlapping disorders, grouping Anxiety, Panic, and PTSD in 1 class and Mood and Somatic symptom disorders in another while keeping Eating and Substance abuse disorders separate, followed by a more specialized classification model to distinguish among the specific disorders within the groups. This prevents the best predictable class (in our case, Eating) from dominating the machine learning process. In addition, because one of the problems was finding (enough) discriminative keywords for some of the disorders, adding additional open questions to the web-based intake procedure to collect more text data may be helpful. Adjusting the questions by focusing less on symptoms (which are found to overlap for some disorders) and focusing instead on aspects possibly more defining for each disorder may also lead to more discriminative keywords and consequently better models.

Second, further uses of text mining and machine learning in mental health care practice should be explored. Text mining can be (and is) used for many more activities during and after treatment; for example, in analyzing patient–physician or patient–carer communication [58] or in evaluating treatments by capturing patients’ opinions from comments on the web [59]. In addition, text mining can also be used to assess factors and processes underlying recovery of, for example, patients with an eating disorder [60]. A new application for text mining in e–mental health practice could be to use it as a tool to support therapists by offering suggestions for patient-specific feedback. The current computerized CBT process as used in this study consists of sequential homework assignments covering common CBT interventions. On the basis of the content of these assignments, therapists offer standardized feedback and instructions, including motivational techniques, adapted to the needs and situation of the patient [14]. It would be interesting to examine whether we could use text mining to automatically highlight sections in the assignments that require attention or that may indicate a positive or negative change in behavior.

Conclusions

This study showed that automatic text classification can improve the current web-based intake and referral procedure of a Dutch mental health clinic by providing an additional outcome score to be used as input for the indicative referral advice and the formal diagnostic interview. Automatically generating an additional indicator based on the textual input may lead to a more efficient and standardized intake process, saving time and resources because the text no longer needs to be processed and interpreted by the therapist. As such, automatic text screening could be a step in the right direction for solving patient, systemic, and provider factors underlying the underdetection of mental health disorders and underuse of available mental health treatments [6]. The overall complaint-discriminating quality of the screener still has to be improved, but the good detection performance with regard to eating disorders in this study (and with regard to PTSD and depression in other studies) shows that text-based screening is a promising technique for psychiatry. This paper contains multiple recommendations for research paths that could improve this complaint-discriminating quality of text screeners (eg, using stratified analysis techniques when symptoms overlap complaints). Altogether, the technique is getting closer to implementation in general practice, where it definitely could be of great value. Especially in areas around the world with a limited number of mental health care workers, automatic text classification could be helpful. It could save time that is now spent on screening and assessment of patients, time that could be used for counseling and treatment.

Conflicts of Interest

The authors MH and BS are affiliated with the organization within which the research was conducted.

Multimedia Appendix 1

The translated Digitale Indicatiehulp Psychische Problemen (DIPP; Digital Indication Aid for Mental Health Problems) questionnaire used for the web-based intake of new patients. The DIPP questionnaire was originally developed and validated in Dutch [24]. The DIPP questionnaire starts with the Dutch version of the 4D Symptom Questionnaire [26-28], followed by additional questions regarding current symptoms, treatment goals, anamnesis, psychosis risk, substance use, and medication.

[DOCX File, 20 KB - mental_v9i4e21111_app1.docx ]

References

1. GBD 2015 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of


47. Smalbergher I. Learning analytics towards identifying the processes of higher order thinking in online discussions of MOOCs (Master’s thesis). University of Twente. 2017. URL: http://essay.utwente.nl/74246/ [accessed 2022-01-25]


Abbreviations

CBT: cognitive behavioral therapy
DIPP: Digitale Indicatiehulp Psychische Problemen (Digital Indication Aid for Mental Health Problems)
FN: false negative
FP: false positive
O-a-A: one-against-all
PTSD: posttraumatic stress disorder
SVM: support vector machine
TN: true negative
TP: true positive

©Sytske Wiegersma, Maurice Hidajat, Bart Schrieken, Bernard Veldkamp, Miranda Olff. Originally published in JMIR Mental Health (https://mental.jmir.org), 11.04.2022. 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 https://mental.jmir.org/, as well as this copyright and license information must be included.
A Serious Game for Young People With First Episode Psychosis (OnTrack>The Game): Qualitative Findings of a Randomized Controlled Trial

Samantha Jankowski1*, MA; Kathleen Ferreira2*, PhD; Franco Mascayano1*, PhD; Effy Donovan2*, BA; Reanne Rahim1, MA; Michael L Birnbaum3, MD; Sabrina Yum-Chan3, BS; Deborah Medoff4, PhD; Bethany Marcogliese2, PhD; Lijuan Fang5, PhD; Terriann Nicholson5, MD; Lisa Dixon1,5, MPH, MD

1Department of Psychiatry, New York State Psychiatric Institute, New York, NY, United States
2Department of Research and Evaluation, C4 Innovations, Needham, MA, United States
3Department of Psychiatry, The Zucker Hillside Hospital, Glen Oaks, NY, United States
4Department of Psychiatry, University of Maryland School of Medicine, Baltimore, MD, United States
5Department of Psychiatry, Columbia University Irving Medical Center, New York, NY, United States

* these authors contributed equally

Corresponding Author:
Lisa Dixon, MPH, MD
Department of Psychiatry
New York State Psychiatric Institute
1051 Riverside Drive
New York, NY, 10032
United States
Phone: 1 212 326 8441
Email: Lisa.Dixon@nyspi.columbia.edu

Abstract

Background: Several studies have shown the benefits of coordinated specialty care (CSC) for individuals with first episode psychosis; however, pathways to care are marred by lack of knowledge, stigma, and difficulties with treatment engagement. Serious games or video interventions may provide a way to address these factors.

Objective: This study focuses on qualitative results of a randomized controlled trial comparing OnTrack>The Game (OTG) with recovery videos (RVs) on engagement, stigma, empowerment, hope, recovery, and understanding of psychosis in clients receiving CSC. Clinicians are also interviewed regarding their perceptions of the interventions and suggestions for improvement.

Methods: A total of 16 clients aged 16-30 years, with first episode psychosis attending a CSC program in New York State, and 9 clinicians participated in the qualitative interviews. Interviews were analyzed using the rapid identification of themes from audio recordings method.

Results: For clients, themes included relatability of game content, an increased sense of hope and the possibility of recovery, decreased self-stigma and public stigma, increased understanding of the importance of social support, and increased empowerment in the OTG group. Clinicians had a preference for RV and provided suggestions for dissemination and implementation.

Conclusions: Themes that may help inform future research in this area, particularly regarding dissemination and implementation of OTG and RV, emerged.

Trial Registration: ClinicalTrials.gov NCT03390491; https://clinicaltrials.gov/ct2/show/NCT03390491

(JMIR Ment Health 2022;9(4):e33526) doi:10.2196/33526

KEYWORDS

video gaming; internet; recovery; schizophrenia; psychosis; clinicians; mobile phone
Introduction

Background

Schizophrenia can be a debilitating illness that affects approximately 20 million people worldwide [1]. Several studies have demonstrated the benefits of intensive early intervention programs, known in the United States as coordinated specialty care (CSC), for individuals with nonaffective first episode psychosis (FEP) [2-4]. However, treatment engagement can be challenging [5,6] and pathways to care are affected by stigma, symptom misattribution, and preference for self-management [7]. There is a need for tools that can help engage young people in treatment; reduce stigma; increase the understanding of psychosis, empowerment, and hope; and promote personal recovery.

A recent review has proposed the role of serious gaming as a way to promote engagement of individuals with psychosis, particularly if the game has a clear goal and involves service users in game development [8]. Serious games are those that have a clear educational purpose, can provide individuals with the opportunity to engage in decision-making for real-world situations in a safe environment, and help them envision positive future events and roles [8]. These games also have the potential to motivate individuals to engage in treatment and provide an avenue for disseminating mental health information in a manner that is nonthreatening or portrayed in a more casual and easily accessible way, rather than through formal clinician–patient contact [9]. A recent study by Ferchaud et al [10] has shown that self-identification with a video game avatar with psychosis reduced the desire for social distance from individuals with mental illness, thereby, providing the opportunity to reduce stigma. However, the impact of serious games on individuals with psychosis needs further examination, particularly owing to a dearth of randomized controlled trials (RCTs) in this area [11].

In a previous pilot study, we developed and tested a prototype of OnTrack®The Game (OTG), a computer-based role-playing game for individuals with FEP. We asked 20 individuals who are enrolled in OnTrackNY, a CSC program for individuals with FEP, to test the game in one 45–60–minute sitting [12]. The game included a customizable character, quests to practice real-world skills, information about FEP, and videos emphasizing stories of hope and recovery with individuals who have experienced FEP (ie, recovery videos [RVs]). The RVs have been previously shown to be effective in reducing public stigma in cross-sectional [13] and longitudinal [14] RCTs. Results from the OTG pilot study suggested a significant increase in positive attitudes toward recovery; however, we did not detect significant differences in hope, empowerment, and engagement. Moreover, qualitative results indicated that individuals found the community setting of OTG, educational components, and RVs to be the most helpful, and participants noted that playing OTG has the potential to translate to real-world decision-making.

In this study, we refine and augment OTG using clinician and client feedback (eg, Youth Advisory Board composed of individuals with FEP) and test it in an RCT comparing the short-term (2 months) and long-term (5 months) effects of OTG versus RVs on engagement, stigma, empowerment, hope, recovery, and understanding of psychosis. RVs are chosen as the comparison group owing to pilot study participants reporting RVs as the most valued game feature [12]. In this study, we want to determine whether the videos alone were producing an effect on our domains of interest or if the videos embedded in the game in addition to other game elements were creating an effect. RCT results will be presented separately. This paper focuses on the qualitative findings from OnTrackNY client and clinician interviews.

Objectives

The study aims to gain a better understanding of clients’ experiences with the game or videos, including technical difficulties, what was helpful, what could be improved, and impact on the study variables of interest, and clinician’s impressions of the game’s impact on clients and ideas for dissemination and implementation.

Methods

Development of OTG

In the pilot study, we created a prototype of OTG, a game based in a fictional town where a player with psychosis has their own apartment and can go to school, work, mental health clinic, park, movie theater, and gym [12]. OTG included quests to practice real-world skills and opportunities to collect coins, information about FEP, and videos with stories of hope and recovery (ie, RVs that are in the comparison condition). Our team used participant feedback from the pilot study and the Youth Advisory Board to refine and augment the game. We held 3 web-based meetings with the Youth Advisory Board. For some members, technology was a challenge; thus, our team also provided the opportunity for members to offer written feedback about the game. On the basis of the feedback, we improved the game’s functionality (changing to a new platform that allows mobile use and developing a smartphone interface), included more quests and interactions with nonplayer characters, expanded the video library, added more customization features to the main character (hair, facial features, and body type), and provided rewards for collecting points such as opportunities to upgrade the avatar’s outfit and decorate their apartment. We also included an in-game computer with links to information about psychosis, treatment options, and wellness strategies.

As modifications were made, the Center for Social Innovation development team used Agile methodologies, working in 2-week sprints. Partners at OnTrackNY were included in weekly planning meetings to enhance and facilitate communication and stakeholder buy-in. This was critical to the development of the final product. Several times, modifications recommended by the Youth Advisory Board or that emerged from the pilot data were overruled by OnTrackNY partners to avoid content that could be traumatizing or viewed as offering treatment within the game. For example, in a scenario where the player had to decide what to do if a barista gives him the wrong drink at a coffee shop, we decided not to include topics such as having the player feel overwhelmed by noise and crowding in the coffee shop. At times, this created a shift away from situations in the
game that were viewed as more realistic by reviewers; however, our team deferred to clinically trained staff at OnTrackNY to make this determination. OnTrackNY staff relied on evidence-based approaches to create materials and dialogues for the game. For example, a stress management handout included deep breathing, progressive muscle relaxation, and visualization. In dialogues with other nonplayer characters, social skills training principles were used, including how to have a casual conversation, how to ask for something, and how to assertively express needs. Other recommendations, such as the Youth Advisory Board’s observation that OTG would be enhanced by playing with other people, were beyond the scope and budget of this project.

Quests provided opportunities to engage in decision-making and view potential consequences of these decisions. For example, in one scenario, the player was given the wrong drink at a coffee shop and had to decide how to respond, and in another, the player had to decide whether to engage in a social interaction with a neighbor. These scenarios provided individuals with the opportunity to practice social skills, see the outcomes of potential responses, and receive immediate feedback in a safe environment mimicking potential real-world situations, designed to promote empowerment, hope, and recovery. Resources available on the in-game computer and the RVs triggered throughout the game (and comprising the RV condition) provided opportunities for psychoeducation and promoted stories of hope and recovery in individuals with psychosis. All these additions were aimed at addressing stigma, engagement, hope, recovery, and knowledge about psychosis.

**RV Condition**

The RVs, present in both comparison conditions, comprised 24 videos, each with duration of 3–5 minutes, featuring individuals who received treatment in OnTrackNY and their relatives. These individuals shared their experiences living with psychosis and the challenges and successes they experienced during the recovery process. For example, in a video, a young man with psychosis and his mother described his symptoms during illness onset, the benefits of connecting with treatment, his current participation in work and school, and positive responses from others when he shared his experience.

These videos are freely available to patients and families on the New York State Psychiatric Institute Center of Practice Innovations website; however, they are not formally part of treatment as usual. A shortened 90-second version of the videos has been tested in previous RCTs and has shown reduced public stigma in an MTurk general population sample immediately after the intervention [13] and at the 30-day follow-up [14]. In the tested video, a young woman with schizophrenia described her symptoms during illness onset; her current daily difficulties, including experiencing attenuated psychotic symptoms; and her ability to participate in work and maintain meaningful relationships. Similar to other contact-based stigma reduction interventions, which provide opportunities to engage in live or video-based contact with a presenter with mental illness, the video provided participants with the opportunity to engage in video-based social contact with an individual who exhibited some attenuated symptoms of psychosis and was able to function well in her social and occupational roles, thereby, potentially disconfirming preconceived notions about individuals with psychosis and reducing stigma [15].

**Recruitment**

Recruitment for the quantitative study of 159 individuals with FEP (ie, clients) occurred between April 18, 2019, and December 30, 2020. Participants completed a baseline assessment and follow-up assessments at 2 months and 5 months after enrollment. Participants were initially recruited from 18 clinical sites in the OnTrackNY network, a CSC program in New York State. Owing to recruitment challenges imposed by the COVID-19 pandemic, 35.8% (57/159) of participants were recruited via the web from various CSC programs across the country. Inclusion criteria were the following: a diagnosis of nonaffective psychosis and receiving services at a CSC program, aged 16–30 years, can speak English, able to give fully informed consent, access to the internet with a PC or tablet (sites provide access to computers or tablets on-site if this is a barrier), and access to email. Participants provided consent, which included a question that asked whether they would give permission to be contacted to participate in a qualitative interview at a later time.

Study participants were randomly assigned to the OTG or RV intervention. Participants were provided basic instructions on how to access and navigate the intervention to which they were assigned and were given 2 months to play the entire game or watch all RVs. They were also encouraged to use OTG or RV on a weekly basis throughout that 2-month period. Our research team collected analytic data on the extent to which each participant accessed OTG or the RV site and the content within each site. Data reflect cumulative use and indicate that approximately 30.2% (48/159) of the participants accessed the sites they were assigned. Inconsistent use, possible reasons for lack of use, and overall implications will be discussed in a subsequent manuscript on outcomes related to the study. For qualitative interviews, the team used purposive sampling, examining use data to identify individuals with a range of exposure to the intervention. Interviewing participants who varied in product use allowed the team to identify barriers to and facilitators of use.

For the qualitative study, a subsample of participants from the OTG and RV groups who agreed to be contacted were recruited between July 22, 2020, and December 1, 2020. We planned to enroll 20 clients in 1:1 ratio (eg, 10/20, 50% OTG participants and 10/20, 50% RV participants). Of the 42 individuals who provided consent to be contacted for the qualitative study, 93% (39/42) were contacted, and 41% (16/39) of them agreed to participate in the interview (5/16, 31% in the RV group and 11/16, 69% in the OTG group). Of the 23 participants who were contacted but did not participate, 13% (3/23) had difficulties with scheduling, 35% (8/23) declined, and 52% (12/23) did not respond or had invalid contact information. Unfortunately, we faced difficulties in engaging participants from the RV group. Some RV participants indicated they did not want to stay involved in the study after finding out they would not be able to play OTG and some participants did not want to complete additional assessments.
A total of 9 clinicians from CSC programs (n=1, 11% with a client in RV group and n=8, 89% with clients in both OTG and RV groups) were enrolled in the qualitative study between June 25, 2020, and August 31, 2020. Some clinicians involved in the study chose not to participate in the qualitative interviews. Inclusion criteria were being a licensed mental professional providing care in a New York CSC program and working with a client participating in the quantitative study.

**Ethics Approval**

All procedures were approved by the New York State Psychiatric Institute Review Board (protocol #7643) and the trial was registered on Clinicaltrials.gov ID # NCT03390491.

**Qualitative Interviews**

Interviews were conducted by 2 trained research assistants (Terriann Nicholson and Sapna Mendon-Plasek) via WebEx (Cisco) and were audio-recorded. Interviews lasted approximately 30 minutes, and OnTrackNY participants were compensated US $25. Clinicians were not compensated. Brief semistructured interview guides were used for participants and clinicians. Client interview questions were intended to elaborate on quantitative measures and included questions about the game’s influence on treatment engagement; recovery; empowerment; hope; understanding of psychosis; and stigma, which was divided into self-stigma and public stigma. Self-stigma encompasses the judgments and negative beliefs people internalize or hold about themselves [16]. Public stigma encompasses the attitudes and feelings expressed by many people in the public toward individuals with psychosis [16]. It involves identifying differences, connecting those differences to stereotypes, and separating groups by those stereotyped differences. Questions corresponding to each domain are included in Textbox 1. Additional questions were related to potential technical difficulties, most favorite and least favorite aspects, frequency of use, relevance to experiences with symptoms and treatment, and recommendations for improvement.

Clinician interviews included questions about the following: whether the game or videos were discussed in the session, most helpful or least helpful aspects, perceived changes in client interactions with the team and their attitudes, whether they would recommend the game or videos to other clients, perceived barriers to client use, ways to encourage individuals to use the products, and recommendations for improvement.

**Textbox 1.** Client qualitative interview questions and domains assessed.

<table>
<thead>
<tr>
<th><strong>Engagement</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Thinking back to before you started playing the game or using the recovery video site and after, has your participation in treatment and other support services changed?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Recovery</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Let’s talk about your feelings about recovery before playing the game or using the videos website and after. When you think back, have there been any changes in how you think or feel about recovery or do you feel the same?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Empowerment</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Did playing the game or using the videos site help you think about ways to speak up for yourself?</td>
<td></td>
</tr>
<tr>
<td>Did you think more about how you can participate in and make decisions about your treatment?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Hope</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Let’s talk about your hope for the future and your treatment before playing the game or using the videos site and after. When you think back, have you felt any changes in your hopefulness for the future or do you feel the same?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Understanding of psychosis</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>How was the content relevant to your own life and experiences with your symptoms and treatment?</td>
<td></td>
</tr>
<tr>
<td>Did you learn anything about your symptoms or treatment?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Self-stigma</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>After playing the game or using the videos site, did you feel differently about yourself?</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Public stigma</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>How do you think the game or videos site portrayed individuals living with mental illness?</td>
<td></td>
</tr>
</tbody>
</table>

**Data Analysis**

For qualitative data, the rapid identification of themes from audio recordings method was used [17]. Similar to procedures in our pilot study [12], a summary template was created by one of the authors (FM) with the predetermined interview codes described above. Codes for the client data included the following: treatment engagement, recovery, stigma, empowerment, hope, understanding of psychosis, how often game or videos were used, likes and dislikes, and technical difficulties. Codes for clinician data included the following: most and least helpful aspects of game or videos, changes
observed in the client, barriers and ways to improve engagement, whether they would recommend the game or videos, and additional feedback. Upon completion of interviews, the authors listened to the recordings and documented key messages and relevant quotes. Then, the data were categorized into common themes across study conditions and differences across study conditions and respondent roles. Some themes reflected a priori topics from the template and others emerged directly from the interviews. Another author (ED) revised the proposed codes, added new themes, and synthesized the results. Both coders used a focused coding approach to determine which topics arose often and which represented unusual or particular concerns. Disagreements between coders were resolved by discussion among all authors.

Table 1. Demographic characteristics for recovery video and OnTrack>The Game groups.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Treatment group</th>
<th>Comparison</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>OnTrack&gt;The Game group (n=11)</td>
<td>Recovery videos group (n=5)</td>
<td>Chi-square test (df)</td>
<td>t test(^a) (df)</td>
</tr>
<tr>
<td>Sex, n (%)</td>
<td></td>
<td></td>
<td>0.04 (1)</td>
<td>N/A(^b)</td>
</tr>
<tr>
<td>Men</td>
<td>6 (55)</td>
<td>3 (60)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>5 (45)</td>
<td>2 (40)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age (years), mean (SD)</td>
<td>21.55 (4.01)</td>
<td>22.40 (1.95)</td>
<td>N/A</td>
<td>0.57 (13.79)</td>
</tr>
<tr>
<td>Race, n (%)</td>
<td></td>
<td></td>
<td>1.9 (3)</td>
<td>N/A</td>
</tr>
<tr>
<td>White</td>
<td>3 (27)</td>
<td>2 (40)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Black or African American</td>
<td>5 (45)</td>
<td>1 (20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>1 (9)</td>
<td>0 (0)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>2 (18)</td>
<td>2 (40)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethnicity, n (%)</td>
<td></td>
<td></td>
<td>0.04 (1)</td>
<td>N/A</td>
</tr>
<tr>
<td>Yes, Hispanic or Latino</td>
<td>5 (45)</td>
<td>2 (40)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not Hispanic or Latino</td>
<td>6 (55)</td>
<td>3 (60)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^a\) A 2-tailed t test.
\(^b\) N/A: not applicable.

Similar Themes for Clients Across Study Conditions

Overview
The analysis generated data related to 6 main themes (Textbox 2). The first 4 themes demonstrate the benefits of client exposure to OTG and RV content. These included (1) access to depictions of relatable experiences that provide individuals with psychosis with models of possibility, (2) increased sense of hope for the future and the ability to advance through recovery, (3) decreased self-stigma and public stigma, and (4) increased understanding of the importance of family and social support and their inclusion in the treatment process. The final 2 themes demonstrate differences between user experiences in the OTG and RV conditions. Improved empowerment was identified only by OTG participants, and clinicians strongly valued use of RVs over OTG with clients.

Results

Sample Characteristics
On the basis of self-reported data from participants in the qualitative subsample, the clients’ mean age was 21.81 years, and 56% (9/16) of the them were men. Participants identified as Black or African American (6/16, 38%), White (5/16, 31%), and Asian (1/16, 6%), and approximately half of them were Hispanic (7/16, 44%). There were no significant differences between RV and OTG groups at baseline (Table 1). All participants were from New York CSC sites. Demographic data were not collected for clinicians.
OnTrack> The Game (OTG) and recovery videos (RVs) are relatable.

- The vast majority of OTG or RV participants report that the content is relatable to them and other individuals with psychosis.
- OTG and RV provide accurate representation of real-life experiences and daily challenges for individuals with psychosis.
- OTG provides helpful tools for self-reflection for individuals with psychosis.

OTG and RVs improve feelings of hope for the future.

- The vast majority of OTG participants report that the game increases their sense of hope for their future and their ability to live, work, and advance their lives as a result of their participation.
- The majority of RV participants report that the RVs increase their sense of hope for their future and their ability to advance through recovery.

OTG and RVs decrease self-stigma and public stigma.

- OTG and RV participants report a slight decrease in both self-stigma and public stigma.
- Although most OTG and RV participants report a decrease in self-stigma and public stigma, there are some mixed results. For example, some OTG participants note that individuals with psychosis are portrayed “too favorably” by the game.

A circle of support is important.

- OTG and RV participants report an increased understanding of the importance of family and social support.
- OTG and RV participants report an increased understanding of the importance of involving the people who are part of your support system in your treatment.

Improved empowerment is identified only by OTG participants.

- Most OTG respondents report that the game increases their feelings of empowerment.
- This theme did not emerge for RV participants.

Clinicians strongly value use of RVs over OTG for clients.

- All clinician respondents would recommend the RVs to future clients.
- Slightly more than half of the clinicians would recommend OTG to future clients.

Relatability

Most study participants reported that both OTG and RVs were relatable. Totally, 64% (7/11) of the OTG participants and 80% (4/5) of the RV participants reported that they related to the content and felt enhanced feelings of connection and understanding after engaging with it. Participants provided many examples of the potential for this content to enhance feelings of connection and understanding. They reported feeling less lonely or isolated. OTG and RV participants also reported that OTG or RV provided accurate representation of real-life experiences and daily challenges for individuals with psychosis. Furthermore, OTG participants reported that the game provided helpful tools for self-reflection for individuals with psychosis:

[The RVs] help identify some of the struggles that come with psychosis and gives helpful tools to help with challenges, especially parts that talk about disclosure and when it’s appropriate to share. [RV participant]

Seeing daily activities in the midst of anxious situations, made conventional life more relatable. Learned a lot about myself, a great tool to be reflective. [OTG participant]

I feel different. I feel hopeful for the future. The goals in the game, such as getting up and taking care of yourself, made me more aware of what I was doing daily. [OTG participant]

Improved Feelings of Hope

Most of both OTG and RV participants (15/16, 94%) reported increased feelings of hope. All the OTG participants (11/11, 100%) noted an increased sense of hope for their future and increased their ability to live, work, and advance their lives as a result of their participation. OTG aided participants in addressing some of their self-doubt and beginning the process of overcoming it. Of the 5 RV participants, 4 (80%) participants reported that the RVs increased their sense of hope for the future and their ability to advance through recovery:

They had a diagnosis and still continued their life; at the beginning, I thought ‘that’s it, I’m not going to live a normal life’; now I feel proactive and am in school; looking at participants in the game helped. [OTG participant]

They’ve been through episodes, but overcame them, continue living their life -- that made me feel like I can live a life. [OTG participant]

I am more hopeful after watching the videos, and pretty much whenever I have connection with participants with mental illness. It made me realize that there is always another side no matter what
Difficulty you are going through...I feel more hopeful when I see success stories and I see the artwork they create. It reinforced the things I knew before. Even if you are in hell there is light at the end of the tunnel. [RV participant]

Decreased Self-Stigma and Public Stigma

Most OTG and RV participants reported increased positivity and a decrease in both self-stigma and public stigma as a result of playing OTG and watching the RVs. In all, 73% (8/11) of the OTG participants and 100% (5/5) of the RV participants noted that exposure to content that portrays individuals with psychosis positively, as complex people with talents and goals, not only helped them to cope with their own diagnosis but also to judge others with mental illness less harshly:

It gave me positivity that mental illness is nothing bad. Whatever happened, I went through it and survived. [OTG participant]

[Individuals with psychosis] were portrayed very positively and honest and genuine. They spoke very openly about their experiences and how they overcame it. When I finished all the videos, I reflected on what I saw and heard on the videos and it was very inspiring. I felt very happy and positive about it, seeing people in my age group going through the same struggles and overcoming it. [RV participant]

Although most OTG and RV participants reported a decrease in self-stigma and public stigma, some OTG participants reported that the game portrayed individuals with psychosis “too favorably.” They suggested that OTG include content where characters are struggling or not feeling well, which would allow those players to see examples of how those characters practice self-care and help themselves come out of difficult situations:

The game didn’t really show what happens in my own mind. It didn’t portray the individual’s feelings and thoughts outside of communicating with others. It was portrayed too positively. Nothing was negative or showed obstacles. [OTG participant]

The Importance of a Circle of Support

Many OTG and RV participants reported an increased understanding of the importance of family and social support for individuals with psychosis. Furthermore, participants reported an increased understanding of the importance of involving the people who are part of your support system in your treatment. Both the OTG and RVs emphasize the pivotal importance of relationships in healing and recovery. The content encourages users to identify family, friends, partners, groups (eg, support groups, study groups, and sports teams), mental health and healing practitioners (eg, primary clinician, peer specialist, and acupuncturist), mentors (eg, leader of a youth group, work supervisor, friend of the family, and teacher or professor), support animals, and any other outlets that are supportive. RVs and an activity sheet helped users consider the people who are part of their circle of support and develop effective ways to strengthen those relationships:

[The RVs] made me realize the importance of a support system. It affected my treatment because prior, when I had symptoms I isolated [myself] and psychosis is hard to understand. So, a support system was important because it helped bring back reality. [RV participant]

Differences Across Study Conditions

Although several similar themes were identified across OTG and RVs, important differences were also evident. Most notable were participants’ perspectives on the impact of the interventions on empowerment. The theme of empowerment did not emerge for RV participants. Of the 11 OTG participants, 10 (91%) participants noted that OTG increased their feelings of empowerment in treatment by showing examples of how to speak up during treatment and elsewhere and by presenting models of possibility in the form of these characters who are dealing with some of the same challenges as our participants and still building rich, fulfilling lives:

Even before I was involved with treatment, the game showed me I have to be involved—helped me become involved in decisions that doctors were making about me. [OTG participant]

I actually have a say in my treatment; the game helped me realize that I have to advocate for myself and have to be honest, I can control what happens in my treatment. [OTG participant]

Differences in Clinician and Client Perspectives

The most notable difference between clinician and client perspectives of the interventions was related to the value of RVs and OTG. Clinicians overwhelmingly stated that they would recommend RVs over OTG to clients; however, client preferences were mixed. All the clinicians interviewed during the study (9/9, 100%) noted that they would recommend the RVs to future clients, but only 56% (5/9) of them stated that they would recommend OTG. They cited issues such as difficulty for clients to access adequate computers and internet and computer skills of clients as barriers to use. They also noted that OTG graphics and technology had become outdated and the graphics were not well-suited for the key audience: “Some clients were uncomfortable with OTG’s graphics and felt they were intended for a younger audience” [clinician participant].

Clients also reported some challenges with gameplay. Some clients had difficulty in understanding how to move from one level to another:

I would complete a task and when I went back it would be erased and I would have to repeat a task that was already completed. [OTG participant]

Sometimes choices weren’t clear, so participants had to choose every option just to advance. [OTG participant]

Although clinicians found more value in the RV’s, results were mixed at the client level, and some participants shared that they found it difficult to find time to watch the RVs and that the videos themselves were “too long.” Furthermore, the interviewed clients had a more positive view of OTG than the interviewed
Clinicians. Clients expressed enthusiasm about many different facets of OTG: being able to create or personalize their own character, the game’s portrayal of the decision-making process and the ability to make decisions within the game, and the game’s messaging around communication skills and how to manage daily-life commitments.

Additional Implementation Considerations

Overview

In addition to themes described above, important considerations should be made for adoption and implementation of OTG and RVs in the field. The research team asked clinicians to provide suggestions regarding how to improve client engagement with the products tested in this study. The suggestions described below will steer future implementation and help our team maximize resources and achieve wider reach with OTG and RVs.

In their suggestions, clinicians were considering two questions: (1) How can OTG and RVs be made more accessible to clients? and (2) How can OTG and the RVs be integrated into clients’ overall treatment plans?

Accessibility

To make OTG and RVs more accessible, clinicians suggested providing access to computers or tablets for clients, hosting a demonstration of OTG and RVs on the OnTrackNY website, having on-site peers disseminate OTG and RVs while providing access to computers or tablets, and making OTG and RV resources available to families of clients using the technology so they can encourage them to use the products. Clinicians also had recommendations for making OTG and RVs easier to find. For example, they suggested setting up an Instagram page for OTG and RVs that clients could follow or, perhaps, advertising with videos that feature individuals who have played the game to build some familiarity. In addition to suggesting the creation of social media accounts for OTG and RVs, clinicians also suggested having OnTrackNY advertise OTG and RVs more in their social media campaigns.

Integration With Services

Clinicians provided suggestions for seamless integration of OTG and RVs into the clients’ treatment. They recommended using OTG and RVs as training tools when onboarding providers; enabling providers to use OTG and RV, so that they can help orient clients to the technology and discuss how it fits into their treatment; playing the game or watching the RVs during home visits; and providing examples of peer feedback from other participants who have used OTG and RV, so that it feels more recognizable. Clinicians also suggested using OTG and RVs to facilitate communication during early sessions, when rapport is still being built: “Sometimes it is hard for participants to talk and it would be nice to be able to play the game” [clinician participant].

Discussion

Principal Findings

This study qualitatively examined the impact of OTG and RVs on engagement, stigma, empowerment, hope, recovery, and understanding of psychosis in clients receiving CSC. Common themes for clients included relatability, increased hope and possibility of recovery, decreased self-stigma and public stigma, and increased understanding of the importance of family and social support, and OTG participants reported increased empowerment. Clinicians strongly preferred RVs and offered suggestions for dissemination and implementation for both OTG and RVs, including providing clients with computers or tablets, advertising the products through social media or a website, having families facilitate dissemination, and using the products during onboarding of providers and during initial sessions with clients.

Relatability of content was an expected finding as a Youth Advisory Board was involved in game development and refinement, and feedback was incorporated from pilot participants. The newest iteration of the game included resources about FEP and more family and friends characters and provided the option for characters to make poor decisions and experience the consequences of those decisions [12]. A notable and interesting finding included that some OTG participants believed that the game portrayed individuals with psychosis “too favorably.” This finding suggests that participants would like to see a balance of positive and negative experiences that individuals face.

The positive impact the games or videos had on improving hope and reducing both self-stigma and public stigma are also notable. These findings are consistent with previous research, which found that the RVs led to reduced public stigma in the general population [13,14]. Previous gaming research also suggests that video game avatars have the potential to reduce stigma through transportation into the story line and identification with the avatar [10]. It is possible that exposure to content that portrayed individuals with psychosis in a positive light and allowed participants to take on the identity of this character equipped them with the ability to cope with their own diagnosis and be less stigmatizing toward others with mental illness. This may have also contributed to the increased sense of hope noted by participants.

A finding that emerged outside our predetermined themes included an increased understanding of the importance of social support and their involvement in treatment. Previous research suggests that social support is particularly important during FEP and can correlate with low levels of positive symptoms and few hospitalizations [18]. In addition, a large number of individuals with psychosis have poor perceived support and are susceptible to feelings of loneliness and anxiety [19]. The game or videos may encourage participants, particularly those who tend to withdraw when they are symptomatic, to stay connected with individuals in their social network and involve them in treatment.
It was interesting to note that the theme of empowerment emerged in OTG but not in the RV group. Although this may have been owing to insufficient sample size in the RV group, it is also possible that specific game elements, such as providing knowledge, supporting behavior, and providing skills training, contributed to increased empowerment [20]. The game provided knowledge about psychosis and provided individuals the opportunity to learn from stories of other individuals with psychosis through the RVs. It also taught participants skills applicable to daily life, such as assertiveness, and supported behavior by providing players with choice in dialogues with nonplayer characters and direct feedback on these choices. Although the RV group also provided knowledge and skills training in the form of handouts and the videos, the supportive behavior component was lacking, as participants did not have the opportunity to engage in a choice and immediately see consequences or receive corrective feedback.

Notable findings also emerged regarding dissemination and implementation; clinicians had a preference for RV over OTG, but clients did not. Clinicians cited access issues, lack of computer skills, outdated and immature graphics, and lack of understanding of how to move from one level to another as common reasons. It was surprising that interviewed clients had a more positive view of the game. Clinicians may have overestimated how much these factors were impacting participants or the participants experiencing difficulties may not have participated in interviews. In the future, as clinicians may be the ones introducing OTG and RV to clients, it is important that they feel the materials are relevant and accessible to participants.

Importantly, many of the issues that were described reflect challenges identified by the Nonadoption, Abandonment, Scale-up, Spread, and Sustainability framework, which aims to help researchers predict and identify challenges in the implementation of technology in health care [21]. Factors to consider when implementing new technologies include the complexity of the health condition being treated, consumers’ sociocultural aspects that influence product use (eg, lack of computer skills, outdated and immature graphics, and lack of ease of use of the product and esthetics (eg, clunkiness), perceived value of the product, whether staff and consumers are willing or able to adopt the product, organizational capacity and readiness, work involved in product implementation, and wider institutional and sociocultural context influencing spread and sustainability [21]. To address concerns identified in our study regarding access and use, clinicians suggested providing participants with technological devices; having demonstration versions or having a peer, family, or clinician aid in dissemination; and using social media to advertise the products. Suggestions to have peers or clinicians aid in dissemination are consistent with calls for digital navigators to aid in the integration of technology into health care [22]. Although providing clients with technological devices may not be the most cost-effective solution for all clinics, these resources are available in OnTrackNY clinics. Clinicians also suggested that OTG and RV could be used to facilitate rapport-building and communication during initial treatment engagement. Perhaps, clients and clinicians could watch the videos or play the game together or they could be assigned as between-session homework and various topics could be discussed during the following session.

Limitations
This study had some limitations. The small sample of clients and clinicians may limit the generalizability and reliability of study findings. The small number of individuals in the RV group made it difficult to determine whether the difference in empowerment observed in the OTG was owing to the intervention or whether this theme would have emerged in the RV group also. Another potential limitation is that as all the participants were enrolled in OnTrackNY and were in different stages of treatment, it was unclear whether findings that emerged were owing to the interventions or being part of a treatment program. In addition, both groups had access to the same RVs (with RVs embedded within OTG); thus, it is difficult to distinguish differences that might have emerged from both groups.

Conclusions
It is important to note that despite these limitations, clear themes emerged that can help generate hypotheses and future research in this area. This study was able to highlight similar and differing themes in the experiences of clients and clinicians in using OTG and RV. Future research should aim to explore the dissemination and implementation of OTG and RV and the impact of these interventions on participants who are not enrolled in FEP programs. Other studies should also explore the differential impact of the intervention on clients at various stages of treatment and the impact of these products on family members or individuals without psychosis. The findings of our quantitative study will be reported separately.

Conflicts of Interest
None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (version 1.6.1).
[PDF File (Adobe PDF File), 1226 KB - mental_v9i4e33526_app1.pdf ]

References


https://mental.jmir.org/2022/4/e33526

JMIR Ment Health 2022 | vol. 9 | iss. 4 | e33526 | p.94
(page number not for citation purposes)
**Abbreviations**

- **CSC**: coordinated specialty care
- **FEP**: first episode psychosis
- **OTG**: OnTrack>The Game
- **RCT**: randomized controlled trial
- **RV**: recovery video

---

©Samantha Jankowski, Kathleen Ferreira, Franco Mascayano, Effy Donovan, Reanne Rahim, Michael L Birnbaum, Sabrina Yum-Chan, Deborah Medoff, Bethany Marcogliese, Lijuan Fang, Terriann Nicholson, Lisa Dixon. Originally published in JMIR Mental Health (https://mental.jmir.org), 06.04.2022. 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 https://mental.jmir.org/, as well as this copyright and license information must be included.
Original Paper

Brief Digital Interventions to Support the Psychological Well-being of NHS Staff During the COVID-19 Pandemic: 3-Arm Pilot Randomized Controlled Trial

Johannes H De Kock1,2, MA, PhD, CPsychol; Helen Ann Latham3, BSc, MBChB; Richard G Cowden4, PhD; Breda Cullen5, DClinPsy, PhD; Katia Narzisi1, PhD; Shaun Jerdan1, BSc, MRes, PhD; Sarah-Anne Munoz1, MA, MSc, PhD; Stephen J Leslie6, BSc, MBChB, PhD; Andreas Stamatis7, PhD; Jude Eze8, BSc, MSc, PhD

1Division of Rural Health and Wellbeing, Institute of Health Research and Innovation, School of Health, University of the Highlands and Islands, Inverness, United Kingdom
2Department of Clinical Psychology, New Craigs Psychiatric Hospital, Inverness, United Kingdom
3Nairn Town and Country Hospital, NHS Highland, United Kingdom
4Human Flourishing Program, Harvard University, Cambridge, MA, United States
5Institute of Health and Wellbeing, University of Glasgow, Glasgow, United Kingdom
6Cardiac Unit, Raigmore Hospital, NHS Highland, Inverness, United Kingdom
7Exercise and Nutrition Sciences, State University of New York, Plattsburgh, NY, United States
8Epidemiology Research Unit, Department of Veterinary and Animal Science, Northern Faculty, Scotland’s Rural College, Inverness, United Kingdom

Corresponding Author:
Johannes H De Kock, MA, PhD, CPsychol
Division of Rural Health and Wellbeing
Institute of Health Research and Innovation
School of Health, University of the Highlands and Islands
12b Ness Walk
Inverness, IV3 5SQ
United Kingdom
Phone: 44 01463 255000
Email: Hannes.De@nhs.scot

Abstract

Background: Health and social care staff are at high risk of experiencing adverse mental health (MH) outcomes during the COVID-19 pandemic. Hence, there is a need to prioritize and identify ways to effectively support their psychological well-being (PWB). Compared to traditional psychological interventions, digital psychological interventions are cost-effective treatment options that allow for large-scale dissemination and transcend social distancing, overcome rurality, and minimize clinician time.

Objective: This study reports MH outcomes of a Consolidated Standards of Reporting Trials (CONSORT)-compliant parallel-arm pilot randomized controlled trial (RCT) examining the potential usefulness of an existing and a novel digital psychological intervention aimed at supporting psychological health among National Health Service (NHS) staff working through the COVID-19 pandemic.

Methods: NHS Highland (NHSH) frontline staff volunteers (N=169) were randomly assigned to the newly developed NHSH Staff Wellbeing Project (NHSWBP), an established digital intervention (My Possible Self [MPS]), or a waitlist (WL) group for 4 weeks. Attempts were made to blind participants to which digital intervention they were allocated. The interventions were fully automated, without any human input or guidance. We measured 5 self-reported psychological outcomes over 3 time points: before (baseline), in the middle of (after 2 weeks), and after treatment (4 weeks). The primary outcomes were anxiety (7-item General Anxiety Disorder), depression (Patient Health Questionnaire), and mental well-being (Warwick-Edinburgh Mental Well-being Scale). The secondary outcomes included mental toughness (Mental Toughness Index) and gratitude (Gratitude Questionnaire-6).

Results: Retention rates mid- and postintervention were 77% (n=130) and 63.3% (n=107), respectively. Postintervention, small differences were noted between the WL and the 2 treatment groups on anxiety (vs MPS: Cohen d=0.07, 95% CI –0.20 to 0.33; vs NHSWBP: Cohen d=0.06, 95% CI –0.19 to 0.31), depression (vs MPS: Cohen d=0.37, 95% CI 0.07-0.66; vs NHSWBP: Cohen d=0.18, 95% CI –0.11 to 0.46), and mental well-being (vs MPS: Cohen d=–0.04, 95% CI –0.62 to –0.08; vs NHSWBP: Cohen
A similar pattern of between-group differences was found for the secondary outcomes. The NHSWBP group generally had larger within-group effects than the other groups and displayed a greater rate of change compared to the other groups on all outcomes, except for gratitude, where the rate of change was greatest for the MPS group.

Conclusions: Our analyses provided encouraging results for the use of brief digital psychological interventions in improving PWB among health and social care workers. Future multisite RCTs, with power to reliably detect differences, are needed to determine the efficacy of contextualized interventions relative to existing digital treatments.

Trial Registration: ISRCTN Registry (ISRCTN) ISRCTN18107122; https://www.isrctn.com/ISRCTN18107122

(JMIR Ment Health 2022;9(4):e34002) doi:10.2196/34002

KEYWORDS
eHealth; public health; depression; anxiety; well-being; mobile health; intervention studies; staff; occupational health; NHS; intervention; support; COVID-19; randomized controlled trial

Introduction

Background

Mental health (MH) has been deteriorating both globally and across the U.K. during the COVID-19 pandemic, with large-scale population studies reporting increased prevalence of depression and anxiety [1]. There are concerns that the public health crisis has disproportionately impacted the well-being of specialized populations, including health and social care workers (HSCWs) who provide valuable health care services. HSCWs exhibited high levels of preexisting MH problems before the COVID-19 pandemic [2-5], and recent evidence suggests that this group is at increased risk of experiencing worsening MH outcomes as a direct result of the COVID-19 pandemic [6-9]. MH problems in this population can affect morale and quality of care [5], which could have particularly devastating consequences for health systems because many parts of the world have been overwhelmed by the burden of COVID-19.

The majority of the general public [10] and health care staff [11] with common MH conditions do not access professional help, despite the existence of effective psychological treatments. Common reasons include a lack of service availability (especially in rural and remote areas), problems recognizing symptoms, treatment cost, and time constraints [11]. For HSCWs, the stigma surrounding mental illness and concerns about confidentiality have been identified as major barriers to accessing treatment and recovery, which can affect the quality of care HSCWs provide to patients [11,12]. Although research is ongoing, these barriers to treatment and downstream consequences for HSCWs, their families, and their patients appear to have been exacerbated by working through the COVID-19 pandemic [13].

Interventions designed to improve MH and psychological well-being (PWB) could help to mitigate the adverse effects of the COVID-19 pandemic on the well-being of HSCWs [8]. Digital psychological interventions overcome social distancing, rurality, and already overburdened clinician time constraints. Furthermore, digital interventions have a low cost relative to traditional psychological interventions, have already been widely used [14], are generally popular with users, and can be accessed anonymously at the user’s convenience. Evidence-based and rigorously tested digital interventions could allow for a rapid, economical, and large-scale dissemination of urgently needed psychological support for frontline staff working through the COVID-19 pandemic and its aftermath.

The past decade has seen digital psychological interventions being tested and validated in controlled and long-term follow-up studies, and the number of mobile MH interventions that are available is increasing rapidly. User reports indicate a significant increase in these apps downloaded during the first year of the COVID-19 pandemic in the U.K. [15] and in the U.S. [14]. Although validated digital interventions have been shown to be clinically efficacious, with effect sizes similar to that of traditional or face-to-face therapy [16], there is little research into the efficacy of such treatment approaches for frontline HSCWs who have been working through the COVID-19 pandemic [17]. Furthermore, the majority of digital psychological interventions during the public health crisis have been focused on decreasing symptoms associated with psychopathology (ie, depression and anxiety); few have been designed with end-user input (patient and public involvement [PPI]) and oriented toward enhancing PWB [17]. Given the unprecedented scale of the COVID-19 pandemic’s burden on HSCWs, specialized and contextual interventions are needed to support the MH of this population [18].

This Study

This study aims to provide preliminary evidence on the use of digital psychological interventions to support frontline staff psychological health in the context of the COVID-19 pandemic. In a pilot randomized controlled trial (RCT), we evaluated the use of 2 smartphone apps designed to support PWB against a control condition (waitlist [WL]): (1) My Possible Self (MPS) [19], which is a well-established validated app with a track record of showing significant improvements in depression, anxiety, and stress in users over a short period [19], together with good user satisfaction rates; (2) the National Health Service (NHS) Highland Wellbeing Project (NHSWBP), which is a PPI-informed, brief, fully automated, and context- (COVID-19 pandemic) and population-specific (frontline staff) digital psychological intervention built on the MPS model and wireframe to promote PWB among HSCWs.

We predicted that symptoms of depression and anxiety would decline among users randomly allocated to receive digital psychological interventions, while mental well-being would increase, relative to the WL group. Two positive psychology concepts shown to mitigate the negative effects of depression...
and anxiety and promote positive adaptation in the face of adversity (eg, what frontline staff are facing while working through a pandemic) that are amenable to change are mental toughness (MT) [20] and gratitude [21]. We also predicted that use of digital psychological interventions would increase MT and gratitude. Although we predicted both digital interventions to yield improvements relative to the WL group, we expected that the NHSWBP group would show greater rates of improvements because it is designed specifically for the COVID-19 context. To the best of our knowledge, this is the first trial to examine fully automated, brief digital psychological interventions aimed to support the psychological health of frontline staff working through the COVID-19 pandemic.

Methods

Eligibility Criteria

Participants were required to meet the following criteria: UK resident, aged 18 years and over, working in the NHSH as a health or social care worker during the COVID-19 pandemic, and owning an internet-enabled mobile phone. Both clinical (doctors, nurses, allied health professionals) and nonclinical (eg, administrators) staff were eligible.

Sample, Setting, and Procedure

Given that this was a pilot trial being conducted in a limited time, the sample size targets were based on pragmatic factors rather than an expectation of having the power to enable detection of the expected effect sizes. Participants were recruited locally and online between July and September 2020. Data collection took place at the beginning, middle, and end of the pilot RCT intervention phase, which ran from September 7 to October 5, 2020. Recruitment was conducted digitally by NHSH human resources, which included emails and electronic newsletters. Further recruitment was conducted via general physician (GP) practice managers, as well as heads of departments in primary and secondary care. A secondary level of recruitment was conducted on social media; a page for the study was created on Twitter, Facebook, and LinkedIn. Paid advertisements were also used on Facebook and LinkedIn to promote the study. Across all recruitment routes, interested individuals were directed to a secure data collection website via a weblink, where they first reviewed information about the study and provided electronic consent to participate. Eligible participants then completed a baseline survey, after which they were randomized to a condition. All participants were asked to complete follow-up surveys after the first 2 weeks of the intervention (middle) and 4 weeks after baseline following completion of the intervention period. At each assessment point, participants accessed the survey via a weblink sent to them in an email message. Demographic and basic clinical information was collected during the baseline survey, which included age, gender, place of work, job type, level of education, years of experience, previous psychiatric diagnosis, and whether the person was working directly with COVID-19 patients.

This study was part of the Scottish Government’s Rapid Research into COVID-19, and time restrictions limited recruitment activities; it was not possible to extend recruitment activities or product development beyond the grant’s funding time frame. Written informed consent was provided by all participants. The RCT was approved by the NHS Health Research Authority (20/SW/0098) and registered at the ISRCTN Registry (ISRCTN18107122). The intervention phase ran from September 7 to October 5, 2020, during the start of the second wave of the COVID-19 pandemic in Scotland.

Design

A mixed factorial repeated measures design was used, with full randomization to 3 parallel groups.

Randomization

A research assistant not involved in the RCT randomized participants after baseline using computerized simple randomization. Allocation was either to the MPS, NHSWBP, or WL group. Participants received advice of their group assignment by email. Participants were blinded to which intervention they received by styling the 2 interventions and communications to participants similarly. Participants downloaded the same app from the iTunes/App Store/Play Store, and a code was sent back to them to initiate the intervention that they received.

Interventions

My Possible Self

MPS (version 2.0.0) is a tried and tested, NHS-approved [22] smartphone well-being app with a validated track record of showing significant improvements in depression, anxiety, and stress in its users over a short period [19]. It is fully automated and freely available to NHS staff. This intervention has modules that cover a variety of topics and can be accessed in any order, including coping effectively with depression and anxiety, enhancing happiness, improving sleep quality, and practicing mindfulness.

NHS Highland Staff Wellbeing Project

The NHSWBP is a PPI-informed, brief, fully automated, and context- (COVID-19) and population-specific (NHSH frontline staff) digital psychological intervention (smartphone app) based on the MPS. It utilizes the tried and tested cognitive behavioral therapeutic (CBT) and positive psychological techniques delivered via the MPS [19] smartphone app’s modules. There were a number of ways in which the NHSWBP app differed from the MPS app. First, the NHSWBP was presented as a coherent narrative with a fictional character, a Scottish nurse named Iona, who guided participants through the linear narrative of the app and its interventions. Participants also received automated text messages from Iona to engage them in the overall narrative and to motivate continued engagement with the intervention. Second, the NHSWBP was designed following PPI feedback, which included input about which MPS modules were most relevant, the duration of the modules, and the coherence and flow of the presentation format. Third, the NHSWBP provided links to local and national 24-hour support services. Similarly to the MPS, participants were able to monitor and record their mood and levels of distress or well-being, add notes, and identify and record triggers for low mood and anxiety. The intervention lasted for 4 weeks and consisted of 2 parts: part 1 (duration 2 weeks) focused on increasing participants’
happiness, resilience, and well-being, and part 2 (duration 2 weeks) focused on managing low mood and anxiety effectively. The NHSWBP was designed using the MPS app platform and participant communication system, owing to its established track record and NHS approval.

**Primary Outcomes**
Postintervention was the primary timepoint for all outcomes.

**Depression**
The Patient Health Questionnaire (PHQ-9) [23] was used to measure depression. The 9 items ask participants to consider how bothered they have been over the past 2 weeks according to each statement (eg, “feeling tired or having little energy”). The questionnaire score ranges from 0 to 27; each question is given a 4-point response (0=not at all to 3=nearly every day). The questionnaire has demonstrated diagnostic validity [23]. This measure has been used extensively in the U.K. [24] and internationally [25] to measure levels of depression in various population settings during the COVID-19 pandemic.

**Anxiety**
The 7-item General Anxiety Disorder (GAD-7) [26] scale was used to measure anxiety. Similar to the PHQ-9, each item asks the respondent to consider the statement based on how much they have been bothered over a 2-week period (eg, “feeling nervous anxious or on edge”). Each item is scaled from 0 (not at all) to 3 (nearly every day), with a total score range of 0-21. A number of studies during the COVID-19 pandemic have used the GAD-7 to measure levels of anxiety in various UK and international population settings, including in frontline staff working through this pandemic [8,24].

**Mental Well-being**
Mental well-being was measured using the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) [27]. The scale consists of 14 items used to measure subjective well-being and psychological functioning. The wording of each item is positive and aimed to address positive aspects of MH. Responses are completed using a 5-point scale (1=none of the time to 5=all of the time); the total score ranges from 14 to 70. The WEMWBS has been validated for use in the U.K. [27] and has been used internationally [28] and in the U.K. [29] to measure the MWB of HSCWs during this pandemic.

**Secondary Outcomes**

**Mental Toughness**
The Mental Toughness Index (MTI) [20] was used to measure MT. The 8 items (eg, “I can find a positive in most situations”) are rated using a 7-point response format (1=false, 100% of the time, to 7=true, 100% of the time), with responses combined for a total MT score. Studies involving samples from different countries (eg, Australia, South Africa) [20,30-32] have added evidence that supports the construct validity (eg, convergent, criterion) of the MTI. In prior studies, internal consistency reliability estimates for the MTI have been ≥0.87 [30,31,33].

**Gratitude**
Participants completed the Gratitude Questionnaire-6 (GQ-6) [34], which is a 6-item measure of dispositional gratitude. Items (eg, “I have so much in life to be thankful for”) are rated on a 7-point response format (1=strongly disagree to 7=strongly agree), 2 of which are reverse-scored. Evidence from studies involving diverse samples [34-36] supports the factorial validity of the GQ-6 as a measure of the grateful disposition that is conceptually distinct from related constructs (eg, hope, optimism). Internal consistency reliability values reported in previous research have been ≥0.82 [34-36].

**Statistical Analyses**
Statistical analyses and data manipulations were implemented using R (R Core Team) [37]. Baseline characteristics of participants randomly allocated to the 3 intervention groups were compared using the chi-square test. The effects of the MPS and NHSWBP interventions on psychological measures were examined using intention-to-treat (ITT) analyses that included data from all participants who completed the baseline assessment and any follow-up assessment. No imputation was used for missing data. Standard regression models assume independent observations. To adequately account for the dependencies in the data, we adopted the linear mixed modeling (LMM) approach [38] for the analyses of the data. This approach is appropriate for studying the relationships and sources of variation in the data set. It uses all available data and efficiently handles missing data, thereby avoiding listwise deletion. LMM models all sources of variation in the data and avoids the need for data imputation. Each psychological outcome was modeled as a function of time, treatment group, and their interaction and adjusting for random effects due to individual differences and repeated observations from each participant. The models allow for each participant to have a different trajectory. Model parameters were estimated using the restricted maximum likelihood. The best model was selected using the likelihood ratio test. Based on the chosen model, marginal means were estimated and multiple comparisons of groups by time interaction tests conducted using sets of Tukey-adjusted interaction contrasts [39]; degrees of freedom were calculated using the Kenward-Roger test [40].

The effects were tested at a significance level of .05, adjusted depending on the number of contrasts in multiple tests. Cohen d was calculated by standardizing the mean difference of within and between groups using the square root of the sum of all the variance components from the mixed models. This is to adequately represent the study design and account for all sources of variation in data [41,42].

Linear regression slopes of each psychological measure were modeled as a function of time, treatment, and time-treatment interaction. Pairs of the slopes were then compared using the lsmeans approach of Lenth (2016) to determine the intervention that brought about a higher rate of change in the mean of the psychological measures [43]. This analysis used data for the 3 time periods and modeled the average trend for each of the measured outcomes. A second analysis adjusted for the baseline by entering the baseline values of the outcome of interest as a
covariate in the mixed effect model that also included group-time intervention as a fixed effect.

**Results**

**Randomization and Study Attrition**

Details of enrolment into the trial, organized according to the Consolidated Standards of Reporting Trials (CONSORT) guidelines [44], are shown in Figure 1. Of the 225 people who expressed an interest in the study, completed eligibility screening information, and provided consent to participate, 54 (24%) did not complete the baseline questionnaire and 2 declined to participate (0.9%). These 56 individuals were excluded from the analyses, leaving a study sample of 169 (75.1%) participants.

The distribution of participant characteristics at baseline and postintervention is reported in Table 1. Participants were mostly female (n=149, 88.2%) and nurses (n=48, 28.4%), doctors (n=39, 23.1%), allied health professionals (n=21, 12.4%), administrative staff (n=16, 9.5%), health care assistants (n=8, 4.7%), carers (n=6, 3.6%), and other HSWCs (n=31, 18.3%).

At baseline, 53 (31.4%) of the 169 participants met the criteria for low mental well-being (WEMWBS score<40), 51 (30.2%) met the criteria for possible depression (PHQ-9>10), and 46 (27.2%) met the criteria for possible anxiety (GAD-7≥10).

The 3 groups did not differ on the professional, demographic, and clinical history variables assessed at baseline (P>.05). The rate of attrition for the total sample was 23.0% in the middle of the intervention and 36.7% at postintervention assessment. Rates of attrition across the demographic, professional, and clinical characteristics of the participants are presented in Table 1. Participants who worked fewer hours per week and were not employed in administrative positions (eg, doctors) were more likely to drop out during the intervention (all P≤.04).

Postintervention, attrition in the MPS (47.1%) and NHSWBP (43.3%) groups was higher (χ²=9.89, P=.01) compared to the WL group (20.7%). In all 3 treatment conditions, there was little evidence of baseline differences on the demographic, employment, and clinical history variables between participants who were retained and those who dropped out of the study (all P>.05).
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Baseline (N=169), n (%)</th>
<th>Postintervention (N=107), n (%)</th>
<th>Attrition rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>149 (88.2)</td>
<td>94 (87.9)</td>
<td>0.37</td>
</tr>
<tr>
<td>Male</td>
<td>20 (11.8)</td>
<td>13 (12.1)</td>
<td>0.35</td>
</tr>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18-25</td>
<td>4 (2.4)</td>
<td>4 (3.7)</td>
<td>N/A</td>
</tr>
<tr>
<td>26-30</td>
<td>10 (5.9)</td>
<td>4 (3.7)</td>
<td>0.60</td>
</tr>
<tr>
<td>31-40</td>
<td>31 (18.3)</td>
<td>16 (15.0)</td>
<td>0.48</td>
</tr>
<tr>
<td>&gt;40</td>
<td>124 (73.4)</td>
<td>83 (77.6)</td>
<td>0.33</td>
</tr>
<tr>
<td>Education level</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Undergraduate or lower</td>
<td>65 (38.5)</td>
<td>40 (37.4)</td>
<td>0.38</td>
</tr>
<tr>
<td>Postgraduate or higher</td>
<td>104 (61.5)</td>
<td>67 (62.6)</td>
<td>0.36</td>
</tr>
<tr>
<td>Type of employment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nurse</td>
<td>48 (28.4)</td>
<td>30 (28.0)</td>
<td>0.38</td>
</tr>
<tr>
<td>Doctor</td>
<td>39 (23.1)</td>
<td>21 (19.6)</td>
<td>0.46</td>
</tr>
<tr>
<td>Allied health professional</td>
<td>21 (12.4)</td>
<td>12 (11.2)</td>
<td>0.43</td>
</tr>
<tr>
<td>Administrative</td>
<td>16 (9.5)</td>
<td>13 (12.1)</td>
<td>0.19</td>
</tr>
<tr>
<td>Carer</td>
<td>6 (3.6)</td>
<td>4 (3.7)</td>
<td>0.33</td>
</tr>
<tr>
<td>Health care assistant</td>
<td>8 (4.7)</td>
<td>4 (3.7)</td>
<td>0.50</td>
</tr>
<tr>
<td>Other</td>
<td>31 (18.3)</td>
<td>23 (21.5)</td>
<td>0.26</td>
</tr>
<tr>
<td>Years of employment experience</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;2</td>
<td>14 (8.3)</td>
<td>11 (10.5)</td>
<td>0.21</td>
</tr>
<tr>
<td>2-5</td>
<td>13 (7.8)</td>
<td>8 (7.6)</td>
<td>0.38</td>
</tr>
<tr>
<td>5-10</td>
<td>21 (12.6)</td>
<td>10 (9.5)</td>
<td>0.52</td>
</tr>
<tr>
<td>&gt;10</td>
<td>119 (71.3)</td>
<td>76 (72.4)</td>
<td>0.36</td>
</tr>
<tr>
<td>Workplace</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Community, GP&lt;sup&gt;b&lt;/sup&gt;, and PC&lt;sup&gt;c&lt;/sup&gt;</td>
<td>73 (43.7)</td>
<td>41 (39.0)</td>
<td>0.44</td>
</tr>
<tr>
<td>Hospital</td>
<td>74 (44.3)</td>
<td>51 (48.6)</td>
<td>0.31</td>
</tr>
<tr>
<td>Other</td>
<td>20 (12.0)</td>
<td>14 (13.3)</td>
<td>0.30</td>
</tr>
<tr>
<td>N/A</td>
<td>N/A</td>
<td>1 (1.0)</td>
<td>N/A</td>
</tr>
<tr>
<td>Hours worked/week</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20</td>
<td>8 (4.7)</td>
<td>3 (2.8)</td>
<td>0.62</td>
</tr>
<tr>
<td>20-30</td>
<td>31 (18.3)</td>
<td>18 (16.8)</td>
<td>0.42</td>
</tr>
<tr>
<td>30-40</td>
<td>100 (59.2)</td>
<td>61 (57.0)</td>
<td>0.39</td>
</tr>
<tr>
<td>&gt;40</td>
<td>30 (17.8)</td>
<td>25 (23.4)</td>
<td>0.17</td>
</tr>
<tr>
<td>Work with COVID-19 patients</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>129 (77.2)</td>
<td>84 (80.0)</td>
<td>0.35</td>
</tr>
<tr>
<td>Yes</td>
<td>38 (22.8)</td>
<td>21 (20.0)</td>
<td>0.45</td>
</tr>
<tr>
<td>Level of disruption</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No disruption</td>
<td>3 (1.8)</td>
<td>1 (0.9)</td>
<td>0.67</td>
</tr>
<tr>
<td>Minor</td>
<td>15 (8.9)</td>
<td>10 (9.3)</td>
<td>0.33</td>
</tr>
<tr>
<td>Moderate</td>
<td>65 (38.0)</td>
<td>41 (38.3)</td>
<td>0.37</td>
</tr>
<tr>
<td>Characteristic</td>
<td>Baseline (N=169), n (%)</td>
<td>Postintervention (N=107), n (%)</td>
<td>Attrition rate</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>-------------------------</td>
<td>----------------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Major</td>
<td>66 (39.0)</td>
<td>40 (37.4)</td>
<td>0.39</td>
</tr>
<tr>
<td>Severe</td>
<td>20 (12.0)</td>
<td>15 (14.0)</td>
<td>0.25</td>
</tr>
<tr>
<td><strong>Shielding</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>145 (85.8)</td>
<td>93 (86.9)</td>
<td>0.36</td>
</tr>
<tr>
<td>Yes</td>
<td>7 (4.1)</td>
<td>5 (4.7)</td>
<td>0.29</td>
</tr>
<tr>
<td>Family member is shielding</td>
<td>17 (10.1)</td>
<td>9 (8.4)</td>
<td>0.47</td>
</tr>
<tr>
<td><strong>Psychiatric disorder</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>131 (77.5)</td>
<td>78 (72.9)</td>
<td>0.40</td>
</tr>
<tr>
<td>Yes</td>
<td>38 (22.5)</td>
<td>29 (27.1)</td>
<td>0.23</td>
</tr>
</tbody>
</table>

\(a\) N/A: not applicable.
\(b\) GP: general physician.
\(c\) PC: primary care.

**Outcomes**

Table 2 reports the observed mean scores for each outcome at baseline, midintervention, and postintervention in the 3 treatment groups. **Figure 2** depicts these scores for the 3 groups on the primary and secondary outcome measures at baseline, midintervention, and postintervention.

**Table 2.** Descriptive statistics for outcomes at baseline, midintervention, and postintervention in each treatment condition.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>MPS(^a), mean (SD)</th>
<th>NHSWBP(^b), mean (SD)</th>
<th>WL(^c), mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Anxiety</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>7.16 (5.60)</td>
<td>7.77 (4.87)</td>
<td>7.43 (5.10)</td>
</tr>
<tr>
<td>Midintervention</td>
<td>6.45 (5.03)</td>
<td>6.74 (4.69)</td>
<td>7.35 (5.23)</td>
</tr>
<tr>
<td>Postintervention</td>
<td>6.89 (5.71)</td>
<td>5.85 (3.66)</td>
<td>6.72 (5.59)</td>
</tr>
<tr>
<td><strong>Depression</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>6.76 (5.04)</td>
<td>7.60 (4.31)</td>
<td>7.80 (5.23)</td>
</tr>
<tr>
<td>Midintervention</td>
<td>5.74 (4.31)</td>
<td>7.23 (5.47)</td>
<td>8.00 (5.06)</td>
</tr>
<tr>
<td>Postintervention</td>
<td>5.18 (3.27)</td>
<td>5.68 (4.39)</td>
<td>7.56 (6.26)</td>
</tr>
<tr>
<td><strong>Mental well-being</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>47.5 (10.2)</td>
<td>45.3 (8.65)</td>
<td>44.3 (10.1)</td>
</tr>
<tr>
<td>Midintervention</td>
<td>50.3 (9.75)</td>
<td>46.9 (8.68)</td>
<td>44.8 (10.4)</td>
</tr>
<tr>
<td>Postintervention</td>
<td>48.7 (10.1)</td>
<td>48.2 (7.38)</td>
<td>46.1 (11.1)</td>
</tr>
<tr>
<td><strong>MT(^d)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>40.7 (8.04)</td>
<td>39.3 (6.84)</td>
<td>37.9 (9.81)</td>
</tr>
<tr>
<td>Midintervention</td>
<td>40.7 (9.10)</td>
<td>39.3 (9.55)</td>
<td>36.8 (9.20)</td>
</tr>
<tr>
<td>Postintervention</td>
<td>39.7 (9.80)</td>
<td>41.3 (8.33)</td>
<td>39.0 (10.5)</td>
</tr>
<tr>
<td><strong>Gratitude</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline</td>
<td>27.3 (3.46)</td>
<td>26.2 (3.35)</td>
<td>26.7 (3.73)</td>
</tr>
<tr>
<td>Midintervention</td>
<td>27.9 (3.63)</td>
<td>27.1 (4.14)</td>
<td>26.2 (4.30)</td>
</tr>
<tr>
<td>Postintervention</td>
<td>28.2 (4.23)</td>
<td>27.1 (4.24)</td>
<td>27.2 (3.72)</td>
</tr>
</tbody>
</table>

\(^a\) MPS: My Possible Self.
\(^b\) NHSWBP: National Health Service Highland Staff Wellbeing Project.
\(^c\) WL: waitlist.
\(^d\) MT: mental toughness.
Mean scores across all outcomes indicated higher levels of functioning in the MPS group compared to the NHSWBP and WL groups at baseline. The NHSWBP group saw the largest increase in mental well-being scores across groups between baseline (mean 45.3 [SD 8.65]) and postintervention (mean 48.2 [SD 7.38]).

Levels of depression decreased from baseline to midintervention for both MPS and NHSWBP groups, while the WL group in contrast saw a rise in depression scores from baseline (mean 7.80 [SD 5.23]) to midintervention (mean 8.00 [SD 5.06]). Mean levels of depression for the MPS and NHSWBP groups continued to decrease over time, with the NHSWBP group showing the largest decrease from 7.60 (SD 4.31) at baseline to 5.68 (SD 4.39) postintervention. The WL group showed a slight decrease in levels of depression at the postintervention measurement (mean 7.56 [SD 6.26]).

Levels of anxiety decreased across all groups from baseline to postintervention. The WL group showed a consistent decrease in anxiety levels from baseline (mean 7.43 [SD 5.10]) to midintervention (mean 7.35 [SD 5.23]) to postintervention (mean 6.72 [SD 5.59]). The MPS group indicated a decrease in levels of anxiety from baseline (mean 7.16 [SD 5.60]) to midintervention (mean 6.45 [SD 5.03]) and a slight increase postintervention when compared to midintervention (mean 6.89 [SD 5.71]). Baseline levels of anxiety were highest in the NHSWBP group (mean 7.77 [SD 4.87]), and this group also evidenced the greatest decrease in anxiety levels postintervention (mean 5.68 [SD 4.39]).

Levels of gratitude in the MPS group increased from baseline to midintervention to postintervention. Mean gratitude scores in the NHSWBP group increased from baseline to midintervention and then remained constant postintervention. The WL group saw a slight decrease in levels of gratitude from baseline (mean 26.7 [SD 3.73]) to midintervention (mean 26.2 [SD 4.3]), with an increase noted postintervention (mean 27.2 [SD 3.72]).

Levels of MT in the NHSWBP group remained constant from baseline (mean 39.3 [SD 6.84]) to midintervention (mean 39.3 [SD 9.55]), before increasing slightly postintervention to 41.3 (SD 8.33). For the MPS group, these levels also remained constant from baseline (mean 40.7 [SD 8.04]) to midintervention (mean 40.7 [SD 9.10]), before decreasing postintervention (mean 39.7 [SD 9.80]). For the WL group, MT levels decreased from baseline (mean 37.9 [SD 9.81]) to midintervention (mean 36.8 [SD 9.20]), before increasing to 39.0 (SD 10.5) postintervention.

**Figure 2.** Effect size plot for the 3 conditions on the primary and secondary outcome measures at baseline, midintervention, and postintervention. MPS: My Possible Self; NHSWBP: National Health Service Highland Staff Wellbeing Project; WL: waitlist.
Standardized Effect Size

The between- and within-group effect sizes (standardized mean difference) on the primary and secondary outcomes calculated using observed means are presented in Tables 3 and 4. Postintervention, between-group effect sizes were small to medium for the primary (NHSWBP vs MPS Cohen $d=0.19$ to $–0.20$; WL vs MPS Cohen $d=–0.04$ to $0.36$; WL vs NHSWBP Cohen $d=0.06$ to $–0.18$) and secondary outcome measures. The results showed a consistent pattern of greater improvements in depression, anxiety, well-being, MT, and gratitude among participants in the digital intervention groups (MPS and NHSWBP) postintervention compared to the WL group.

Postintervention, a small difference was noted between the WL and the 2 treatment groups on anxiety (vs MPS: Cohen $d=0.07$, 95% CI $–0.20$ to $0.33$; vs NHSWBP: Cohen $d=0.06$, 95% CI $–0.19$ to $0.31$), depression (vs MPS: Cohen $d=0.37$, 95% CI $0.07$ to $0.66$; vs NHSWBP: Cohen $d=0.18$, 95% CI $–0.11$ to $0.46$), and mental well-being (vs MPS: Cohen $d=–0.04$, 95% CI $–0.62$ to $–0.08$; vs NHSWBP: Cohen $d=–0.15$, 95% CI $–0.41$ to $0.10$). The NHSWBP group generally had larger within-group effects than the other groups. Within-group effects for both MPS and NHSWBP groups ranged from small to medium based on observed means (MPS Cohen $d=–0.31$ to $0.25$, NHSWBP Cohen $d=–0.38$ to $0.24$). For the WL group, within-group effects were generally small for the primary outcomes (Cohen $d=–0.12$ to $0.16$) and small to medium for the secondary outcome measures (Cohen $d=0.13$ to $0.27$).

### Table 3. Between-group effects calculated using observed means.

<table>
<thead>
<tr>
<th>Between-group effects</th>
<th>Anxiety, Cohen $d$ (95% CI)</th>
<th>Depression, Cohen $d$ (95% CI)</th>
<th>Mental well-being, Cohen $d$ (95% CI)</th>
<th>MT, Cohen $d$ (95% CI)</th>
<th>Gratitude, Cohen $d$ (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NHSWBP vs MPS$^b$</td>
<td>0.01 (–0.26 to 0.28)</td>
<td>0.19 (–0.12 to 0.50)</td>
<td>–0.20 (–0.48 to 0.08)</td>
<td>–0.07 (–0.41 to 0.27)</td>
<td>–0.26 (–0.57 to 0.06)</td>
</tr>
<tr>
<td>WL vs MPS$^c$</td>
<td>0.07 (–0.20 to 0.33)</td>
<td>0.37 (0.07–0.66)</td>
<td>–0.04 (–0.62 to –0.08)</td>
<td>–0.31 (–0.64 to 0.02)</td>
<td>–0.28 (–0.58 to 0.02)</td>
</tr>
<tr>
<td>WL vs NHSWBP</td>
<td>0.06 (–0.19 to 0.31)</td>
<td>0.18 (–0.11 to 0.46)</td>
<td>–0.15 (–0.41 to 0.10)</td>
<td>–0.24 (–0.55 to 0.07)</td>
<td>–0.02 (–0.32 to 0.27)</td>
</tr>
</tbody>
</table>

$^a$ MT: mental toughness.

$^b$ NHSWBP: National Health Service Highland Staff Wellbeing Project.

$^c$ MPS: My Possible Self.

$^d$ WL: waitlist.

### Table 4. Within-group effects calculated using observed means.

<table>
<thead>
<tr>
<th>Within-groups effects</th>
<th>Anxiety, Cohen $d$ (95% CI)</th>
<th>Depression, Cohen $d$ (95% CI)</th>
<th>Mental well-being, Cohen $d$ (95% CI)</th>
<th>MT, Cohen $d$ (95% CI)</th>
<th>Gratitude, Cohen $d$ (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MPS$^b$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postintervention vs. baseline</td>
<td>–0.05 (–0.72 to 0.63)</td>
<td>–0.31 (–1.08 to 0.46)</td>
<td>0.11 (–0.58 to 0.80)</td>
<td>0.13 (–0.97 to 0.72)</td>
<td>0.25 (–0.54 to 1.04)</td>
</tr>
<tr>
<td>Postintervention vs. midintervention</td>
<td>0.07 (–0.60 to 0.75)</td>
<td>–0.11 (–0.87 to 0.66)</td>
<td>–0.14 (–0.83 to 0.55)</td>
<td>0.13 (–0.97 to 0.72)</td>
<td>0.08 (–0.70 to 0.87)</td>
</tr>
<tr>
<td>NHSWBP$^c$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postintervention vs. baseline</td>
<td>–0.32 (–0.94 to 0.29)</td>
<td>–0.38 (–1.08 to .32)</td>
<td>0.27 (–0.36 to 0.90)</td>
<td>0.24 (–0.53 to 1.01)</td>
<td>0.22 (–0.50 to 0.94)</td>
</tr>
<tr>
<td>Postintervention vs. midintervention</td>
<td>–0.15 (–0.77 to 0.47)</td>
<td>–0.30 (–1.00 to 0.39)</td>
<td>0.12 (–0.51 to 0.75)</td>
<td>0.24 (–0.53 to 1.01)</td>
<td>–0.01 (–0.73 to 0.70)</td>
</tr>
<tr>
<td>WL$^d$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postintervention vs. baseline</td>
<td>–0.12 (–0.67 to 0.43)</td>
<td>–0.05 (–0.67 to 0.58)</td>
<td>0.16 (–0.40 to 0.72)</td>
<td>0.13 (–0.56 to 0.81)</td>
<td>0.13 (–0.52 to 0.77)</td>
</tr>
<tr>
<td>Postintervention vs. midintervention</td>
<td>–0.11 (–0.66 to 0.45)</td>
<td>–0.09 (–0.71 to 0.54)</td>
<td>0.12 (–0.44 to 0.68)</td>
<td>0.25 (–0.44 to 0.94)</td>
<td>0.27 (–0.37 to 0.91)</td>
</tr>
</tbody>
</table>

$^a$ MT: mental toughness.

$^b$ MPS: My Possible Self.

$^c$ NHSWBP: National Health Service Highland Staff Wellbeing Project.

$^d$ WL: waitlist.

Comparing the Rate of Change per Condition

Table 5 shows the rate of change observed due to the interventions by comparing the trends in the effect size plot. When the gradient of the slopes of linear regression of each psychological outcome was estimated as a function of time, treatment, and time-treatment interaction, each group demonstrated improvements in average scores on all the 3 outcomes over the study period. Although the test for differences between the slopes of each group did not reach statistical significance ($P>.05$), the rate of improvement in anxiety, depression, and mental well-being was largest among those in...
the NHSWBP group. The WL group evidenced the smallest rate of change on each of the 3 outcomes.

For the secondary outcome measures, the average scores for MT increased in the NHSWBP and WL groups, whereas a slight decline was found in the MPS group. The rate of increase in MT was greatest in the NHSWBP group. We also observed the average scores for gratitude to increase for all 3 groups, with the greatest rate of increase in the MPS group. The smallest rate of increase in gratitude was found in the WL group.

Table 5. Response trends comparing baseline to postintervention in each treatment condition.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Effect estimate (SE)</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anxiety</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPS</td>
<td>-0.05 (0.66)</td>
<td>-1.35 to 1.26</td>
</tr>
<tr>
<td>NHSWBP</td>
<td>-0.79 (0.61)</td>
<td>-1.98 to 0.41</td>
</tr>
<tr>
<td>WL</td>
<td>-0.35 (0.55)</td>
<td>-1.42 to 0.72</td>
</tr>
<tr>
<td>Depression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPS</td>
<td>-0.76 (0.62)</td>
<td>-1.99 to 0.48</td>
</tr>
<tr>
<td>NHSWBP</td>
<td>-0.94 (0.57)</td>
<td>-2.06 to 0.18</td>
</tr>
<tr>
<td>WL</td>
<td>-0.18 (0.51)</td>
<td>-1.19 to 0.83</td>
</tr>
<tr>
<td>Mental well-being</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPS</td>
<td>1.11 (1.22)</td>
<td>-1.29 to 3.50</td>
</tr>
<tr>
<td>NHSWBP</td>
<td>1.62 (1.11)</td>
<td>-0.57 to 3.81</td>
</tr>
<tr>
<td>WL</td>
<td>0.88 (0.99)</td>
<td>-1.07 to 2.83</td>
</tr>
<tr>
<td>MT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPS</td>
<td>-0.54 (1.15)</td>
<td>-2.80 to 1.71</td>
</tr>
<tr>
<td>NHSWBP</td>
<td>0.97 (1.04)</td>
<td>-1.08 to 3.02</td>
</tr>
<tr>
<td>WL</td>
<td>0.52 (0.94)</td>
<td>-1.33 to 2.37</td>
</tr>
<tr>
<td>Gratitude</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MPS</td>
<td>0.59 (0.50)</td>
<td>-0.39 to 1.57</td>
</tr>
<tr>
<td>NHSWBP</td>
<td>0.39 (0.45)</td>
<td>-0.49 to 1.28</td>
</tr>
<tr>
<td>WL</td>
<td>0.23 (0.40)</td>
<td>-0.56 to 0.84</td>
</tr>
</tbody>
</table>

aMPS: My Possible Self.  
bNHSWBP: National Health Service Highland Staff Wellbeing Project.  
cWL: waitlist.  
dMT: mental toughness.

Program Adherence

Adherence, defined as the extent to which participants engaged with the intervention, was examined for both the NHSWBP and MPS groups with respect to average interactions per user. Adherence was deemed to be good for both digital interventions, with participants in the NHSWBP group interacting, on average, 37.4 times with the intervention (more than once per day, on average) during the month-long intervention, while those in the MPS group interacted, on average, 37.5 times. None of the adherence indices correlated with demographic, clinical history, and primary and secondary outcome data obtained at baseline. No harmful or unintended effects were reported by the participants.

Post hoc Power Calculation

Instead of using the observed effect size to calculate the post hoc study power (which could introduce bias) [45], we used the observed sample size and a fixed threshold for power and significance and calculated the smallest effect size that could be reliably detected with our sample size. By using this approach, together with our study design, we found that our study could detect an effect size of at least $f=0.27$ at 80% power.

Discussion

Principal Findings

We conducted a novel pilot RCT to evaluate 2 brief, fully automated digital health interventions in a sample of frontline staff working through the second wave of the COVID-19 pandemic. The trial proceeded successfully during challenging circumstances in the shadow of the second wave of the COVID-19 pandemic in the U.K. Our low-cost study demonstrated that it was possible to recruit 169 people working in a small NHS board within a short duration and deliver a technically innovative intervention on a modest financial budget.
The NHSWBP app was designed with end-user (PPI) input and worked well throughout, with good adherence and no major flaws or bugs, nor evidence of harm reported by the participants. Furthermore, the WL control design was effective at retaining participants (who otherwise might have lost interest in the study and dropped out if it was just a no-treatment control rather than a WL). We also accumulated rich background data that could assist in identifying the possible drivers of dropout, which could be used to modify the design of the intervention to improve retention in a larger future trial. Although this was a pilot trial that we conducted during a prescribed limited time with a relatively small sample size, the findings of this study provide encouraging results for future full trials of digital psychological interventions that are designed to support the MH and PWB of HSCWs who are working under conditions of extreme stress.

There are 3 key findings of interest in this study. First, the primary outcomes investigated showed decreases in levels of depression (NHSWBP and MPS groups) and anxiety (NHSWBP group) when compared to the WL group. The rate of decrease in depression and anxiety symptoms was the greatest among those exposed to the NHSWBP intervention. Our results also indicate that the individuals exposed to the digital interventions and WL conditions experienced an increase in mental well-being, with the rate of increase again shown to be the greatest for those exposed to the NHSWBP intervention.

Second, for the secondary outcomes investigated, our results showed increases in MT for the NHSWBP and WL groups, with the rate of increase in MT again being the greatest for those exposed to the NHSWBP intervention. All groups experienced an increase in gratitude over the treatment period, with the rate of change being the greatest for those exposed to the MPS intervention. Overall, our results show greater rates of symptom improvements in the digital intervention groups than in the WL condition. Concerning gratitude, our efforts add on the new research direction that investigates how it can be enhanced to be favorable to MH outcomes [46]. With regard to MT, our study adds to a growing field suggesting that MT could buffer the negative effects of depression and anxiety and promote adaptive MH outcomes [47]. Our research adds an interventional design and avoids the commonly used cross-sectional designs used to investigate gratitude [46] and MT [48]. Our pilot intervention showed promise in terms of both the traditional view of clinical psychology (ie, anxiety, depression) and also with regard to the science of positive psychology and character strengths [49].

Third, our results also provide preliminary support for the development or modification of digital interventions to be context specific, as of the 2 interventions tested, the NHSWBP showed greater rates of symptom improvement. Future trials assessing context-specific digital interventions for specialized populations in larger samples are warranted, as there is good reason to believe those larger studies will demonstrate efficacy. The digital nature of these interventions was seen to be safe, cost effective, and rapidly modifiable to context. The future application of similar, context-specific, robustly tested interventions could be scalable to other contexts with MH human resource needs [50].

**Limitations and Future Research Directions**

There are several limitations of this study that need to be acknowledged. First, participants included a small sample of HSCWs from a single NHS site. Although the majority of respondents were female, this does not differ dramatically from the gender composition of the whole HSCW workforce in NHS Highland [51]. As our objective was to gather preliminary evidence on the potential benefits of 2 digital interventions in this population, the study was not powered as an efficacy trial, and so CIs around estimated effects were wide (indicating the small sample may have contributed to statistical uncertainty) and the findings may not be generalizable to other populations and HSCWs living in other contexts. Second, the treatment period was restricted to 4 weeks, and it is possible that changes in MH and PWB require more engagement in the intervention materials. In addition, some outcomes may change more gradually and require a longer period to improve. For example, gratitude exercises can orientate a person to experience more grateful emotions, but it could take more than 4 weeks for changes in dispositional gratitude to emerge. Future research would do well to track and monitor whether gains that are made during treatment are maintained or change over time. Third, the MPS app was publicly available for download throughout the duration of our study, and participants were not restricted to use other modalities or medications during this pilot, which raises the possibility that treatment effects might be cross-contaminated. Fourth, the attrition rate postintervention was 36.7%. The dropout rate was lowest in the WL group, which is likely attributable to participants waiting to receive either of the digital interventions. Although we did not find any substantial evidence of attrition bias, it is possible that participants who dropped out from the intervention groups were less satisfied with the program or experienced less than positive outcomes. Additional research is needed to explore the mechanisms underlying the effects that emerged in this study and to identify the relative contributions of the components that constituted each of the digital interventions. There may also be value in taking a broader approach to outcome assessment by examining other domains of well-being that extend beyond the domain of PWB. For example, previous research along these lines has reported postintervention improvements in social relationships [52].

**Conclusion**

The results of this study provide preliminary support for efforts to invest in refining of existing digital interventions for specialized populations and assessing their efficacy in larger samples, as there is good reason to believe that larger studies will demonstrate efficacy. Robust testing of efficacious digital interventions could allow for rapid, economical, safe, and large-scale dissemination of urgently needed psychological support for frontline staff who are working through the COVID-19 pandemic and its aftermath.
Acknowledgments
The authors thank the Chief Science Office of Scotland for providing the funding for this research. The authors also thank the MPS for being our technical partner in this project.

Authors’ Contributions
JDK, HAL, SJL, and SAM designed the study. HDK and HL designed the intervention. JDK, SAM, SJL, SJ, and KN conducted the study. JDK, BC, RGC, and AS drafted the manuscript. All authors provided revisions, and JE conducted the statistical analysis.

Conflicts of Interest
None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).

References


37. R Core Team. The R Project for Statistical Computing. URL: https://www.r-project.org/ [accessed 2022-03-14]


51. NHS Highland. Welcome to NHS Careers. URL: https://www.nhshighland.scot.nhs.uk/CAREERS/Pages/Home.aspx [accessed 2022-03-14]


Abbreviations

- GAD-7: 7-item General Anxiety Disorder
- GP: general physician
- GQ-6: Gratitude Questionnaire-6
- HSCW: health and social care worker
- LMM: linear mixed modeling
- MH: mental health
- MPS: My Possible Self
- MT: mental toughness
- MTI: Mental Toughness Index
- NHS: National Health Service
- NHSH: NHS Highland
- NHSWBP: NHSH Staff Wellbeing Project
- PHQ-9: Patient Health Questionnaire
- PPI: patient and public involvement
- PWB: psychological well-being
- RCT: randomized controlled trial
- WEMWBS: Warwick-Edinburgh Mental Well-being Scale
- WL: waitlist

Edited by J Torous; submitted 03.10.21; peer-reviewed by D Madhusudhan, J Torous; comments to author 06.01.22; revised version received 13.01.22; accepted 16.01.22; published 04.04.22.

Please cite as:

Brief Digital Interventions to Support the Psychological Well-being of NHS Staff During the COVID-19 Pandemic: 3-Arm Pilot Randomized Controlled Trial
JMIR Ment Health 2022;9(4):e34002
URL: https://mental.jmir.org/2022/4/e34002
doi:10.2196/34002
PMID:35044927
Predicting Uptake of the COVID Coach App Among US Military Veterans: Funnel Analysis Using a Probability-Based Panel

Beth K Jaworski¹, PhD; Katherine Taylor¹, MPH, PsyD; Kelly M Ramsey¹; Adrienne J Heinz¹, PhD; Sarah Steinmetz¹, PhD; Jason E Owen¹, MPH, PhD; Jack Tsai²,³, PhD; Robert H Pietrzak⁴,⁵, MPH, PhD

¹National Center for Posttraumatic Stress Disorder, Dissemination & Training Division, US Department of Veterans Affairs, Menlo Park, CA, United States
²National Center on Homelessness Among Veterans, US Department of Veterans Affairs, Tampa, FL, United States
³School of Public Health, San Antonio Campus, University of Texas Health Science Center at Houston, San Antonio, TX, United States
⁴National Center for Posttraumatic Stress Disorder, VA Connecticut Healthcare System, US Department of Veterans Affairs, New Haven, CT, United States
⁵Department of Psychiatry, Yale School of Medicine, New Haven, CT, United States

Corresponding Author:
Beth K Jaworski, PhD
National Center for Posttraumatic Stress Disorder
Dissemination & Training Division
US Department of Veterans Affairs
795 Willow Road
Menlo Park, CA, 94025
United States
Phone: 1 18312781567
Email: beth.jaworski@nih.gov

Abstract

Background: Although the COVID-19 pandemic has not led to a uniform increase of mental health concerns among older adults, there is evidence to suggest that some older veterans did experience an exacerbation of preexisting mental health conditions, and that mental health difficulties were associated with a lack of social support and increasing numbers of pandemic-related stressors. Mobile mental health apps are scalable, may be a helpful resource for managing stress during the pandemic and beyond, and could potentially provide services that are not accessible due to the pandemic. However, overall comfort with mobile devices and factors influencing the uptake and usage of mobile apps during the pandemic among older veterans are not well known.

COVID Coach is a free, evidence-informed mobile app designed for pandemic-related stress. Public usage data have been evaluated; however, the uptake and usage of the app among older veterans have not been explored.

Objective: The purpose of this study was to characterize smartphone ownership rates among US veterans, identify veteran characteristics associated with downloading and use of COVID Coach, and characterize key content usage within the app.

Methods: Data were analyzed from the 2019-2020 National Health and Resilience in Veterans Study (NHRVS), which surveyed a nationally representative, prospective cohort of 3078 US military veterans before and 1 year into the pandemic. The NHRVS sample was drawn from KnowledgePanel, a research panel of more than 50,000 households maintained by Ipsos, Inc. The median time to complete the survey was nearly 32 minutes. The research version of COVID Coach was offered to all veterans who completed the peri-pandemic follow-up assessment on a mobile device (n=814; weighted 34.2% of total sample). App usage data from all respondents who downloaded the app (n=34; weighted 3.3% of the mobile completers sample) were collected between November 14, 2020, and November 7, 2021.

Results: We found that most US veterans (81.5%) own smartphones, and that veterans with higher education, greater number of adverse childhood experiences, higher extraversion, and greater severity of pandemic-related posttraumatic stress disorder symptoms were more likely to download COVID Coach. Although uptake and usage of COVID Coach were relatively low (3.3% of eligible participants, n=34), 50% of the participants returned to the app for more than 1 day of use. The interactive tools for managing stress were used most frequently.

Conclusions: The COVID-19 pandemic has increased the need for and creation of digital mental health tools. However, these resources may require tailoring for older veteran populations. Future research is needed to better understand how to optimize
digital mental health tools such as apps to ensure uptake and usage among older adults, particularly those who have experienced traumas across the lifespan.

**Introduction**

**Background**

*Mental Health Impact of the COVID-19 Pandemic Among Older Veterans*

The coronavirus pandemic has taken an extraordinary toll on health and well-being globally. In the United States, over 900,000 people have died due to COVID-19. Vaccines against the coronavirus are now available to anyone over the age of 5 years in the United States. However, distribution has been uneven within and across states, and many individuals have expressed unwillingness to be vaccinated [1]. The lack of vaccination prolongs the pandemic and in turn its negative impact on society. In addition to physical health consequences, the pandemic and its mitigation strategies have led to a host of negative mental health consequences, including increased symptoms of anxiety, depression, posttraumatic stress disorder (PTSD), psychological distress, increased substance use to cope with pandemic-related stressors, and increased suicidal ideation [2-5]. Unhealthy alcohol use, increased tobacco and cannabis use, and potential misuse of prescription opioids and benzodiazepines were identified as key areas of clinical concern for mental health providers [6]. Recent evidence suggests that these concerns were well-founded, as drug overdose deaths rose 27.2% between April 2020 and April 2021 [7].

However, research indicates a more nuanced, multifaceted relationship between age and pandemic-related mental health concerns [8]. Researchers have documented a range of pandemic impacts on mental health, including negative outcomes, no change in mental health symptomology, and even some evidence of mental health gains. For example, among a sample of older adults with chronic PTSD, researchers did not find that the pandemic significantly increased adverse mental health outcomes, and in fact found that PTSD symptoms among individuals with a PTSD diagnosis decreased relative to those of a trauma-exposed comparison group [9]. Additionally, some veterans have demonstrated increased posttraumatic growth (PTG) and resilience associated with the pandemic. In a probability-based sample of older US military veterans, Pietrzak and colleagues [10] found that 43.3% of the sample endorsed PTG, and veterans who screened positive for COVID-19–associated PTSD were more likely to endorse all aspects of PTG compared to veterans who screened negative for COVID-19–associated PTSD symptoms.

Although the pandemic may not have led to a uniform increase of mental health concerns among older adults, those experiencing loneliness and social isolation may have been particularly vulnerable to increased psychological distress and an exacerbation of preexisting mental health conditions. Loneliness is often a concern among older adults, particularly among individuals with low socioeconomic status [11], and social isolation and loneliness are associated with increased morbidity and mortality [12], even outside the context of a global pandemic. Older veterans indicated that the pandemic increased loneliness and sorrow due to the isolation and disruption of their ordinary routines [13], and researchers also found that a lack of social support and increasing numbers of pandemic-related stressors were associated with mental health difficulties [14]. In a national sample of US veterans, the pre-to peripandemic prevalence of generalized anxiety disorder increased from 7.1% to 9.4%, with the most pronounced increase observed in veterans aged 45-64 years (8.2% to 13.5%) [15]. Results of this study further indicated that prepandemic loneliness and pandemic-related social stressors were associated with an increase in psychological distress [15]. Although suicide among veterans over the course of the pandemic did not increase, low social support and worsening of social relationships were among the risk factors present for veterans who did develop new-onset suicide ideation during the pandemic [16]. Additionally, at the beginning of the pandemic, it was hypothesized that there might be a mental health crisis among older adults due in part to complications from the difficulty in adopting technologies useful during quarantine (eg, software to facilitate telehealth visits or stay connected with loved ones) and lack of contact with friends, family, and caregivers [8]. Evaluation of video mental health visits among US veterans during the COVID-19 pandemic suggests this may be the case. Researchers found that older patients and those with low socioeconomic status had lower odds of completing greater than 50% of their visits via video compared to in-person visits or phone calls [17]. Taken together, these findings suggest that older veterans with low socioeconomic status, and those experiencing loneliness and social isolation were more likely to face negative mental health consequences of the pandemic. Thus, promoting positive social connections and utilizing effective coping strategies for aging veterans have been identified as key suggestions for helping with pandemic-related stressors [18].

**Potential of Mental Health Apps to Support Older Veterans**

Very early in the pandemic, public health scholars called for prevention and early intervention efforts to help promote individual and population mental health [19]. Digital mental health options have been identified as one possible solution with great potential for the pandemic and beyond [20]. However, an important precursor to the adoption of digital mental health technologies, and particularly mental health apps, is mobile
device ownership. Although older adults in the United States may be somewhat less likely than their younger counterparts to use mobile devices, the majority (61%) own smartphones [21]. Older adults express interest in using mobile devices to support health [22], and in some cases their engagement with digital health products may exceed that of younger adults [23]. In general, prior research suggests that older adults are interested in apps for health, but uptake and usage continue to be relatively low [24]. Among older veterans, many are interested in apps to support health [25], but tend to have mixed opinions about mental health apps based on sociodemographics such as rurality [26]. Connolly and colleagues [17] found that rural veterans are more likely to oppose app usage, describe smartphones as hard to navigate, and cite barriers to usage compared to urban-dwelling populations. However, educational materials and training programs can be successfully implemented to increase older veterans’ comfort with using apps and overcome barriers to utilization [27,28]. With additional efforts to promote comfort with and confidence in mental health apps, they may be a scalable, accessible, and helpful resource for older veterans.

Need for Studying COVID Coach Among Older Veterans

To address the mental health needs of veterans in the wake of the pandemic, the US Department of Veterans Affairs (VA) rapidly implemented several virtual resources and practices to support telemental health [29]. In addition to resources for the delivery of telemental health care, the VA’s National Center for PTSD also created COVID Coach, a free public mental health app designed to help individuals manage stress and anxiety resulting from the COVID-19 pandemic. Mobile and internet-based interventions, including COVID Coach, have been identified as a population-level, primary prevention resource for pandemic-related mental health impacts [30]. The app was created in 6 weeks and released on both the Android and iOS platforms at the end of April 2020. Between the app launch and the end of October 2021, the app has been downloaded over 200,000 times (Android, n=27,082 downloads; iOS, n=188,224 downloads).

The COVID Coach app contains four key content areas: (1) Manage Stress (interactive coping tools); (2) Learn (psychoeducational topics covering ways to stay well, stay balanced, navigate relationships, stay safe, and stay healthy from COVID-19); (3) Mood Check (for tracking personal goals and tracking well-being [Warwick Edinburgh Mental Well-Being Scale], and symptoms of anxiety [Generalized Anxiety Disorder-7 questionnaire, GAD-7], depression [Patient Health Questionnaire-9, PHQ-9], and PTSD [PTSD Checklist for DSM-5, PCL-5]); and (4) Find Resources (a comprehensive repository of resources and supports across a range of topics, including mental health crisis support, ways to meet basic needs, and local information about COVID-19). Screenshots of COVID Coach are provided in Multimedia Appendix 1. COVID Coach has been well-received in the Apple App Store (average star rating=4.8 out of 5; n=871 ratings) and Google Play Store (average star rating=4.8 out 5; n=308 ratings). An initial evaluation of the anonymous public usage data revealed that app users have primarily utilized the “Manage Stress” section of the app (with interactive and audio-guided tools for coping with stress and anxiety), but that collectively, thousands of users have accessed the psychoeducational information, assessments, and resources [31]. However, similar to many other mental health apps, engagement and retention were low (eg, [32]). Additionally, understanding for whom and under what conditions COVID Coach was utilized was not possible because usage data from the publicly available version of the app are completely anonymous.

Several factors may be associated with the decision to download and use mobile mental health apps during the COVID-19 pandemic. These include sociodemographic characteristics such as age and education [33]; personality characteristics such as extraversion and conscientiousness [34]; preexisting mental health difficulties such as trauma exposure [35]; event-specific stressors such as pandemic-related social restriction stress [31]; and characteristics of the app itself, such as usability, perceived utility, and trustworthiness [36]. To date, however, no known study has examined factors associated with US military veterans’ uptake and usage of mobile apps specific to addressing pandemic-related stressors, such as COVID Coach.

Current Study

There are estimated to be 19 million US veterans, which accounts for approximately 10% of the US population [37]. Gulf War–era veterans comprise the largest segment, followed by Vietnam-era veterans [38]. Thus, older veterans represent a significant proportion of the overall veteran population [39], and many are open to exploring the usage of mental health apps [25,26,40]. To explore smartphone ownership as well as the uptake and usage of an app for pandemic-related stress and anxiety specifically among older veterans, this study was guided by the following four aims: (1) understand the relationship between sociodemographic characteristics and survey completion on a mobile device compared to a laptop or desktop computer; (2) identify sociodemographic, prepandemic, and pandemic-related variables associated with downloading COVID Coach; (3) explore differences among veterans who used the app for only 1 day compared to those who returned to the app for 2 or more days; and (4) characterize key content usage within the app.

Methods

Study Design and Participants

Data were analyzed from the 2019–2020 National Health and Resilience in Veterans Study (NHRVS), which surveyed a nationally representative, prospective cohort of US military veterans. The NHRVS sample was drawn from KnowledgePanel, a research panel of more than 50,000 households maintained by Ipsos, Inc. KnowledgePanel is a probability-based, online, nonvolunteer access survey panel of US households. To permit generalizability of the study results to the entire population of US veterans, Ipsos statisticians computed poststratification weights using the benchmark distributions of the following sociodemographic characteristics of US military veterans from the most recent (August 2019) Current Veteran Population Supplemental Survey of the US Census Bureau’s American Community Survey [41]: age, gender, race/ethnicity, Census
Region, metropolitan status, education, household income, branch of service, and years in service. With exception of the app usage data, all percentages and inferential statistics displayed in the Results section reflect weighted percentages, utilizing these poststratification weights.

A total of 4069 veterans completed the prepandemic survey (median completion date November 21, 2019) prior to the first documented COVID-19 cases in the United States and 3078 (75.6%) completed a 1-year peripandemic follow-up assessment (median completion date November 14, 2020). The median completion time was 31.9 minutes (IQR 19.9 minutes). These veterans were 22 to 99 years old (mean age 63.2, SD 14.7 years). The sample was mostly male (91.6%); 79.3% were non-Hispanic Caucasian, 10.3% non-Hispanic African American, 6.0% Hispanic, and 4.4% bi/multiracial or other racial identity. The sample included all branches of the US military (majority Army [47.3%], Navy [20.8%], or Air Force [18.9%]), 35.0% were combat veterans, 79.6% reported having enlisted in the military, and 20.1% reported utilizing VA as their primary source of health care.

A total of 814 (34.2%) of the sample completed the survey on a tablet or smartphone and were eligible to download the research version of the COVID Coach app. On the final page of the online questionnaire, these participants saw a brief description of COVID Coach, as well as a link to download the app (see Figure 1 for the message and format of the invitation). Of this sample, 34 (3.3%) downloaded the COVID Coach app. Relative to veterans who did not download the app, those who did spent significantly more time on the landing page (mean 118.8, SD 107.7 seconds vs mean 33.8, SD 49.7 seconds; \(t_{812}=9.10, P<.001\)), but did not differ with respect to whether they used devices with iOS (52.9% vs 46.5%) or Android (47.1% vs 53.5%) operating systems (\(\chi^2_{1}=0.55, P=.46\)).

Figure 1. Invitation to download COVID Coach Explorer.
Ethical Considerations

All participants provided informed consent and the Human Subjects Committee of the VA Connecticut Healthcare System approved the study. Prior to completing the survey, Ipsos obtains informed consent from all participants to participate in the NHRVS, and participants received 15,000 points (equivalent to US $15) as compensation for their participation in this study.

COVID Coach Explorer and Mobile Analytics Data

The research-enabled version of COVID Coach (COVID Coach Explorer, version 1.0) is available for Android [42] and iOS [43]. COVID Coach Explorer contains the exact same features and content as COVID Coach, which includes interactive, evidence-informed tools for coping with stress and anxiety; information about how to stay well, stay connected, and navigate challenges; self-monitoring mental health symptoms and goals; and resources to discover and connect with various types of verified and vetted support. However, COVID Coach Explorer allows researchers to assign a unique code to each research participant, and the app usage data can only be associated with the participant identifiers by approved members of the research team. The app only collects information about app use such as screens selected, button presses, and other nonidentifying patterns, but these usage data are only linked to the unique alphanumeric code. Fully nonidentifying and encrypted event sequences were stored using JavaScript object notation (JSON) format on a remote GovCloud server that meets VA security and privacy requirements. Data are accessible from VA App Connect software, which has been approved for use under the VA’s Technical Reference Model [44]. Each in-app event contains a timestamp (in Coordinated Universal Time) that corresponds to when the event actually occurred, but data are only transmitted to the server when the app is in use and connected to WiFi or utilizing a data plan.

For the purpose of this study, mobile analytics data with timestamps between November 14, 2020, and November 7, 2021, were extracted from the research server on November 7, 2021. Between November 14, 2020, and November 7, 2021, 1752 in-app related events were captured (Android, n=1063; iOS, n=689) across the 34 participants who downloaded COVID Coach Explorer (Android, n=16; iOS, n=18).

Measures

Survey Assessments

The questionnaire administered to panelists included a range of assessments, including demographic information (eg, age, gender, education, military branch, mobile device ownership), prepandemic psychosocial risk factors (eg, adverse childhood experiences, PTSD symptoms, alcohol use disorder symptoms), changes in psychosocial risk factors from pre- to peripandemic, COVID-19 infection stressors, and COVID-19 pandemic stressors. For a detailed overview of each instrument, please see Table 1.
Table 1. Measures of psychiatric, sociodemographic, military, and psychosocial variables and COVID-19 infection and pandemic stressors.

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sociodemographic characteristics</strong></td>
<td>Age (continuous), sex (male, female), race/ethnicity (white, nonwhite), education (college graduate or higher, up to high school diploma), marital status (married/living with partner, not), household income (US $60,000 or more, less than $60,000), employment status (working, not), region of country (south, west, midwest, northeast), metropolitan versus nonmetropolitan residence</td>
</tr>
<tr>
<td><strong>Military characteristics</strong></td>
<td>Combat veteran status (combat exposure, not), military branch (Army, Navy, Air Force, Marine Corps, other); combat veteran status (yes, no), years of military service</td>
</tr>
<tr>
<td><strong>Prepandemic psychosocial risk factors</strong></td>
<td>Adverse childhood experiences Adverse Childhood Experiences Questionnaire (ACEQ) score [45] Lifetime MDD(^b) or PTSD(^c) Lifetime AUD(^e) or DUD(^f) Number of medical conditions Personality characteristics COVID-19 infection stressors COVID-19 pandemic stressors</td>
</tr>
<tr>
<td>Total traumas</td>
<td>Items endorsed on Life Events Checklist for DSM-5(^a) (LEC-5) [46] Lifetime MDD was assessed according to DSM-5 diagnostic criteria using the Mini International Neuropsychiatric Interview (MINI) [47]. Lifetime PTSD was defined as a score of 33+ [48] on the PCL-5(^d) [49], which was modified to include lifetime ratings of all PTSD symptoms in relation to veterans' self-reported “worst” Criterion A trauma on the LEC-5 [46]. Veterans who met criteria for either disorder were coded positive for lifetime MDD or PTSD Lifetime AUD and DUD were defined as meeting DSM-5 diagnostic criteria for AUD or DUD, respectively, as assessed using the MINI [47]; veterans who met criteria for either disorder were coded positive for lifetime AUD or DUD Sum of number of medical conditions adapted from the Alcohol Use Disorders and Associated Disabilities Interview Schedule [50]: “Has a doctor or health care professional ever told you that you have any of the following medical conditions?” (eg, arthritis, cancer, diabetes, heart disease, asthma, kidney disease); range 0-24 conditions Ten-Item Personality Inventory (TIPI) [51] scores, which yield measures of the “Big 5” personality constructs of extraversion, agreeableness, conscientiousness, emotional stability, and openness to experiences COVID-19 infection stressors (endorsement of self infected, know someone in household who was infected, know someone not in household who was infected, and know someone who died of COVID-19) COVID-19 pandemic stressors (Questions from the National Institute of Mental Health Coronavirus Health Impact Survey [52] were used to assess COVID-19–associated worries and concerns at the peripandemic assessment. Factor analysis revealed that these items loaded on five factors (eigen values=1.01-4.94): COVID-19–related disease worries (eg, “In the past month, how worried have you been about being infected with coronavirus?”); COVID-19 social restriction stress (eg, “How stressful have these changes in social contacts been for you?”); COVID-19–associated socioeconomic stress (eg, “In the past month, to what degree have changes associated to the pandemic created financial problems for you or your family?”); COVID-19–associated relationship difficulties (eg, “Has the quality of the relationships between you and members of your family changed?”); and COVID-19–associated social engagement (eg, “In the past month, how many people, from outside of your household, have you had an in-person conversation with?”).</td>
</tr>
</tbody>
</table>

\(^{a}\)DSM-5: Diagnostic and Statistical Manual of Mental Disorders, 5th edition.  
\(^{b}\)MDD: major depressive disorder.  
\(^{c}\)PTSD: posttraumatic stress disorder.  
\(^{d}\)PCL-5: PTSD Checklist for DSM-5.  
\(^{e}\)AUD: alcohol use disorder.  
\(^{f}\)DUD: drug use disorder.

**App Use Metrics**

App use metrics are similar to those outlined in the evaluation of the public use (“in the wild”) data [31]. Overall frequencies for key content usage were computed for each of the four key sections in the app: Manage Stress (tried a tool), Learn (viewed a learn topic), Mood Check (created and rate a goal or completed an assessment), and Find Resources (viewed at least one specific subsection within Find Resources). Because of the variability within and across mobile device operating systems with respect to reliably in capturing unique app sessions (eg, [31,53]), we decided to assess frequency of use by unique days of use rather than by sessions or visits. Distinct days of use within the observation window were calculated, as well as retention days (the number of days between the first day of use and last day of use). Activities within each of the key content areas were totaled as well as tabulated by each distinct day of use.
Analyses

Overview

Broadly, we utilized a funnel or conversion analytic strategy (eg, [54,55]) to characterize how many veterans tried the COVID Coach app, among those who downloaded it, from within the population of veterans who were offered the opportunity to download it.

Survey Data

Data analyses proceeded in five steps. First, we performed independent-samples $t$ tests and $\chi^2$ analyses to compare sociodemographic characteristics of veterans who did and did not complete the survey on a smartphone or tablet. Second, we performed independent-samples $t$ tests and $\chi^2$ analyses to compare sociodemographic, prepandemic, and pandemic-related variables in the subset of veterans who completed the survey on a smartphone or tablet and between those who did and did not download/use the COVID Coach app. Third, we performed multivariable binary regression analyses to identify sociodemographic, prepandemic, and pandemic-related variables that were independently associated with (1) completing the survey on a device compared to a desktop or personal computer and (2) downloading the COVID Coach app. Variables that were associated with completing the survey on a device versus desktop or personal computer and downloading the app in bivariate analyses ($P<.05$) were entered into these analyses. Planned secondary logistic regression analyses were performed to identify aspects of multidimensional measures (eg, PTSD symptoms) that were associated with downloading the app. Finally, to compare veterans who used the app once versus more than once, we performed independent-samples $t$ tests and $\chi^2$ analyses; these analyses were unweighted.

App Usage Data

SQLPro Studio (Hankinsoft Development, Inc) was used for all data preprocessing and extraction. SAS OnDemand for Academics (SAS Institute, Cary, NC) was used for statistical analyses of the app usage data. We calculated descriptive statistics for key content usage, unique days of app use, and retention. $\chi^2$ analyses were performed to understand differences in returning to the app for a second day of use based on key content usage on the first day of app use. We ran separate $\chi^2$ analyses for each predictor.

Results

Smartphone Ownership and Survey Completion Overview

Overall, the majority of participants ($n=2262$, 65.84%) completed the survey on a desktop or personal computer. Notably, regardless of the device type used to complete the survey, many veterans reported owning a smartphone ($n=2443$, 81.50%). Desktop or personal computer completers differed significantly across a number of dimensions compared to those who completed the survey on a mobile device (smartphone or tablet). Participants who completed the survey on a mobile device tended to be younger, identify as female or nonwhite race/ethnicity, be currently employed, and live in a household with an annual income of US $60,000 or greater compared with those who completed the survey on a desktop or personal computer. They were also more likely to have lifetime histories of major depressive disorder or PTSD and alcohol use disorder or drug use disorder, but reported fewer medical conditions. Importantly, the majority of participants indicated owning a smartphone (83.0% among veterans who did not complete the survey on a mobile device and 94.6% among veterans who completed their survey on a mobile device). See Table 2 for additional information about participant characteristics.

The multivariable logistic regression analysis revealed that lower age (odds ratio [OR] 0.97, 95% CI 0.96-0.97), female gender (OR 2.10, 95% CI 1.62-2.74), nonwhite race/ethnicity (OR 1.74, 95% CI 1.43-2.11), lower than college education (OR 1.53, 95% CI 1.27-1.85), and higher household income (OR 1.34, 95% CI 1.12-1.61) were independently associated with completing the survey on a mobile device, while residing in the west of the country was associated with a lower likelihood of doing so (OR 0.60, 95% CI 0.47-0.76). None of the other variables differentiated these groups (all $P>0.05$).
Table 2. Characteristics of US veterans by device type used for survey completion.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Completed survey on a desktop or personal computer (n=2264; weighted 65.8%)</th>
<th>Completed survey on a mobile device (n=814; weighted 34.2%)</th>
<th>t (df=3070) or χ²</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, weighted mean (SD)</td>
<td>64.8 (14.6)</td>
<td>55.5 (15.3)</td>
<td>16.53</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Female gender, n (weighted %)</td>
<td>183 (6.0)</td>
<td>161 (17.0)</td>
<td>92.28 (df=1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td><strong>Race/ethnicity, n (weighted %)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White, non-Hispanic</td>
<td>1932 (83.5)</td>
<td>609 (70.5)</td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Black, non-Hispanic</td>
<td>131 (8.7%)</td>
<td>81 (13.7%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>130 (4.3)</td>
<td>86 (9.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bi/Multiracial or Other</td>
<td>71 (3.6)</td>
<td>38 (6.3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>College graduate or higher education, n (weighted %)</td>
<td>1072 (33.9)</td>
<td>336 (28.4)</td>
<td>9.47 (df=1)</td>
<td>.002</td>
</tr>
<tr>
<td>Married/partnered, n (weighted %)</td>
<td>1634 (73.8)</td>
<td>590 (75.0)</td>
<td>0.54 (df=1)</td>
<td>.46</td>
</tr>
<tr>
<td>Currently employed, n (weighted %)</td>
<td>787 (45.0)</td>
<td>408 (61.2)</td>
<td>72.55 (df=1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Annual household income US $60,000 or above, n (weighted %)</td>
<td>1328 (60.3)</td>
<td>525 (64.4)</td>
<td>4.96 (df=1)</td>
<td>.03</td>
</tr>
<tr>
<td><strong>Region of country, n (weighted %)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South</td>
<td>835 (38.9)</td>
<td>353 (47.6)</td>
<td></td>
<td>&lt;.001</td>
</tr>
<tr>
<td>West</td>
<td>597 (25.2)</td>
<td>174 (17.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Midwest</td>
<td>528 (20.8)</td>
<td>187 (21.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>304 (15.1)</td>
<td>100 (12.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-metro residence</td>
<td>331 (15.8)</td>
<td>143 (17.5)</td>
<td>1.41</td>
<td>.24</td>
</tr>
<tr>
<td><strong>Military branch, n (weighted %)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Army</td>
<td>900 (47.9)</td>
<td>298 (45.3)</td>
<td>10.93 (df=4)</td>
<td>.03</td>
</tr>
<tr>
<td>Navy</td>
<td>467 (19.5)</td>
<td>206 (21.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Air Force</td>
<td>557 (19.5)</td>
<td>176 (17.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marine Corps</td>
<td>128 (4.9)</td>
<td>61 (7.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>212 (8.2)</td>
<td>73 (8.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combat veteran, n (weighted %)</td>
<td>781 (33.3)</td>
<td>271 (37.5)</td>
<td>5.60 (df=1)</td>
<td>.02</td>
</tr>
<tr>
<td>10+ years of military service, n (weighted %)</td>
<td>809 (35.0)</td>
<td>323 (39.5)</td>
<td>6.23 (df=1)</td>
<td>.01</td>
</tr>
<tr>
<td><strong>Health variables, n (weighted %)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lifetime MDD and/or PTSD</td>
<td>356 (16.7)</td>
<td>215 (27.8)</td>
<td>51.83 (df=1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Lifetime AUD and/or DUD</td>
<td>924 (41.0)</td>
<td>343 (48.2)</td>
<td>14.16 (df=1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Number of medical conditions, weighted mean (SD)</td>
<td>2.9 (2.2)</td>
<td>2.7 (2.0)</td>
<td>3.02</td>
<td>.001</td>
</tr>
<tr>
<td>Own any type of cell phone, n (weighted %)</td>
<td>2,103 (91.9%)</td>
<td>793 (97.2%)</td>
<td>32.87 (df=1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Own a smartphone, n (weighted %)</td>
<td>1,698 (76.6%)</td>
<td>745 (91.1%)</td>
<td>96.55 (df=1)</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

*aMDD: major depressive disorder.

bPTSD: posttraumatic stress disorder.

cAUD: alcohol use disorder.

dDUD: drug use disorder.
Predicting COVID Coach Explorer Download

Table 3 shows characteristics for veterans who did and did not download the COVID Coach app. Bivariate analyses revealed that, relative to those who did not download the COVID Coach app, those who did were more likely to have completed college or higher education reported more adverse childhood experiences and potentially traumatic events and scored higher on a measure of extraversion. They also reported a greater number of hours of daily exposure to pandemic-related media and greater severity of pandemic-related worsening of relationships and PTSD symptoms.

Results of the multivariable binary logistic regression analyses revealed that college graduate or higher education (OR 3.67, 95% CI 1.73-7.79), greater number of adverse childhood experiences (OR 1.22, 95% CI 1.05-1.41), higher extraversion (OR 1.68, 95% CI 1.31-2.14), and greater severity of pandemic-related PTSD symptoms (OR 2.46, 95% CI 1.10-5.53) were associated with downloading the COVID Coach app. None of the other variables was significant (all $P > .19$). A posthoc analysis of COVID-related PTSD symptoms revealed that greater severity of exaggerated startle symptoms drove the association with downloading/using the COVID Coach App (OR 1.62, 95% CI 1.11-2.36); none of the other PTSD symptoms was significant (all $P > .14$).
Table 3. Characteristics of US veterans who completed the survey on a tablet or smartphone and did and did not download the COVID Coach app.

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Did not download COVID Coach (n=780, weighted 65.8%)</th>
<th>Downloaded COVID Coach (n=34, weighted 3.3%)</th>
<th>t (df=1049) or $\chi^2$</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years), weighted mean (SD)</td>
<td>55.6 (15.3)</td>
<td>51.8 (15.5)</td>
<td>1.44</td>
<td>.08</td>
</tr>
<tr>
<td>Female gender, n (weighted %)</td>
<td>153 (16.8)</td>
<td>8 (22.9)</td>
<td>0.87 (df=1)</td>
<td>.35</td>
</tr>
<tr>
<td>Nonwhite race/ethnicity, n (weighted %)</td>
<td>196 (30.6)</td>
<td>9 (32.4)</td>
<td>0.05 (df=1)</td>
<td>.83</td>
</tr>
<tr>
<td>College graduate or higher education, n (weighted %)</td>
<td>313 (27.5)</td>
<td>23 (55.9)</td>
<td>13.02 (df=1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Married/partnered, n (weighted %)</td>
<td>561 (74.7)</td>
<td>29 (82.4)</td>
<td>1.02 (df=1)</td>
<td>.31</td>
</tr>
<tr>
<td>Currently employed, n (weighted %)</td>
<td>389 (61.2)</td>
<td>19 (61.8)</td>
<td>0.01 (df=1)</td>
<td>.94</td>
</tr>
<tr>
<td>Annual household income US $60,000 or higher, n (weighted %)</td>
<td>503 (64.6)</td>
<td>22 (61.8)</td>
<td>0.12 (df=1)</td>
<td>.73</td>
</tr>
<tr>
<td>Region of country, n (weighted %)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South</td>
<td>338 (47.3)</td>
<td>15 (58.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>West</td>
<td>167 (18.2)</td>
<td>7 (5.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Midwest</td>
<td>180 (21.8)</td>
<td>7 (20.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Northeast</td>
<td>95 (12.7)</td>
<td>5 (14.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nonmetro residence, n (weighted %)</td>
<td>643 (17.9)</td>
<td>28 (8.8)</td>
<td>1.87 (df=1)</td>
<td>.17</td>
</tr>
<tr>
<td>Military branch, n (weighted %)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Army</td>
<td>285 (45.5)</td>
<td>13 (41.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Navy</td>
<td>200 (21.3)</td>
<td>6 (17.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Air Force</td>
<td>168 (17.6)</td>
<td>8 (23.5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Marine Corps</td>
<td>57 (7.4)</td>
<td>4 (11.8)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>70 (8.3)</td>
<td>3 (5.9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combat veteran, n (weighted %)</td>
<td>258 (37.4)</td>
<td>13 (41.2)</td>
<td>0.20 (df=1)</td>
<td>.66</td>
</tr>
<tr>
<td>10+ years of military service, n (weighted %)</td>
<td>310 (39.6)</td>
<td>13 (35.3)</td>
<td>0.26 (df=1)</td>
<td>.61</td>
</tr>
<tr>
<td>Prepandemic variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adverse childhood experiences, weighted mean (SD)</td>
<td>1.8 (2.2)</td>
<td>2.6 (3.0)</td>
<td>2.14</td>
<td>.02</td>
</tr>
<tr>
<td>Number of traumas, weighted mean (SD)</td>
<td>9.4 (9.2)</td>
<td>12.7 (10.2)</td>
<td>2.05</td>
<td>.02</td>
</tr>
<tr>
<td>Lifetime MDD(^a) and/or PTSD(^b), n (weighted %)</td>
<td>209 (31.6)</td>
<td>15 (35.3)</td>
<td>0.21 (df=1)</td>
<td>.65</td>
</tr>
<tr>
<td>Lifetime AUD(^c) and/or DUD(^d), n (weighted %)</td>
<td>327 (48.3)</td>
<td>16 (44.1)</td>
<td>0.23 (df=1)</td>
<td>.63</td>
</tr>
<tr>
<td>Extraversion, weighted mean (SD)</td>
<td>3.7 (1.6)</td>
<td>4.9 (1.3)</td>
<td>4.56</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Agreeableness, weighted mean (SD)</td>
<td>4.9 (1.4)</td>
<td>5.0 (1.2)</td>
<td>0.26</td>
<td>.40</td>
</tr>
<tr>
<td>Conscientiousness, weighted mean (SD)</td>
<td>5.8 (1.1)</td>
<td>5.7 (1.4)</td>
<td>0.35</td>
<td>.36</td>
</tr>
<tr>
<td>Emotional stability, weighted mean (SD)</td>
<td>5.1 (1.4)</td>
<td>5.1 (1.5)</td>
<td>0.07</td>
<td>.47</td>
</tr>
<tr>
<td>Openness to experiences, weighted mean (SD)</td>
<td>4.8 (1.2)</td>
<td>5.0 (1.1)</td>
<td>0.86</td>
<td>.20</td>
</tr>
<tr>
<td>Pandemic-related variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>COVID-19 infection to self, n (weighted %)</td>
<td>70 (8.5)</td>
<td>6 (17.6)</td>
<td>3.41 (df=1)</td>
<td>.07</td>
</tr>
</tbody>
</table>
Characterizing Usage of COVID Coach Explorer

Distinct Days of Use
Most participants who downloaded the COVID Coach app used it for 4 distinct days or fewer. Half (17/34, 50%) used the app on exactly 1 day and an additional 24% (8/34) used the app on 2 distinct days, followed by one person who used the app for 3 days, and just over 10% of the sample (4/34, 12%) used the app on 4 distinct days. Distinct days of use ranged from 1 to 16 (mean 2.53, SD 2.86). Retention over the course of the observation period varied slightly more. Among participants who used the app for 2 days or more, the mean number of days retained was 97.41 (SD 97.07) with a range of 1 day (meaning the app was used on 2 consecutive days) to 300 days (ie, the time between the first day of app use and the last day of app use spanned approximately 10 months).

Overall Key Content Usage
Nearly half of the participants who downloaded the app (15/34, 44%) completed at least one key event within one of the four sections of the app (Manage Stress, Learn, Mood Check, or Find Resources). The manage stress tools were tried 119 times by 11/34 participants (mean 10.82 tools, SD 19.81; mode 2, range 1-67). The most frequently accessed tool was “Finding Meaning” (12/119, 10.1% of all tool usage). Learn topics were viewed 35 times by 3/34 participants (range 1-67 topics per participant). Within the Mood Check section, 31 assessments were completed by 5/34 unique participants and one participant added a personal goal. Lastly, resources within the Find Resources section were viewed 20 times among 4/34 participants (range 2-14 resources).

Comparison of Veterans Who Used the COVID Coach App Only One Day Versus More Than One Day
Among the 34 veterans who downloaded/used the COVID Coach app, we additionally examined characteristics of veterans who used the app on one day only (n=17) versus 2 or more distinct days (n=17; mean 3.6, SD 2.0) on all of the characteristics shown in Table 2. Results of these analyses revealed that, relative to veterans who used the app one daily only, those who used the app 2 or more days scored lower on prepandemic measures of agreeableness (mean 4.3, SD 1.1 vs mean 5.6, SD 1.0; \( t_{1049} = 3.30, P = .003; d = 1.18 \)) and emotional stability (mean 4.54, SD 1.3 vs mean 6.7, SD 1.5; \( t_{1049} = 2.24, P = .003; d = 0.80 \)), and higher on measures of pandemic-related financial stressors (mean 0.9, SD 1.8 vs mean –0.2, SD 0.6; \( t_{1049} = 2.31, P = .03; d = 0.83 \)) and relationship difficulties (mean 1.0, SD 1.1 vs mean –0.1, SD 0.9; \( t_{1049} = 3.02, P = .005, d = 1.08 \)). Usage of the four key content areas within the app (Manage Stress, Mood Check, Learn, Find Resources) on the first day of use did not predict returning to the app for a second day (all \( P > .10 \)).

Discussion
Principal Findings
This study is among the first to provide population-level estimates for smartphone ownership among US veterans, identify predictors of uptake and usage for a mental health app
focused on pandemic-related stressors, and examine objective app usage data. We found that the vast majority of participants in this study owned smartphones and many had a lifetime history of mental health concerns, yet relatively few individuals were willing to try a free mental health app for pandemic-related stress, and among those who did, overall app usage was fairly low. Among older veterans who completed their KnowledgePanel survey on a mobile device (n=814/3078, mean age 56 years), 91.1% owned smartphones. Smartphone ownership among veterans who completed their survey on a laptop or desktop computer was also high (76.6%) and relatively close to national estimates for the 50-64–year age group [21]. This is notable because the mean age of participants in this group was nearly 65 years, and smartphone ownership estimates for adults aged 65 and older is only 61% of the US population [21]. This finding suggests that members of the KnowledgePanel sample may be more open to adopting mobile technologies than the general population, but it could also suggest that older veterans may be more willing to adopt smartphones than the general older adult population. Additionally, participants who completed their KnowledgePanel survey on a mobile device were more likely to identify as women and nonwhite, be college- or higher-educated, be currently employed, and more likely to have a lifetime history of mental health concerns than those who did not complete their survey on a mobile device. The relationship between mobile device ownership and socioeconomic status is consistent with prior research [21,33]. These findings also highlight that digital health interventions such as mobile apps may be a way to reach older women veterans, veterans of color, and older veterans with mental health concerns.

The results further revealed that smartphone adoption may not necessarily equate to mobile mental health uptake and usage. Among the 800 participants who were offered the opportunity to download and try COVID Coach, only 34 (3.3%) downloaded and tried the app, 50% of whom (n=17/34) returned to the app after their first day of use, and 44% (n=15/34) engaged with content in at least one of the four key content areas within the app. Nonetheless, there were some encouraging findings in the data. Although only 3.3% of eligible participants downloaded and tried the app, one of the strengths of utilizing the KnowledgePanel sample is it being comprised of a population-based, nationally representative sample of US adults. Using population benchmarks, if 3.3% of the US veteran population were to download an app for mental health, that would equate to more than 600,000 veterans. Thus, the potential reach for an app is large, and even if only some veterans use the app for an extended period of time, that may potentially translate to thousands of users and ultimately an important public mental health impact [56]. It is important to note that the majority of veterans do not use the VA as their primary source of health care, and veterans that do use the VA are more likely to be black, younger, female, unmarried, have lower household incomes, and have a lifetime history of psychopathology [57]. Therefore, a multipronged approach is needed to help promote app awareness and uptake among veterans being served both outside and within the VA. For example, national communication strategies that reach veterans wherever they are, such as features in popular media [58,59], promotion from organizations that serve veterans (eg, [60]), virtual veteran communities (eg, Women Veterans Network [WoVEN] [61]), resources and tools for community-based providers (eg, Community Provider Toolkit [62]), and social media campaigns that are tailored to specific veteran communities may be best suited to reach veterans not receiving care within the VA. Programs such as Tech Into Care [63], which train a wide variety of VA staff, including doctors, nurses, psychologists, social workers, audiologists, and chaplains, to be mobile health (mHealth) ambassadors and spread the word about apps for mental health may help raise awareness among both VA employees and the veterans they serve. mHealth ambassadors, who are trained in how to use, offer, and implement mobile mental health apps, may also help develop dissemination approaches that take veteran characteristics, local or regional factors, and Veterans Integrated Services Networks context into account.

Additionally, veterans who downloaded COVID Coach reported a greater number of adverse childhood experiences, greater extraversion, and greater severity of pandemic-related PTSD symptoms (ie, exaggerated startle response) than those who did not download the app. Furthermore, among those who did download the app, veterans with lower levels of emotional stability and who experienced greater pandemic-related financial stressors and relationship difficulties were more likely to return to the app for a second day of use than those who only used the app for a single day. Because COVID Coach was specifically designed to provide tools and resources for coping with pandemic-related stress and concerns, the app was preferentially downloaded by veterans in the sample with greater mental health needs during the pandemic. Indeed, the interactive coping tools in the Manage Stress section were the most popular among veterans who utilized the app, a finding that is consistent with previous work [31]. The “Finding Meaning” tool was the most frequently utilized coping tool within the app. This is notable and contrasts with usage among the general population where “Ambient Sounds,” the first tool in the list due to alphabetization, was the most utilized tool. Among older adults, finding purpose and meaning in life is associated with better health outcomes, including cognitive health [64,65]. Digital mental health interventions that target helping older adults cultivate purpose and meaning in life could have an important impact on mental health as well as promote cognitive health.

A notable strength of this study is that it analyzed data from a contemporary, nationally representative, probability-based sample of US veterans. We were able to estimate the prevalence of smartphone ownership, as well as closely examine the characteristics of veterans who did and did not download the COVID Coach mobile mental health app during the height of the COVID-19 pandemic. Furthermore, we were also able to explore veteran characteristics associated with app download and app usage, which were measured automatically via captured analytics data. Collectively, the results of this study provide an important contribution to understanding veteran smartphone ownership rates, characteristics associated with downloading (or not downloading) an app for mental health, and predictors of return app usage.
Limitations
There are several key limitations of this study. First, the opportunity to download the COVID Coach app was presented as the last screen of a survey with a median completion time of nearly 32 minutes. Following survey completion that assessed a broad range of factors, including pandemic stressors and psychiatric symptoms, participants may have been fatigued and reluctant to take on yet another optional task, particularly since downloading and trying the app did not impact their study compensation. Second, the app was only offered to veterans who completed the survey on a mobile device, regardless of whether they owned a smartphone or tablet, thus limiting the potential number of potential app downloads. Nearly two-thirds of the sample did not complete their survey on a mobile device, yet the vast majority of those participants (76.6%) indicated that they did own a smartphone. Thus, participants who may have been interested in exploring the app were never offered the chance to download it. Third, for those who did receive the information about how to download COVID Coach, the perceived benefits and utility could have been specified more clearly. Previous research has indicated that older veterans, especially those living in rural areas, may be less likely to see the benefits of using mobile apps [26]. It is possible that download instructions were not sufficiently clear and the rationale for downloading the app was not sufficiently compelling for some individuals. Further qualitative work would be helpful in identifying how to optimally present this information to encourage high uptake of the app. Finally, we only have objective use data to understand engagement with COVID Coach. More in-depth qualitative information is needed to explore older veterans’ experiences with the app, including usability of the design, appropriateness of the content, and other factors that may have influenced if, when, and how often they used the app.

Future Research Directions
Results of this study underscore the importance of research that addresses the needs and preferences of veterans to help ensure that the tremendous digital health innovation fueled by the pandemic does not reinforce or exacerbate existing inequities [66]. Digital health tools, including apps, can be part of the solution to help promote better mental health and health care outcomes for older adults [67], as long as they address fundamental issues of digital health equity, such as digital health literacy and inclusive design [68]. To best meet the needs of older veterans, and older adults more generally, future research should utilize qualitative methods and co-design processes to ensure that interventions are solving mental health challenges in usable, meaningful, and engaging ways. Co-design can help address barriers that are specific to older adults (eg, [69]); ensure that the product is findable, accessible, usable, desirable, credible, useful, and valuable (eg, [70]); improve the overall quality of the product [71]; and promote equity and inclusion (eg, [72,73]). More research focused on effective dissemination and implementation strategies is also needed. The number of veterans receiving their health care from the VA has increased over time, and those receiving their care within VA tend to be from populations that are more likely to experience health disparities [57]. These historically underserved groups may benefit from additional digital health supports, and receiving their care within an integrated health care system could potentially facilitate the dissemination and implementation of digital health resources. However, one of the biggest barriers to app uptake identified among veterans receiving care within the VA is app awareness [40]. Furthermore, many veterans receive part or all of their health care outside the VA system. Developing effective strategies for raising awareness of mental health apps and facilitating their usage are crucial for adoption and in turn increased impact.

Conclusions
The COVID-19 pandemic has accelerated the creation and use of digital mental health resources across a variety of settings. As the veteran population in the United States is aging and their smartphone ownership is growing, they are also more successfully engaging with digital health products [23]. To our knowledge, this study is one of the first to document current smartphone ownership rates in the US veteran population, the majority of whom were older, and to examine predictors of uptake and usage of a mental health app focused on COVID-19–related stressors. Although adoption of COVID Coach was relatively low, the app was more likely to be utilized by individuals facing pandemic-related stressors and associated psychiatric symptoms, suggesting that apps may be a way to reach veterans with mental health needs during the pandemic and beyond. Collectively, results of this research suggest that mental health apps have the potential to reach a significant minority of older veterans, although continued efforts are needed to identify strategies to bolster uptake in more naturalistic settings. More work is needed to ensure uptake and meaningful engagement with mental health apps. As the pandemic continues to impact mental health globally, digital mental health resources have an important role to play in meeting the needs of veterans, and the general population, during the pandemic and beyond.

Acknowledgments
The National Health and Resilience in Veterans Study is supported by the US Department of Veterans Affairs National Center for Posttraumatic Stress Disorder. The authors would also like to acknowledge the design and development support from Vertical Design and Mumble in creating the COVID Coach and COVID Coach Explorer apps.

Conflicts of Interest
None declared.
References


59. Women Veterans Network WoVEN. URL: https://www.wovenwomenvets.org/ [accessed 2022-03-23]

60. VA Community Provider Toolkit. URL: https://www.mentalhealth.va.gov/communityproviders/ [accessed 2022-03-23]


Abbreviations
- GAD-7: Generalized Anxiety Disorder-7
- JSON: JavaScript object notation
- mHealth: mobile health
- NHRVS: National Health and Resilience in Veterans Study
- OR: odds ratio
- PCL-5: PTSD Checklist for DSM-5
- PHQ-9: Patient Health Questionnaire-9
- PTG: post traumatic growth
- PTSD: posttraumatic stress disorder
- VA: US Department of Veterans Affairs
- WoVEN: Women Veterans Network

©Beth K Jaworski, Katherine Taylor, Kelly M Ramsey, Adrienne J Heinz, Sarah Steinmetz, Jason E Owen, Jack Tsai, Robert H Pietrzak. Originally published in JMIR Mental Health (https://mental.jmir.org), 05.04.2022. 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 https://mental.jmir.org/, as well as this copyright and license information must be included.