JMIR Mental Health

Journal Impact Factor (JIF) (2023): 4.8
Volume 9 (2022), Issue 4 ISSN 2368-7959 Editor in Chief: John Torous, MD, MBI

Contents

Original Papers

Reviews

Original Paper

Assessment of Population Well-being With the Mental Health Quotient: Validation Study

Jennifer Jane Newson[1], BSc, DPhil; Vladyslav Pastukh[1], BSc; Tara C Thiagarajan[1], BSc, PhD

Sapien Labs, Arlington, VA, United States

Corresponding Author:
Jennifer Jane Newson, BSc, DPhil
Sapien Labs
1201 Wilson Blvd, 27th floor
Arlington, VA, 22209
United States
Phone: 1 7039976657
Email: jennifer@sapienlabs.org

Abstract

Background: The Mental Health Quotient (MHQ) is an anonymous web-based assessment of mental health and well-being that comprehensively covers symptoms across 10 major psychiatric disorders, as well as positive elements of mental function. It uses a novel life impact scale and provides a score to the individual that places them on a spectrum from Distressed to Thriving along with a personal report that offers self-care recommendations. Since April 2020, the MHQ has been freely deployed as part of the Mental Health Million Project.

Objective: This paper demonstrates the reliability and validity of the MHQ, including the construct validity of the life impact scale, sample and test-retest reliability of the assessment, and criterion validation of the MHQ with respect to clinical burden and productivity loss.

Methods: Data were taken from the Mental Health Million open-access database (N=179,238) and included responses from English-speaking adults (aged≥18 years) from the United States, Canada, the United Kingdom, Ireland, Australia, New Zealand, South Africa, Singapore, India, and Nigeria collected during 2021. To assess sample reliability, random demographically matched samples (each 11,033/179,238, 6.16%) were compared within the same 6-month period. Test-retest reliability was determined using the subset of individuals who had taken the assessment twice ≥3 days apart (1907/179,238, 1.06%). To assess the construct validity of the life impact scale, additional questions were asked about the frequency and severity of an example symptom (*feelings of sadness, distress, or hopelessness*; 4247/179,238, 2.37%). To assess criterion validity, elements rated as having a highly negative life impact by a respondent (equivalent to experiencing the symptom ≥5 days a week) were mapped to clinical diagnostic criteria to calculate the clinical burden (174,618/179,238, 97.42%). In addition, MHQ scores were compared with the number of workdays missed or with reduced productivity in the past month (7625/179,238, 4.25%).

Results: Distinct samples collected during the same period had indistinguishable MHQ distributions and MHQ scores were correlated with r=0.84 between retakes within an 8- to 120-day period. Life impact ratings were correlated with frequency and severity of symptoms, with a clear linear relationship (R^2>0.99). Furthermore, the aggregate MHQ scores were systematically related to both clinical burden and productivity. At one end of the scale, 89.08% (8986/10,087) of those in the Distressed category mapped to one or more disorders and had an average productivity loss of 15.2 (SD 11.2; SEM [standard error of measurement] 0.5) days per month. In contrast, at the other end of the scale, 0% (1/24,365) of those in the Thriving category mapped to any of the 10 disorders and had an average productivity loss of 1.3 (SD 3.6; SEM 0.1) days per month.

Conclusions: The MHQ is a valid and reliable assessment of mental health and well-being when delivered anonymously on the web.

(JMIR Ment Health 2022;9(4):e34105) doi:10.2196/34105

KEYWORDS

psychiatry; public health; methods; mental health; population health; social determinants of health; global health; behavioral symptoms; diagnosis; symptom assessment; psychopathology; mental disorders; mHealth; depression; anxiety; attention deficit disorder with hyperactivity; autistic disorder; internet

Introduction

Background

The World Health Organization defines mental health as "a state of well-being in which the individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively and fruitfully, and is able to make a contribution to his or her community" [1]. On the basis of this definition, assessments of mental health should reflect the presence of dysfunction and also provide insight into the positive aspects of mental functioning [2-5]. However, the clinical heritage of mental health assessment means that most assessment tools are built around specific psychiatric disorder categories taken from the clinical classification systems of the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) [6], or the International Classification of Diseases [7] and, therefore, are not designed to provide a perspective on the continuum of mental health and well-being across the general population. In contrast, assessments of mental health that are more relevant across the spectrum of the general population can support the early identification of at-risk individuals before symptoms escalate, improve uptake in help-seeking behaviors, and reveal relevant social determinants to support the active management of mental health and well-being through self-care behaviors and preventative strategies and interventions at various scales, from organizations to countries [8-10].

Population-level assessments also provide an opportunity for understanding the scale of mental health challenges at a global level. For example, in 2017, 792 million people were estimated to be living with a mental health disorder worldwide [11], whereas depression is the leading cause of disability as measured by years lived with disability [12]. In addition, suicide was the fourth leading cause of death among people aged 15-29 years worldwide in 2019 [13] and is still poorly understood [14,15]. However, presently, there are few reliable and valid tools that can provide an aggregate and measurable view across the full spectrum of a global population from distressed to thriving as well as estimating clinical burden in an aggregate, disorder-agnostic way. Furthermore, as mental health and well-being can change substantially based on external circumstances, as evidenced by the COVID-19 pandemic [16-18], it is important to have metrics that track the extent and nature of these changes and their impact on clinical burden as well as on the productive capacity of a population. For example, evidence suggests a clear relationship between mental health and well-being and productivity, resulting in both absenteeism and presenteeism [19-25], an increased prevalence of burnout [26,27], and significant personal and economic loss [28-32]. This further highlights the importance of preventative measures for actively managing the mental health and well-being of working-age adults in the general population.

To address this need for a population-based, disorder-agnostic assessment that spans the spectrum of mental health and well-being and for a single measurable metric of mental health and well-being, we developed a new web-based assessment tool delivering a metric called the Mental Health Quotient (MHQ) [33]. The MHQ assesses the complete breadth of mental health elements spanning the range from symptoms to positive mental assets using a unique life impact scale and aims to enable a paradigm for managing and improving the lives and well-being of all people, not just those with a clinical disorder.

The MHQ Assessment

The MHQ was developed based on a comprehensive review of symptoms by coding questions across 126 commonly used psychiatric assessment tools spanning depression, anxiety, bipolar disorder, attention-deficit/hyperactivity disorder, posttraumatic stress disorder, obsessive-compulsive disorder (OCD), addiction, schizophrenia, eating disorders, and autism spectrum disorder as well as cross-disorder tools (see Newson et al [34] for a complete list of assessment tools). These disorders were selected based on a review of the disorders included in the Structured Clinical Interview for the DSM-5, Clinician Version [35]. In addition, autism spectrum disorder and eating disorders were included because of both their prevalence and their broad public and scientific interest. Symptoms from these 126 assessments were consolidated into a set of 43 symptom categories and reviewed and expanded in the context of the Research Domain Criteria constructs put forward by the National Institute of Mental Health [36-38]. The resultant 47 elements were then split into two formats: mental functions that could manifest as a spectrum from positive to negative (spectrum questions) and those symptoms that purely represented detractions from overall mental health (problem questions).

Spectrum and problem questions within the MHQ are answered using a 9-point scale reflecting the consequences on one's life functioning and impact on their ability to carry out tasks and activities in their daily life. Therefore, the scale is different from traditional mental health assessments, which typically focus on the frequency, severity, duration, or timing of symptoms [34]. An aggregate MHQ score developed using an algorithm that nonlinearly transforms the life impact scale based on different categories of symptom seriousness is provided on completion of the assessment [33]. This score is intended as a representation of the overall mental health and well-being of the individual and is categorized from Distressed (−100) to Thriving (+200).

The MHQ is currently used on the web as part of an open data project called the Mental Health Million Project, which is a web-based platform that monitors the status of population mental health across the globe and currently spans 30 countries and 4 languages (English, French, Spanish, and Arabic). In this paper, we evaluate the potential of the MHQ to be used as a valid and reliable measure of mental health and well-being both at the individual and population levels to determine how mental health and well-being evolves over time across the globe and the impact of these changes on clinical burden and productivity. We aim to address (1) how the unique MHQ life impact scale relates to more commonly used metrics of frequency and severity; (2) whether an anonymous web-based assessment serves as a true measure of the population by demonstrating the population and test-retest reliability of the MHQ; and (3) how well the composite MHQ score relates to functional criteria such as clinical diagnostic criteria, workdays missed, and overall life productivity. We hypothesize that the MHQ will show good

validity and reliability as an assessment of mental health and well-being when delivered anonymously on the web.

Methods

Recruitment of Participants

The data were taken from the Mental Health Million open-access database [39] and included responses from 179,238 English-speaking individuals from the United States, Canada, the United Kingdom, Ireland, Australia, New Zealand, South Africa, Singapore, India, and Nigeria collected during 2021. Participants were recruited via outreach campaigns on Facebook and using Google Ads by targeting a broad cross-section of adults aged 18-85 years across a wide geographic and socioeconomic demographic. The anonymous assessment was freely available on the web for anyone to complete, and individuals took the assessment for the purpose of obtaining their personalized mental health and well-being report on completion. The provision of a personal report aimed to ensure greater interest of the respondent in answering questions thoughtfully and accurately. Only respondents who found the assessment easy to understand (ie, responded *Yes* to the question *Did you find this assessment easy to understand?*) were included in the analysis. This resulted in the exclusion of 2.58% (4620/179,238) of respondents, leaving 174,618 for the full analysis.

Ethics Approval

The study received ethics approval from the Health Media Lab Institutional Review Board (Office for Human Research Protections Institutional Review Board #00001211, Federal Wide Assurance #00001102, IORG #0000850).

Assessment of Reliability

Reliability Across Randomly Selected Samples

All respondents from the United States, India, Australia, and the United Kingdom between January 2021 and June 2021 were pooled together (44,132/174,618, 25.27%). A total of 4 randomly selected and nonoverlapping samples of 11,033 people with similar demographic composition were selected. The average rating (1-9) of each MHQ-scored element for each sample, the average MHQ score for each sample, and the statistical differences between the samples were then computed.

Internal Consistency Analysis

The MHQ is designed to be as parsimonious as possible without repetition. However, the internal consistency of the MHQ was evaluated by looking at the relative correlations between elements that would be expected to be correlated compared with those that would not. First, the correlation between 2 questions about sleep quality within the MHQ was computed (N=174,618). Sleep question 1 asked respondents to *Assess your sleep quality*, and sleep question 2 asked respondents: *In general, I get as much sleep as I need*. The 1-9 rating score from sleep question 1 was correlated with the transformed answers to sleep question 2, where each answer option was assigned a number that was roughly equivalent to the text description: *all the time*=7, *most of the time*=5, *some of the time*=3, and *hardly ever*=1. Second, the correlation between 2 questions about mood was computed. Mood question 1 asked respondents to *Assess your feelings of sadness, distress, or hopelessness* and was rated on a 1-9 life impact scale. Mood question 2 asked respondents *How would you describe your overall mood right now?* and was rated on a 1-9 scale from *Very negative* to *Very positive*. Comparisons were also made between responses to the related MHQ elements of *Self-worth and confidence* and *Self-image*. In addition, comparisons were made between the elements of *Physical intimacy* and *Memory*, *Emotional control* and *Coordination*, and *Memory* and *Emotional control*, which would not be expected to have a significant correlation.

Test-Retest Reliability

The MHQ is designed to measure changes in the mental health and well-being of the population and, therefore, in individual mental health and well-being status over time. Therefore, the MHQ scores of individuals could change over time. However, over short time frames of less than a year, most individuals would not be expected to change significantly. Within the sample of 174,618 respondents, email addresses were provided by 80,955 (46.36%) to receive their MHQ report. These email addresses were automatically converted into anonymous unique identifiers to identify repeat respondents. Of these 80,955 respondents, 2231 (2.76%) had taken the MHQ twice at varying time intervals up to 15 months from the time of the first assessment. Those who took the MHQ twice within the same day or immediately the next day were excluded as they were more likely to be experimenting with answer choices than evaluating their own change over time in an honest way. Thus, only those who had at least 3 days between attempts were included in the analysis (1907/2231, 85.48%). We examined the test-retest reliability of the MHQ in this sample by looking at the correlation between the element ratings on the first and second attempts as well as the correlation between MHQ scores across both attempts.

Validation of the Life Impact Scale

Clinical assessments are heterogeneous in their evaluation of the frequency and severity of symptoms. For example, a review of 126 assessment tools found that, across 19 commonly used depression scales, 51% of questions asked about frequency of symptoms and 32% asked about severity, whereas, across 9 posttraumatic stress disorder assessment tools, 17% of questions asked about frequency and 53% asked about severity [34]. Given the lack of a clear understanding of which aspect (eg, frequency or duration) of a symptom matters most, the MHQ uses a 9-point life impact scale reflecting the impact of a particular mental aspect on one's ability to function [33]. For example, for questions pertaining to mental health challenges, 1 referred to *Never causes me any problems*, 9 referred to *Has a constant and severe impact on my ability to function*, and 5 referred to *Sometimes causes me difficulties or distress but I can manage*. For the purpose of validation, for the question that asked individuals to rate the impact of their *Feelings of sadness, distress, or hopelessness* on this 9-point scale, two additional questions were asked when a value of ≥5 was selected: (1) *How many days in the last week did you experience these feelings?* with options for selection from 0-7 (similar to the format in depression screening tools such as the Center for Epidemiologic

Studies–Depression scale [40,41]) and (2) *On these days, how did these feelings impact your ability to function in life?* with five options of increasing severity—1=*They would come and go while I went about my life as normal*; 2=*I did what I had to do, but they were always there in the back of my mind*; 3=*I managed but it took extreme effort*; 4=*They stopped me doing the things I usually do, or would want to do*; and 5=*They consumed me so much I was unable to get out of bed*. The average frequency and severity and the standard error of measurement (SEM) were then computed for each selection from 5 to 9 on the life impact scale.

Relationship Between MHQ Score and Clinical Burden

The computation of the MHQ score takes into account the number of severe symptoms (ie, those scored as having a highly negative life impact). Thus, the number of elements with a rating that signifies a highly negative life impact decreases as the MHQ score increases, although the nonlinear weighting differs for different types of symptoms [33]. To assess how effectively the MHQ score relates to clinical burden, we mapped elements of the MHQ to the diagnostic criteria for each of the 10 major DSM-5 disorders on which the MHQ is based (see *The MHQ Assessment* section) and examined (1) the percentage of individuals meeting the diagnostic criteria for at least one disorder and (2) the average number of diagnoses per person for each MHQ score bin of 25. Full details of the thresholds used to determine the presence of a clinical symptom and the mapping to the DSM-5 disorder criteria are described in Newson et al [42]. In brief, MHQ elements were first mapped to the symptoms described within the diagnostic criteria for each of the 10 DSM-5 disorders based on the closest semantic match. For each of the 47 MHQ elements, responses were determined to be clinically significant symptoms if they met a particular threshold of impact on the individual's ability to function, approximately equivalent to experiencing the symptom 5 days a week (≥8 for problem elements and ≤1 for spectrum elements). The specific diagnostic criteria rules of the DSM-5 (eg, must be experiencing ≥5 symptoms) were then applied to arrive at a disorder diagnostic mapping. A set of rules using combinations of the DSM-5–mapped MHQ elements was then developed to align with these criteria descriptions for each of the 10 disorders. For each respondent (N=174,618), these rules were applied to their MHQ clinical symptom profile to determine the diagnostic match to each of the 10 disorders. See Newson et al [42] and the *Limitations* section below for further discussion of this approach.

Relationship Between MHQ Score and Productivity Criterion

To assess the relationship between the MHQ score and measures of functional productivity, a subset of participants (7625/174,618, 4.37%) were asked two additional questions: (1) *How many days during the past month were you totally unable to work or carry out your normal activities because of problems with your physical or mental health* and (2) *How many days during the past month were you able to work and carry out your normal activities, but could not get as much done because of problems with your physical or mental health?* with options to select a number between 0 and 31. Individuals were then grouped by MHQ score in bins of 25, and the average and SEM of days of work missed (*M*) and days with reduced productivity (*R*) were then computed for each bin. We then computed the overall loss of life productivity for each individual as $M + n * R$, where n represented an assumed loss of productivity on those days ranging from 20% to 50%. Data were examined for all respondents together and for a subset of respondents who answered *Employed/Self-Employed* to the MHQ question *Please select which best describes your occupational status?* (alternative answer options included *Homemaker*, *Unemployed*, *Retired*, *Studying*, and *Not able to work*).

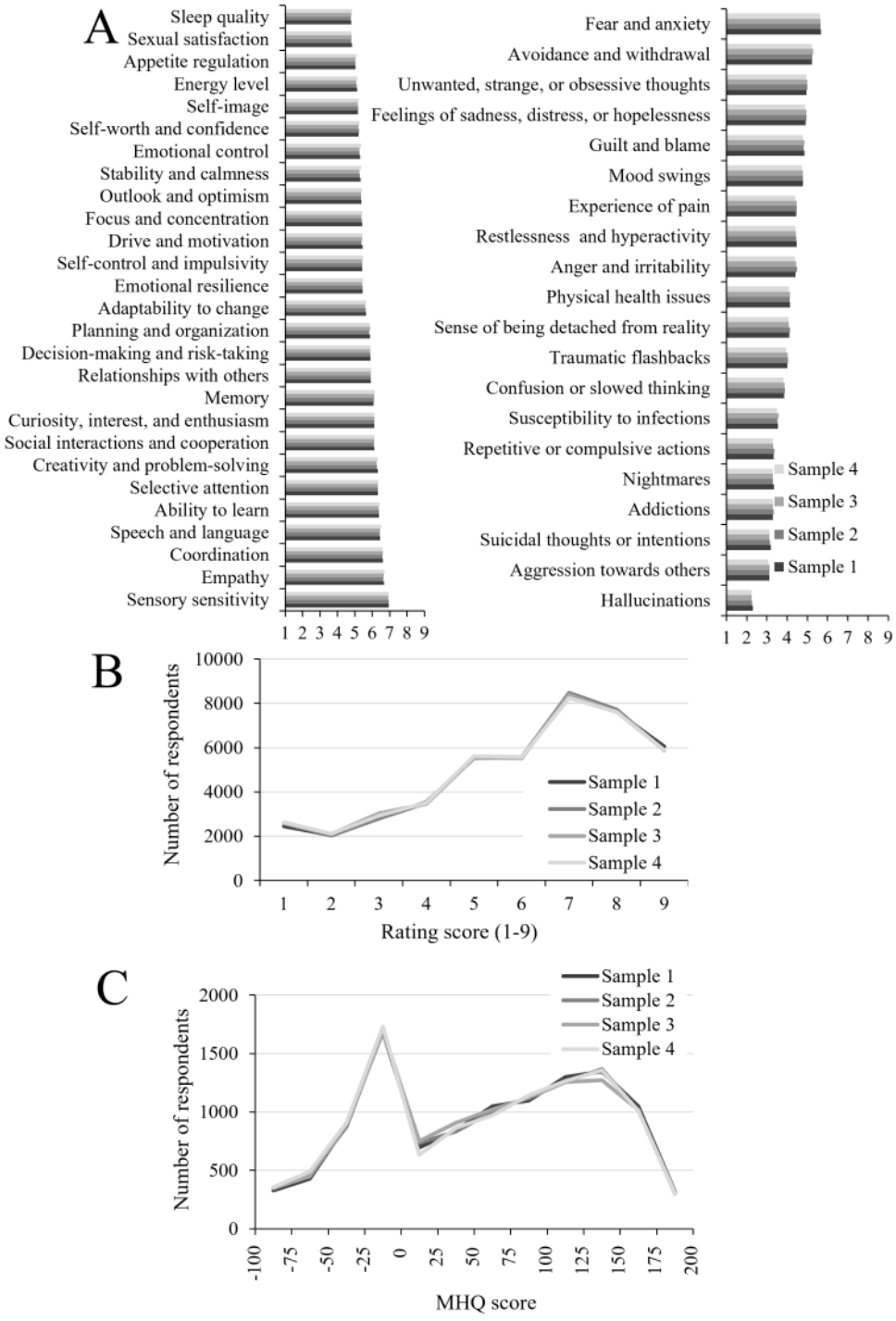
Results

Reliability and Internal Consistency in the MHQ

Assessment Reliability

Figure 1A shows the average rating for each element of the assessment for spectrum (left) and problem (right) elements across the 4 randomly selected demographically matched samples. Across all samples, the ratings were correlated with $r>0.8$ for all pairs, and the distributions of ratings for individual elements were highly similar and statistically indistinguishable (analysis of variance; $P=.99$), with an example from the element *Self-image* shown in Figure 1B. Similarly, the distributions of the resulting MHQ scores for each of these 4 samples were highly similar (Figure 1C, analysis of variance; $P=.18$). These results confirm that the MHQ, when offered anonymously and on the web, produces similar results across similar samples. Should responses have been randomly generated (eg, by bots) or if individuals had highly inconsistent interpretations of the life impact scale, this would not have been the case.

Figure 1. Reliability of the Mental Health Quotient (MHQ). (A) The average ratings of each of the 27 MHQ spectrum elements (left) and 20 MHQ problem elements (right) in 4 separate samples of the MHQ obtained over a similar period were indistinguishable (each bar is a sample). (B) Distribution of ratings for an example MHQ element (Self-image) in each of the 4 samples (each line is a sample). (C) Distribution of MHQ scores in each of the 4 samples (each line is a sample) for an example MHQ element (Self-image).



Internal Consistency

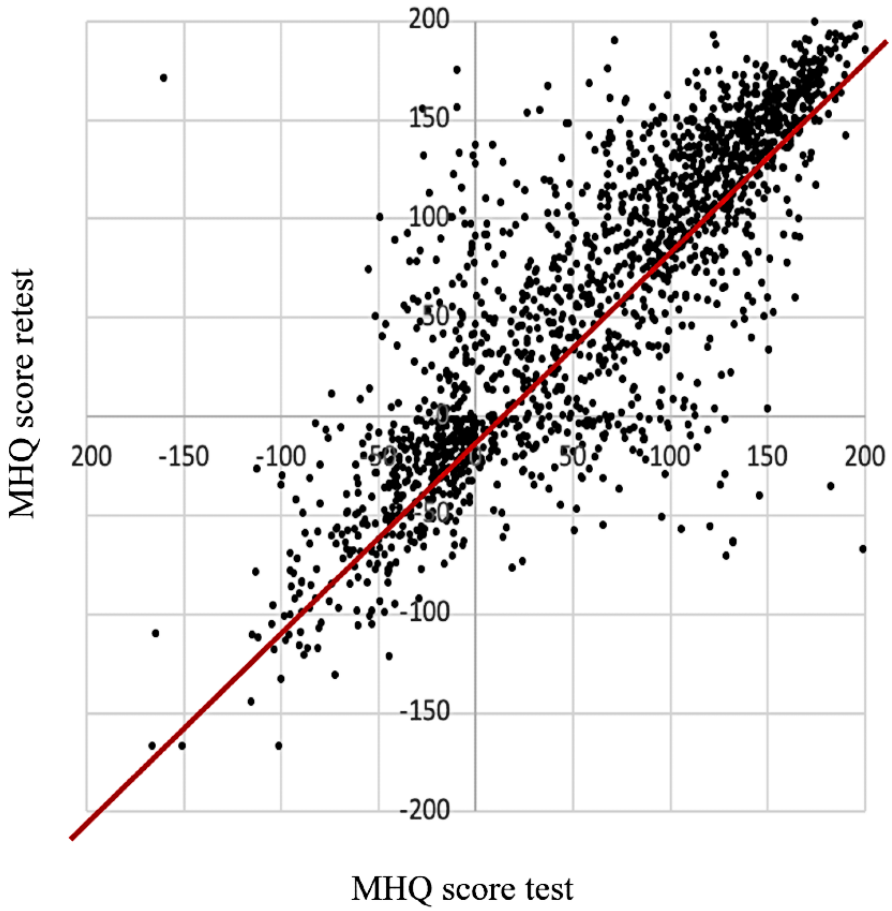
Among the related elements, the 2 questions relating to sleep quality and sleep sufficiency had a 0.63 correlation. Thus, those who had challenges with sleep quality were also likely to have fewer days of sufficient sleep. Similarly, the 2 questions relating to mood had a 0.64 correlation, indicating that those with a more significant impact of *Feelings of sadness, distress, or hopelessness* were also more likely to have a negative mood at the time of taking the assessment. Finally, the life impact rating of the MHQ element *Self-image* had a 0.77 correlation with the rating of the element *Self-worth and confidence*. In contrast, ratings of unrelated elements had lower correlations. For example, *Memory* and *Physical intimacy* had a correlation of

0.35, *Emotional control* and *Coordination* had a correlation of 0.36, and *Memory* and *Emotional control* had a correlation of 0.39. Therefore, related elements within the MHQ were more highly correlated than the unrelated elements examined.

Test-Retest Reliability

Among all those who could be identified as having taken the MHQ twice at least 3 days apart, MHQ scores were correlated with r=0.84 (P<.001). Figure 2 shows the MHQ scores for the test plotted against the MHQ scores for the retest, demonstrating that points fall around the line $y=x$. Furthermore, this correlation did not change significantly as the interval between attempts increased, although correlations were as high as r=0.88 for retest intervals of 8-120 days. The correlations (r) were as follows for MHQ scores and MHQ items ratings respectively: 3 to 7 days=0.7, 0.58; 8 to 30 days=0.88, 0.73; 31 to 60 days=0.88, 0.72; 61 to 120 days=0;.83, 0.7; 121 to 450 days=0.79, 0.68). Finally, the correlation between the ratings of individual elements on each attempt was r=0.70 (P<.001) and did not change as the interval between attempts increased. Thus, the MHQ had high test-retest reliability but also reflected changes that can occur in mental health and well-being over time.

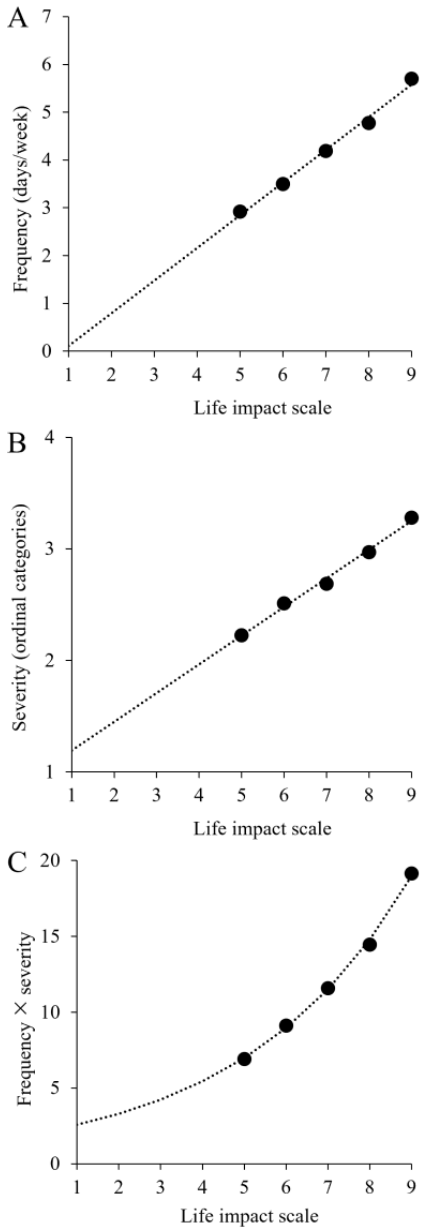
Figure 2. Test-retest reliability of the Mental Health Quotient (MHQ). MHQ score for the assessment retake (retest) versus MHQ score for the first take (test). The red line represents y=x. MHQ: Mental Health Quotient.



Relationship Between Life Impact Score and Frequency and Severity

MHQ life impact ratings ≥5 (corresponding to the negative end of the scale) for the MHQ element *Feelings of sadness, distress, or hopelessness* (data subset of 4247/174,618, 2.43%) were correlated with symptom frequency measured as the number of days in the past week where they experienced the symptom, with r=0.5. At the aggregate level, the mean and SEM of frequency for each rating on the life impact scale were linearly related, with R^2=0.99 (Figure 3A). Extrapolation of this function to life impact ratings <5 shows that those selecting 1 (the lowest end of the scale, indicating no impact) would have experienced that symptom at a frequency of <1 day in the previous week. Across all data, life impact was similarly positively correlated with the level of severity selected (where levels of severity were coded from 1 to 5) but less so than with frequency (r=0.32). However, the aggregate mean and SEM of severity for each life impact rating were also linearly related (Figure 3B; R^2=0.99). Finally, in the aggregate, a composite measure of frequency × severity was nonlinearly related to the life impact rating, with R^2=0.98 (Figure 3C). Therefore, we demonstrated a strong relationship between the rating on the life impact scale and both frequency and severity of symptoms.

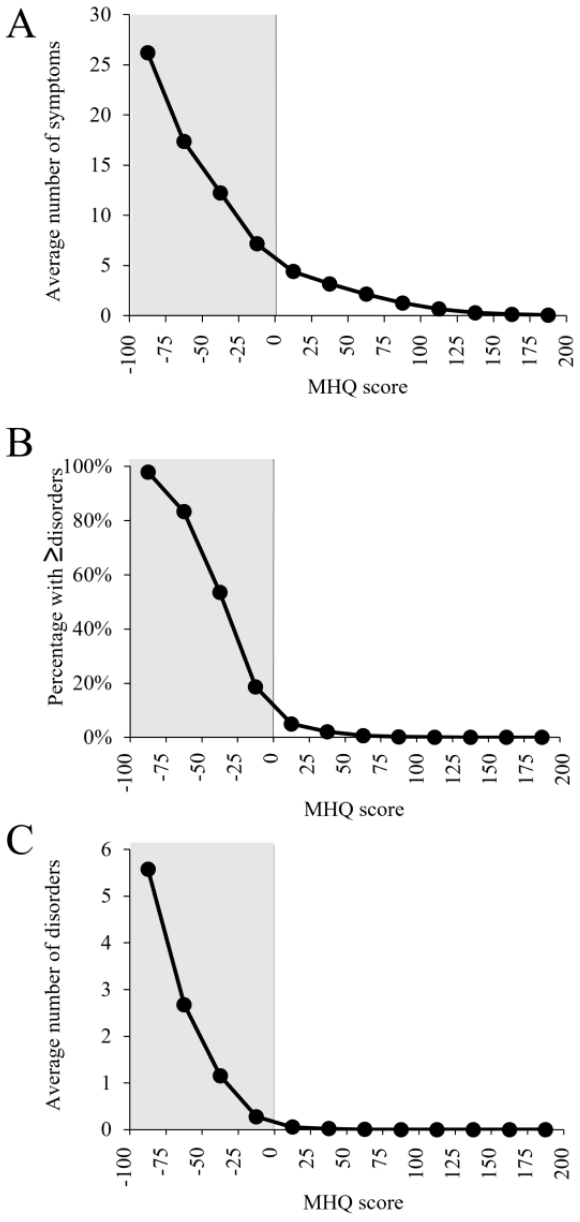
Figure 3. Example relationship between life impact and frequency and severity. (A) The selection on the Mental Health Quotient (MHQ) life impact scale for Feelings of sadness, distress, and hopelessness was linearly related to the frequency of these feelings. (B) The selection on the MHQ life impact scale was similarly linearly related to the severity of the symptom. (C) Frequency × severity was nonlinearly related to the MHQ life impact selection.



Relationship Between MHQ Score and Clinical Burden

First, as would be expected, the average number of clinical symptoms increased as the MHQ score decreased given the nonlinear weighting of symptom severity within the MHQ score (Figure 4A). Beyond this, the percentage of people with clinical symptom profiles that aligned with any of the 10 DSM-5–defined disorder criteria increased as the MHQ score decreased such that 89.08% (8986/10,087) of those in the Distressed category (MHQ score <−50) had symptom profiles that aligned with at least one of the 10 DSM-5–defined disorders, whereas 0.03% (21/70,367) in the categories of Succeeding and Thriving (MHQ score >100) had profiles that aligned with at least one disorder (Figure 4B). Similarly, the number of disorders per individual decreased systematically as MHQ scores increased, with the average number of disorders per person at 3.8 (SD 2.7) for those in the Distressed group and 0.0 (SD 0.02) for those in the Succeeding and Thriving groups (Figure 4C). Thus, the MHQ score is also reflective of the overall clinical burden of mental health.

Figure 4. Relationship between Mental Health Quotient (MHQ) score and clinical symptoms and diagnosis. (A) The average number of symptoms (life impact ≥8 for problem elements and ≤1 for spectrum elements) decreased as the MHQ score increased. The grey area represents the negative side of the scale or the MHQ score categories of Distressed and Struggling (all panels). (B) Approximately 97.83% (3926/4013) of those with the lowest MHQ scores (−100 to −75) mapped to at least one of 10 major clinical disorders and decreased systematically. (C) The average number of disorders per person decreased with MHQ score.

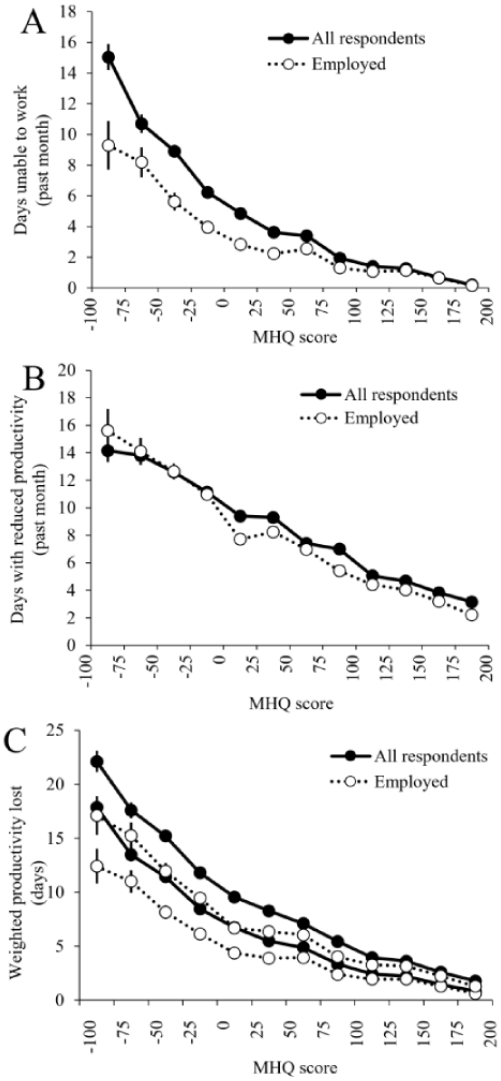


Loss of Function Criterion Validation

As MHQ scores increased, the average number of days of work missed in the past month (Figure 5A) decreased systematically and was best fit by an exponential function, with R^2=0.98. Those in the lowest MHQ score bin (−75 to −100) were unable to work or carry out their daily activities 15.0 (SD 11.3; SEM 0.9) days on average, whereas those who were employed (as opposed to studying, unable to work, unemployed, retired, or occupied with household work; 3306/174,618, 1.89%) were unable to work 9.3 (SD 10.0; SEM 1.6) days on average in the last month. In contrast, those who were in the highest MHQ bin (175-200) lost only an average of 0.2 (SD 1.6; SEM 0.1) days, whereas those who were employed lost an average of 0.2 (SD 0.54; SEM 0.07) days. Furthermore, as MHQ scores increased, the average number of days where people reported not being as productive as usual at work (presenteeism) or in their daily activities decreased linearly (Figure 5B; R^2=0.98 for all respondents and employed respondents alone). Here, those in the lowest MHQ bin (−75 to −100) were not productive an average of 14.2 (SD 11.4; SEM 0.9) days, whereas those employed alone were not productive an average of 15.6 (SD 10.2; SEM 1.6) days. This decreased to an average of 3.2 (SD 9.5; SEM 0.8) and 2.2 (SD 8.5; SEM 1.1) days for all respondents and employed respondents, respectively, in the highest MHQ score bin (175-200). Figure 5C shows the total loss of life productivity as a function of the MHQ score considering both days of work missed and days that were less productive, assuming a range of

20% to 50% loss of productivity on less productive days. Altogether, those with the lowest MHQ scores had an overall reduction in life productivity of anywhere from 18 to 23 days per month on average. Although those with the highest MHQ scores did not often miss a day of work, even this group reported a few unproductive days per month. Thus, MHQ scores are a good representation of behavioral loss of function.

Figure 5. Relationship between Mental Health Quotient (MHQ) score and productivity. (A) The days unable to work in the past month decreased nonlinearly as the MHQ score increased (closed circles, exponential fit, R^2=0.98). Employed people (open circles) with low MHQ scores missed fewer days of work or productive activity. (B) The days in the past month with reduced productivity (presenteeism) decreased linearly as the MHQ score increased (closed circles, exponential fit, R^2=0.98). Employed people (open circles) with low MHQ scores had more days of presenteeism. (C) Total productivity loss for employed (dotted line) and all respondents together (solid line) as a function of the MHQ score (calculated as days missed + n * days with reduced productivity, where n is assumed to be a range between 0.2 [lower dotted or solid line] and 0.5 [upper dotted or solid line]).



Discussion

Principal Findings

In this study, we have demonstrated that the MHQ taken anonymously on the web has excellent sample reliability, internal consistency, and test-retest reliability; that the life impact scale used in the MHQ reflects a combination of both severity and frequency of symptoms; and that the MHQ score relates systematically to clinical burden in the population as well as loss of function from the perspective of days of work missed and loss of productivity. Specifically, the results showed that (1) the MHQ scores were highly similar and statistically indistinguishable between multiple randomly selected, demographically matched samples of respondents; (2) MHQ scores were correlated between retakes with an r=0.84; (3) the life impact rating scale of the MHQ was systematically related to both symptom frequency and severity (R^2=0.99); (4) ratings on related elements were more correlated than unrelated elements; and, finally, (5) MHQ scores decreased systematically with clinical burden and productivity (both R^2=0.98). Thus, the MHQ provides a valid and reliable estimation of population mental health and well-being that, in turn, reflects the clinical burden of mental health and the productive capacity of a

population. Therefore, it is also well-suited to measure changes in the status of mental health and well-being of a population. In addition, the test-retest reliability establishes the MHQ as a useful tool for individuals to track their mental health and well-being trajectory over time.

MHQ and Clinical Burden

Surveillance of population mental health requires assessments that are reliable, valid, and accessible to the general population and that provide a comprehensive profile of mental health and well-being that has clinical and real-world relevance. Currently, many of the assessments used in epidemiological studies evaluate the prevalence of individual disorders rather than overall mental health and well-being. These tools typically consider mental illness as an all-or-nothing phenomenon, raising challenges relating to where the *border* between normal and disordered should lie [42-44] and leading to wide-ranging prevalence estimates that are dependent on the tool used and the thresholds considered as well as on geography and time period [18,45-51]. In addition, focusing only on individual disorders creates a siloed landscape of clinical burden that is at odds with the real-life heterogeneous and comorbid nature of symptomatic experiences and profiles [42,52-62]. Thus, generally, the aggregate burden of clinical-level mental health challenges beyond the domain of individual disorders is unknown in the general population. In this study, we have established the MHQ as a valid and reliable measure of mental health and well-being that can provide a view of overall mental distress and clinical burden. Rolled out at scale as it is currently being actioned as part of the Mental Health Million Project, the MHQ thus provides a solid foundation for the global surveillance of population mental health across different countries. This will help identify relevant risk factors to support the rollout of preventative strategies and the development of interventions or policies that could induce large-scale shifts in population well-being [8-10].

MHQ and Productivity

Over the past few decades, there has been mounting evidence supporting the relationship between mental health and well-being and productivity [19-25] as well as the resultant economic loss to society as a consequence of days lost and unproductive days (eg, presenteeism) [28-31]. With the increased prevalence of mental distress as a result of the COVID-19 pandemic [16-18] and increased levels of burnout in the population [26,27], there is a need to better understand the relationship between mental health and well-being and productivity in the general population. The systematic relationship observed between the spectrum of MHQ scores from Distressed to Thriving and productivity loss along with the general reliability and validity of the MHQ support its use as an assessment of the productive capacity of a population independent of any disorder classification. It also positions the MHQ as an important tool for companies to assess the mental health and well-being of their workforce, providing relevant metrics that can help them address challenges such as employee burnout and work-home imbalance [26,27], as well as for university student bodies, where young adults are disproportionally affected by mental health challenges [63-65]. This will allow and encourage organizations and institutions to be more strategic in their management of mental health and well-being.

Limitations and Future Directions

It is important that we acknowledge some limitations of these data and study. First, the validation of the life impact rating against symptom frequency and severity was performed for a single MHQ element (*Feelings of sadness, distress, or hopelessness*). However, it is possible that the correspondence between frequency or severity and life impact rating may differ from element to element. Furthermore, these results were used to select an appropriate threshold value for clinical significance [42] to determine clinical burden, indicating that a threshold of 8 was equivalent, on average, to experiencing the symptom 5 days per week. However, it could be the case that other threshold values may have been more appropriate for other elements.

Second, the mapping of MHQ elements to DSM-5 diagnostic criteria was constrained by the presence of broad or imperfect matches for certain symptoms pertaining to OCD and bipolar disorder that could have affected the accuracy of the mapping [42]. For example, for bipolar disorder, symptoms denoting extreme versions of positive assets (eg, *grandiosity and decreased need for sleep*) were not fully articulated within the MHQ, whereas, for OCD, the MHQ elements were broader (eg, obsessive thoughts were incorporated within a general element reflecting strange, unwanted, and obsessive thoughts). Furthermore, a specific criterion of symptom timing was not included as this is not included in the MHQ, which assesses an individual's current perception.

Third, in the future, it will be important to compare the MHQ outcomes with more commonly used assessments (eg, mapping against the 9-item Patient Health Questionnaire [66] and the 7-item Generalized Anxiety Disorder scale [67]) to determine the alignment between DSM-5–mapped MHQ symptom profiles for depression and anxiety and the scores from these 2 questionnaires, respectively. Triangulating data arising from the Mental Health Million Project against other external data metrics that provide insight into clinical burden (eg, quality-adjusted life years or disability-adjusted life years) will also be important.

Altogether, the MHQ supports a valid and reliable monitoring of population mental health and well-being. As the MHQ continues to underpin large-scale initiatives such as the Mental Health Million Project, it will provide deeper insights into social determinants and the societal impact of changes in mental health and well-being. These insights can, in turn, enable preventative strategies for better management of global mental health and well-being.

Acknowledgments

This work was supported by internal funding from Sapien Labs.

Authors' Contributions

JJN and TCT conceptualized, designed, and led the study. JJN and VP performed the analysis. JJN and TCT interpreted the data and drafted the manuscript. JJN created the figures. TCT and JJN made critical revisions. JJN, TCT, and VP approved the final version of the manuscript and agreed to be accountable for all aspects of the work.

Conflicts of Interest

TCT received a grant award from the National Institute of Mental Health to develop a commercial version of the Mental Health Quotient tool referenced herein.

References

1. Promoting mental health: concepts, emerging evidence, practice: a report of the World Health Organization, Department of Mental Health and Substance Abuse in collaboration with the Victorian Health Promotion Foundation and the University of Melbourne. World Health Organization. 2005. URL: https://www.who.int/mental_health/evidence/en/promoting_mhh.pdf [accessed 2021-09-01]
2. Cloninger CR. The science of well being: an integrated approach to mental health and its disorders. Psychiatr Danub 2006;18(3-4):218-224. [Medline: 17183779]
3. Slade M. Mental illness and well-being: the central importance of positive psychology and recovery approaches. BMC Health Serv Res 2010;10:26 [FREE Full text] [doi: 10.1186/1472-6963-10-26] [Medline: 20102609]
4. Bohlmeijer E, Westerhof G. The model for sustainable mental health: future directions for integrating positive psychology into mental health care. Front Psychol 2021;12:747999 [FREE Full text] [doi: 10.3389/fpsyg.2021.747999] [Medline: 34744925]
5. Schotanus-Dijkstra M, Ten Have M, Lamers SM, de Graaf R, Bohlmeijer ET. The longitudinal relationship between flourishing mental health and incident mood, anxiety and substance use disorders. Eur J Public Health 2017;27(3):563-568. [doi: 10.1093/eurpub/ckw202] [Medline: 27818372]
6. American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th edition. Washington, D.C: American Psychiatric Association; 2013.
7. International statistical classification of diseases and related health problems. 11th revision. World Health Organization. 2018. URL: https://www.who.int/standards/classifications/classification-of-diseases [accessed 2021-09-01]
8. Purtle J, Nelson KL, Counts NZ, Yudell M. Population-based approaches to mental health: history, strategies, and evidence. Annu Rev Public Health 2020;41:201-221 [FREE Full text] [doi: 10.1146/annurev-publhealth-040119-094247] [Medline: 31905323]
9. Fuhrer R, Keyes KM. Population mental health in the 21st century: time to act. Am J Public Health 2019;109(S3):S152-S153. [doi: 10.2105/AJPH.2019.305200] [Medline: 31242014]
10. Huppert FA. A new approach to reducing disorder and improving well-being. Perspect Psychol Sci 2009;4(1):108-111. [doi: 10.1111/j.1745-6924.2009.01100.x] [Medline: 26158844]
11. Dattani S, Ritchie H, Roser M. Mental health. Our World In Data. 2021. URL: https://ourworldindata.org/mental-health [accessed 2021-09-01]
12. Friedrich MJ. Depression is the leading cause of disability around the world. JAMA 2017;317(15):1517. [doi: 10.1001/jama.2017.3826] [Medline: 28418490]
13. Suicide worldwide in 2019: global health estimates. World Health Organization. 2021. URL: https://www.who.int/publications/i/item/9789240026643 [accessed 2021-09-01]
14. Orsolini L, Latini R, Pompili M, Serafini G, Volpe U, Vellante F, et al. Understanding the complex of suicide in depression: from research to clinics. Psychiatry Investig 2020;17(3):207-221 [FREE Full text] [doi: 10.30773/pi.2019.0171] [Medline: 32209966]
15. De Berardis D, Fornaro M, Valchera A, Cavuto M, Perna G, Di Nicola M, et al. Eradicating suicide at its roots: preclinical bases and clinical evidence of the efficacy of ketamine in the treatment of suicidal behaviors. Int J Mol Sci 2018;19(10):2888 [FREE Full text] [doi: 10.3390/ijms19102888] [Medline: 30249029]
16. Pierce M, Hope H, Ford T, Hatch S, Hotopf M, John A, et al. Mental health before and during the COVID-19 pandemic: a longitudinal probability sample survey of the UK population. Lancet Psychiatry 2020;7(10):883-892 [FREE Full text] [doi: 10.1016/S2215-0366(20)30308-4] [Medline: 32707037]
17. Serafini G, Parmigiani B, Amerio A, Aguglia A, Sher L, Amore M. The psychological impact of COVID-19 on the mental health in the general population. QJM 2020;113(8):531-537 [FREE Full text] [doi: 10.1093/qjmed/hcaa201] [Medline: 32569360]
18. Xiong J, Lipsitz O, Nasri F, Lui LM, Gill H, Phan L, et al. Impact of COVID-19 pandemic on mental health in the general population: a systematic review. J Affect Disord 2020;277:55-64 [FREE Full text] [doi: 10.1016/j.jad.2020.08.001] [Medline: 32799105]
19. Kessler RC, Frank RG. The impact of psychiatric disorders on work loss days. Psychol Med 1997;27(4):861-873. [doi: 10.1017/s0033291797004807] [Medline: 9234464]

20. Alonso J, Petukhova M, Vilagut G, Chatterji S, Heeringa S, Üstün TB, et al. Days out of role due to common physical and mental conditions: results from the WHO World Mental Health surveys. Mol Psychiatry 2011;16(12):1234-1246 [FREE Full text] [doi: 10.1038/mp.2010.101] [Medline: 20938433]
21. Burton WN, Schultz AB, Chen C, Edington DW. The association of worker productivity and mental health: a review of the literature. Int J Workplace Health Manag 2008;1(2):78-94. [doi: 10.1108/17538350810893883]
22. Hemp P. Presenteeism: at work--but out of it. Harv Bus Rev 2004;82(10):49-58. [Medline: 15559575]
23. Lim D, Sanderson K, Andrews G. Lost productivity among full-time workers with mental disorders. J Ment Health Policy Econ 2000;3(3):139-146. [doi: 10.1002/mhp.93] [Medline: 11967449]
24. Evans-Lacko S, Knapp M. Global patterns of workplace productivity for people with depression: absenteeism and presenteeism costs across eight diverse countries. Soc Psychiatry Psychiatr Epidemiol 2016;51(11):1525-1537 [FREE Full text] [doi: 10.1007/s00127-016-1278-4] [Medline: 27667656]
25. Bubonya M, Cobb-Clark D, Wooden M. Mental health and productivity at work: does what you do matter? Labour Econ 2017;46(C):150-165. [doi: 10.1016/j.labeco.2017.05.001]
26. Threlkeld K. Employee burnout report: COVID-19's impact and 3 strategies to curb it. Indeed. 2021. URL: https://www.indeed.com/lead/preventing-employee-burnout-report [accessed 2022-01-12]
27. Employee burnout is ubiquitous, alarming—and still underreported. McKinsey. 2021. URL: https://www.mckinsey.com/featured-insights/coronavirus-leading-through-the-crisis/charting-the-path-to-the-next-normal/employee-burnout-is-ubiquitous-alarming-and-still-underreported [accessed 2021-01-12]
28. Marcotte DE, Wilcox-Gök V. Estimating the employment and earnings costs of mental illness: recent developments in the United States. Soc Sci Med 2001;53(1):21-27. [doi: 10.1016/s0277-9536(00)00312-9] [Medline: 11386306]
29. Stewart WF, Ricci JA, Chee E, Hahn SR, Morganstein D. Cost of lost productive work time among US workers with depression. JAMA 2003;289(23):3135-3144. [doi: 10.1001/jama.289.23.3135] [Medline: 12813119]
30. Trautmann S, Rehm J, Wittchen HU. The economic costs of mental disorders: do our societies react appropriately to the burden of mental disorders? EMBO Rep 2016;17(9):1245-1249 [FREE Full text] [doi: 10.15252/embr.201642951] [Medline: 27491723]
31. Whiteford H. The productivity commission inquiry into mental health. Aust N Z J Psychiatry 2021:48674211031159. [doi: 10.1177/00048674211031159] [Medline: 34259051]
32. Chisholm D, Sweeny K, Sheehan P, Rasmussen B, Smit F, Cuijpers P, et al. Scaling-up treatment of depression and anxiety: a global return on investment analysis. Lancet Psychiatry 2016;3(5):415-424 [FREE Full text] [doi: 10.1016/S2215-0366(16)30024-4] [Medline: 27083119]
33. Newson JJ, Thiagarajan TC. Assessment of population well-being with the mental health quotient (MHQ): development and usability study. JMIR Ment Health 2020;7(7):e17935 [FREE Full text] [doi: 10.2196/17935] [Medline: 32706730]
34. Newson JJ, Hunter D, Thiagarajan TC. The heterogeneity of mental health assessment. Front Psychiatry 2020;11:76 [FREE Full text] [doi: 10.3389/fpsyt.2020.00076] [Medline: 32174852]
35. First MB, Williams JB, Karg RS, Spitzer RL. Structured clinical interview for DSM-5 disorders, clinician version (SCID-5-CV). Arlington, VA: American Psychiatric Association; 2016.
36. Cuthbert BN. The role of RDoC in future classification of mental disorders . Dialogues Clin Neurosci 2020;22(1):81-85 [FREE Full text] [doi: 10.31887/DCNS.2020.22.1/bcuthbert] [Medline: 32699508]
37. Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. Am J Psychiatry 2010;167(7):748-751. [doi: 10.1176/appi.ajp.2010.09091379] [Medline: 20595427]
38. Insel TR. The NIMH research domain criteria (RDoC) project: precision medicine for psychiatry. Am J Psychiatry 2014;171(4):395-397. [doi: 10.1176/appi.ajp.2014.14020138] [Medline: 24687194]
39. Newson JJ, Thiagarajan T. Dynamic dataset of global population mental wellbeing. 2021. URL: https://sapienlabs.org/wp-content/uploads/2021/06/NewsonThiagarajan_2021_Dynamic-Dataset-of-Global-Population-Mental-Wellbeing.pdf [accessed 2021-09-01]
40. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Meas 1977;1(3):385-401. [doi: 10.1177/014662167700100306]
41. Eaton WM, Smith C, Ybarra M, Muntaner C, Tien A. Center for Epidemiologic Studies Depression Scale: review and revision (CESD and CESD-R). In: Maruish ME, editor. The use of psychological testing for treatment planning and outcomes assessment: volume 3: instruments for adults. 3rd edition. Mahwah, NJ: Lawrence Erlbaum Associates; 2004:363-377.
42. Newson JJ, Pastukh V, Thiagarajan TC. Poor separation of clinical symptom profiles by DSM-5 disorder criteria. Front Psychiatry 2021;12:775762 [FREE Full text] [doi: 10.3389/fpsyt.2021.775762] [Medline: 34916976]
43. Rössler W. What is normal? The impact of psychiatric classification on mental health practice and research. Front Public Health 2013;1:68 [FREE Full text] [doi: 10.3389/fpubh.2013.00068] [Medline: 24350236]
44. Wakefield JC. Diagnostic issues and controversies in DSM-5: return of the false positives problem. Annu Rev Clin Psychol 2016;12:105-132. [doi: 10.1146/annurev-clinpsy-032814-112800] [Medline: 26772207]

45. Wu T, Jia X, Shi H, Niu J, Yin X, Xie J, et al. Prevalence of mental health problems during the COVID-19 pandemic: a systematic review and meta-analysis. J Affect Disord 2021;281:91-98 [FREE Full text] [doi: 10.1016/j.jad.2020.11.117] [Medline: 33310451]
46. Cénat JM, Blais-Rochette C, Kokou-Kpolou CK, Noorishad PG, Mukunzi JN, McIntee SE, et al. Prevalence of symptoms of depression, anxiety, insomnia, posttraumatic stress disorder, and psychological distress among populations affected by the COVID-19 pandemic: a systematic review and meta-analysis. Psychiatry Res 2021;295:113599 [FREE Full text] [doi: 10.1016/j.psychres.2020.113599] [Medline: 33285346]
47. Baxter AJ, Scott KM, Vos T, Whiteford HA. Global prevalence of anxiety disorders: a systematic review and meta-regression. Psychol Med 2013;43(5):897-910. [doi: 10.1017/S003329171200147X] [Medline: 22781489]
48. Steel Z, Marnane C, Iranpour C, Chey T, Jackson JW, Patel V, et al. The global prevalence of common mental disorders: a systematic review and meta-analysis 1980-2013. Int J Epidemiol 2014;43(2):476-493 [FREE Full text] [doi: 10.1093/ije/dyu038] [Medline: 24648481]
49. Lim GY, Tam WW, Lu Y, Ho CS, Zhang MW, Ho RC. Prevalence of depression in the community from 30 countries between 1994 and 2014. Sci Rep 2018;8(1):2861 [FREE Full text] [doi: 10.1038/s41598-018-21243-x] [Medline: 29434331]
50. McGrath J, Saha S, Welham J, El Saadi O, MacCauley C, Chant D. A systematic review of the incidence of schizophrenia: the distribution of rates and the influence of sex, urbanicity, migrant status and methodology. BMC Med 2004;2:13 [FREE Full text] [doi: 10.1186/1741-7015-2-13] [Medline: 15115547]
51. Atwoli L, Stein DJ, Koenen KC, McLaughlin KA. Epidemiology of posttraumatic stress disorder: prevalence, correlates and consequences. Curr Opin Psychiatry 2015;28(4):307-311 [FREE Full text] [doi: 10.1097/YCO.0000000000000167] [Medline: 26001922]
52. Angold A, Costello EJ, Erkanli A. Comorbidity. J Child Psychol Psychiatry 1999;40(1):57-87. [Medline: 10102726]
53. Kessler RC, Nelson CB, McGonagle KA, Liu J, Swartz M, Blazer DG. Comorbidity of DSM-III-R major depressive disorder in the general population: results from the US National Comorbidity Survey. Br J Psychiatry Suppl 1996(30):17-30. [Medline: 8864145]
54. Maj M. "Psychiatric comorbidity": an artefact of current diagnostic systems? Br J Psychiatry 2005;186:182-184. [doi: 10.1192/bjp.186.3.182] [Medline: 15738496]
55. Feczko E, Miranda-Dominguez O, Marr M, Graham AM, Nigg JT, Fair DA. The heterogeneity problem: approaches to identify psychiatric subtypes. Trends Cogn Sci 2019;23(7):584-601 [FREE Full text] [doi: 10.1016/j.tics.2019.03.009] [Medline: 31153774]
56. Fried EI, Nesse RM. Depression is not a consistent syndrome: an investigation of unique symptom patterns in the STAR*D study. J Affect Disord 2015;172:96-102 [FREE Full text] [doi: 10.1016/j.jad.2014.10.010] [Medline: 25451401]
57. Caspi A, Houts RM, Ambler A, Danese A, Elliott ML, Hariri A, et al. Longitudinal assessment of mental health disorders and comorbidities across 4 decades among participants in the Dunedin birth cohort study. JAMA Netw Open 2020;3(4):e203221 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.3221] [Medline: 32315069]
58. McElroy E, Patalay P. In search of disorders: internalizing symptom networks in a large clinical sample. J Child Psychol Psychiatry 2019;60(8):897-906 [FREE Full text] [doi: 10.1111/jcpp.13044] [Medline: 30900257]
59. Plana-Ripoll O, Pedersen CB, Holtz Y, Benros ME, Dalsgaard S, de Jonge P, et al. Exploring comorbidity within mental disorders among a Danish national population. JAMA Psychiatry 2019;76(3):259-270 [FREE Full text] [doi: 10.1001/jamapsychiatry.2018.3658] [Medline: 30649197]
60. Buch AM, Liston C. Dissecting diagnostic heterogeneity in depression by integrating neuroimaging and genetics. Neuropsychopharmacology 2021;46(1):156-175 [FREE Full text] [doi: 10.1038/s41386-020-00789-3] [Medline: 32781460]
61. Luo Y, Weibman D, Halperin JM, Li X. A review of heterogeneity in attention deficit/hyperactivity disorder (ADHD). Front Hum Neurosci 2019;13:42 [FREE Full text] [doi: 10.3389/fnhum.2019.00042] [Medline: 30804772]
62. Masi A, DeMayo MM, Glozier N, Guastella AJ. An overview of autism spectrum disorder, heterogeneity and treatment options. Neurosci Bull 2017;33(2):183-193 [FREE Full text] [doi: 10.1007/s12264-017-0100-y] [Medline: 28213805]
63. Kessler RC, Amminger GP, Aguilar-Gaxiola S, Alonso J, Lee S, Ustün TB. Age of onset of mental disorders: a review of recent literature. Curr Opin Psychiatry 2007;20(4):359-364 [FREE Full text] [doi: 10.1097/YCO.0b013e32816ebc8c] [Medline: 17551351]
64. Twenge JM, Cooper AB, Joiner TE, Duffy ME, Binau SG. Age, period, and cohort trends in mood disorder indicators and suicide-related outcomes in a nationally representative dataset, 2005-2017. J Abnorm Psychol 2019;128(3):185-199. [doi: 10.1037/abn0000410] [Medline: 30869927]
65. Newson JJ, Pastukh V, Sukhoi O, Taylor J, Thiagarajan TC. Mental state of the world 2020. Mental Health Million Project, Sapien Labs. 2021. URL: https://sapienlabs.org/wp-content/uploads/2021/03/Mental-State-of-the-World-Report-2020-1.pdf [accessed 2021-09-01]
66. Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary care evaluation of mental disorders. Patient health questionnaire. JAMA 1999;282(18):1737-1744. [doi: 10.1001/jama.282.18.1737] [Medline: 10568646]
67. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]

Abbreviations

DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
MHQ: Mental Health Quotient
OCD: obsessive-compulsive disorder
SEM: standard error of measurement

Edited by J Torous; submitted 07.10.21; peer-reviewed by G Serafini, S Six, L Ospina-Pinillos; comments to author 29.11.21; revised version received 14.01.22; accepted 24.01.22; published 20.04.22.

Please cite as:
Newson JJ, Pastukh V, Thiagarajan TC
Assessment of Population Well-being With the Mental Health Quotient: Validation Study
JMIR Ment Health 2022;9(4):e34105
URL: https://mental.jmir.org/2022/4/e34105
doi:10.2196/34105
PMID:35442210

Original Paper

Characterizing Use of a Multicomponent Digital Intervention to Predict Treatment Outcomes in First-Episode Psychosis: Cluster Analysis

Shaunagh O'Sullivan[1,2], BA, MPsychSc; Lianne Schmaal[1,2], BA, MSc, PhD; Simon D'Alfonso[1,3], BASc (Hons), PhD; Yara Jo Toenders[1,2], MSc; Lee Valentine[1,2], BA, MSocWk; Carla McEnery[1,2], BPsychSc (Hons), PhD; Sarah Bendall[1,2], BA, MA (Hons), PGDipClinPsych, PhD; Barnaby Nelson[1,2], MPsych, PhD; John F Gleeson[4,5], BA (Hons), MPsych, PhD; Mario Alvarez-Jimenez[1,2], BSc (Hons), MAResearchMeth, PhD, DClinPsy

[1]Orygen, Parkville, Australia

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

[3]School of Computing and Information Systems, University of Melbourne, Melbourne, Australia

[4]Health Brain and Mind Research Centre, Australian Catholic University, Melbourne, Australia

[5]School 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 and 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].

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.

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

Therapy-related variables

- Number of steps started (*steps* refer to the intervention modules)
- Number of actions done (*actions* refer to the activities that comprise a step)
- Visited suggested content (*suggested content* refers to the therapeutic content recommended by clinical moderators)
- Visited therapy (*visiting therapy* refers to visiting the homepage of the therapy component of the intervention)

Social networking–related variables

- Number of newsfeed posts (*newsfeed* refers to the social network)
- Number of newsfeed comments
- Number of Talk it Out posts (*Talk it Out* refers to a problem-solving forum run by peer moderators)
- Number of likes made
- Number of reactions made (*reactions* refer to short support messages in response to a post, eg, "thinking of you")
- Visited messages (*messages* refer to a private message section where moderators could contact participants directly)
- Visited notifications
- Visited newsfeed
- Visited Talk it Out

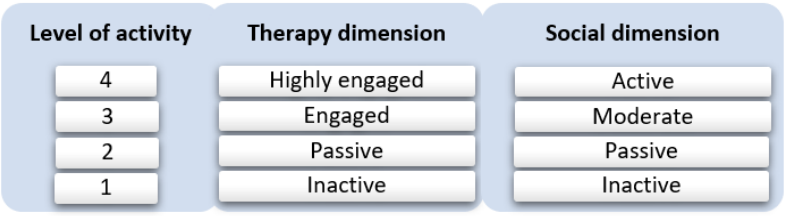
Therapeutic *Pathways* were divided into themes including understanding psychosis, identifying early warning signs to prevent relapse, identifying and exercising personal strengths, promoting social connections and positive emotions, and managing stress, anxiety, and depression. To increase usability, *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." Furthermore, participants could use the *Talk It Out* function to nominate relevant issues to discuss in a moderated forum, informed by an evidence-based problem-solving framework [43]. Participants received notifications when other users communicated on the social network. Participants also received private messages when a moderator contacted them directly via the platform. See Multimedia Appendix 2 for an example of a newsfeed post with *likes* and *reactions* on the Horyzons social network.

Daily Activity Categories

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.

Figure 1. Hierarchical 2D daily activity categories.



Statistical Analyses

Identifying User Profiles

K-means clustering for joint trajectories was implemented using the *R* package *kml3d* (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 (χ^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

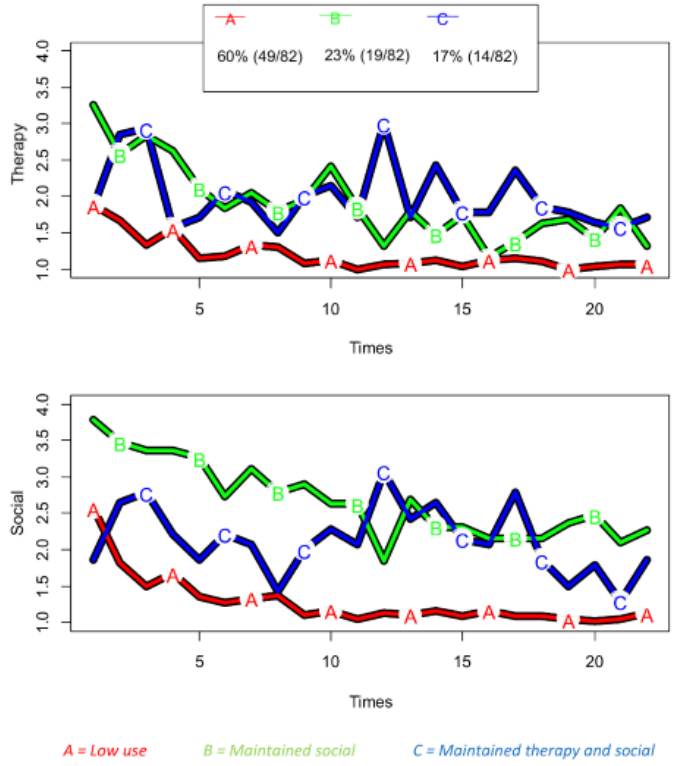
Clusters Based on Joint Trajectories of System Use

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.

	User profiles			Group × time interaction, *F* test (*df*)	*P* value
	Low use (n=49), mean (SD)	Maintained social (n=19), mean (SD)	Maintained therapy and social (n=14), mean (SD)		
PSP[a]					
Baseline	67.09 (1.97)	70.17 (3.15)	65.31 (3.64)	N/A[b]	N/A
6 months	68.64 (2.62)	67.11 (3.15)	76.31 (3.64)	3.58 (2, 51)	.04
FESFS[c] independent living skills					
Baseline	13.66 (0.33)	14.07 (0.52)	13.11 (0.60)	N/A	N/A
6 months	13.59 (0.41)	14.04 (0.54)	13.03 (0.62)	0.003 (2, 45)	.99
FESFS interacting with people					
Baseline	12.96 (0.35)	12.52 (0.56)	11.90 (0.64)	N/A	N/A
6 months	13.26 (0.44)	12.09 (0.58)	12.27 (0.66)	0.87 (2, 45)	.42
FESFS friends and activities					
Baseline	18.96 (0.47)	17.50 (0.75)	17.61 (0.86)	N/A	N/A
6 months	19.53 (0.62)	17.70 (0.78)	18.06 (0.89)	0.08 (2, 46)	.92
FESFS intimacy					
Baseline	15.37 (0.45)	15.22 (0.75)	.52 (0.85)	N/A	N/A
6 months	15.24 (0.58)	14.77 (0.75)	13.52 (0.90)	0.43 (2, 40)	.65
PANSS[d] total					
Baseline	45.15 (1.77)	43.69 (2.84)	48.91 (3.27)	N/A	N/A
6 months	44.96 (2.22)	46.80 (2.84)	41.73 (3.27)	3.23 (2, 50)	.048
PANSS positive					
Baseline	10.49 (0.54)	10.38 (0.86)	10.31 (1.00)	N/A	N/A
6 months	10.29 (0.70)	11.31 (0.86)	9.89 (1.00)	0.60 (2, 52)	.55
PANSS negative					
Baseline	10.73 (0.55)	11.24 (0.89)	13.62 (1.03)	N/A	N/A
6 months	10.02 (0.72)	10.55 (0.89)	9.28 (1.03)	4.45 (2, 51)	.02
PANSS general psychopathology					
Baseline	23.93 (1.06)	22.07 (1.70)	24.99 (1.96)	N/A	N/A
6 months	24.65 (1.33)	24.92 (1.70)	22.56 (1.96)	2.34 (2, 50)	.11
CDSS[e]					
Baseline	3.91 (0.67)	3.31 (1.08)	2.78 (1.25)	N/A	N/A
6 months	4.41 (0.84)	3.84 (1.08)	2.20 (1.25)	0.33 (2, 50)	.72
DASS[f] anxiety					
Baseline	12.14 (1.35)	6.79 (2.02)	10.15 (2.45)	N/A	N/A
6 months	12.02 (1.77)	10.58 (2.18)	9.42 (2.53)	1.35 (2, 42)	.27

[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.

Figure 3. Mean trends in PSP social functioning scores for user profiles (95% CIs). PSP: Personal and Social Performance Scale.

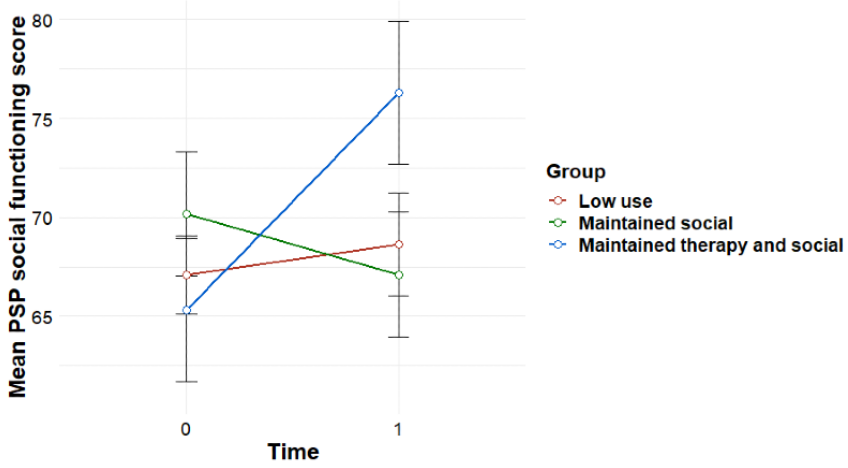


Figure 4. Mean trends in PANSS overall psychiatric symptom severity scores for user profiles (95% CIs). PANSS: Positive and Negative Syndrome Scale.

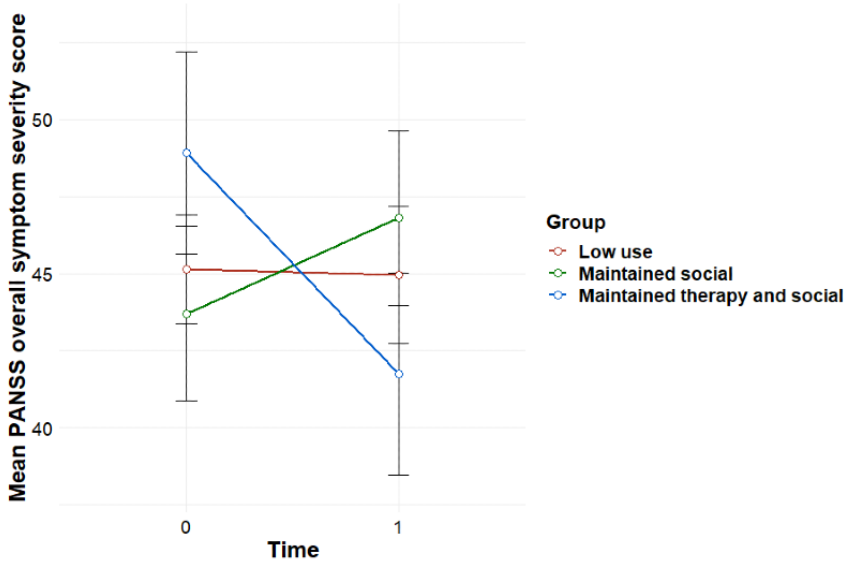
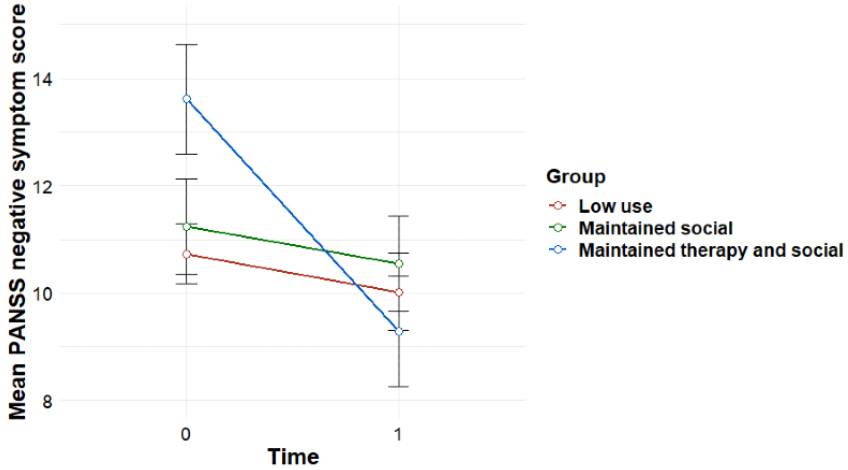


Figure 5. Mean trends in PANSS negative symptom scores for user profiles (95% CIs). PANSS: Positive and Negative Syndrome Scale.



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[a] groups.

	TAU (n=84), mean (SD)	Maintained therapy and social (n=19), mean (SD)	Group × time interaction, *F* test (*df*)	*P* value
PSP[b]				
Baseline	65.19 (1.57)	65.79 (3.82)	N/A[c]	N/A
6 months	66.89 (1.80)	76.79 (3.82)	4.68 (1, 62)	.03
FESFS[d] independent living skills				
Baseline	13.71 (0.21)	13.12 (0.50)	N/A	N/A
6 months	13.71 (0.24)	13.04 (0.51)	0.02 (1, 60)	.88
FESFS interacting with people				
Baseline	12.78 (0.24)	11.94 (0.58)	N/A	N/A
6 months	12.62 (0.27)	12.31 (0.59)	0.99 (1, 57)	.32
FESFS friends and activities				
Baseline	18.47 (0.43)	17.74 (1.00)	N/A	N/A
6 months	18.25 (0.47)	18.20 (1.02)	0.60 (1, 55)	.44
FESFS intimacy				
Baseline	14.82 (0.39)	14.56 (0.92)	N/A	N/A
6 months	14.77 (0.42)	13.47 (0.96)	1.63 (1, 52)	.20
PANSS[e] total				
Baseline	44.26 (1.37)	48.57 (3.32)	N/A	N/A
6 months	45.72 (1.54)	41.40 (3.32)	5.66 (1, 63)	.02
PANSS positive				
Baseline	9.47 (0.43)	10.24 (1.05)	N/A	N/A
6 months	9.72 (0.50)	9.82 (1.05)	0.24 (1, 64)	.62
PANSS negative				
Baseline	11.00 (0.43)	13.42 (1.05)	N/A	N/A
6 months	10.64 (0.48)	9.09 (1.05)	14.61 (1, 62)	<.001
PANSS general psychopathology				
Baseline	23.78 (0.83)	24.92 (2.01)	N/A	N/A
6 months	25.39 (0.95)	22.50 (2.01)	2.76 (1, 63)	.10
CDSS[f]				
Baseline	2.72 (0.41)	2.86 (0.99)	N/A	N/A
6 months	3.46 (0.46)	2.29 (0.99)	1.56 (1, 62)	.22
DASS[g] anxiety				
Baseline	12.35 (1.16)	8.85 (2.80)	N/A	N/A
6 months	9.00 (1.32)	8.21 (2.89)	0.76 (1, 52)	.39

[a]TAU: treatment as usual.

[b]PSP: Personal and Social Performance Scale.

[c]N/A: not applicable.

[d]FESFS: First Episode Social Functioning Scale.

[e]PANSS: Positive and Negative Syndrome Scale.

[f]CDSS: Calgary Depression Scale for Schizophrenia.

[g]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.

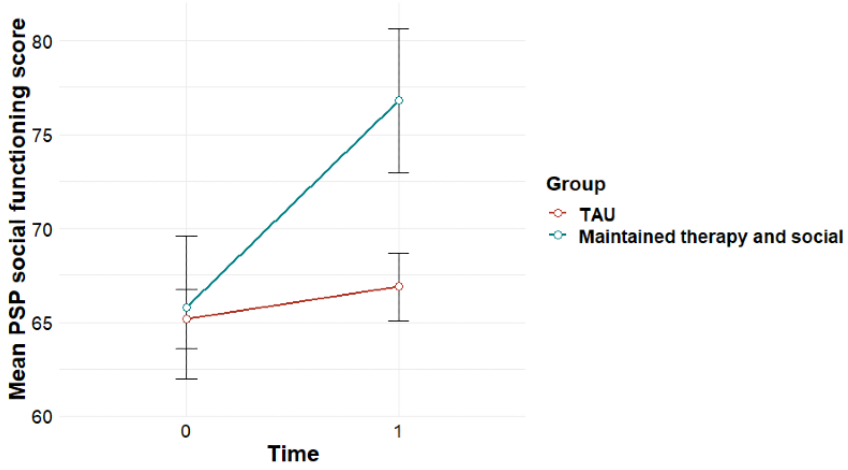


Figure 7. Mean trends in PANSS overall psychiatric symptom severity scores for the maintained therapy and social and TAU groups (95% CIs). PANSS: Positive and Negative Syndrome Scale; TAU: treatment as usual.

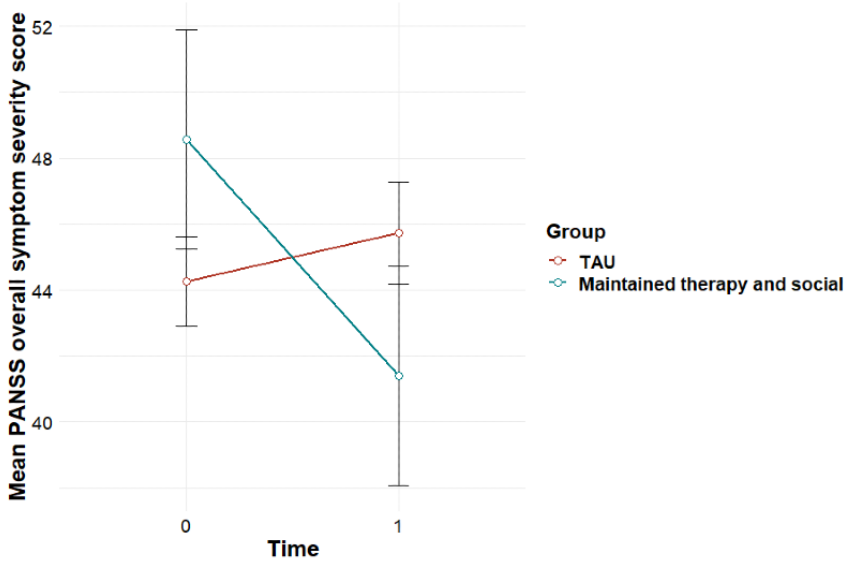


Figure 8. Mean trends in PANSS negative symptom scores for the maintained therapy and social and TAU groups (95% CIs). PANSS: Positive and Negative Syndrome Scale; TAU: treatment as usual.

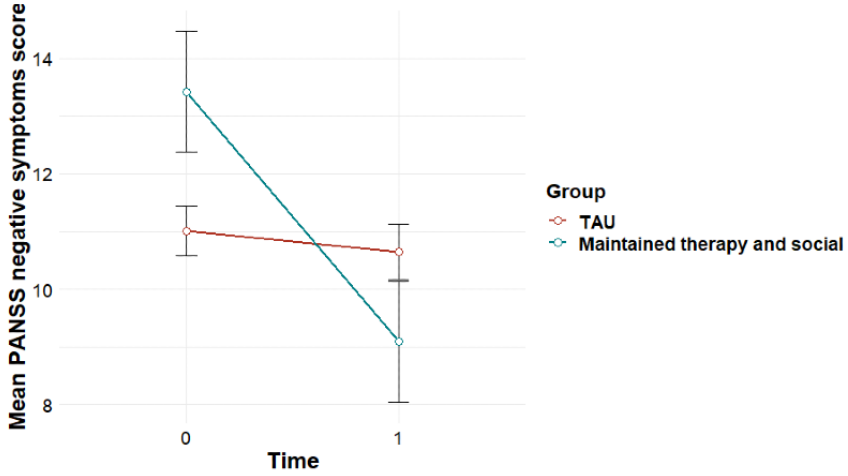
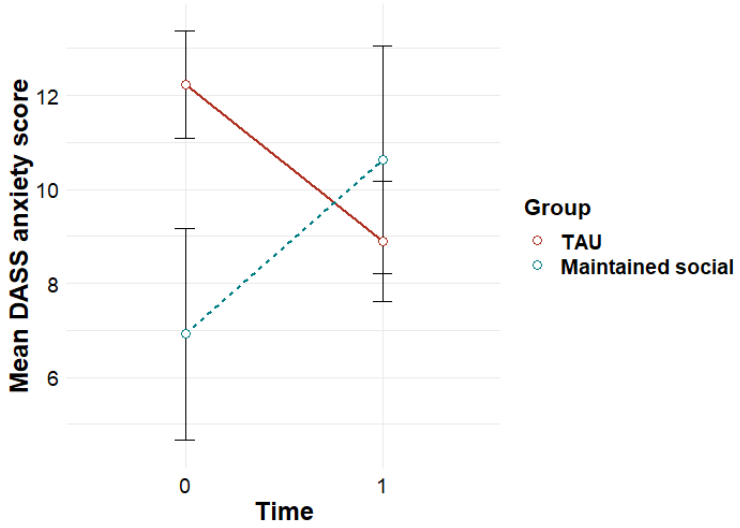


Figure 9. Mean trends in DASS anxiety scores for the maintained social and TAU groups (95% CIs). DASS: Depression Anxiety and Stress Scale; TAU: treatment as usual.



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

1. Alvarez-Jimenez M, Bendall S, Lederman R, Wadley G, Chinnery G, Vargas S, et al. On the HORYZON: moderated online social therapy for long-term recovery in first episode psychosis. Schizophr Res 2013 Jan;143(1):143-149. [doi: 10.1016/j.schres.2012.10.009] [Medline: 23146146]
2. Nordentoft M, Rasmussen JO, Melau M, Hjorthøj CR, Thorup AA. How successful are first episode programs? A review of the evidence for specialized assertive early intervention. Curr Opin Psychiatry 2014 May;27(3):167-172. [doi: 10.1097/YCO.0000000000000052] [Medline: 24662959]
3. Addington D, Addington MD, Patten S. Relapse rates in an early psychosis treatment service. Acta Psychiatr Scand 2007 Feb;115(2):126-131. [doi: 10.1111/j.1600-0447.2006.00879.x] [Medline: 17244176]
4. Alvarez-Jimenez M, Priede A, Hetrick S, Bendall S, Killackey E, Parker A, et al. Risk factors for relapse following treatment for first episode psychosis: a systematic review and meta-analysis of longitudinal studies. Schizophr Res 2012 Aug;139(1-3):116-128. [doi: 10.1016/j.schres.2012.05.007] [Medline: 22658527]

5. Bertelsen M, Jeppesen P, Petersen L, Thorup A, Øhlenschlaeger J, Le Quach P, et al. Course of illness in a sample of 265 patients with first-episode psychosis--five-year follow-up of the Danish OPUS trial. Schizophr Res 2009 Feb;107(2-3):173-178. [doi: 10.1016/j.schres.2008.09.018] [Medline: 18945593]
6. Gafoor R, Nitsch D, McCrone P, Craig TK, Garety PA, Power P, et al. Effect of early intervention on 5-year outcome in non-affective psychosis. Br J Psychiatry 2010 May 02;196(5):372-376. [doi: 10.1192/bjp.bp.109.066050] [Medline: 20435962]
7. Malla A, Joober R, Iyer S, Norman R, Schmitz N, Brown T, et al. Comparing three-year extension of early intervention service to regular care following two years of early intervention service in first-episode psychosis: a randomized single blind clinical trial. World Psychiatry 2017 Oct 21;16(3):278-286 [FREE Full text] [doi: 10.1002/wps.20456] [Medline: 28941098]
8. Albert N, Melau M, Jensen H, Emborg C, Jepsen JR, Fagerlund B, et al. Five years of specialised early intervention versus two years of specialised early intervention followed by three years of standard treatment for patients with a first episode psychosis: randomised, superiority, parallel group trial in Denmark (OPUS II). Br Med J 2017 Feb 24;356:j1015 [FREE Full text] [doi: 10.1136/bmj.j1015] [Medline: 28235902]
9. Alvarez-Jimenez M, O'Donoghue B, Thompson A, Gleeson JF, Bendall S, Gonzalez-Blanch C, et al. Beyond clinical remission in first episode psychosis: thoughts on antipsychotic maintenance vs. guided discontinuation in the functional recovery era. CNS Drugs 2016 May 22;30(5):357-368. [doi: 10.1007/s40263-016-0331-x] [Medline: 27106296]
10. Fowler D, Hodgekins J, French P, Marshall M, Freemantle N, McCrone P, et al. Social recovery therapy in combination with early intervention services for enhancement of social recovery in patients with first-episode psychosis (SUPEREDEN3): a single-blind, randomised controlled trial. Lancet Psychiatry 2018 Jan;5(1):41-50. [doi: 10.1016/s2215-0366(17)30476-5]
11. Stouten LH, Veling W, Laan W, van der Helm M, van der Gaag M. Psychosocial functioning in first-episode psychosis and associations with neurocognition, social cognition, psychotic and affective symptoms. Early Interv Psychiatry 2017 Feb 13;11(1):23-36. [doi: 10.1111/eip.12210] [Medline: 25585960]
12. Alvarez-Jimenez M, Bendall S, Koval P, Rice S, Cagliarini D, Valentine L, et al. HORYZONS trial: protocol for a randomised controlled trial of a moderated online social therapy to maintain treatment effects from first-episode psychosis services. BMJ Open 2019 Feb 19;9(2):e024104 [FREE Full text] [doi: 10.1136/bmjopen-2018-024104] [Medline: 30782893]
13. Chang WC, Kwong VW, Lau ES, So HC, Wong CS, Chan GH, et al. Sustainability of treatment effect of a 3-year early intervention programme for first-episode psychosis. Br J Psychiatry 2017 Jul 02;211(1):37-44. [doi: 10.1192/bjp.bp.117.198929] [Medline: 28385705]
14. Alvarez-Jimenez M, Alcazar-Corcoles M, González-Blanch C, Bendall S, McGorry P, Gleeson J. Online, social media and mobile technologies for psychosis treatment: a systematic review on novel user-led interventions. Schizophr Res 2014 Jun;156(1):96-106. [doi: 10.1016/j.schres.2014.03.021] [Medline: 24746468]
15. Torous J, Woodyatt J, Keshavan M, Tully LM. A new hope for early psychosis care: the evolving landscape of digital care tools. Br J Psychiatry 2019 May 11;214(5):269-272 [FREE Full text] [doi: 10.1192/bjp.2019.8] [Medline: 30739613]
16. Alvarez-Jimenez M, Gleeson J, Bendall S, Penn D, Yung A, Ryan R, et al. Enhancing social functioning in young people at Ultra High Risk (UHR) for psychosis: a pilot study of a novel strengths and mindfulness-based online social therapy. Schizophr Res 2018 Dec;202:369-377. [doi: 10.1016/j.schres.2018.07.022] [Medline: 30031616]
17. Rotondi AJ, Anderson CM, Haas GL, Eack SM, Spring MB, Ganguli R, et al. Web-based psychoeducational intervention for persons with schizophrenia and their supporters: one-year outcomes. Psychiatr Serv 2010 Nov;61(11):1099-1105. [doi: 10.1176/ps.2010.61.11.1099] [Medline: 21041348]
18. Bucci S, Barrowclough C, Ainsworth J, Machin M, Morris R, Berry K, et al. Actissist: proof-of-concept trial of a theory-driven digital intervention for psychosis. Schizophr Bull 2018 Aug 20;44(5):1070-1080 [FREE Full text] [doi: 10.1093/schbul/sby032] [Medline: 29566206]
19. Gottlieb JD, Romeo KH, Penn DL, Mueser KT, Chiko BP. Web-based cognitive-behavioral therapy for auditory hallucinations in persons with psychosis: a pilot study. Schizophr Res 2013 Apr;145(1-3):82-87. [doi: 10.1016/j.schres.2013.01.002] [Medline: 23410709]
20. Alvarez-Jimenez M, Koval P, Schmaal L, Bendall S, O'Sullivan S, Cagliarini D, et al. The Horyzons project: a randomized controlled trial of a novel online social therapy to maintain treatment effects from specialist first-episode psychosis services. World Psychiatry 2021 Jun 18;20(2):233-243 [FREE Full text] [doi: 10.1002/wps.20858] [Medline: 34002511]
21. Granholm E, Ben-Zeev D, Link PC, Bradshaw KR, Holden JL. Mobile Assessment and Treatment for Schizophrenia (MATS): a pilot trial of an interactive text-messaging intervention for medication adherence, socialization, and auditory hallucinations. Schizophr Bull 2012 May 10;38(3):414-425 [FREE Full text] [doi: 10.1093/schbul/sbr155] [Medline: 22080492]
22. Glynn SM, Randolph ET, Garrick T, Lui A. A proof of concept trial of an online psychoeducational program for relatives of both veterans and civilians living with schizophrenia. Psychiatr Rehabil J 2010;33(4):278-287. [doi: 10.2975/33.4.2010.278.287] [Medline: 20374986]
23. de Beurs D, van Bruinessen I, Noordman J, Friele R, van Dulmen S. Active involvement of end users when developing web-based mental health interventions. Front Psychiatry 2017 May 03;8:72 [FREE Full text] [doi: 10.3389/fpsyt.2017.00072] [Medline: 28515699]

24. Berry N, Lobban F, Emsley R, Bucci S. Acceptability of interventions delivered online and through mobile phones for people who experience severe mental health problems: a systematic review. J Med Internet Res 2016 May 31;18(5):e121 [FREE Full text] [doi: 10.2196/jmir.5250] [Medline: 27245693]
25. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? A meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
26. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. J Med Internet Res 2019 Jul 22;21(7):e12869 [FREE Full text] [doi: 10.2196/12869] [Medline: 31333198]
27. Alvarez-Jimenez M, Rice S, D'Alfonso S, Leicester S, Bendall S, Pryor I, et al. A novel multimodal digital service (moderated online social therapy+) for help-seeking young people experiencing mental ill-health: pilot evaluation within a national youth e-mental health service. J Med Internet Res 2020 Aug 13;22(8):e17155 [FREE Full text] [doi: 10.2196/17155] [Medline: 32788151]
28. Garland EL, Fredrickson B, Kring AM, Johnson DP, Meyer PS, Penn DL. Upward spirals of positive emotions counter downward spirals of negativity: insights from the broaden-and-build theory and affective neuroscience on the treatment of emotion dysfunctions and deficits in psychopathology. Clin Psychol Rev 2010 Nov;30(7):849-864. [doi: 10.1016/j.cpr.2010.03.002]
29. Vaskinn A, Ventura J, Andreassen OA, Melle I, Sundet K. A social path to functioning in schizophrenia: from social self-efficacy through negative symptoms to social functional capacity. Psychiatry Res 2015 Aug 30;228(3):803-807. [doi: 10.1016/j.psychres.2015.05.019] [Medline: 26089018]
30. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000;55(1):68-78. [doi: 10.1037/0003-066x.55.1.68]
31. Ellis RE, Seal ML, Simmons JG, Whittle S, Schwartz OS, Byrne ML, et al. Longitudinal trajectories of depression symptoms in adolescence: psychosocial risk factors and outcomes. Child Psychiatry Hum Dev 2017 Aug 12;48(4):554-571. [doi: 10.1007/s10578-016-0682-z] [Medline: 27619221]
32. Mas S, Gassó P, Rodríguez N, Cabrera B, Mezquida G, Lobo A, PEPs group. Personalized medicine begins with the phenotype: identifying antipsychotic response phenotypes in a first-episode psychosis cohort. Acta Psychiatr Scand 2020 Jun 05;141(6):541-552. [doi: 10.1111/acps.13131] [Medline: 31746462]
33. Hall M, Holton KM, Öngür D, Montrose D, Keshavan MS. Longitudinal trajectory of early functional recovery in patients with first episode psychosis. Schizophr Res 2019 Jul;209:234-244 [FREE Full text] [doi: 10.1016/j.schres.2019.02.003] [Medline: 30826261]
34. Tang A, Van Lieshout RJ, Lahat A, Duku E, Boyle MH, Saigal S, et al. Shyness trajectories across the first four decades predict mental health outcomes. J Abnorm Child Psychol 2017 Nov 25;45(8):1621-1633. [doi: 10.1007/s10802-017-0265-x] [Medline: 28120251]
35. Cotton SM, Filia KM, Ratheesh A, Pennell K, Goldstone S, McGorry PD. Early psychosis research at Orygen, The National Centre of Excellence in Youth Mental Health. Soc Psychiatry Psychiatr Epidemiol 2016 Jan 23;51(1):1-13. [doi: 10.1007/s00127-015-1140-0] [Medline: 26498752]
36. McGorry PD, Edwards J, Mihalopoulos C, Harrigan SM, Jackson HJ. EPPIC: an evolving system of early detection and optimal management. Schizophr Bull 1996 Jan 01;22(2):305-326. [doi: 10.1093/schbul/22.2.305] [Medline: 8782288]
37. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (4th Edition). Washington, DC: American Psychiatric Association; 1994.
38. Kay SR, Fiszbein A, Opler LA. The positive and negative syndrome scale (PANSS) for schizophrenia. Schizophr Bull 1987 Jan 01;13(2):261-276. [doi: 10.1093/schbul/13.2.261] [Medline: 3616518]
39. Morosini PL, Magliano L, Brambilla L, Ugolini S, Pioli R. Development, reliability and acceptability of a new version of the DSM-IV Social and Occupational Functioning Assessment Scale (SOFAS) to assess routine social functioning. Acta Psychiatr Scand 2000 Apr;101(4):323-329. [Medline: 10782554]
40. Lecomte T, Corbière M, Ehmann T, Addington J, Abdel-Baki A, Macewan B. Development and preliminary validation of the First Episode Social Functioning Scale for early psychosis. Psychiatry Res 2014 May 30;216(3):412-417. [doi: 10.1016/j.psychres.2014.01.044] [Medline: 24613006]
41. Addington D, Addington J, Maticka-tyndale E. Assessing sepression in schizophrenia: the Calgary depression scale. Br J Psychiatry 2018 Aug 06;163(S22):39-44. [doi: 10.1192/s0007125000292581]
42. Lovibond SH, Lovibond PF. Manual for the Depression Anxiety Stress Scales. Sydney, N.S.W: Psychology Foundation of Australia; 1996.
43. D'Zurilla TJ, Goldfried MR. Problem solving and behavior modification. Journal of Abnormal Psychology 1971 Aug;78(1):107-126. [doi: 10.1037/h0031360]
44. Genolini C, Pingault J, Driss T, Côté S, Tremblay R, Vitaro F, et al. KmL3D: a non-parametric algorithm for clustering joint trajectories. Comput Methods Programs Biomed 2013 Jan;109(1):104-111. [doi: 10.1016/j.cmpb.2012.08.016] [Medline: 23127283]

45. Genolini C, Alacoque X, Sentenac M, Arnaud C. kml and kml3d: R Packages to cluster longitudinal data. J Stat Soft 2015;65(4):1-34. [doi: 10.18637/jss.v065.i04]
46. RStudio Team. RStudio: Integrated Development Environment for R. 2021. URL: http://www.rstudio.com/ [accessed 2021-03-22]
47. Kahkoska AR, Nguyen CT, Adair LA, Aiello AE, Burger KS, Buse JB, et al. Longitudinal phenotypes of type 1 diabetes in youth based on weight and glycemia and their association with complications. J Clin Endocrinol Metabol 2019 Jul 10;104(12):6003-6016. [doi: 10.1210/jc.2019-00734] [Medline: 31290977]
48. Calinski T, Harabasz J. A dendrite method for cluster analysis. Commun Stat Theory Methods 1974;3(1):1-27. [doi: 10.1080/03610927408827101]
49. Ray S, Turi R. Determination of number of clusters in k-means clustering and application in colour image segmentation. In: Proceedings of the 4th International Conference on Advances in Pattern Recognition and Digital Techniques. 1999 Presented at: 4th International Conference on Advances in Pattern Recognition and Digital Techniques; 1999; Calcutta, India URL: https://www.semanticscholar.org/paper/Determination-of-Number-of-Clusters-in-K-Means-and-Ray-Turi/0ec132fce9971d1e0e670e650b58176dc7bf36da
50. Davies DL, Bouldin DW. A cluster separation measure. IEEE Trans Pattern Anal Mach Intell 1979 Apr;PAMI-1(2):224-227. [doi: 10.1109/tpami.1979.4766909]
51. Rand WM. Objective criteria for the evaluation of clustering methods. J Am Stat Assoc 1971 Dec;66(336):846-850. [doi: 10.1080/01621459.1971.10482356]
52. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Soft 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
53. Chan SK, Hui CL, Chang WC, Lee EH, Chen EY. Ten-year follow up of patients with first-episode schizophrenia spectrum disorder from an early intervention service: predictors of clinical remission and functional recovery. Schizophr Res 2019 Feb;204:65-71. [doi: 10.1016/j.schres.2018.08.022] [Medline: 30126816]
54. Marshall M, Lewis S, Lockwood A, Drake R, Jones P, Croudace T. Association Between Duration of Untreated Psychosis and Outcome in Cohorts of First-Episode Patients. Arch Gen Psychiatry 2005 Sep 01;62(9):975. [doi: 10.1001/archpsyc.62.9.975]
55. Penttilä M, Jääskeläinen E, Hirvonen N, Isohanni M, Miettunen J. Duration of untreated psychosis as predictor of long-term outcome in schizophrenia: systematic review and meta-analysis. Br J Psychiatry 2014 Aug 02;205(2):88-94. [doi: 10.1192/bjp.bp.113.127753] [Medline: 25252316]
56. Emsley R, Rabinowitz J, Medori R, Early Psychosis Global Working Group. Remission in early psychosis: rates, predictors, and clinical and functional outcome correlates. Schizophr Res 2007 Jan;89(1-3):129-139. [doi: 10.1016/j.schres.2006.09.013] [Medline: 17095194]
57. Álvarez-Jiménez M, Gleeson JF, Henry LP, Harrigan SM, Harris MG, Killackey E, et al. Road to full recovery: longitudinal relationship between symptomatic remission and psychosocial recovery in first-episode psychosis over 7.5 years. Psychol Med 2011 Aug 19;42(3):595-606. [doi: 10.1017/s0033291711001504]
58. Alvarez-Jimenez M, Gleeson J, Henry L, Harrigan S, Harris M, Amminger G, et al. Prediction of a single psychotic episode: a 7.5-year, prospective study in first-episode psychosis. Schizophr Res 2011 Feb;125(2-3):236-246. [doi: 10.1016/j.schres.2010.10.020] [Medline: 21081266]
59. Sanatkar S, Baldwin PA, Huckvale K, Clarke J, Christensen H, Harvey S, et al. Using cluster analysis to explore engagement and e-attainment as emergent behavior in electronic mental health. J Med Internet Res 2019 Nov 28;21(11):e14728 [FREE Full text] [doi: 10.2196/14728] [Medline: 31778115]
60. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
61. Fleming T, Bavin L, Lucassen M, Stasiak K, Hopkins S, Merry S. Beyond the trial: systematic review of real-world uptake and engagement with digital self-help interventions for depression, low mood, or anxiety. J Med Internet Res 2018 Jun 06;20(6):e199 [FREE Full text] [doi: 10.2196/jmir.9275] [Medline: 29875089]
62. Ryan RM, Deci EL. Self-Determination Theory: Basic Psychological Needs in Motivation, Development, and Wellness. New York: Guilford Publications; 2018.
63. Thomas EC, Zisman-Ilani Y, Salzer MS. Self-determination and choice in mental health: qualitative insights from a study of self-directed care. Psychiatr Serv 2019 Sep 01;70(9):801-807 [FREE Full text] [doi: 10.1176/appi.ps.201800544] [Medline: 31109262]
64. Hanley T. The working alliance in online therapy with young people: preliminary findings. Br J Guid Counsel 2009 Aug;37(3):257-269. [doi: 10.1080/03069880902956991]
65. Laws KR, Darlington N, Kondel TK, McKenna PJ, Jauhar S. Cognitive Behavioural Therapy for schizophrenia - outcomes for functioning, distress and quality of life: a meta-analysis. BMC Psychol 2018 Jul 17;6(1):32 [FREE Full text] [doi: 10.1186/s40359-018-0243-2] [Medline: 30016999]
66. Polese D, Fornaro M, Palermo M, De Luca V, de Bartolomeis A. Treatment-resistant to antipsychotics: a resistance to everything? Psychotherapy in treatment-resistant schizophrenia and nonaffective psychosis: a 25-year systematic review

and exploratory meta-analysis. Front Psychiatry 2019 Apr 17;10:210 [FREE Full text] [doi: 10.3389/fpsyt.2019.00210] [Medline: 31057434]
67. Thorisdottir IE, Sigurvinsdottir R, Asgeirsdottir BB, Allegrante JP, Sigfusdottir ID. Active and passive social media use and symptoms of anxiety and depressed mood among Icelandic adolescents. Cyberpsychol Behav Soc Netw 2019 Aug;22(8):535-542. [doi: 10.1089/cyber.2019.0079] [Medline: 31361508]
68. Pung A, Fletcher SL, Gunn JM. Mobile app use by primary care patients to manage their depressive symptoms: qualitative study. J Med Internet Res 2018 Sep 27;20(9):e10035 [FREE Full text] [doi: 10.2196/10035] [Medline: 30262449]
69. Zilcha-Mano S, Roose SP, Barber JP, Rutherford BR. Therapeutic alliance in antidepressant treatment: cause or effect of symptomatic levels? Psychother Psychosom 2015 Mar 28;84(3):177-182 [FREE Full text] [doi: 10.1159/000379756] [Medline: 25832111]
70. D'Alfonso S, Santesteban-Echarri O, Rice S, Wadley G, Lederman R, Miles C, et al. Artificial intelligence-assisted online social therapy for youth mental health. Front Psychol 2017 Jun 02;8:796 [FREE Full text] [doi: 10.3389/fpsyg.2017.00796] [Medline: 28626431]

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

Edited by J Torous; submitted 31.03.21; peer-reviewed by L Balcombe, R Lundin; comments to author 01.06.21; revised version received 18.06.21; accepted 10.01.22; published 07.04.22.

Please cite as:
O'Sullivan S, Schmaal L, D'Alfonso S, Toenders YJ, Valentine L, McEnery C, Bendall S, Nelson B, Gleeson JF, Alvarez-Jimenez M Characterizing Use of a Multicomponent Digital Intervention to Predict Treatment Outcomes in First-Episode Psychosis: Cluster Analysis
JMIR Ment Health 2022;9(4):e29211
URL: https://mental.jmir.org/2022/4/e29211
doi:10.2196/29211
PMID:35389351

Original Paper

Telehealth-Based Psychoeducation for Caregivers: The Family Intervention in Recent-Onset Schizophrenia Treatment Study

Kim T Mueser[1], PhD; Eric D Achtyes[2,3], MD, MS; Jagadish Gogate[4], PhD; Branislav Mancevski[5], MD, NCPsyA; Edward Kim[5], MD, MBA; H Lynn Starr[5], MD

[1]Center for Psychiatric Rehabilitation, Boston University, Boston, MA, United States

[2]Michigan State University College of Human Medicine, East Lansing, MI, United States

[3]Cherry Health, Grand Rapids, MI, United States

[4]Janssen Research & Development, LLC, Titusville, NJ, United States

[5]Janssen 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

(JMIR Ment Health 2022;9(4):e32492) doi:10.2196/32492

KEYWORDS

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.

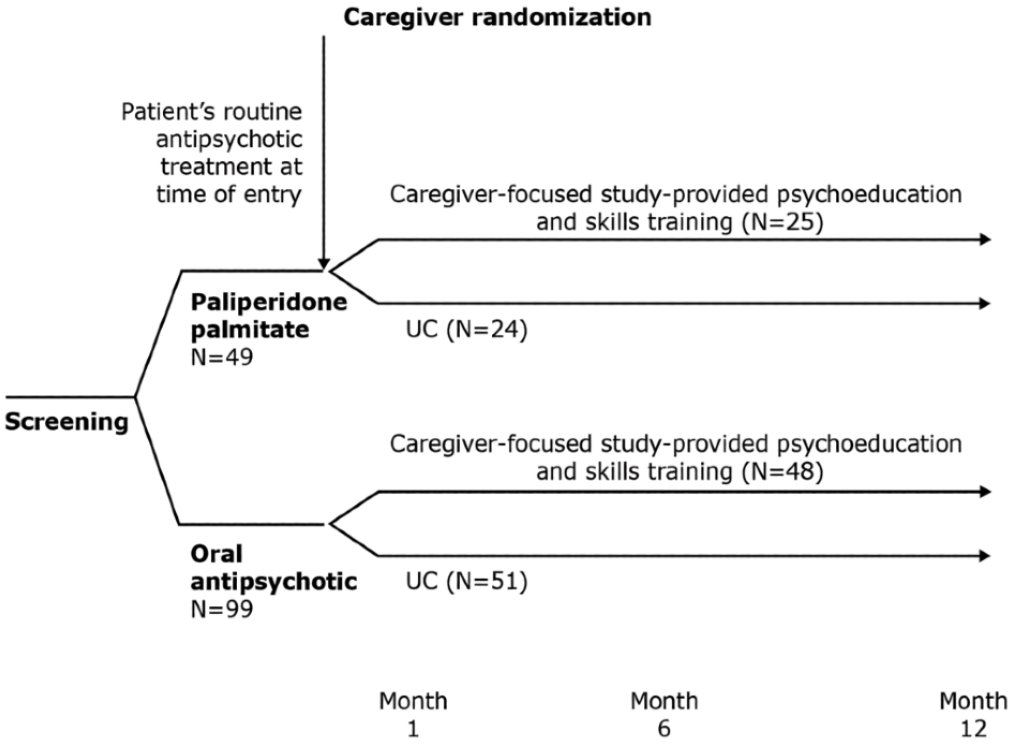
The Family Intervention in Recent-Onset Schizophrenia Treatment (FIRST) study was designed to evaluate the impact of FP given specifically to caregivers on the outcomes of patients with schizophrenia spectrum disorder under their care and family burden. In the FIRST study, FP was delivered using Healios 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 antipsychotic; 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).

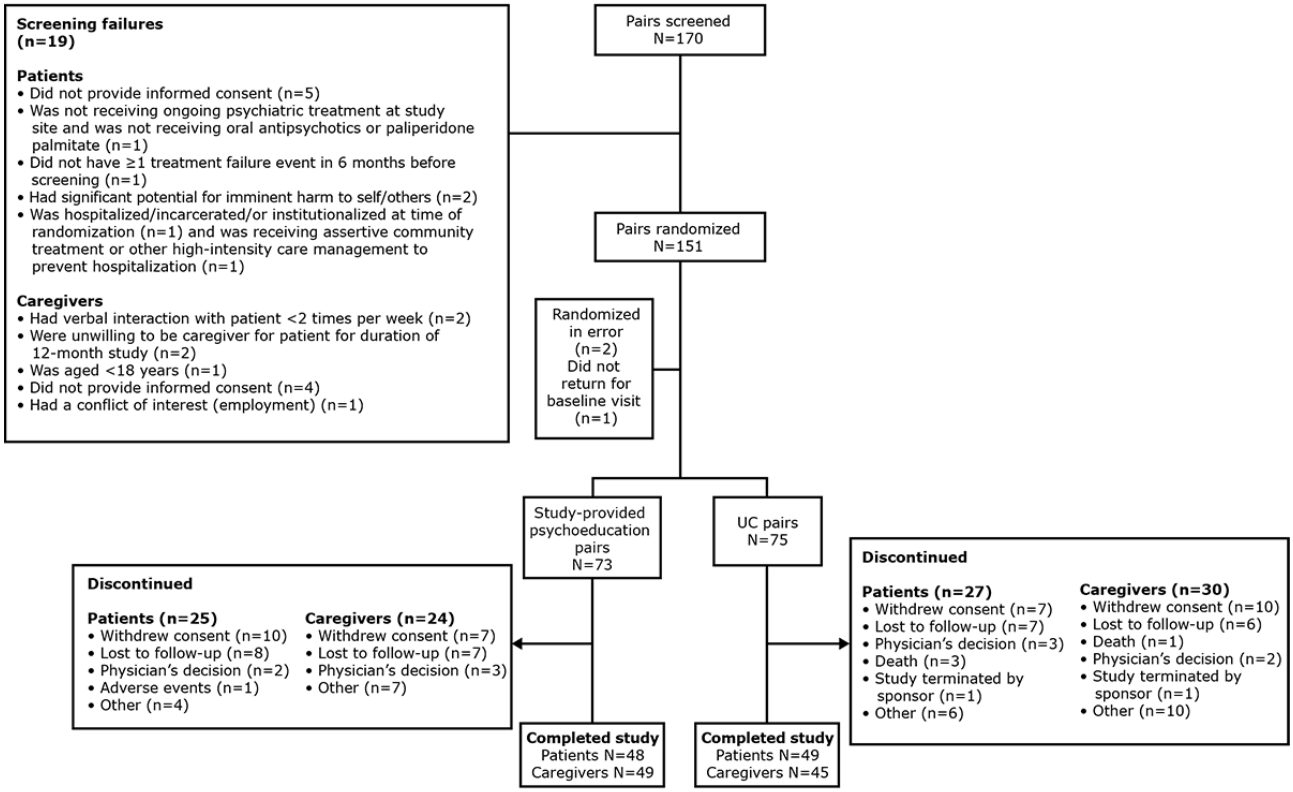
The TF rate in the control group was assumed to be 0.50 based on a previous study with a similar end point [30]. The effect size in terms of a risk ratio of 0.60 was obtained from a meta-analysis of 18 randomized controlled studies examining the effect of face-to-face psychoeducation for caregivers on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

Parameter	Intervention[a]		
	Study-provided psychoeducation (n=73)	Usual care (n=75)	Total
Patients			
Age (years), mean (SD)	25.3 (4.7)	25.1 (5.0)	25.2 (4.8)
Sex, n (%)			
Male	54 (74)	57 (76)	111 (75)
Female	19 (26)	18 (24)	37 (25)
Race, n (%)			
White	45 (61.6)	39 (52)	84 (56.8)
Black or African American	21 (28.8)	31 (41.3)	52 (35.1)
Multiple or other	7 (9.6)	4 (5.3)	11 (7.4)
Unknown or not reported	0 (0)	1 (1.3)	1 (0.7)
Ethnicity, n (%)			
Hispanic or Latino	13 (17.8)	17 (22.7)	30 (20.3)
Not Hispanic or Latino	60 (82.2)	57 (76)	117 (79.1)
Unknown or not reported	0 (0)	1 (1.3)	1 (0.7)
Living status, n (%)			
At home with family or friends	63 (86.3)	68 (90.7)	131 (88.5)
At home alone	5 (6.8)	5 (6.7)	10 (6.8)
Sheltered living	2 (2.7)	0 (0)	2 (1.4)
Other	3 (4.1)	1 (1.3)	4 (2.7)
Diagnosis, n (%)			
Schizophrenia	44 (60.3)	38 (50.7)	82 (55.4)
Schizoaffective disorder	31 (42.5)	34 (45.3)	65 (43.9)
Schizophreniform disorder	0 (0)	5 (6.7)	5 (3.4)
Functioning, mean (SD)			
IMR[b] total score	48.3 (6.8)	49.2 (7.1)	48.8 (7.0)
CGI-S[c] score	4.1 (1.1)	4.3 (1.1)	4.2 (1.1)
Caregivers			
Age (years), mean (SD)	52.1 (11.2)	49.0 (12.5)	50.5 (11.9)
Sex, n (%)			
Male	14 (19.2)	18 (24)	32 (21.6)
Female	59 (80.8)	57 (76)	116 (78.4)
Race, n (%)			
White	47 (64.4)	40 (53.3)	87 (58.8)
Black or African American	18 (24.7)	31 (41.3)	49 (33.1)
Multiple or other	7 (9.6)	2 (2.7)	9 (6.1)
Not reported or unknown	1 (1.4)	2 (2.7)	3 (2.0)
Ethnicity, n (%)			
Hispanic or Latino	9 (12.3)	16 (21.3)	25 (16.9)
Not Hispanic or Latino	64 (87.7)	58 (77.3)	122 (82.4)
Not reported	0 (0)	1 (1.3)	1 (0.7)

Parameter	Intervention[a]		
	Study-provided psychoeducation (n=73)	Usual care (n=75)	Total
Relationship with patient, n (%)			
Parent[d]	56 (76.7)	56 (74.7)	112 (75.7)
Sibling	2 (2.7)	6 (8.0)	8 (5.4)
Other relative	5 (6.8)	4 (5.3)	9 (6.1)
Spouse or partner	6 (8.2)	5 (6.7)	11 (7.4)
Friend	3 (4.1)	2 (2.7)	5 (3.4)
Other	1 (1.4)	2 (2.7)	3 (2.0)
IEQ[e] score, mean (SD)			
Total	30.8 (16.4)	29.7 (17.3)	30.2 (16.8)
Tension	8.1 (5.4)	6.8 (6.0)	7.5 (5.7)
Supervision	3.1 (3.3)	3.8 (3.8)	3.5 (3.6)
Worrying	11.3 (6.2)	9.7 (5.8)	10.5 (6.0)
Urging	10.7 (5.9)	11.9 (7.1)	11.3 (6.6)
SF-12[f] PCS[g] score, mean (SD)[h]	50.3 (10.3)	51.7 (9.4)	51.0 (9.8)
SF-12 MCS[i] score, mean (SD)[h]	45.3 (10.4)	48.7 (9.5)	47.0 (10.1)

[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 2. Demographics and baseline characteristics of patients with schizophrenia spectrum disorders and caregivers by caregiver discontinuation status (N=148).

Parameter	Caregiver discontinuation status[a]		
	Discontinued early (n=45)	Completed study (n=103)	Total
Patients			
Age (years), mean (SD)	25.2 (4.3)	25.2 (5.1)	25.2 (4.8)
Sex, n (%)			
Male	31 (68.9)	80 (77.7)	111 (75)
Female	14 (31.1)	23 (22.3)	37 (25)
Race, n (%)			
White	26 (57.8)	58 (56.3)	84 (56.8)
Black or African American	17 (37.8)	35 (34.0)	52 (35.1)
Multiple or other	1 (2.2)	10 (9.7)	11 (7.4)
Unknown or not reported	1 (2.2)	0 (0)	1 (0.7)
Ethnicity, n (%)			
Hispanic or Latino	8 (17.8)	22 (21.4)	30 (20.3)
Not Hispanic or Latino	37 (82.2)	80 (77.7)	117 (79.1)
Unknown or not reported	0 (0)	1 (1.0)	1 (0.7)
Living status, n (%)			
At home with family or friends	41 (91.1)	90 (87.4)	131 (88.5)
At home alone	3 (6.7)	7 (6.8)	10 (6.8)
Sheltered living	0 (0)	2 (1.9)	2 (1.4)
Other	1 (2.2)	3 (2.9)	4 (2.7)
Diagnosis, n (%)			
Schizophrenia	29 (64.4)	53 (51.5)	82 (55.4)
Schizoaffective disorder	19 (42.2)	46 (44.7)	65 (43.9)
Schizophreniform disorder	0 (0)	5 (4.9)	5 (3.4)
Functioning, mean (SD)			
IMR[b] total score, mean (SD)	46.0 (6.2)	50.0 (7.0)	48.8 (7.0)
CGI-S[c] score, mean (SD)	4.2 (1.0)	4.2 (1.2)	4.2 (1.1)
Caregivers			
Age (years), mean (SD)	47.8 (13.3)	51.7 (11.2)	50.5 (11.9)
Sex, n (%)			
Male	9 (20)	23 (22.3)	32 (21.6)
Female	36 (80)	80 (77.7)	116 (78.4)
Race, n (%)			
White	28 (62.2)	59 (57.3)	87 (58.8)
Black or African American	16 (35.6)	33 (32)	49 (33.1)
Multiple or other	0 (0)	9 (8.7)	9 (6.1)
Not reported or unknown	1 (2.2)	2 (1.9)	3 (2)
Ethnicity, n (%)			
Hispanic or Latino	7 (15.6)	18 (17.5)	25 (16.9)
Not Hispanic or Latino	38 (84.4)	84 (81.6)	122 (82.4)
Not reported	0 (0)	1 (1)	1 (0.7)

Parameter	Caregiver discontinuation status[a]		
	Discontinued early (n=45)	Completed study (n=103)	Total
Relationship with patient, n (%)			
Parent[d]	29 (64.4)	83 (80.6)	112 (75.7)
Sibling	4 (8.9)	4 (3.9)	8 (5.4)
Other relative	5 (11.1)	4 (3.9)	9 (6.1)
Spouse or partner	4 (8.9)	7 (6.8)	11 (7.4)
Friend	2 (4.4)	3 (2.9)	5 (3.4)
Other	1 (2.2)	2 (1.9)	3 (2)
IEQ[e] score, mean (SD)			
Total	33.2 (18.2)	29.0 (16.1)	30.2 (16.8)
Tension	8.6 (6.1)	7.0 (5.6)	7.5 (5.7)
Supervision	4.2 (4.6)	3.1 (3.0)	3.5 (3.6)
Worrying	10.7 (6.2)	10.5 (6.0)	10.5 (6.0)
Urging	12.1 (7.2)	10.9 (6.2)	11.3 (6.6)
SF-12[f] PCS[g] score, mean (SD)[h]	48.5 (11.4)	52.0 (8.9)	51.0 (9.8)
SF-12 MCS[i] score, mean (SD)[h]	46.4 (10.2)	47.2 (10.0)	47.0 (10.1)

[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].

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

[a]All-randomized analysis set (all caregivers who were randomly assigned and entered the study).

[b]A 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].

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

[a]All-randomized analysis set (all caregivers who were randomly assigned and entered the study).

[b]NAMI: 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=.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.

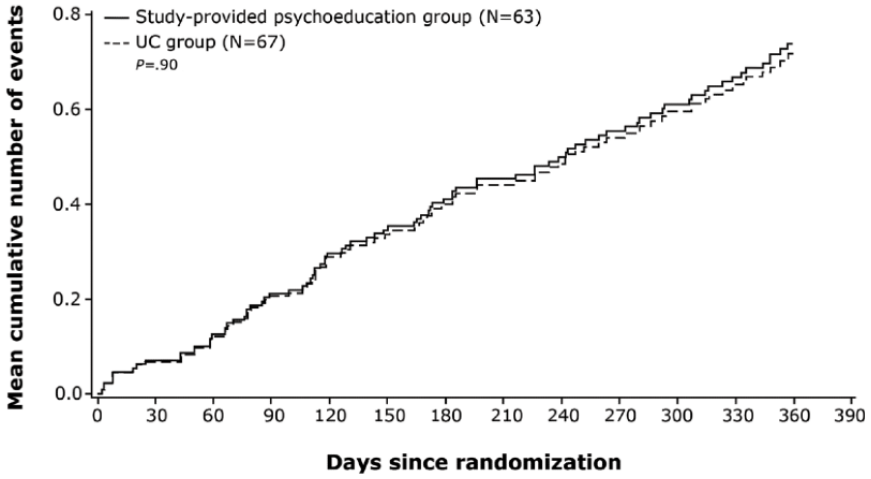


Figure 4. Treatment failure rates by antipsychotic treatment strata (post hoc analysis). Efficacy analysis set (n=130, all patients who entered the study and had at least one postbaseline efficacy assessment). UC: usual care.

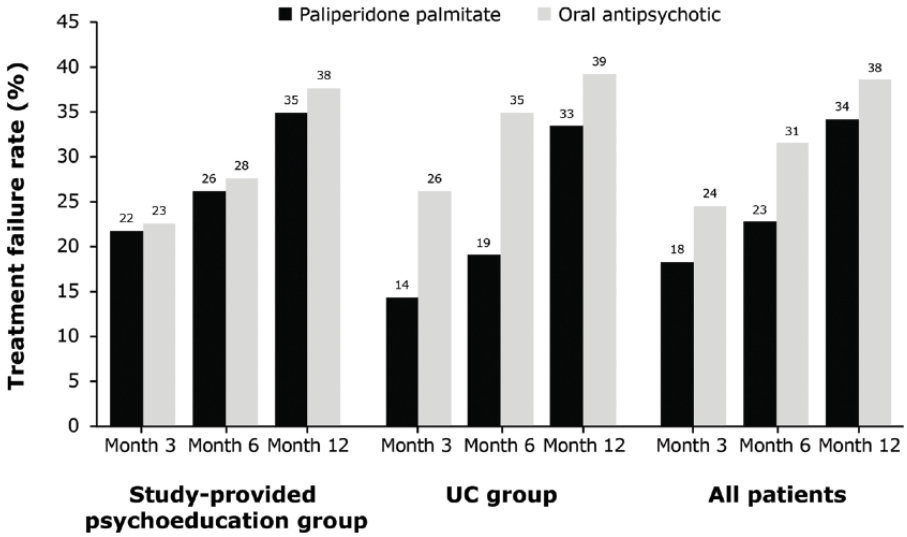


Table 5. Summary of secondary end points: caregiver and patient secondary outcomes.

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

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.

Outcomes	3 months		6 months		12 months	
	Study-provided psychoeducation	UC[a]	Study-provided psychoeducation	UC	Study-provided psychoeducation	UC
Caregiver outcomes						
IEQ[b,c] change from baseline						
LS[d] mean (SE)	−2.95 (1.59)	−5.08 (1.54)	−6.96 (1.59)	−7.62 (1.58)	−4.15 (1.96)	−5.61 (2.08)
Difference (study-provided psychoeducation vs UC)						
LS mean (95% CI)	2.13 (−2.25 to 6.52)	2.13 (−2.25 to 6.52)	0.66 (−3.79 to 5.10)	0.66 (−3.79 to 5.10)	1.46 (−4.21 to 7.13)	1.46 (−4.21 to 7.13)
P value	.34	.34	.77	.77	.61	.61
SF-12[e] MCS[f,g] change from baseline						
LS mean (SE)	0.63 (1.16)	1.27 (1.12)	2.86 (1.07)	2.92 (1.07)	1.94 (1.34)	3.01 (1.42)
Difference (study-provided psychoeducation vs UC)						
LS mean (95% CI)	−0.64 (−3.84 to 2.56)	−0.64 (−3.84 to 2.56)	−0.05 (−3.07 to 2.96)	−0.05 (−3.07 to 2.96)	−1.08 (−4.96 to 2.80)	−1.08 (−4.96 to 2.80)
P value	.69	.69	.97	.97	.58	.58
SF-12 PCS[h] change from baseline						
LS mean (SE)	−0.59 (0.89)	−0.75 (0.86)	−2.38 (0.87)	−1.59 (0.87)	−2.08 (1.17)	−2.82 (1.26)
Difference (study-provided psychoeducation vs UC)						
LS mean (95% CI)	0.16 (−2.29 to 2.60)	0.16 (−2.29 to 2.60)	−0.79 (−3.23 to 1.64)	−0.79 (−3.23 to 1.64)	0.74 (−2.67 to 4.15)	0.74 (−2.67 to 4.15)
P value	.90	.90	.52	.52	.67	.67
Patient outcomes						
IMR[i,j] change from baseline						
LS mean (SE)	1.12 (0.75)	1.04 (0.72)	4.34 (0.89)	3.40 (0.85)	4.47 (0.81)	3.61 (0.83)
Difference (study-provided psychoeducation vs UC)						
LS mean (95% CI)	0.09 (−1.97 to 2.14)	0.09 (−1.97 to 2.14)	0.94 (−1.50 to 3.38)	0.94 (−1.50 to 3.38)	0.86 (−1.45 to 3.16)	0.86 (−1.45 to 3.16)
P value	.93	.93	.45	.45	.46	.46
CGI-S[k,l] change from baseline						
LS mean (SE)	−0.18 (0.10)	−0.12 (0.10)	−0.24 (0.11)	−0.30 (0.11)	−0.30 (0.13)	−0.44 (0.13)
Difference (study-provided psychoeducation vs UC)						
LS mean (95% CI)	−0.06 (−0.34 to 0.22)	−0.06 (−0.34 to 0.22)	0.06 (−0.25 to 0.37)	0.06 (−0.25 to 0.37)	0.13 (−0.22 to 0.49)	0.13 (−0.22 to 0.49)
P value	.66	.66	.69	.69	.46	.46

[a]UC: usual care.

[b]IEQ: Involvement Evaluation Questionnaire.

[c]Higher scores on the IEQ indicate a higher caregiver burden.

[d]LS: least squares.

[e]SF-12: 12-item Short Form Health Survey.

[f]MCS: mental component summary.

[g]Higher scores on the SF-12 indicate better health.

[h]PCS: physical component summary.

[i]IMR: Illness Management and Recovery.

[j]Higher scores on the IMR indicate better recovery status.

[k]CGI-S: Clinical Global Impression of Severity.

[l]Higher scores on the CGI-S indicate higher symptom severity.

Table 7. Summary of secondary end points: decreases in health resource utilization (N=148).

Health resource use	Baseline, n (%)	6 months, n (%)	12 months, n (%)
Hospitalizations	46 (31)	10 (7)	6 (4)
Emergency department visits	72 (49)	17 (12)	17 (12)
Intensive outpatient treatment	11 (7)	3 (2)	0 (0)

Table 8. Decreases in health resource utilization.

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

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 multicaregiver 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).
[PDF File (Adobe PDF File), 1268 KB - mental_v9i4e32492_app1.pdf]

Multimedia Appendix 2
Number of psychoeducation sessions provided to caregivers in the study-provided psychoeducation group (N=148).
[PNG File , 34 KB - mental_v9i4e32492_app2.png]

Multimedia Appendix 3
Caregiver demographics and baseline characteristics by number of study-provided psychoeducation sessions (≤8 vs >8; safety analysis set).
[DOCX File , 17 KB - mental_v9i4e32492_app3.docx]

Multimedia Appendix 4
Summary of safety data.
[DOCX File , 14 KB - mental_v9i4e32492_app4.docx]

References

1. Häfner H, Maurer K, Löffler W, Riecher-Rössler A. The influence of age and sex on the onset and early course of schizophrenia. Br J Psychiatry 1993 Jan 03;162(1):80-86. [doi: 10.1192/bjp.162.1.80] [Medline: 8425144]
2. Awad AG, Voruganti LN. The burden of schizophrenia on caregivers: a review. Pharmacoeconomics 2008;26(2):149-162. [doi: 10.2165/00019053-200826020-00005] [Medline: 18198934]

3. Addington J, Coldham EL, Jones B, Ko T, Addington D. The first episode of psychosis: the experience of relatives. Acta Psychiatr Scand 2003 Oct;108(4):285-289. [doi: 10.1034/j.1600-0447.2003.00153.x] [Medline: 12956829]
4. Hsiao CY, Lu HL, Tsai YF. Caregiver burden and health-related quality of life among primary family caregivers of individuals with schizophrenia: a cross-sectional study. Qual Life Res 2020 Oct;29(10):2745-2757. [doi: 10.1007/s11136-020-02518-1] [Medline: 32394137]
5. Millier A, Schmidt U, Angermeyer MC, Chauhan D, Murthy V, Toumi M, et al. Humanistic burden in schizophrenia: a literature review. J Psychiatr Res 2014 Jul;54:85-93 [FREE Full text] [doi: 10.1016/j.jpsychires.2014.03.021] [Medline: 24795289]
6. Camacho-Gomez M, Castellvi P. Effectiveness of family intervention for preventing relapse in first-episode psychosis until 24 months of follow-up: a systematic review with meta-analysis of randomized controlled trials. Schizophr Bull 2020 Jan 04;46(1):98-109 [FREE Full text] [doi: 10.1093/schbul/sbz038] [Medline: 31050757]
7. Yesufu-Udechuku A, Harrison B, Mayo-Wilson E, Young N, Woodhams P, Shiers D, et al. Interventions to improve the experience of caring for people with severe mental illness: systematic review and meta-analysis. Br J Psychiatry 2015 Apr;206(4):268-274. [doi: 10.1192/bjp.bp.114.147561] [Medline: 25833867]
8. Ma CF, Chien WT, Bressington DT. Family intervention for caregivers of people with recent-onset psychosis: a systematic review and meta-analysis. Early Interv Psychiatry 2018 Aug;12(4):535-560. [doi: 10.1111/eip.12494] [Medline: 29076263]
9. Pharoah F, Mari J, Rathbone J, Wong W. Family intervention for schizophrenia. Cochrane Database Syst Rev 2010 Dec 08(12):CD000088 [FREE Full text] [doi: 10.1002/14651858.CD000088.pub2] [Medline: 21154340]
10. American Psychiatric Association. Practice Guideline for the Treatment of Patients with Schizophrenia. Virginia, USA: American Psychiatric Association; 1997.
11. Psychosis and Schizophrenia in Adults: The NICE Guideline on Treatment and Management. London, UK: National Collaborating Centre for Mental Health; 2014.
12. Ince P, Haddock G, Tai S. A systematic review of the implementation of recommended psychological interventions for schizophrenia: rates, barriers, and improvement strategies. Psychol Psychother 2016 Sep;89(3):324-350. [doi: 10.1111/papt.12084] [Medline: 26537838]
13. Leff J, Berkowitz R, Shavit N, Strachan A, Glass I, Vaughn C. A trial of family therapy v. a relatives group for schizophrenia. Br J Psychiatry 1989 Jan;154:58-66. [doi: 10.1192/bjp.154.1.58] [Medline: 2673479]
14. Selick A, Durbin J, Vu N, O'Connor K, Volpe T, Lin E. Barriers and facilitators to implementing family support and education in Early Psychosis Intervention programmes: a systematic review. Early Interv Psychiatry 2017 Oct;11(5):365-374. [doi: 10.1111/eip.12400] [Medline: 28418227]
15. Kavanagh DJ, Piatkowska O, Clark D, O'Halloran P, Manicavasagar V, Rosen A, et al. Application of cognitive-behavioural family intervention for schizophrenia in multidisciplinary teams: what can the matter be? Australian Psychol 1993;28(3):181-188. [doi: 10.1080/00050069308258899]
16. Blom MM, Zarit SH, Groot Zwaaftink RB, Cuijpers P, Pot AM. Effectiveness of an internet intervention for family caregivers of people with dementia: results of a randomized controlled trial. PLoS One 2015 Feb 13;10(2):e0116622 [FREE Full text] [doi: 10.1371/journal.pone.0116622] [Medline: 25679228]
17. Hedman E, Ljótsson B, Rück C, Bergström J, Andersson G, Kaldo V, et al. Effectiveness of internet-based cognitive behaviour therapy for panic disorder in routine psychiatric care. Acta Psychiatr Scand 2013 Dec;128(6):457-467. [doi: 10.1111/acps.12079] [Medline: 23406572]
18. Barak A, Hen L, Boniel-Nissim M, Shapira N. A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. J Technol Human Serv 2008;26(2-4):109-160. [doi: 10.1080/15228830802094429]
19. Rotondi AJ, Anderson CM, Haas GL, Eack SM, Spring MB, Ganguli R, et al. Web-based psychoeducational intervention for persons with schizophrenia and their supporters: one-year outcomes. Psychiatr Serv 2010 Nov;61(11):1099-1105. [doi: 10.1176/ps.2010.61.11.1099] [Medline: 21041348]
20. Rotondi AJ, Haas GL, Anderson CM, Newhill CE, Spring MB, Ganguli R, et al. A clinical trial to test the feasibility of a telehealth psychoeducational intervention for persons with schizophrenia and their families: intervention and 3-month findings. Rehabil Psychol 2005 Nov;50(4):325-336 [FREE Full text] [doi: 10.1037/0090-5550.50.4.325] [Medline: 26321774]
21. Purba J, Suttharangsee W, Chaowalit A. Effects of a caregiver coaching program on family caregivers' competencies for persons with schizophrenia: a pilot study. J Health Res 2016 Aug;30(4):267-274. [doi: 10.14456/jhr.2016.36]
22. Beauchamp N, Irvine AB, Seeley J, Johnson B. Worksite-based internet multimedia program for family caregivers of persons with dementia. Gerontologist 2005 Dec;45(6):793-801. [doi: 10.1093/geront/45.6.793] [Medline: 16326661]
23. Devaramane V, Pai NB, Vella SL. The effect of a brief family intervention on primary carer's functioning and their schizophrenic relatives levels of psychopathology in India. Asian J Psychiatr 2011 Sep;4(3):183-187. [doi: 10.1016/j.ajp.2011.06.004] [Medline: 23051115]
24. Ashcroft K, Kim E, Elefant E, Benson C, Carter JA. Meta-analysis of caregiver-directed psychosocial interventions for schizophrenia. Community Ment Health J 2018 Oct;54(7):983-991. [doi: 10.1007/s10597-018-0289-x] [Medline: 29948624]
25. Family Intervention in Recent Onset Schizophrenia Treatment (FIRST). ClinicalTrials. URL: https://clinicaltrials.gov/ct2/show/NCT02600741 [accessed 2022-03-29]

26. Mueser K, Gingerich S, Salyers MP, Mcguire AB, Reyes RU, Cunningham H. Illness Management and Recovery Scales. Virginia, United States: APA PsycTests; 2005.
27. Guy W. ECDEU assessment manual for psychopharmacology. Open Library. URL: https://openlibrary.org/books/OL24341821M/ECDEU_assessment_manual_for_psychopharmacology [accessed 2022-08-03]
28. van Wijngaarden B, Schene AH, Koeter M, Vázquez-Barquero JL, Knudsen HC, Lasalvia A, et al. Caregiving in schizophrenia: development, internal consistency and reliability of the Involvement Evaluation Questionnaire--European Version. EPSILON Study 4. European Psychiatric Services: inputs linked to outcome domains and needs. Br J Psychiatry Suppl 2000;177(S39):s21-s27. [doi: 10.1192/bjp.177.39.s21] [Medline: 10945074]
29. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care 1996 Mar;34(3):220-233. [doi: 10.1097/00005650-199603000-00003] [Medline: 8628042]
30. Alphs L, Benson C, Cheshire-Kinney K, Lindenmayer JP, Mao L, Rodriguez SC, et al. Real-world outcomes of paliperidone palmitate compared to daily oral antipsychotic therapy in schizophrenia: a randomized, open-label, review board-blinded 15-month study. J Clin Psychiatry 2015 May;76(5):554-561 [FREE Full text] [doi: 10.4088/JCP.14m09584] [Medline: 25938474]
31. INVEGA SUSTENNA® (paliperidone palmitate). Titusville, NJ: Janssen Pharmaceuticals Inc; 2021.
32. INVEGA TRINZA® (paliperidone palmitate). Titusville, NJ: Janssen Pharmaceuticals Inc; 2021.
33. Savitz AJ, Xu H, Gopal S, Nuamah I, Ravenstijn P, Janik A, et al. Efficacy and safety of paliperidone palmitate 3-month formulation for patients with schizophrenia: a randomized, multicenter, double-blind, noninferiority study. Int J Neuropsychopharmacol 2016 Jul;19(7):pyw018 [FREE Full text] [doi: 10.1093/ijnp/pyw018] [Medline: 26902950]
34. McFarlane WR. Family interventions for schizophrenia and the psychoses: a review. Fam Process 2016 Sep;55(3):460-482. [doi: 10.1111/famp.12235] [Medline: 27411376]
35. Hogarty GE, Anderson CM, Reiss DJ, Kornblith SJ, Greenwald DP, Javna CD, et al. Family psychoeducation, social skills training, and maintenance chemotherapy in the aftercare treatment of schizophrenia. I. One-year effects of a controlled study on relapse and expressed emotion. Arch Gen Psychiatry 1986 Jul;43(7):633-642. [doi: 10.1001/archpsyc.1986.01800070019003] [Medline: 2872870]
36. Harvey C. Family psychoeducation for people living with schizophrenia and their families. BJPsych Advances 2018 Jan;24(1):9-19. [doi: 10.1192/bja.2017.4]
37. Falloon IR, Boyd JL, McGill CW, Williamson M, Razani J, Moss HB, et al. Family management in the prevention of morbidity of schizophrenia. Clinical outcome of a two-year longitudinal study. Arch Gen Psychiatry 1985 Sep;42(9):887-896. [doi: 10.1001/archpsyc.1985.01790320059008] [Medline: 2864032]
38. McFarlane WR, Lukens E, Link B, Dushay R, Deakins SA, Newmark M, et al. Multiple-family groups and psychoeducation in the treatment of schizophrenia. Arch Gen Psychiatry 1995 Aug;52(8):679-687. [doi: 10.1001/archpsyc.1995.03950200069016] [Medline: 7632121]
39. Dunning D, Johnson K, Ehrlinger J, Kruger J. Why people fail to recognize their own incompetence. Curr Dir Psychol Sci 2003 Jun 1;12(3):83-87. [doi: 10.1111/1467-8721.01235]
40. Williams CC. Insight, stigma, and post-diagnosis identities in schizophrenia. Psychiatry 2008;71(3):246-256. [doi: 10.1521/psyc.2008.71.3.246] [Medline: 18834275]
41. Mann DM, Chen J, Chunara R, Testa PA, Nov O. COVID-19 transforms health care through telemedicine: evidence from the field. J Am Med Inform Assoc 2020 Jul 01;27(7):1132-1135 [FREE Full text] [doi: 10.1093/jamia/ocaa072] [Medline: 32324855]
42. Guinart D, Marcy P, Hauser M, Dwyer M, Kane JM. Patient attitudes toward telepsychiatry during the COVID-19 pandemic: a nationwide, multisite survey. JMIR Ment Health 2020 Dec 22;7(12):e24761 [FREE Full text] [doi: 10.2196/24761] [Medline: 33302254]
43. Kopec K, Janney CA, Johnson B, Spykerman K, Ryskamp B, Achtyes ED. Rapid transition to telehealth in a community mental health service provider during the COVID-19 pandemic. Prim Care Companion CNS Disord 2020;22(5):20br02787. [doi: 10.4088/pcc.20br02787]
44. Kozloff N, Mulsant BH, Stergiopoulos V, Voineskos AN. The COVID-19 global pandemic: implications for people with schizophrenia and related disorders. Schizophr Bull 2020 Jul 08;46(4):752-757 [FREE Full text] [doi: 10.1093/schbul/sbaa051] [Medline: 32343342]
45. Stoll J, Sadler JZ, Trachsel M. The ethical use of telepsychiatry in the Covid-19 pandemic. Front Psychiatry 2020;11:665 [FREE Full text] [doi: 10.3389/fpsyt.2020.00665] [Medline: 32760299]
46. Avalone L, Barron C, King C, Linn-Walton R, Lau J, McQuistion HL, et al. Rapid telepsychiatry implementation during COVID-19: increased attendance at the largest health system in the United States. Psychiatr Serv 2021 Jun;72(6):708-711. [doi: 10.1176/appi.ps.202000574] [Medline: 33730881]
47. Sin J, Henderson C, Woodham LA, Sesé Hernández A, Gillard S. A multicomponent eHealth intervention for family carers for people affected by psychosis: a coproduced design and build study. J Med Internet Res 2019 Aug 06;21(8):e14374 [FREE Full text] [doi: 10.2196/14374] [Medline: 31389333]

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

Edited by J Torous; submitted 03.08.21; peer-reviewed by A Kopelowicz, E Moran; comments to author 02.10.21; revised version received 22.12.21; accepted 20.01.22; published 15.04.22.

Please cite as:
Mueser KT, Achtyes ED, Gogate J, Mancevski B, Kim E, Starr HL
Telehealth-Based Psychoeducation for Caregivers: The Family Intervention in Recent-Onset Schizophrenia Treatment Study
JMIR Ment Health 2022;9(4):e32492
URL: https://mental.jmir.org/2022/4/e32492
doi:10.2196/32492
PMID:35436231

Review

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

Holly Shannon[1,2], BSc; Katie Bush[1], BA; Paul J Villeneuve[2,3*], PhD; Kim GC Hellemans[2*], PhD; Synthia Guimond[1,2,4*], PhD

[1]Department of Psychiatry, The Royal's Institute of Mental Health Research, University of Ottawa, Ottawa, ON, Canada

[2]Department of Neuroscience, Carleton University, Ottawa, ON, Canada

[3]School of Mathematics and Statistics, Carleton University, Ottawa, ON, Canada

[4]Department 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, Andreassen 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.

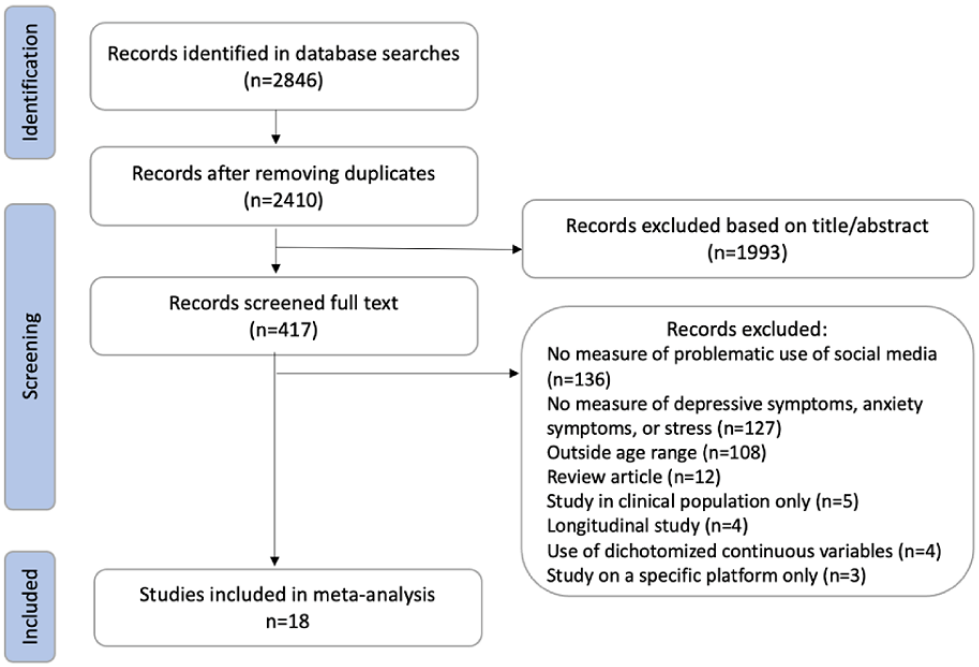


Table 1. Summary of included studies on the relationship between social media use and outcome variables (note that not all studies measured all three outcomes. Giordano et al [33] assessed anxiety and depression combined and was therefore only included in the meta-regression analyses).

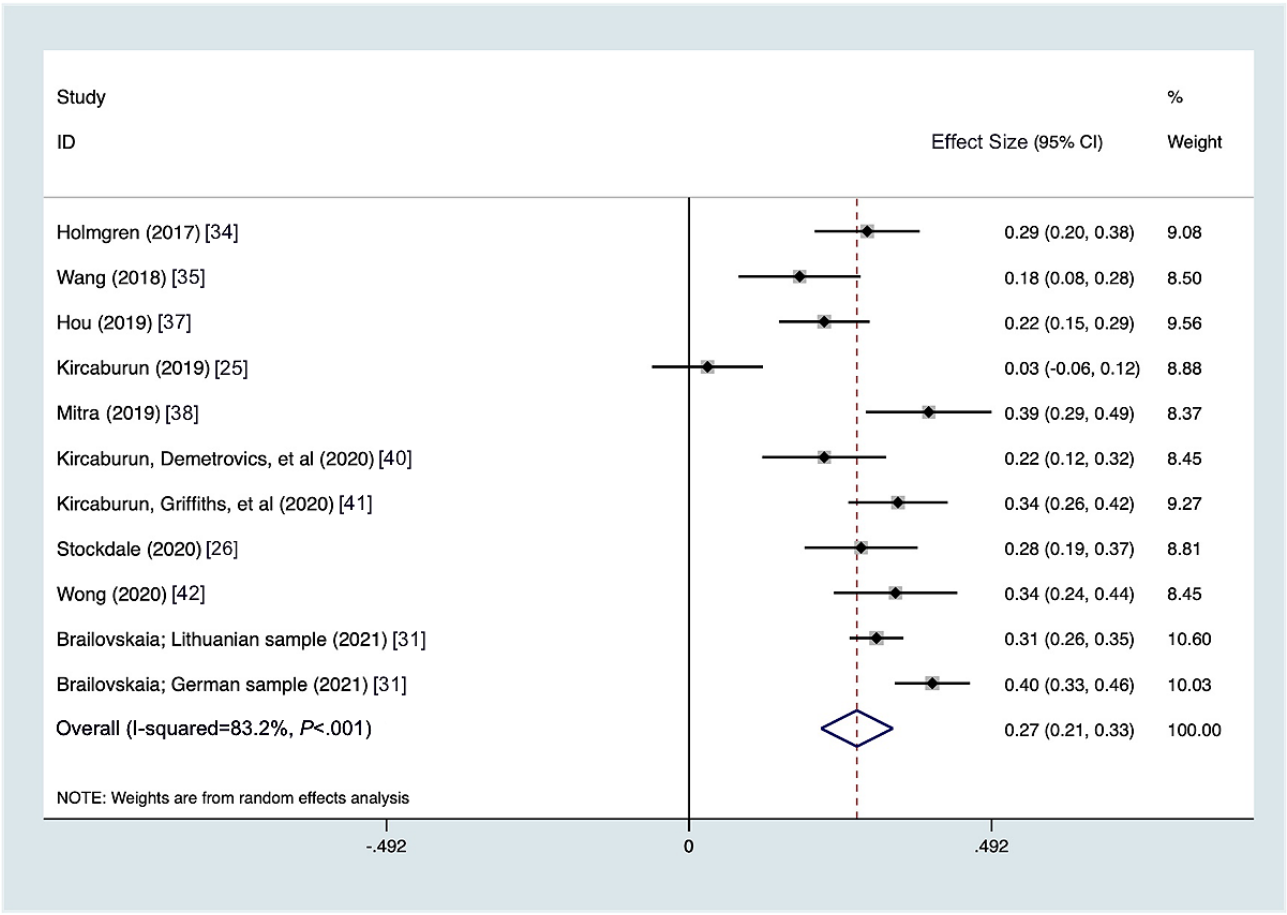
First author (year)	Sample size	Female, n (%)	Male, n (%)	Age (years) range (mean)	Country	Problematic use and depression (*r*)	Problematic use and anxiety (*r*)	Problematic use and stress (*r*)	Problematic use and depression and anxiety combined (*r*)
Holmgren (2017) [34]	442	228 (51.6)	214 (48.4)	18-21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

[a]N/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 I^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).

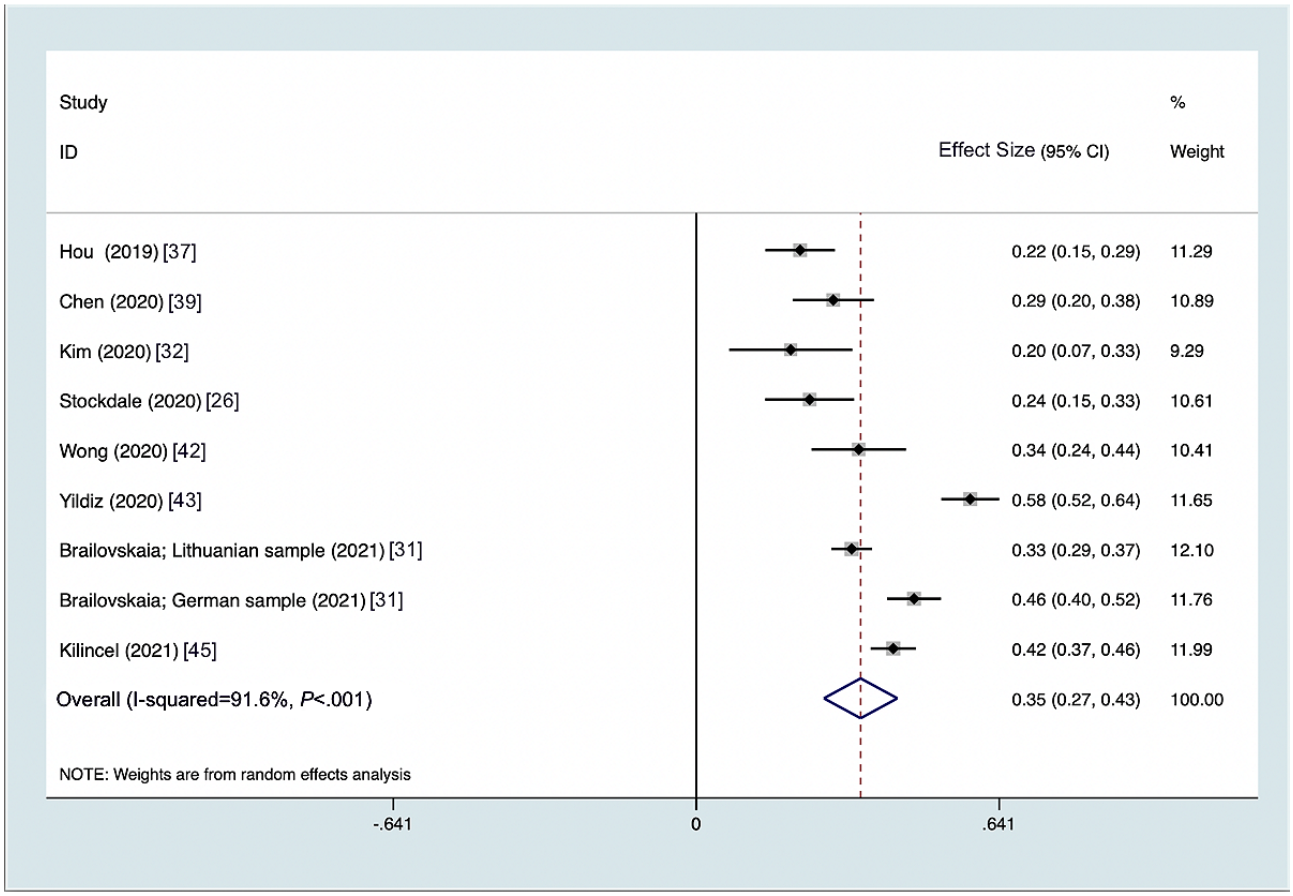
Figure 2. Forest plot of depressive symptoms and problematic social media use by year.



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 I^2=91.6%, Q^2=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).

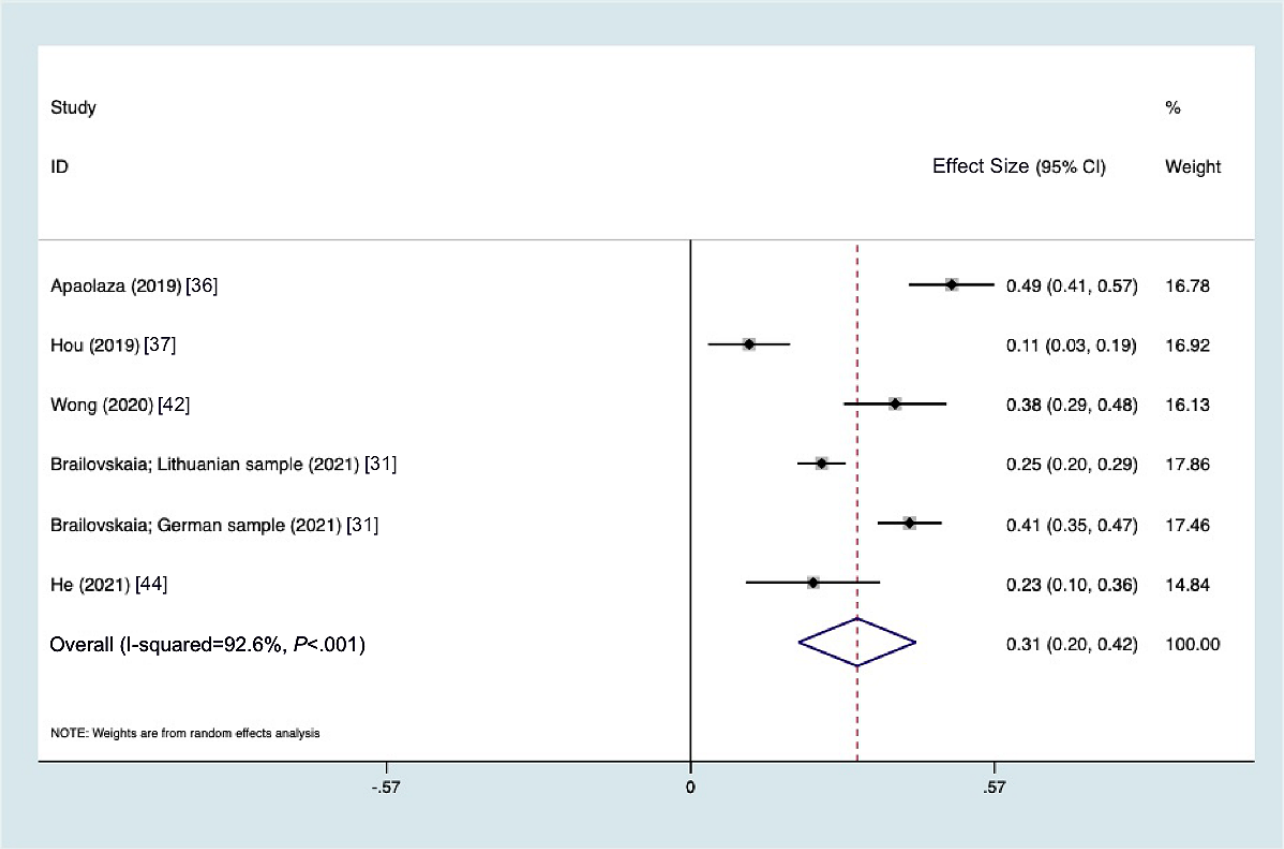
Figure 3. Forest plot of anxiety symptoms and problematic social media use by year.



Problematic Social Media Use and Stress

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).

Figure 4. Forest plot of stress and problematic social media use by year.



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.
[DOCX File , 399 KB - mental_v9i4e33450_app1.docx]

References

1. Andreassen CS, Pallesen S, Griffiths MD. The relationship between addictive use of social media, narcissism, and self-esteem: Findings from a large national survey. Addict Behav 2017 Jan;64:287-293. [doi: 10.1016/j.addbeh.2016.03.006] [Medline: 27072491]
2. Villanti AC, Johnson AL, Ilakkuvan V, Jacobs MA, Graham AL, Rath JM. Social Media Use and Access to Digital Technology in US Young Adults in 2016. J Med Internet Res 2017 Jun 07;19(6):e196 [FREE Full text] [doi: 10.2196/jmir.7303] [Medline: 28592394]
3. Woods HC, Scott H. #Sleepyteens: Social media use in adolescence is associated with poor sleep quality, anxiety, depression and low self-esteem. J Adolesc 2016 Aug;51:41-49. [doi: 10.1016/j.adolescence.2016.05.008] [Medline: 27294324]
4. Vannucci A, Flannery KM, Ohannessian CM. Social media use and anxiety in emerging adults. J Affect Disord 2017 Jan 01;207:163-166. [doi: 10.1016/j.jad.2016.08.040] [Medline: 27723539]
5. Griffiths MD, Kuss DJ, Demetrovics Z. Chapter 6 - Social Networking Addiction: An Overview of Preliminary Findings. In: Behavioral Addictions Criteria, Evidence, and Treatment. London, UK: Academic Press; 2014:119-141.
6. Pantic I, Damjanovic A, Todorovic J, Topalovic D, Bojovic-Jovic D, Ristic S, et al. Association between online social networking and depression in high school students: behavioral physiology viewpoint. Psychiatr Danub 2012 Mar;24(1):90-93 [FREE Full text] [Medline: 22447092]
7. Lin LY, Sidani JE, Shensa A, Radovic A, Miller E, Colditz JB, et al. Association Between Social Media Use and Depression Among U.S. Young Adults. Depress Anxiety 2016 Apr;33(4):323-331 [FREE Full text] [doi: 10.1002/da.22466] [Medline: 26783723]
8. McDougall MA, Walsh M, Wattier K, Knigge R, Miller L, Stevermer M, et al. The effect of social networking sites on the relationship between perceived social support and depression. Psychiatry Res 2016 Dec 30;246:223-229. [doi: 10.1016/j.psychres.2016.09.018] [Medline: 27721061]
9. Twenge JM, Campbell WK. Media Use Is Linked to Lower Psychological Well-Being: Evidence from Three Datasets. Psychiatr Q 2019 Jun;90(2):311-331. [doi: 10.1007/s11126-019-09630-7] [Medline: 30859387]
10. Sampasa-Kanyinga H, Lewis RF. Frequent Use of Social Networking Sites Is Associated with Poor Psychological Functioning Among Children and Adolescents. Cyberpsychol Behav Soc Netw 2015 Jul;18(7):380-385. [doi: 10.1089/cyber.2015.0055] [Medline: 26167836]
11. Chen W, Lee K. Sharing, liking, commenting, and distressed? The pathway between Facebook interaction and psychological distress. Cyberpsychol Behav Soc Netw 2013 Oct;16(10):728-734. [doi: 10.1089/cyber.2012.0272] [Medline: 23745614]
12. Stronge S, Mok T, Ejova A, Lee C, Zubielevitch E, Yogeeswaran K, et al. Social Media Use Is (Weakly) Related to Psychological Distress. Cyberpsychol Behav Soc Netw 2019 Sep 01;22(9):604-609. [doi: 10.1089/cyber.2019.0176] [Medline: 31526298]
13. Ivie EJ, Pettitt A, Moses LJ, Allen NB. A meta-analysis of the association between adolescent social media use and depressive symptoms. J Affect Disord 2020 Oct 01;275:165-174. [doi: 10.1016/j.jad.2020.06.014] [Medline: 32734903]
14. Nabi RL, Prestin A, So J. Facebook friends with (health) benefits? Exploring social network site use and perceptions of social support, stress, and well-being. Cyberpsychol Behav Soc Netw 2013 Oct;16(10):721-727. [doi: 10.1089/cyber.2012.0521] [Medline: 23790356]
15. Campisi J, Folan D, Diehl G, Kable T, Rademeyer C. Social media users have different experiences, motivations, and quality of life. Psychiatry Res 2015 Aug 30;228(3):774-780. [doi: 10.1016/j.psychres.2015.04.042] [Medline: 26054935]
16. Bányai F, Zsila Á, Király O, Maraz A, Elekes Z, Griffiths MD, et al. Problematic Social Media Use: Results from a Large-Scale Nationally Representative Adolescent Sample. PLoS One 2017 Jan 9;12(1):e0169839 [FREE Full text] [doi: 10.1371/journal.pone.0169839] [Medline: 28068404]
17. Andreassen CS, Torsheim T, Brunborg GS, Pallesen S. Development of a Facebook Addiction Scale. Psychol Rep 2012 Apr;110(2):501-517. [doi: 10.2466/02.09.18.PR0.110.2.501-517] [Medline: 22662404]
18. Schou Andreassen C, Billieux J, Griffiths MD, Kuss DJ, Demetrovics Z, Mazzoni E, et al. The relationship between addictive use of social media and video games and symptoms of psychiatric disorders: A large-scale cross-sectional study. Psychol Addict Behav 2016 Mar;30(2):252-262. [doi: 10.1037/adb0000160] [Medline: 26999354]
19. Hanprathet N, Manwong M, Khumsri J, Yingyeun R, Phanasathit M. Facebook Addiction and Its Relationship with Mental Health among Thai High School Students. J Med Assoc Thai 2015 Apr;98 Suppl 3:S81-S90. [Medline: 26387393]
20. Brailovskaia J, Schillack H, Margraf J. Facebook Addiction Disorder in Germany. Cyberpsychol Behav Soc Netw 2018 Jul;21(7):450-456. [doi: 10.1089/cyber.2018.0140] [Medline: 29995531]
21. Jeri-Yabar A, Sanchez-Carbonel A, Tito K, Ramirez-delCastillo J, Torres-Alcantara A, Denegri D, et al. Association between social media use (Twitter, Instagram, Facebook) and depressive symptoms: Are Twitter users at higher risk? Int J Soc Psychiatry 2019 Feb 30;65(1):14-19. [doi: 10.1177/0020764018814270] [Medline: 30497315]

22. O'Reilly M, Dogra N, Whiteman N, Hughes J, Eruyar S, Reilly P. Is social media bad for mental health and wellbeing? Exploring the perspectives of adolescents. Clin Child Psychol Psychiatry 2018 Oct 20;23(4):601-613. [doi: 10.1177/1359104518775154] [Medline: 29781314]
23. Cunningham S, Hudson CC, Harkness K. Social Media and Depression Symptoms: a Meta-Analysis. Res Child Adolesc Psychopathol 2021 Feb 06;49(2):241-253. [doi: 10.1007/s10802-020-00715-7] [Medline: 33404948]
24. Crone EA, Konijn EA. Media use and brain development during adolescence. Nat Commun 2018 Feb 21;9(1):588 [FREE Full text] [doi: 10.1038/s41467-018-03126-x] [Medline: 29467362]
25. Kircaburun K, Griffiths MD, Billieux J. Trait emotional intelligence and problematic online behaviors among adolescents: The mediating role of mindfulness, rumination, and depression. Personality and Individual Differences 2019 Mar;139:208-213. [doi: 10.1016/j.paid.2018.11.024]
26. Stockdale LA, Coyne SM. Bored and online: Reasons for using social media, problematic social networking site use, and behavioral outcomes across the transition from adolescence to emerging adulthood. J Adolesc 2020 Feb;79:173-183. [doi: 10.1016/j.adolescence.2020.01.010] [Medline: 31978836]
27. Kormas G, Critselis E, Janikian M, Kafetzis D, Tsitsika A. Risk factors and psychosocial characteristics of potential problematic and problematic internet use among adolescents: a cross-sectional study. BMC Public Health 2011 Jul 27;11:595 [FREE Full text] [doi: 10.1186/1471-2458-11-595] [Medline: 21794167]
28. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
29. Herzog R, Álvarez-Pasquin MJ, Díaz C, Del Barrio JL, Estrada JM, Gil Á. Are healthcare workers' intentions to vaccinate related to their knowledge, beliefs and attitudes? A systematic review. BMC Public Health 2013 Mar 19;13:154 [FREE Full text] [doi: 10.1186/1471-2458-13-154] [Medline: 23421987]
30. Stata Statistical Software: Release 16. StataCorp LLC. 2019. URL: https://www.stata.com/support/faqs/resources/citing-software-documentation-faqs/ [accessed 2020-11-08]
31. Brailovskaia J, Truskauskaite-Kuneviciene I, Kazlauskas E, Margraf J. The patterns of problematic social media use (SMU) and their relationship with online flow, life satisfaction, depression, anxiety and stress symptoms in Lithuania and in Germany. Curr Psychol 2021 Apr 15:w. [doi: 10.1007/s12144-021-01711-w]
32. Kim M, Oh J, Huh B. Analysis of Factors Related to Social Network Service Addiction Among Korean High School Students. J Addict Nurs 2020;31(3):203-212. [doi: 10.1097/JAN.0000000000000350] [Medline: 32868612]
33. Giordano AL, Prosek EA, Watson JC. Understanding Adolescent Cyberbullies: Exploring Social Media Addiction and Psychological Factors. Journal of Child and Adolescent Counseling 2021 Feb 03;7(1):42-55. [doi: 10.1080/23727810.2020.1835420]
34. Holmgren HG, Coyne SM. Can’t stop scrolling!: pathological use of social networking sites in emerging adulthood. Addiction Research & Theory 2017 Mar 07;25(5):375-382 [FREE Full text] [doi: 10.1080/16066359.2017.1294164]
35. Wang P, Wang X, Wu Y, Xie X, Wang X, Zhao F, et al. Social networking sites addiction and adolescent depression: A moderated mediation model of rumination and self-esteem. Personality and Individual Differences 2018 Jun;127:162-167. [doi: 10.1016/j.paid.2018.02.008]
36. Apaolaza V, Hartmann P, D'Souza C, Gilsanz A. Mindfulness, Compulsive Mobile Social Media Use, and Derived Stress: The Mediating Roles of Self-Esteem and Social Anxiety. Cyberpsychol Behav Soc Netw 2019 Jun;22(6):388-396. [doi: 10.1089/cyber.2018.0681] [Medline: 31070455]
37. Hou X, Wang H, Hu T, Gentile DA, Gaskin J, Wang J. The relationship between perceived stress and problematic social networking site use among Chinese college students. J Behav Addict 2019 Jun 01;8(2):306-317 [FREE Full text] [doi: 10.1556/2006.8.2019.26] [Medline: 31172814]
38. Mitra R, Rangaswamy DM. Excessive Social Media use and its Association with Depression and Rumination in an Indian Young Adult Population: A Mediation Model. Journal of Psychosocial Research 2019 Jul 25;14(1):223-231 [FREE Full text] [doi: 10.32381/jpr.2019.14.01.24]
39. Chen Y, Li R, Zhang P, Liu X. The Moderating Role of State Attachment Anxiety and Avoidance Between Social Anxiety and Social Networking Sites Addiction. Psychol Rep 2020 Jun 06;123(3):633-647. [doi: 10.1177/0033294118823178] [Medline: 30612521]
40. Kircaburun K, Demetrovics Z, Király O, Griffiths MD. Childhood Emotional Trauma and Cyberbullying Perpetration Among Emerging Adults: A Multiple Mediation Model of the Role of Problematic Social Media Use and Psychopathology. Int J Ment Health Addiction 2018 May 23;18(3):548-566. [doi: 10.1007/s11469-018-9941-5]
41. Kircaburun K, Griffiths MD, Şahin F, Bahtiyar M, Atmaca T, Tosuntaş Ş. The Mediating Role of Self/Everyday Creativity and Depression on the Relationship Between Creative Personality Traits and Problematic Social Media Use Among Emerging Adults. Int J Ment Health Addiction 2018 May 14;18(1):77-88. [doi: 10.1007/s11469-018-9938-0]
42. Wong HY, Mo HY, Potenza MN, Chan MNM, Lau WM, Chui TK, et al. Relationships between Severity of Internet Gaming Disorder, Severity of Problematic Social Media Use, Sleep Quality and Psychological Distress. Int J Environ Res Public Health 2020 Mar 13;17(6):1879 [FREE Full text] [doi: 10.3390/ijerph17061879] [Medline: 32183188]

43. Yıldız Durak H. Modeling of variables related to problematic internet usage and problematic social media usage in adolescents. Curr Psychol 2018 Apr 3;39(4):1375-1387. [doi: 10.1007/s12144-018-9840-8]
44. He Z, Li M, Ma X, Liu C. Family Socioeconomic Status and Social Media Addiction in Female College Students: The Mediating Role of Impulsiveness and Inhibitory Control. J Genet Psychol 2021 Dec 09;182(1):60-74. [doi: 10.1080/00221325.2020.1853027] [Medline: 33292090]
45. Kılınçel S, Muratdağı G. Evaluation of factors affecting social media addiction in adolescents during the COVID-19 pandemic. Ann Clin Anal Med 2021;12(04):447-451. [doi: 10.4328/acam.20541]
46. Yoon S, Kleinman M, Mertz J, Brannick M. Is social network site usage related to depression? A meta-analysis of Facebook-depression relations. J Affect Disord 2019 Apr 01;248:65-72. [doi: 10.1016/j.jad.2019.01.026] [Medline: 30711871]
47. Vahedi Z, Zannella L. The association between self-reported depressive symptoms and the use of social networking sites (SNS): A meta-analysis. Curr Psychol 2019 Jan 26;40(5):2174-2189 [FREE Full text] [doi: 10.1007/s12144-019-0150-6]
48. Orben A. Teenagers, screens and social media: a narrative review of reviews and key studies. Soc Psychiatry Psychiatr Epidemiol 2020 Apr 10;55(4):407-414. [doi: 10.1007/s00127-019-01825-4] [Medline: 31925481]
49. Thorisdottir IE, Sigurvinsdottir R, Asgeirsdottir BB, Allegrante JP, Sigfusdottir ID. Active and Passive Social Media Use and Symptoms of Anxiety and Depressed Mood Among Icelandic Adolescents. Cyberpsychol Behav Soc Netw 2019 Aug;22(8):535-542. [doi: 10.1089/cyber.2019.0079] [Medline: 31361508]
50. Aalbers G, McNally RJ, Heeren A, de Wit S, Fried EI. Social media and depression symptoms: A network perspective. J Exp Psychol Gen 2019 Aug;148(8):1454-1462. [doi: 10.1037/xge0000528] [Medline: 30507215]
51. Barry CT, Sidoti CL, Briggs SM, Reiter SR, Lindsey RA. Adolescent social media use and mental health from adolescent and parent perspectives. J Adolesc 2017 Dec;61:1-11. [doi: 10.1016/j.adolescence.2017.08.005] [Medline: 28886571]
52. Marino C, Mazzieri E, Caselli G, Vieno A, Spada MM. Motives to use Facebook and problematic Facebook use in adolescents. J Behav Addict 2018 Jun 01;7(2):276-283 [FREE Full text] [doi: 10.1556/2006.7.2018.32] [Medline: 29846087]
53. Orben A, Przybylski AK. Screens, Teens, and Psychological Well-Being: Evidence From Three Time-Use-Diary Studies. Psychol Sci 2019 May;30(5):682-696 [FREE Full text] [doi: 10.1177/0956797619830329] [Medline: 30939250]
54. Przybylski AK, Weinstein N. A Large-Scale Test of the Goldilocks Hypothesis. Psychol Sci 2017 Feb;28(2):204-215 [FREE Full text] [doi: 10.1177/0956797616678438] [Medline: 28085574]
55. Tsitsika AK, Tzavela EC, Janikian M, Ólafsson K, Iordache A, Schoenmakers TM, et al. Online social networking in adolescence: patterns of use in six European countries and links with psychosocial functioning. J Adolesc Health 2014 Jul;55(1):141-147. [doi: 10.1016/j.jadohealth.2013.11.010] [Medline: 24618179]
56. Banjanin N, Banjanin N, Dimitrijevic I, Pantic I. Relationship between internet use and depression: Focus on physiological mood oscillations, social networking and online addictive behavior. Computers in Human Behavior 2015 Feb;43:308-312. [doi: 10.1016/j.chb.2014.11.013]
57. Spies Shapiro LA, Margolin G. Growing up wired: social networking sites and adolescent psychosocial development. Clin Child Fam Psychol Rev 2014 Mar;17(1):1-18 [FREE Full text] [doi: 10.1007/s10567-013-0135-1] [Medline: 23645343]
58. Twenge JM, Joiner TE, Rogers ML, Martin GN. Increases in Depressive Symptoms, Suicide-Related Outcomes, and Suicide Rates Among U.S. Adolescents After 2010 and Links to Increased New Media Screen Time. Clinical Psychological Science 2017 Nov 14;6(1):3-17. [doi: 10.1177/2167702617723376]
59. Brooks S, Longstreet P. Social networking's peril: Cognitive absorption, social networking usage, and depression. Cyberpsychology 2015 Dec 01;9(4):CP2015-4-5. [doi: 10.5817/cp2015-4-5]
60. Lenhart A. Pew Research Center. 2015. URL: https://www.pewresearch.org/internet/2015/04/09/teens-social-media-technology-2015/ [accessed 2022-03-30]
61. Van Zalk N. Social anxiety moderates the links between excessive chatting and compulsive Internet use. Cyberpsychology 2016 Oct 01;10(3):CP2016-3-3. [doi: 10.5817/cp2016-3-3]
62. Vuorre M, Orben A, Przybylski AK. There Is No Evidence That Associations Between Adolescents' Digital Technology Engagement and Mental Health Problems Have Increased. Clinical Psychological Science 2021 May 03;9(5):823-835. [doi: 10.1177/2167702621994549]

Abbreviations

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

Edited by J Torous; submitted 08.09.21; peer-reviewed by A Daros, N van Zalk; comments to author 19.11.21; revised version received 01.02.22; accepted 05.02.22; published 14.04.22.

Please cite as:
Shannon H, Bush K, Villeneuve PJ, Hellemans KGC, Guimond S
Problematic Social Media Use in Adolescents and Young Adults: Systematic Review and Meta-analysis
JMIR Ment Health 2022;9(4):e33450
URL: https://mental.jmir.org/2022/4/e33450
doi:10.2196/33450
PMID:35436240

Review

Social Robot Interventions in Mental Health Care and Their Outcomes, Barriers, and Facilitators: Scoping Review

Imane Guemghar[1,2]; Paula Pires de Oliveira Padilha[1], MD; Amal Abdel-Baki[1,3], MD, MSc; Didier Jutras-Aswad[1,3], MD, MSc; Jesseca Paquette[1], MSc; Marie-Pascale Pomey[1,2,4,5], MD, PhD

[1]Centre de Recherche du Centre Hospitalier de l'Université de Montréal, Montreal, QC, Canada

[2]Faculté de Médecine, Université de Montréal, Montreal, QC, Canada

[3]Département de Psychiatrie et d'Addictologie, Université de Montréal, Montreal, QC, Canada

[4]Centre d'Excellence pour le Partenariat avec les Patients et le Public, Montreal, QC, Canada

[5]Département de Gestion, Évaluation et Politique de Santé, École de Santé Publique de l'Université de Montréal, Montreal, QC, Canada

Corresponding Author:
Marie-Pascale Pomey, MD, PhD
Centre de Recherche du Centre Hospitalier de l'Université de Montréal
850 rue Saint-Denis
Montreal, QC, H2X 0A9
Canada
Phone: 1 514 343 6111 ext 1364
Email: marie-pascale.pomey@umontreal.ca

Abstract

Background: The use of social robots as innovative therapeutic tools has been increasingly explored in recent years in an effort to address the growing need for alternative intervention modalities in mental health care.

Objective: The aim of this scoping review was to identify and describe social robot interventions in mental health facilities and to highlight their outcomes as well as the barriers and facilitators to their implementation.

Methods: A scoping review of the literature published since 2015 was conducted using the Arksey and O'Malley's framework. The MEDLINE, Embase, Cochrane Central Register of Controlled Trials, and PsycINFO databases were searched, and 2239 papers were retrieved. The papers included were primary empirical studies published in peer-reviewed literature. Eligible studies were set in mental health facilities and they included participants with a known mental health disorder. The methodological quality of the included papers was also assessed using the Mixed Methods Appraisal Tool.

Results: A total of 30 papers met the eligibility criteria for this review. Studies involved participants with dementia, cognitive impairment, schizophrenia, depression, autism spectrum disorder, attention-deficit hyperactivity disorder, and an intellectual disability. The outcomes studied included engagement, social interaction, emotional state, agitation, behavior, and quality of life.

Conclusions: The methodological weaknesses of the studies conducted this far and the lack of diversity in the conditions studied limit the generalizability of the results. However, despite the presence of certain barriers to their implementation (eg, technical problems, unsuitable environment, staff resistance), social robot interventions generally show positive effects on patients with mental health disorders. Studies of stronger methodological quality are needed to further understand the benefits and the place of social robots in mental health care.

(JMIR Ment Health 2022;9(4):e36094) doi:10.2196/36094

KEYWORDS

social robots; socially assistive robots; SARs; mental health; mental health services; dementia; autism spectrum disorder; schizophrenia; depression; scoping review

Introduction

Health care needs are on the rise. Faced with a shortage of staff, equipment, and funding, the quest for innovative solutions to address these needs is thriving. Among emerging solutions, the use of robots is increasingly popular. Indeed, robots are becoming more prominent in the health industry, where they are already employed as surgery, drug delivery, and diagnosis devices [1]. Lately, the use of socially assistive robots (SARs) is attracting the interest of many researchers.

SARs (or social robots) are robots meant to provide assistance through social interaction [2]. Their built-in sound, image, and motion sensors enable them to respond autonomously to a user and his environment [3-5]. SARs can be classified into 2 categories: companion and service-type robots. Although companion robots offer psychological support to the patient, service-type robots provide functional assistance to complete daily tasks [5]. It is worth noting that while this distinction may be found in the literature, many social robots can be featured in both categories. SARs, often in animal or humanoid forms, have a variety of functionalities to engage a user's attention [6,7]. Animal-like robots are created to reproduce the physiological, psychological, cognitive, and socioemotional benefits of animal-assisted therapy without the associated inconveniences [8-11]. Real animals can cause allergies and evoke fear in some patients [12]. Pet robots require much less maintenance and are considered a safer choice for therapy in a care setting [13]. The reduced noise level, the diminished workload requirement, and the lower costs are the additional benefits [14,15]. Pet robots generally fall into the category of companion robots. Conversely, SARs embodied in a humanoid appearance show the highest levels of acceptability and usability among participants. These robots, with humanlike facial features, communication modalities, and motion patterns, seem to create a more natural interaction [16-22]. Some can converse, play music, and display images or videos. Others may even perform movements to demonstrate a set of physical exercises to an audience. Humanoid robots are usually considered to be service-type robots.

Although research is still in its early stages, SAR interventions have been carried out in a number of areas in health care. In pediatric research, studies suggest that social robots could contribute to the reduction of pain and distress in hospitalized children [23,24]. Other studies, including participants with autism spectrum disorders (ASDs), also showed that social robots could be used to teach certain behaviors and communication skills [25-27]. More often, studies with social robots are conducted with a geriatric population. With this population, it has been determined that SARs could be used to improve physical exercise and monitor health status [28,29]. In this respect, a recent randomized clinical trial found that social robots improved the adherence to medications and rehabilitation exercises in older adults with chronic obstructive pulmonary disease [30].

Currently, research with SARs is focused on their use in mental health care. In a paper published in 2015, Rabbitt et al [7] discussed social robots' applications in mental health care. Among numerous observations, it was pointed out that the clinical application of social robots was limited to few diagnoses. Indeed, numerous studies showed beneficial effects of social robots' intervention on the quality of life and well-being of people with dementia [5]. Other mental health conditions were not given the same degree of consideration. It was also noted that the quality and amount of evidence available lacked strength. These elements were also raised in other reviews [31,32]. To improve understanding of how social robots have been used to help people in mental health care in recent years, we conducted a scoping review to identify the outcomes, barriers, and facilitators of SAR interventions. Although there have been reviews of SAR use in other health care contexts, reviews solely focused on mental health care are lacking [33-35]. Furthermore, recent reviews on SARs have either limited the scope of their review to a precise diagnosis, to an exact type of robot, or to a population of certain age [36-39]. To fully understand how social robots could be used in mental health care, we chose to avoid such limits.

Methods

The PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist was used as a guideline to ensure the methodological transparency of this review [40]. This scoping review was conducted following the Arksey and O'Malley's framework [41]. The framework consists of the following 5 stages.

Stage 1: Identifying the Research Questions

This scoping review addresses the following research questions:

1. What types of social robots have been used in mental health care in the past years?
2. What were the outcomes of social robot interventions?
3. What were the barriers and facilitators of their implementation?
4. Based on the results of our scoping review, what aspects require further research?

Stage 2: Identifying Relevant Studies

Eligibility Criteria

The following eligibility criteria were established to guide the literature review:

1. Date of publication: The field of robotics is ever changing, and improvements are made at an astonishing speed. The limitations identified several years ago are not the same as those currently encountered. Since we wanted an up-to-date portrait of the use of social robots in mental health care, we reviewed all publications only from 2015 to the present.
2. Language of publication: The language of the studies was restricted to English.
3. Study design: Included papers were restricted to primary empirical studies (eg, quantitative, qualitative, or mixed methods) published in the peer-reviewed literature. Publications were excluded if they were considered gray literature (eg, reports, theses, newsletters).
4. Setting: Eligible studies were set in mental health facilities. Hospitals and nursing homes were included. Studies set in patients' own homes were excluded. Studies set in schools were also excluded.
5. Population: Participants of eligible studies had a mental health disorder. A mental health disorder was defined as the existence of a clinically recognizable set of symptoms or behavior associated in most cases with distress and with interference with personal functions according to the International Classification of Diseases, 10th edition [42]. No restrictions were applied on the population age.
6. Program of care or intervention: Eligible studies implemented an evidence-based social robot program or

intervention in mental health facilities. Teaching programs involving social robots were excluded. For instance, interventions using robots to teach communication skills to participants with ASDs were excluded. Brain training programs where robots provided exercises to improve cognition and memory in people with dementia were also excluded for the same reason.

Search Strategy

The following electronic databases were searched using the Ovid research platform: MEDLINE, Embase, Cochrane Central Register of Controlled Trials, and PsycINFO. The search strategy was developed in Ovid MEDLINE. It consisted of keywords and subject headings (Textbox 1). It was subsequently adapted for other databases. The final search strategy was validated by an experienced librarian to ensure that the literature was covered in a comprehensive manner. The electronic databases were first searched on March 26, 2021 and then searched again on November 2, 2021.

Textbox 1. Search strategy in Ovid MEDLINE.

[psychology.fs. AND Robotics/] OR Self-Help Devices/px [Psychology] OR companion robot* OR social robot* OR human* robot* OR robopet* OR (Social* adj2 robot*) OR (pet* adj1 robot*) OR (therap* adj1 robot*) OR (animal* adj1 robot*) OR non-human* robot* OR (interacti* adj1 robot*)

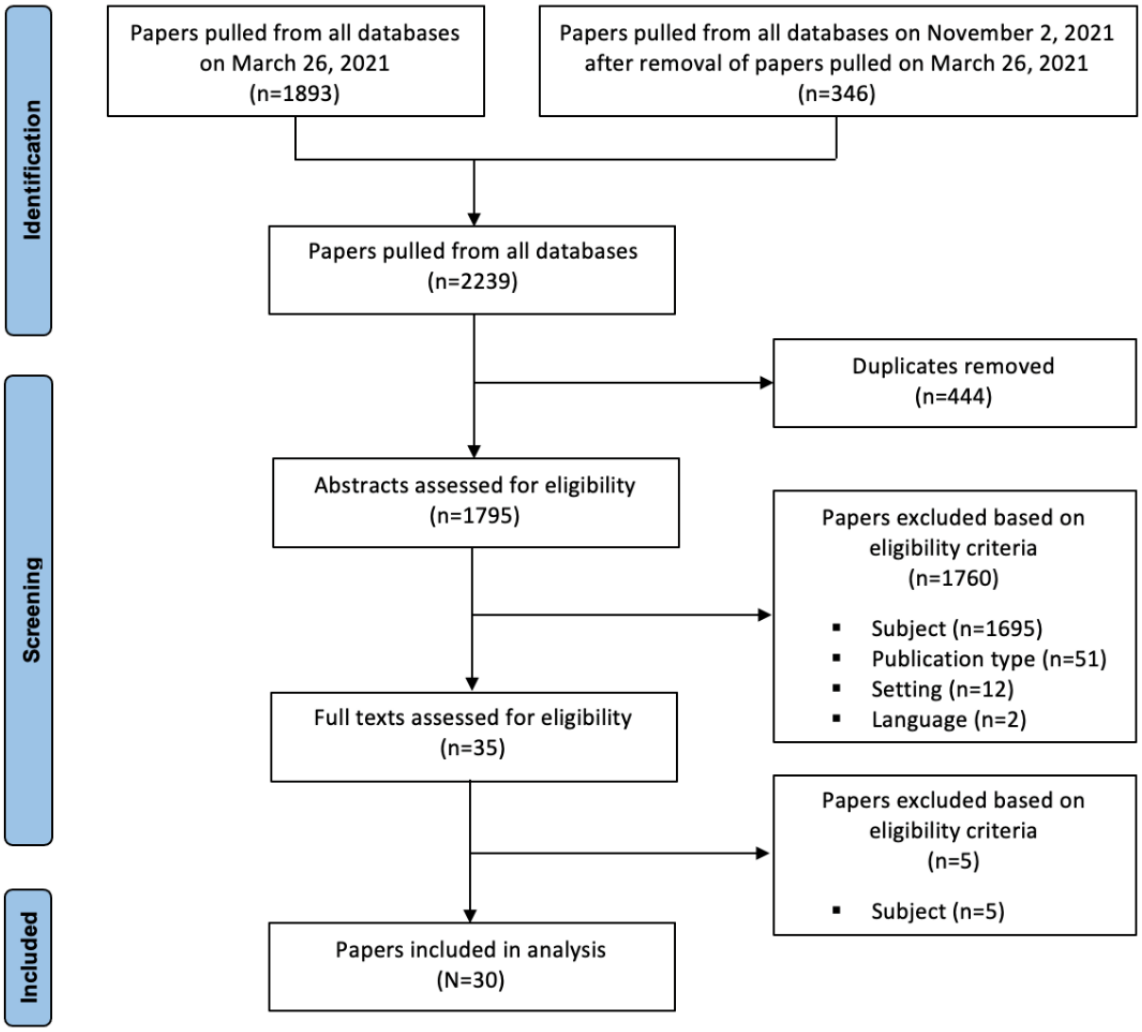
AND

psychiatr* OR dementia OR schizophreni* OR autis* OR depress* OR isolat* OR solitude OR alzheimer* OR mental* OR psycholog* OR anxiet* OR Neurodegenerative Diseases/px [Psychology] OR exp Mental Health/OR exp Mental Disorders/OR exp Mental Health Services/OR Mental Healing/px [Psychology] OR exp Psychiatry/OR psychology/OR psychology, positive/OR psychology, adolescent/OR psychology, child/OR cognitive science/OR psychology, developmental/OR psychology, clinical/OR psychology, comparative/OR psychology, educational/OR psychology, experimental/OR psychology, medical/OR psychology, social/OR exp Autism Spectrum Disorder/OR exp Anxiety/OR exp Schizophrenia/OR exp Psychotic Disorders/OR exp Neurocognitive Disorders/OR exp Dementia/OR Hospitals, Isolation/OR Depression/OR exp Anxiety Disorders/OR Cognitive Dysfunction/

Stage 3: Study Selection

Screening was carried out using the Rayyan reference management tool. After duplicates were removed, 2239 titles and abstracts were assessed for eligibility by 2 independent reviewers (IG and JP). To confirm understanding of the eligibility criteria, screening of the first 50 papers was pilot tested. If necessary, the criteria were redefined to ensure consistency between the reviewers. Subsequently, the full texts were evaluated to confirm inclusion. A senior reviewer (MPP) was consulted when consensus could not be achieved through discussions, and all exclusions were documented. Thirty papers were included in the scoping review. The screening process is detailed in the PRISMA flow diagram shown in Figure 1.

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



Stage 4: Charting the Data

The 30 papers selected for this review were tabulated in Microsoft Excel. A data extraction grid was developed, and 2 reviewers collected the data. The following data were extracted from the selected papers: (1) descriptive characteristics (eg, author, year, country, publication date, setting, study design, participants' characteristics), (2) social robot interventions and outcomes, (3) implementation strategies as defined by the Effective Practice and Organization of Care (EPOC) taxonomy [43], and (4) barriers and facilitators encountered during the implementation as defined by the Consolidated Framework for Implementation Research (CFIR) [44]. The methodological quality of the included studies was assessed using the 2018 version of the Mixed Method Appraisal Tool [45]. Each publication was assessed independently by 2 raters (IG and PPdOP). Differences in appraisal were discussed until consensus was reached.

Stage 5: Collating, Summarizing, and Reporting the Results

The characteristics of the included studies (eg, author, year, publication date, study design/method, participants' characteristics) were described. Interventions and their outcomes were summarized and tabulated. Tables were also used to present implementation strategies as well as barriers and facilitators.

Results

Characteristics of the Included Studies

Thirty papers were included in this scoping review [8-10,16,18-22,46-66]. Studies used 15 different social robots. Eighteen studies used animal-shaped robots, among which the PARO seal robot was used most often (n=12), followed distantly by the AIBO dog robot (n=2). Two studies also used cat robots of different brands: JustoCat (n=1) and Joy For All (n=1). One study used both cat and dog Hasbro robots. Finally, 1 study used a robotic sheep. Seven studies used humanoid robots: NAO (n=2), CommU (n=1), Kabochan (n=1), MARIO KOMPAÏ (n=1), Pepper (n=1), and Telenoid (n=1). Three studies used other types of robots: Chapit (n=1), CuDDler (n=1), and PaPeRo (n=1). The included papers were published between 2015 and 2021 in a variety of peer-reviewed journals (5 were published in 2015, 3 in 2016, 5 in 2017, 3 in 2018, 3 in 2019, 9 in 2020, and 2 in 2021).

Ten publications were quantitative nonrandomized studies, 8 were designed as randomized controlled trials, 6 were qualitative studies, 5 were mixed method studies, and 1 was a quantitative

descriptive study. Studies were set in Australia (n=5), Japan (n=5), Netherlands (n=3), United States (n=3), Norway (n=1), Taiwan (n=2), Canada (n=1), China (n=1), France (n=1), New Zealand (n=1), Spain (n=1), Sweden (n=1), and Kazakhstan (n=1). Three publications reported on a multicenter study set in Ireland, Italy, and in the United Kingdom.

The sample size of the included studies ranged from between 1 and 415 participants. In 2 studies, the participants were children. In another study, participants were adults of various ages. The other 27 publications reported on studies conducted on a geriatric population. In accordance with the eligibility criteria, all studies involved participants with a mental health disorder. Twenty-four papers reported on participants with dementia. Other studies involved participants with a cognitive impairment (n=2), schizophrenia (n=2), depression (n=1), ASD (n=2), attention-deficit hyperactivity disorder (n=1), and intellectual disability (n=1). A catalog of the included papers [8-10,16,18-22,46-66] describing studies, samples, interventions, and main findings collected in our review is available in Multimedia Appendix 1. The EPOC implementation strategies discussed in each paper are compiled in Table 1. Note that the terms “education” and “educational” do not refer to the intervention but rather to the implementation. For example, educational meetings may refer to training sessions during which the functions of the robots are explained to the staff involved in the intervention.

Table 1. Effective Practice and Organization of Care implementation strategies discussed in the included papers.

Implementation	References for the included papers
Communities of practice	[10,18,56]
Continuous quality improvement	[47]
Educational games	[59,62,63]
Educational materials	[9,16,49,51,55,56,62-64,66]
Educational meetings	[8,9,16,46,49,51-54,57,60,64-66]
Educational outreach visits or academic detailing	[16]
Interprofessional education	[66]
Local consensus processes	[8,9,16,22,46,48,49,52,55,57,60,61,65,66]
Managerial supervision	[51,53-56]
Patient-mediated interventions	[10,16,18,21,48,54,57,59,61,63-65]
Routine patient-reported outcome measures	[10]
Tailored interventions	[18,21,22,48,49,53,56,57,59,62,65,66]

Mental Health Outcomes

In most cases, studies assessed the impact of social robots on engagement, social interaction, emotional state, agitation, behavior, and quality of life. The majority reported positive results on patients’ quality of life, including reduced loneliness and isolation [18,48,51,59] and improvements in mood and anxiety [9,18-20,48,53,56,60,61,66] and agitated behaviors [9,47,52]. Feelings of comfort or reduced stress following social robot interventions were also described [51,52,66], although 1 study including participants with cognitive decline showed changes in the electroencephalogram, which were indicative of increased stress [50]. In some studies that focused on participants with dementia, SARs appeared to increase social engagement between patients, caregivers, and family members [8,16,18,20,21,47,49,52,53,58]. Further, SARs were emphasized as an alternative to alleviate the burden of caregivers, since they could free up time allowing carers to partake in other professional or daily tasks [16,20,47,63]. Furthermore, the use of social robots could enhance communication skills and improvements in joint attention among children with ASD, as described in the study by Kumazaki et al [22].

Barriers and Facilitators to the Implementation of Social Robots in Mental Health Facilities

Some barriers and facilitators were identified in the 30 included publications, using the CFIR as a guide to present the results in an adapted form. Three of the 5 domains in the CFIR were identified in this study: intervention characteristics, which refers to the key attributes of the intervention, called by the authors as “technical category;” inner setting, which refers to the features of the implementing organization, called by the authors as “organizational category;” and the characteristics of the individuals involved in the implementation, called by the authors as “clinical category.” A summary of these is presented in Table 2.

Table 2. Barriers and facilitators to the implementation discussed in the included papers.

Factors	References of the papers
Barriers	
Organizational	
Noisy environment during interaction	[16,18,48]
Storage area necessary	[46]
Charging necessary	[46]
Hygiene measures necessary	[46]
Staff/caregivers resistant to implementation	[18,66]
Increased workload for staff/caregivers	[8,66]
Frequency of sessions not adapted to patients' needs	[61]
Clinical	
Participants with an advanced cognitive decline	[18-20,54,66]
Participants with a hearing impairment	[20,63]
Difficult disengagement after the robot's removal	[55]
Risk of deception	[49,51,66]
Participants with a language impairment	[50]
Interaction with the robot seemed infantilizing	[8,51]
Participants feared the robot	[22]
Participants misunderstood the purposes of the study	[8]
Frustrating interruption of activities	[57]
Technical	
Robot was difficult to understand	[16,21,63]
Robot's touchscreen was difficult to use	[16,48]
Robot's voice recognition system was deficient	[18,48]
Limited visibility of the robot's screen display	[21]
Robot's speech rhythm deficient (too fast, long pauses, etc)	[21,62,63]
Robot was too noisy	[8,56,61]
Connection between devices was unstable	[8]
Robot was fragile	[8]
Robot was heavy	[8,61]
Robot was too big	[8]
Robot interrupted conversations	[63]
Robot spoke a limited number of languages	[19,48]
Facilitators	
Organizational	
Staff/caregivers had a positive perception of the robot	[18,46,66]
Staff/caregivers received training	[8,9,16,22,46,49,51-55,57,60,64-66]
Staff/caregivers promoted the use of the robot	[46,66]
Robot was easily available	[10,47]
Low cost	[10,19,47]
Robot was named by participants	[55]
Demonstration at the beginning of the intervention	[21,53,56]
Intervention did not replace usual activities	[49]

Factors	References of the papers
Hygiene measures were easily applicable	[51]
Participants were given ownership of their robot	[10]
Cleaning protocol was developed	[66]
Sessions were carried out in a quiet separate room	[8,53,54,56]
Exclusion of patients uninterested by the robot	[53,54]
Activities with the robot were organized (eg, bingo, listening to music)	[53]
Verbal/written instructions for staff/caregivers	[53,56]
Length of sessions were flexible	[16,56]
Facilitator was present during sessions	[8,49,56,62,63]
Technical	
Robot's appearance was pleasing	[10,16,18,21,22,47,51,61]
Addition of stylus pen to facilitate the use of the robot's touchscreen	[16,48]
Robot was easy to use, little training required	[21,47]
Robot was responsive to patients' touch	[9,47,66]
Robot's speech modalities were adequate	[61,66]
Robot was voice- and face-activated	[21]
Robot's sound was clear	[10,21,22,62]
Robot's voice/face recognition feature was adequate	[21]
Contextual interaction (intervention within augmented reality display)	[49]
Robot had entertaining features (apps, images, music)	[16,18,21,48,63]

Barriers to implementation were primarily related to the characteristics of social robots, such as their physical attributes (eg, weight, size, sound, overall appearance) [8,56,61]. Technical issues (eg, connection instability, fragility and susceptibility to damages, deficient speech recognition, complexity of operating the touchscreen and preprogrammed functions, limited visibility of the robot's screen display) were mentioned as barriers [8,16,18,21,48]. Furthermore, organizational, and institutional barriers in mental health facilities such as space allocation (eg, lack of an adequate space for interactions with SARs, lack of storage area), background noises during participants' interaction, and uncertainty on how to delineate hygiene concerns were reported [16,18,46,48]. Negative attitudes toward social robots by staff and caregivers (eg, fear of job replacements by robots) were emphasized in 2 studies in our review [18,66]. However, some stakeholders developed positive perceptions toward social robots after witnessing their positive impacts, as reported by Bemelmans et al [67].

Most of the identified facilitators correspond with the identified barriers. For instance, the characteristics of the social robots, such as the robot's appearance, ease of use, and technical functions (eg, the robot's adequate speech modalities, the robot's responsiveness to patients touch, the robot's clear sound, the robot's appropriate voice and face recognition) were seen as enablers [9,10,16,18,21,22,47,51,61,66]. Less noisy robots were less likely to distress the interlocutor, notably in children with ASD [22,62]. Further, the ability to adapt the robot's functions to participants' preferences and customize the modes of robot interaction through apps were identified as implementation facilitators [16,18,21,48,49,59,63]. An introduction phase with training and familiarization also facilitated greater acceptance to social robots [16,21,22,53,56,65]. Organizational and institutional facilitators such as easily applicable hygiene measures, flexibility in the number and duration of sessions to match users' needs, and appropriate and quiet spaces for interactions were also identified as facilitators [8,16,51,53,54,56].

Discussion

Principal Findings

Overall, our review aimed to evaluate how social robots have been used to influence clinical outcomes in mental health care and the main barriers and facilitators encountered during their implementation. Our review includes 15 different social robots, and interventions ranged generally from positive to mixed results, although the statistical significance was not considered in some of the studies [8,10,20,22,51,59,62,64]. Most of the studies had very small sample sizes, a very brief duration, and had no follow-up measurements, which might make it difficult to conclude about the efficacy of the interventions [9,10,16,18,46-51,57,61,66]. In 2 of them, the intervention was not clearly described [47,63]. These methodological limitations were also highlighted in previous reviews of SAR use in mental health services [3,36].

Almost all of the studies included in this review focused on providing comfort, well-being, and companionship to the study

participants. Only a minority used SARs to implement a specific intervention to improve patients' self-management abilities or to address psychoeducation strategies. It must be considered that our scoping review excluded papers involving social robots in teaching-learning scenarios; thus, relevant studies might be missed. Further, most papers in our review (24 of 30) reported on interventions with participants with dementia. With this population, the main priorities in using SARs were the reduction of neuropsychiatric symptoms as well as the feeling of isolation and loneliness [9,51]. Therefore, the possibility to address companionship and improvement in daily support might be seen as a more relevant therapeutic benefit than self-managing treatment, as previously described in the literature [36,68].

One study in our scoping review raised the issue of the possible use of social robots to reduce loneliness during the COVID-19 pandemic [18]. Three different roles of SARs (ie, social utility, social identity, and social connectivity) allow social robots to create a supportive relationship capable of mitigating feelings of loneliness during quarantine and lockdown contexts [69]. Their role in promoting well-being was also highlighted as a promising avenue for those who are more vulnerable during the pandemic, particularly older adults and children [70]. Moreover, it must be considered that the demographics and the clinical characteristics of the participants influenced their needs. As most of the selected papers included older people with dementia, some particularities of this population must be raised. Some studies reported on participants with an advanced cognitive decline or with language and hearing impairments, which made it difficult to interact with the robot [18-20,50,54,63,66]. In addition, the complexity of operating the touchscreen and preprogrammed functions were also highlighted in this population [16,48]. Thus, tailoring an intervention to patients' needs by using a personalized approach is identified as an important enabler that was also previously highlighted [36].

Assessing staff, family, stakeholders, and caregivers' perspectives about SAR use in mental health services is another relevant aspect that should be considered. Consistent with our review, negative reactions were primarily described in some studies [71,72], but other studies also recorded how some stakeholders developed positive perceptions toward social robots after witnessing their positive impacts on patients [67,73-76]. Positive attitudes of care professionals toward SARs were reported as key facilitators to acceptability among users [73]. All these findings are consistent with those reported in a recent scoping review by Koh and colleagues [77]. SARs might potentially integrate traditional mental health care apps in an interactive social companion, providing a more engaging and dynamic platform for users [3]. In our review, some studies reported that the presence of different applications adapted and personalized to participants facilitated and sustained their engagement with the robot as well as their interactive behaviors [16,18,48,61]. Combining these capabilities with active user interaction allows SARs to deliver different interventions (eg, psychoeducation, techniques of cognitive-behavioral therapy), which can help users take greater ownership of their own health and well-being [3,78,79].

Although SARs have emerged as a promising approach across the field of mental health, they should be treated as an additional and complementary resource in mental health services and thus, poor substitutes for human contact [8,18,52,80]. Ethical concerns such as reduced human contact, emotional deception, and issues surrounding data security, confidentiality, and information privacy must be considered during the implementation of SARs in mental health services [80]. Most of the robots cannot assess a patient's emotional state with great accuracy, and the absence of a human professional can have a negative impact on a patient's adherence to a program [18,63,80]. Ideally, social robots should remain under the supervision of trained mental health professionals and should be used as a means of providing comfort, quality of life, and purposeful engagements [52,80].

Strengths and Limitations

There are some strengths used in sustaining this work. First, the methodological framework was transparent and rigorous, and we searched multiple databases. Second, we consulted experts in the field of social robots as well as mental health researchers and professionals to emphasize the main points in each area. Finally, the Mixed Methods Appraisal Tool was employed to evaluate the quality of studies included in this review and a scientific and valuable implementation tool, CFIR, was used to guide the presentation of results. Nevertheless, this review has several limitations. Papers that were not published in English were excluded in this review and, as a result, relevant studies might be missed. In addition, the review aggregated only studies set in mental health facilities, and studies set in patients' homes were excluded. This fact seemed to form a bias regarding the severity of mental health disorders that were included in this review. Although we did not limit our search to specifically a mental health diagnosis and did not define a specific age range, the bulk of our sample consisted of interventions with older adults with dementia. Therefore, the generalizability of our findings is limited by study characteristics. Moreover, most of the studies had small sample sizes, with brief and sometimes unclear interventions and poor and heterogeneous methodology, which might make it difficult to preclude conclusions about the efficacy of the interventions.

Future Research and Practical Implications

Overall, our review has shown that the potential of social robots in mental health care is broad. However, there are still many gaps in this field. Since previous works on SAR interventions have mainly focused on older adults (ie, for the treatment of dementia) and children (ie, for the treatment of ASD), expanding the diagnosis would be a relevant option for the next steps in the research. As an illustration, attention is warranted for major depressive disorder, which has the highest lifetime prevalence among psychiatric disorders and is associated with high costs for the society [81]. In our scoping review, only 1 study had addressed the use of SARs for patients with major depressive disorder, and it found a statistically significant reduction in depression and loneliness and improvement in the quality of life [55]. However, both the small sample size and the relatively short duration of the intervention limit the generalizability of their results. Further research studies with larger samples, assessing long-term follow-up and with clear intervention protocols are needed in this field.

Furthermore, the use of SARs in different settings should be raised, notably for individuals with mental health needs living in remote areas and for those who feel stigmatized in traditional mental health care settings. In the context of rural communities and other resource-scarce areas, SARs would allow patients to receive health care remotely, thereby enabling such patients to avoid potential barriers to care such as travel or scheduling and thus improving patient outcomes. Further, the recent COVID-19 pandemic has highlighted telehealth's potential in almost all health care settings. The possible use of social robots to reduce social isolation during the pandemic is a significant issue that could be explored [69,70]. Other psychotherapeutic strategies (eg, self-care tactics) combined with SARs should be raised in future research, which could be helpful in facilitating engagement with self-help treatment programs and users' autonomy. Rather than developing a novel program treatment, SARs could be integrated into existing psychosocial approaches to improve the effectiveness of the intervention, such as an adjuvant in cognitive behavioral therapy. In this context, SARs could improve in real-time the monitoring and feedback for users through dynamic applications [7].

Tailored interventions aiming to fulfill the specific needs of a well-defined population should be explored, and further qualitative research (entirely user-centered) should investigate what people expect from the social robots' roles played in mental health care. Single-case experimental designs might be a very useful starting point to ensure that different needs are met (ie, clinical, users', engineers', and roboticists' goals). Improvements in methodology and study design, beyond pilot studies, and the use of psychometrically validated measures should also be taken into consideration.

Research that evaluates the implementation of SARs in mental health programs and that identifies their barriers and facilitators is also relevant when it comes to guiding the successful implementation of social robots in a real-world setting, particularly in an organizational context (eg, policy and government regulations for project planning and evaluation, the expense of robotics and the cost-versus-benefit relationship within services). The cost of mental disorders is already placing a high financial burden on individuals with mental health problems, their families, and the society in general, and creating cost-efficient robots seems to be a good opportunity in greatly reducing the cost in mental health care [82-84]. Further research in these areas, using an implementation framework, is needed. In all these aspects, it is essential for mental health professionals to work closely with patients and with robotics experts (ie, computer scientists, programmers, and engineers) to provide critical feedback on what tasks robots can reasonably do and which ones should be considered in the design of future interventions.

As the demand for mental health services increases, it is becoming imperative to find solutions to meet the growing needs. The use of social robots is a viable solution. Despite some technical flaws, advances in robotics now make it possible to offer a quality service for users. Our scoping review has highlighted the therapeutic effects of social robots in a variety of contexts. However, the methodological weakness of the studies often limits the generalizability of their results. Further studies should go beyond the framework of the pilot study in order to target the use of social robots for a well-defined case and to further potentiate the attributes of these technologies.

Acknowledgments

This research was funded by the Canadian Institutes of Health Research (CIHR funding VR4-172769). MPP has a Senior Career Award financed by the Que´bec Health Research Fund (FRQS), the Centre de Recherche du Centre Hospitalier de l'Université de Montre´al, and the Ministe` re de la Sante´ et des Services Sociaux du Que´bec.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Catalog of the included studies and extracted data, sorted alphabetically by study author.
[DOCX File , 72 KB - mental_v9i4e36094_app1.docx]

References

1. Boubaker O. Chapter 7: Medical robotics. In: Control Theory in Biomedical Engineering. Cambridge, Massachusetts: Academic Press; Jan 01, 2020:153-204.
2. Feil-Seifer D, Mataric MJ. Socially assistive robotics. 2005 Presented at: 9th International Conference on Rehabilitation Robotics; June 28-July 1; Chicago, IL, USA. [doi: 10.1109/icorr.2005.1501143]
3. Scoglio AA, Reilly ED, Gorman JA, Drebing CE. Use of social robots in mental health and well-being research: systematic review. J Med Internet Res 2019 Jul 24;21(7):e13322. [doi: 10.2196/13322] [Medline: 31342908]
4. Flandorfer P. Population ageing and socially assistive robots for elderly persons: the importance of sociodemographic factors for user acceptance. International Journal of Population Research 2012 Apr 22;2012:1-13. [doi: 10.1155/2012/829835]
5. Broekens J, Heerink M, Rosendal H. Assistive social robots in elderly care: A review. Gerontechnology 2009 Apr 01;8:94-103. [doi: 10.4017/gt.2009.08.02.002.00]

6. Pu L, Moyle W, Jones C, Todorovic M. The effectiveness of social robots for older adults: a systematic review and meta-analysis of randomized Controlled Studies. Gerontologist 2019 Jan 09;59(1):e37-e51. [doi: 10.1093/geront/gny046] [Medline: 29897445]
7. Rabbitt SM, Kazdin AE, Scassellati B. Integrating socially assistive robotics into mental healthcare interventions: applications and recommendations for expanded use. Clin Psychol Rev 2015 Feb;35:35-46. [doi: 10.1016/j.cpr.2014.07.001] [Medline: 25462112]
8. Moyle W, Jones C, Sung B, Bramble M, O'Dwyer S, Blumenstein M, et al. What effect does an animal robot called CuDDler have on the engagement and emotional response of older people with dementia? a pilot feasibility study. Int J of Soc Robotics 2015 Oct 24;8(1):145-156. [doi: 10.1007/s12369-015-0326-7]
9. Demange M, Lenoir H, Pino M, Cantegreil-Kallen I, Rigaud AS, Cristancho-Lacroix V. Improving well-being in patients with major neurodegenerative disorders: differential efficacy of brief social robot-based intervention for 3 neuropsychiatric profiles. CIA 2018 Jul;Volume 13:1303-1311. [doi: 10.2147/cia.s152561]
10. Hammarlund RA, Whatley KL, Zielinski MH, Jubert JC. Benefits of affordable robotic pet ownership in older adults with dementia. J Gerontol Nurs 2021 Mar 01;47(3):18-22. [doi: 10.3928/00989134-20210209-03] [Medline: 33626160]
11. Krueger F, Mitchell KC, Deshpande G, Katz JS. Human-dog relationships as a working framework for exploring human-robot attachment: a multidisciplinary review. Anim Cogn 2021 Mar;24(2):371-385 [FREE Full text] [doi: 10.1007/s10071-021-01472-w] [Medline: 33486634]
12. Peluso S, De Rosa A, De Lucia N, Antenora A, Illario M, Esposito M, et al. Animal-assisted therapy in elderly patients: evidence and controversies in dementia and psychiatric disorders and future perspectives in other neurological diseases. J Geriatr Psychiatry Neurol 2018 May;31(3):149-157. [doi: 10.1177/0891988718774634] [Medline: 29764282]
13. Hung L, Liu C, Woldum E, Au-Yeung A, Berndt A, Wallsworth C, et al. The benefits of and barriers to using a social robot PARO in care settings: a scoping review. BMC Geriatr 2019 Aug 23;19(1):1-10 [FREE Full text] [doi: 10.1186/s12877-019-1244-6] [Medline: 31443636]
14. Fujita M. On activating human communications with pet-type robot AIBO. Proceedings of the IEEE 2004 Nov;92(11):1804-1813. [doi: 10.1109/jproc.2004.835364]
15. Miklósi A, Gácsi M. On the utilization of social animals as a model for social robotics. Front Psychol 2012;3:75 [FREE Full text] [doi: 10.3389/fpsyg.2012.00075] [Medline: 22457658]
16. Barrett E, Burke M, Whelan S, Santorelli A, Oliveira BL, Cavallo F, et al. Evaluation of a companion robot for individuals with dementia: quantitative findings of the MARIO project in an Irish residential care setting. J Gerontol Nurs 2019 Jul 01;45(7):36-45. [doi: 10.3928/00989134-20190531-01] [Medline: 31237660]
17. Broadbent E. Interactions with robots: the truths we reveal about ourselves. Annu Rev Psychol 2017 Jan 03;68:627-652. [doi: 10.1146/annurev-psych-010416-043958] [Medline: 27648986]
18. Casey D, Barrett E, Kovacic T, Sancarlo D, Ricciardi F, Murphy K, et al. The perceptions of people with dementia and key stakeholders regarding the use and impact of the social robot MARIO. IJERPH 2020 Nov 20;17(22):8621. [doi: 10.3390/ijerph17228621]
19. Chen K, Lou VW, Tan KC, Wai M, Chan L. Effects of a humanoid companion robot on dementia symptoms and caregiver distress for residents in long-term care. J Am Med Dir Assoc 2020 Nov;21(11):1724-1728.e3. [doi: 10.1016/j.jamda.2020.05.036] [Medline: 32713772]
20. Chen L, Sumioka H, Ke L, Shiomi M, Chen L. Effects of teleoperated humanoid robot application in older adults with neurocognitive disorders in Taiwan: a report of three cases. Aging Med Healthc 2020 Jun 29;11(2):67-71. [doi: 10.33879/amh.2020.066-2001.003]
21. Chu M, Khosla R, Khaksar SMS, Nguyen K. Service innovation through social robot engagement to improve dementia care quality. Assist Technol 2017;29(1):8-18. [doi: 10.1080/10400435.2016.1171807] [Medline: 27064692]
22. Kumazaki H, Yoshikawa Y, Yoshimura Y, Ikeda T, Hasegawa C, Saito DN, et al. The impact of robotic intervention on joint attention in children with autism spectrum disorders. Mol Autism 2018;9:46 [FREE Full text] [doi: 10.1186/s13229-018-0230-8] [Medline: 30202508]
23. Beran TN, Ramirez-Serrano A, Vanderkooi OG, Kuhn S. Reducing children's pain and distress towards flu vaccinations: a novel and effective application of humanoid robotics. Vaccine 2013 Jun 07;31(25):2772-2777. [doi: 10.1016/j.vaccine.2013.03.056] [Medline: 23623861]
24. Shibata T, Mitsui T, Wada K, Touda A, Kumasaka T, Tagami K, et al. Mental commit robot and its application to therapy of children. 2001 Presented at: IEEE/ASME International Conference on Advanced Intelligent Mechatronics; July 8; Como, Italy. [doi: 10.1109/AIM.2001.936838]
25. Yun S, Choi J, Park S, Bong G, Yoo H. Social skills training for children with autism spectrum disorder using a robotic behavioral intervention system. Autism Res 2017 Jul;10(7):1306-1323. [doi: 10.1002/aur.1778] [Medline: 28464438]
26. Sartorato F, Przybylowski L, Sarko DK. Improving therapeutic outcomes in autism spectrum disorders: Enhancing social communication and sensory processing through the use of interactive robots. J Psychiatr Res 2017 Jul;90:1-11. [doi: 10.1016/j.jpsychires.2017.02.004] [Medline: 28213292]

27. Huskens B, Verschuur R, Gillesen J, Didden R, Barakova E. Promoting question-asking in school-aged children with autism spectrum disorders: effectiveness of a robot intervention compared to a human-trainer intervention. Dev Neurorehabil 2013 Oct;16(5):345-356. [doi: 10.3109/17518423.2012.739212] [Medline: 23586852]
28. Olde Keizer RACM, van Velsen L, Moncharmont M, Riche B, Ammour N, Del Signore S, et al. Using socially assistive robots for monitoring and preventing frailty among older adults: a study on usability and user experience challenges. Health Technol 2019 Apr 9;9(4):595-605. [doi: 10.1007/s12553-019-00320-9]
29. Boumans R, van Meulen F, Hindriks K, Neerincx M, Olde Rikkert MGM. Robot for health data acquisition among older adults: a pilot randomised controlled cross-over trial. BMJ Qual Saf 2019 Oct;28(10):793-799 [FREE Full text] [doi: 10.1136/bmjqs-2018-008977] [Medline: 30894423]
30. Broadbent E, Garrett J, Jepsen N, Li Ogilvie V, Ahn HS, Robinson H, et al. Using robots at home to support patients with chronic obstructive pulmonary disease: pilot randomized controlled trial. J Med Internet Res 2018 Feb 13;20(2):e45 [FREE Full text] [doi: 10.2196/jmir.8640] [Medline: 29439942]
31. Mordoch E, Osterreicher A, Guse L, Roger K, Thompson G. Use of social commitment robots in the care of elderly people with dementia: a literature review. Maturitas 2013 Jan;74(1):14-20. [doi: 10.1016/j.maturitas.2012.10.015] [Medline: 23177981]
32. Bemelmans R, Gelderblom GJ, Jonker P, de Witte L. Socially assistive robots in elderly care: a systematic review into effects and effectiveness. J Am Med Dir Assoc 2012 Feb;13(2):114-120.e1. [doi: 10.1016/j.jamda.2010.10.002] [Medline: 21450215]
33. Abdi J, Al-Hindawi A, Ng T, Vizcaychipi MP. Scoping review on the use of socially assistive robot technology in elderly care. BMJ Open 2018 Feb 12;8(2):e018815 [FREE Full text] [doi: 10.1136/bmjopen-2017-018815] [Medline: 29440212]
34. Dawe J, Sutherland C, Barco A, Broadbent E. Can social robots help children in healthcare contexts? A scoping review. BMJ Paediatr Open 2019;3(1):e000371 [FREE Full text] [doi: 10.1136/bmjpo-2018-000371] [Medline: 30815587]
35. Papadopoulos I, Koulouglioti C, Lazzarino R, Ali S. Enablers and barriers to the implementation of socially assistive humanoid robots in health and social care: a systematic review. BMJ Open 2020 Jan 09;10(1):e033096 [FREE Full text] [doi: 10.1136/bmjopen-2019-033096] [Medline: 31924639]
36. Abbott R, Orr N, McGill P, Whear R, Bethel A, Garside R, et al. How do "robopets" impact the health and well-being of residents in care homes? A systematic review of qualitative and quantitative evidence. Int J Older People Nurs 2019 Sep;14(3):e12239 [FREE Full text] [doi: 10.1111/opn.12239] [Medline: 31070870]
37. Leng M, Liu P, Zhang P, Hu M, Zhou H, Li G, et al. Pet robot intervention for people with dementia: A systematic review and meta-analysis of randomized controlled trials. Psychiatry Res 2019 Jan;271:516-525. [doi: 10.1016/j.psychres.2018.12.032] [Medline: 30553098]
38. Ong YC, Tang A, Tam W. Effectiveness of robot therapy in the management of behavioural and psychological symptoms for individuals with dementia: A systematic review and meta-analysis. J Psychiatr Res 2021 Aug;140:381-394. [doi: 10.1016/j.jpsychires.2021.05.077] [Medline: 34144442]
39. Szymona B, Maciejewski M, Karpiński R, Jonak K, Radzikowska-Büchner E, Niderla K, et al. Robot-Assisted Autism Therapy (RAAT). Criteria and types of experiments using anthropomorphic and zoomorphic robots. review of the research. Sensors (Basel) 2021 May 27;21(11):3720 [FREE Full text] [doi: 10.3390/s21113720] [Medline: 34071829]
40. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
41. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
42. The ICD-10 Classification of Mental And Behavioral Disorders: Diagnostic Criteria for Research. Geneva, Switzerland: World Health Organization; 1993.
43. Effective practice and organisation of care (EPOC) taxonomy. Cochrane Effective Practice and Organisation of Care. 2015. URL: https://epoc.cochrane.org/epoc-taxonomy [accessed 2021-01-18]
44. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
45. Hong Q, Pluye P, Fàbregues S, Bartlett G, Boardman F, Cargo M, et al. Mixed Methods Appraisal Tool (MMAT). 2018. URL: http://mixedmethodsappraisaltoolpublic.pbworks.com/w/file/fetch/146002140/MMAT_2018_criteria-manual_2018-08-08c.pdf [accessed 2021-01-18]
46. Bemelmans R, Gelderblom GJ, Jonker P, de Witte L. Effectiveness of robot Paro in intramural psychogeriatric care: a multicenter quasi-experimental study. J Am Med Dir Assoc 2015 Nov 01;16(11):946-950. [doi: 10.1016/j.jamda.2015.05.007] [Medline: 26115817]
47. Brecher DB. Use of a robotic cat to treat terminal restlessness: a case study. J Palliat Med 2020 Mar;23(3):432-434. [doi: 10.1089/jpm.2019.0157] [Medline: 31328997]

48. D'Onofrio G, Sancarlo D, Raciti M, Burke M, Teare A, Kovacic T, et al. MARIO project: validation and evidence of service robots for older people with dementia. J Alzheimers Dis 2019;68(4):1587-1601. [doi: 10.3233/JAD-181165] [Medline: 30958360]
49. Feng Y, Barakova EI, Yu S, Hu J, Rauterberg GWM. Effects of the level of interactivity of a social robot and the response of the augmented reality display in contextual interactions of people with dementia. Sensors 2020 Jul 05;20(13):3771. [doi: 10.3390/s20133771]
50. Goda A, Shimura T, Murata S, Kodama T, Nakano H, Ohsugi H. Psychological and neurophysiological effects of robot assisted activity in elderly people with cognitive decline. Gerontol Geriatr Med 2020;6:2333721420969601 [FREE Full text] [doi: 10.1177/2333721420969601] [Medline: 33241078]
51. Gustafsson C, Svanberg C, Müllersdorf M. Using a robotic cat in dementia care: a pilot study. J Gerontol Nurs 2015 Oct;41(10):46-56. [doi: 10.3928/00989134-20150806-44] [Medline: 26488255]
52. Jones C, Moyle W, Murfield J, Draper B, Shum D, Beattie E, et al. Does cognitive impairment and agitation in dementia influence intervention effectiveness? findings from a cluster-randomized-controlled trial with the therapeutic robot, PARO. J Am Med Dir Assoc 2018 Jul;19(7):623-626. [doi: 10.1016/j.jamda.2018.02.014] [Medline: 29656838]
53. Jøranson N, Pedersen I, Rokstad AMM, Ihlebæk C. Effects on symptoms of agitation and depression in persons with dementia participating in robot-assisted activity: a cluster-randomized controlled trial. J Am Med Dir Assoc 2015 Oct 01;16(10):867-873. [doi: 10.1016/j.jamda.2015.05.002] [Medline: 26096582]
54. Jøranson N, Pedersen I, Rokstad AMM, Aamodt G, Olsen C, Ihlebæk C. Group activity with Paro in nursing homes: systematic investigation of behaviors in participants. Int. Psychogeriatr 2016 Mar 28;28(8):1345-1354. [doi: 10.1017/s1041610216000120]
55. Chen S, Moyle W, Jones C, Petsky H. A social robot intervention on depression, loneliness, and quality of life for Taiwanese older adults in long-term care. Int. Psychogeriatr 2020 Apr 14;32(8):981-991. [doi: 10.1017/s1041610220000459]
56. Liang A, Piroth I, Robinson H, MacDonald B, Fisher M, Nater UM, et al. A pilot randomized rrial of a companion robot for people with dementia living in the community. J Am Med Dir Assoc 2017 Oct 01;18(10):871-878. [doi: 10.1016/j.jamda.2017.05.019] [Medline: 28668664]
57. Moyle W, Jones CJ, Murfield JE, Thalib L, Beattie ERA, Shum DKH, et al. Use of a robotic seal as a therapeutic tool to improve dementia symptoms: a cluster-randomized controlled trial. J Am Med Dir Assoc 2017 Sep 01;18(9):766-773 [FREE Full text] [doi: 10.1016/j.jamda.2017.03.018] [Medline: 28780395]
58. Naganuma M, Ohkubo E, Kato N. Use of robotic pets in providing stimulation for nursing home residents with dementia. Stud Health Technol Inform 2015;217:1009-1012. [Medline: 26294602]
59. Narita S, Ohtani N, Waga C, Ohta M, Ishigooka J, Iwahashi K. A pet-type robot AIBO-assisted therapy as a day care program for chronic schizophrenia patients. AMJ 2016;09:7. [doi: 10.21767/amj.2016.2666]
60. Petersen S, Houston S, Qin H, Tague C, Studley J. The utilization of robotic pets in dementia care. J Alzheimers Dis 2017;55(2):569-574 [FREE Full text] [doi: 10.3233/JAD-160703] [Medline: 27716673]
61. Pu L, Moyle W, Jones C. How people with dementia perceive a therapeutic robot called PARO in relation to their pain and mood: A qualitative study. J Clin Nurs 2020 Feb;29(3-4):437-446. [doi: 10.1111/jocn.15104] [Medline: 31738463]
62. Rakhymbayeva N, Amirova A, Sandygulova A. A long-term engagement with a social robot for autism therapy. Front Robot AI 2021;8:669972 [FREE Full text] [doi: 10.3389/frobt.2021.669972] [Medline: 34222353]
63. Sato M, Yasuhara Y, Osaka K, Ito H, Dino MJS, Ong IL, et al. Rehabilitation care with Pepper humanoid robot: A qualitative case study of older patients with schizophrenia and/or dementia in Japan. Enfermería Clínica 2020 Feb;30:32-36. [doi: 10.1016/j.enfcli.2019.09.021]
64. Valentí Soler M, Agüera-Ortiz L, Olazarán Rodríguez J, Mendoza Rebolledo C, Pérez Muñoz A, Rodríguez Pérez I, et al. Social robots in advanced dementia. Front Aging Neurosci 2015;7:133. [doi: 10.3389/fnagi.2015.00133] [Medline: 26388764]
65. Wagemaker E, Dekkers TJ, Agelink van Rentergem JA, Volkers KM, Huizenga HM. Advances in mental health care: Five N = 1 studies on the effects of the robot seal paro in adults with severe intellectual disabilities. Journal of Mental Health Research in Intellectual Disabilities 2017 May 16;10(4):309-320. [doi: 10.1080/19315864.2017.1320601]
66. Hung L, Gregorio M, Mann J, Wallsworth C, Horne N, Berndt A, et al. Exploring the perceptions of people with dementia about the social robot PARO in a hospital setting. Dementia (London) 2021 Feb;20(2):485-504 [FREE Full text] [doi: 10.1177/1471301219894141] [Medline: 31822130]
67. Bemelmans R, Gelderblom GJ, Jonker P, de Witte L. How to use robot interventions in intramural psychogeriatric care; A feasibility study. Appl Nurs Res 2016 May;30:154-157. [doi: 10.1016/j.apnr.2015.07.003] [Medline: 27091271]
68. Góngora Alonso S, Hamrioui S, de la Torre Díez I, Motta Cruz E, López-Coronado M, Franco M. Social robots for people with aging and dementia: a systematic review of literature. Telemed J E Health 2019 Jul;25(7):533-540. [doi: 10.1089/tmj.2018.0051] [Medline: 30136901]
69. Odekerken-Schröder G, Mele C, Russo-Spena T, Mahr D, Ruggiero A. Mitigating loneliness with companion robots in the COVID-19 pandemic and beyond: an integrative framework and research agenda. JOSM 2020 Aug 21;31(6):1149-1162. [doi: 10.1108/josm-05-2020-0148]
70. Henkel AP, Čaić M, Blaurock M, Okan M. Robotic transformative service research: deploying social robots for consumer well-being during COVID-19 and beyond. JOSM 2020 Aug 15;31(6):1131-1148. [doi: 10.1108/josm-05-2020-0145]

71. Birks M, Bodak M, Barlas J, Harwood J, Pether M. Robotic seals as therapeutic tools in an aged care facility: a qualitative study. J Aging Res 2016;2016:8569602 [FREE Full text] [doi: 10.1155/2016/8569602] [Medline: 27990301]
72. Niemelä M, Määttä H, Ylikauppila M. Expectations and experiences of adopting robots in elderly care in Finland. 2016 Presented at: 4th International Conference on Serviceology; September 6-8; Tokyo, Japan URL: https://cris.vtt.fi/ws/portalfiles/portal/19514418/OA_Expectations_and_experiences_of_adopting_robots_in.pdf
73. Kolstad M, Yamaguchi N, Babic A, Nishihara Y. Integrating socially assistive robots into Japanese nursing care. Stud Health Technol Inform 2020 Jun 26;272:183-186. [doi: 10.3233/SHTI200524] [Medline: 32604631]
74. Melkas H, Hennala L, Pekkarinen S, Kyrki V. Impacts of robot implementation on care personnel and clients in elderly-care institutions. Int J Med Inform 2020 Feb;134:104041 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.104041] [Medline: 31838291]
75. Pike J, Picking R, Cunningham S. Robot companion cats for people at home with dementia: A qualitative case study on companotics. Dementia (London) 2021 May;20(4):1300-1318. [doi: 10.1177/1471301220932780] [Medline: 32668978]
76. Portugal D, Alvito P, Christodoulou E, Samaras G, Dias J. A study on the deployment of a service robot in an elderly care center. Int J of Soc Robotics 2018 Nov 10;11(2):317-341. [doi: 10.1007/s12369-018-0492-5]
77. Koh WQ, Felding SA, Budak KB, Toomey E, Casey D. Barriers and facilitators to the implementation of social robots for older adults and people with dementia: a scoping review. BMC Geriatr 2021 Jun 09;21(1):351 [FREE Full text] [doi: 10.1186/s12877-021-02277-9] [Medline: 34107876]
78. Lima MR, Wairagkar M, Natarajan N, Vaitheswaran S, Vaidyanathan R. Robotic telemedicine for mental health: a multimodal approach to improve human-robot engagement. Front Robot AI 2021;8:618866 [FREE Full text] [doi: 10.3389/frobt.2021.618866] [Medline: 33816568]
79. Rathbone AL, Prescott J. The use of mobile apps and SMS messaging as physical and mental health interventions: systematic review. J Med Internet Res 2017 Aug 24;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]
80. Fiske A, Henningsen P, Buyx A. Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, Psychology, and Psychotherapy. J Med Internet Res 2019 May 09;21(5):e13216 [FREE Full text] [doi: 10.2196/13216] [Medline: 31094356]
81. Depression and other common mental disorders: global health estimates. World Health Organization. URL: https://apps.who.int/iris/handle/10665/254610 [accessed 2022-03-13]
82. Mental disorders - Level 2 cause. Global Health Metrics. 2020 Oct. URL: https://www.thelancet.com/pb-assets/Lancet/gbd/summaries/diseases/mental-disorders.pdf [accessed 2022-04-12]
83. Rehm J, Shield KD. Global burden of disease and the impact of mental and addictive disorders. Curr Psychiatry Rep 2019 Feb 07;21(2):10. [doi: 10.1007/s11920-019-0997-0] [Medline: 30729322]
84. Chisholm D, Sweeny K, Sheehan P, Rasmussen B, Smit F, Cuijpers P, et al. Scaling-up treatment of depression and anxiety: a global return on investment analysis. The Lancet Psychiatry 2016 May;3(5):415-424. [doi: 10.1016/S2215-0366(16)30024-4]

Abbreviations

ASD: autism spectrum disorder
CFIR: Consolidated Framework for Implementation Research
EPOC: Effective Practice and Organization of Care
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
SAR: socially assistive robot

Edited by J Torous, G Eysenbach; submitted 31.12.21; peer-reviewed by M Franco-Martin, B Teachman, M Franco; comments to author 26.02.22; revised version received 21.03.22; accepted 25.03.22; published 19.04.22.

Please cite as:
Guemghar I, Pires de Oliveira Padilha P, Abdel-Baki A, Jutras-Aswad D, Paquette J, Pomey MP
Social Robot Interventions in Mental Health Care and Their Outcomes, Barriers, and Facilitators: Scoping Review
JMIR Ment Health 2022;9(4):e36094
URL: https://mental.jmir.org/2022/4/e36094
doi:10.2196/36094
PMID:35438639

Review

Natural Language Processing Methods and Bipolar Disorder: Scoping Review

Daisy Harvey[1], MA; Fiona Lobban[1], PhD; Paul Rayson[2], PhD; Aaron Warner[1], MSc; Steven Jones[1], PhD

[1]Spectrum Centre for Mental Health Research, Division of Health Research, School of Health and Medicine, Lancaster University, Lancaster, United Kingdom

[2]Department of Computing and Communications, Lancaster University, Lancaster, United Kingdom

Corresponding Author:
Daisy Harvey, MA
Spectrum Centre for Mental Health Research, Division of Health Research
School of Health and Medicine
Lancaster University
Health Innovation One
Sir John Fisher Drive
Lancaster, LA1 4YG
United Kingdom
Phone: 44 152465201
Email: d.harvey4@lancaster.ac.uk

Abstract

Background: Health researchers are increasingly using natural language processing (NLP) to study various mental health conditions using both social media and electronic health records (EHRs). There is currently no published synthesis that relates specifically to the use of NLP methods for bipolar disorder, and this scoping review was conducted to synthesize valuable insights that have been presented in the literature.

Objective: This scoping review explored how NLP methods have been used in research to better understand bipolar disorder and identify opportunities for further use of these methods.

Methods: A systematic, computerized search of index and free-text terms related to bipolar disorder and NLP was conducted using 5 databases and 1 anthology: MEDLINE, PsycINFO, Academic Search Ultimate, Scopus, Web of Science Core Collection, and the ACL Anthology.

Results: Of 507 identified studies, a total of 35 (6.9%) studies met the inclusion criteria. A narrative synthesis was used to describe the data, and the studies were grouped into four objectives: prediction and classification (n=25), characterization of the language of bipolar disorder (n=13), use of EHRs to measure health outcomes (n=3), and use of EHRs for phenotyping (n=2). Ethical considerations were reported in 60% (21/35) of the studies.

Conclusions: The current literature demonstrates how language analysis can be used to assist in and improve the provision of care for people living with bipolar disorder. Individuals with bipolar disorder and the medical community could benefit from research that uses NLP to investigate risk-taking, web-based services, social and occupational functioning, and the representation of gender in bipolar disorder populations on the web. Future research that implements NLP methods to study bipolar disorder should be governed by ethical principles, and any decisions regarding the collection and sharing of data sets should ultimately be made on a case-by-case basis, considering the risk to the data participants and whether their privacy can be ensured.

(JMIR Ment Health 2022;9(4):e35928) doi:10.2196/35928

KEYWORDS

bipolar disorder; mental health; mental illness; natural language processing; computational linguistics

Introduction

Mental Health and Bipolar Disorder

In 2018, the Lancet Commission on global mental health and sustainable development reported that the global burden of disease related to mental health disorders has risen in all countries and that mental health services are frequently of a lower quality than those provided for physical health [1]. The 2013 Global Burden of Disease study [2] described depression as the predominant mental health problem worldwide, followed by anxiety, schizophrenia, and bipolar disorder, and the 2019 Global Burden of Disease study suggested that 1.2% (>815,000 cases) of the UK population has been diagnosed with bipolar disorder [3]. Bipolar disorder is a mood disorder associated with recurring episodes of extreme moods, ranging from severe depression to mania and with episodes lasting up to weeks at a time. Bipolar disorder has been shown to affect psychosocial functioning in areas of work, finance, cognition, and relationships [4], and people living with bipolar disorder are at a high risk for self-harm [5]. Of those diagnosed with bipolar disorder, 10%-20% will die by suicide, and therefore, the prevention of future episodes and the management of depressive and manic episodes are the major goals of treatment in bipolar disorder [6]. It is difficult to understand the lived experience of bipolar disorder through clinical practice alone, primarily because clinicians may only see their patients under a restricted set of conditions and those are likely to frame any discussion about the experiences of the patients.

The clinical diagnosis of bipolar disorder is a lengthy and costly process that takes an average of 9 years to complete [7]. A delayed diagnosis can have major implications for misdiagnosed individuals and may lead to inadequate or inappropriate treatments, a greater severity and frequency of mood episodes, and an increased risk for suicide among individuals who are later diagnosed with bipolar disorder [8]. Considering the economic implications, it is estimated that the total costs associated with bipolar disorder in the United Kingdom, including service costs and lost employment costs, could reach £8.2 billion (US $10.6 billion) by 2026 [9].

Natural Language Processing and Bipolar Disorder

The World Health Organization states that health systems must do more to respond to the burden of mental health disorders and that many people living with mental illness do not receive the care that they need. The development of *strengthened information systems* to provide evidence for population health monitoring and mental health surveillance is 1 of the 4 major objectives of the World Health Organization Mental Health Action Plan 2013-2020 [10]. The Lancet Commission also stated that digital technology can be used both to provide support and tools to people living with mental illness and to facilitate the screening and diagnosis of mental disorders using big data approaches. The increasing use of social media combined with the computational infrastructure of health care systems in the advent of the maturation of natural language processing (NLP) and machine learning (ML) technologies [11] provides exciting possibilities to investigate large amounts of data at the population and individual level.

Le Glaz et al [12] explained how language plays an important role in mental health technologies and how NLP uses the language resources available to analyze text both qualitatively and quantitatively to provide deeper insights into these data. NLP methods can focus on various features, including lexical choices, syntax, and semantics, to perform tasks such as topic modeling, clustering, and classification. Le Glaz et al [12] described that NLP in mental health research comprises the following four main stages: (1) corpus creation—the most common corpora include electronic health records (EHRs), social media data (eg, Reddit and Twitter posts), and transcribed patient interviews; (2) corpus processing—extracting medical terms or processing blocks of language using specific searches; (3) classification methods—ML techniques including deep learning; and (4) goal—the ultimate goal of validating a hypothesis or studying the behavior of a specific population.

Mental health research related to bipolar disorder can benefit from NLP methods in several ways. First, large amounts of longitudinal data from health records can be analyzed to provide population-level insights and to contribute to the creation of semiautomated systems, for example, to improve the specificity and speed of diagnosis [13,14]. Second, NLP methods can also be used for more fine-grained analyses at an individual level by analyzing lived experience accounts of bipolar disorder. This could include monitoring the sentiment and effect of web-based interactions over time [15], using textual cues in web-based communication to shed light on language features that relate to a bipolar disorder diagnosis [16], or using emotion detection methods to learn more about how emotions fluctuate over time [17]. Using NLP methods in the study of bipolar disorder could contribute to greater personalization of care through in-depth analysis of large amounts of textual data [18] and may yield insights that would be difficult to obtain in a formal health care setting owing to financial and time constraints. Analyzing the language used in nonclinical settings also provides an opportunity to learn more about what people with bipolar disorder say unprompted in situations that are not framed by clinicians or researchers. Becker et al [19] suggested that there is a need for a common language between the data science community and the health care community. This common language would enable data scientists to understand the technologies that are needed and how these can be implemented with clients, and enable health care workers to understand technical capabilities and the type of data that is most useful in developing automated systems. Carr [20] also explained that patient and public involvement and the incorporation of knowledge of domain experts (such as people with personal experience of bipolar disorder) are vital for ethical decision-making, because it enables a more robust understanding of language and context.

Objectives

To understand how NLP methodologies have been used to better understand bipolar disorder, we conducted a scoping review. Scoping reviews enable the researcher to present an overview of a diverse body of literature and allow for the synthesis of a range of study designs and methodologies without narrowing it down to a focused research question as in a systematic review [21]. The goal of this scoping review is in line with the definition

by Daudt et al [22], which states that a scoping review aims "to map the literature on a particular topic or research area and provide an opportunity to identify key concepts; gaps in the research; and types and sources of evidence to inform practice, policy making, and research."

An initial broad search of the published literature suggested that no previous studies have systematically reviewed the literature describing how NLP has been used to better understand bipolar disorder. Although there are reviews that have focused on the use of ML methods or big data in the study of bipolar disorder [23,24] or the use of ML and NLP for mental health more widely [12,25,26], this scoping review focused specifically on bipolar disorder and the application of NLP methods to this condition.

This scoping review explored how NLP methods have been used in research to better understand bipolar disorder and also to identify which aspects of bipolar disorder are underresearched and could be aided by computational linguistic methods (a definition of terms can be found in Multimedia Appendix 1 [27,28].

The four research questions that were used to guide this scoping review were as follows:

1. What trends can be observed in literature? (eg, What does the literature talk about? Where are the data sourced from?)
2. Which NLP methods have been used in the literature?
3. What are the clinical and practical applications reported in the literature?
4. What ethical considerations are present in the literature?

Methods

Overview

This scoping review was conducted with reference to the framework proposed by Arksey and O'Malley [29], expanded by Levac et al [30] and Daudt et al [22], and was informed by the guidance provided in the Joanna Briggs Institute (JBI) manual for evidence synthesis in scoping reviews [31]. This scoping review has been reported according to the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist [32] and was also informed by the guidance provided in the JBI manual for evidence synthesis in scoping reviews [33].

Search Strategy

A systematic and computerized search was conducted using 5 databases and 1 anthology: MEDLINE, PsycINFO, Academic Search Ultimate, Scopus, Web of Science Core Collection, and the ACL Anthology. The search was conducted between January 25, 2021, and August 27, 2021, and the search strategy was developed with informed advice from a topic librarian. There were no restrictions on the date of publication.

The search strategy used index terms and free-text terms to cover two core themes: (1) bipolar disorder and (2) NLP. Adjacency operators were used when incorporating free-text terms to ensure the specificity of the returned results. The full search terms are shown in Figure S1 in Multimedia Appendix 2 [13-17,34-66].

The final list of studies that were eligible for screening was imported into the Mendeley Reference Manager for duplicate removal before it was uploaded to Covidence [67], which was used for abstract and full-text screening and data extraction. Citation chaining was conducted on the final set of full-text papers used in this review.

Inclusion and Exclusion Criteria

Only papers written in English and published as peer-reviewed papers, full-text workshops, or conference proceedings were included in this review. It should be noted that the need for a faster review process has made conference proceedings the dominant form of published research in Computer Science and NLP [68]. To be included, studies needed to explicitly describe the application of an NLP method to the study of bipolar disorder (including those studies in which bipolar disorder was one of multiple psychological disorders being studied, but only when the data for bipolar disorder were separable). Studies that described quantitative, qualitative, or mixed method designs were eligible for inclusion, and papers were only included if they described completed research. Study designs and protocols were excluded from the study.

Papers were excluded from the scoping review if they only included an abstract and if the methodology described ML, deep learning, or big data approaches that did not rely on language features, for example, using magnetic resonance imaging data for bipolar disorder classification. Systematic and scoping review papers were also excluded from this study.

Study Selection and Screening Process

Initial screening of the titles and abstracts was conducted independently by the lead reviewer (DH) and the second reviewer (AW) using Covidence [67] to assess the suitability of the studies identified during the search for inclusion in the review. The eligibility criteria were tested in a pilot of 25 studies to ensure that the criteria were suitable for the review. The JBI [33] recommends that an agreement of 75% demonstrates that the inclusion and exclusion criteria performed well, and the agreement between the first and second reviewers for the pilot screening was 84%. After establishing that the eligibility criteria were valid, the reviewers screened the remaining papers independently and resolved any conflicts through discussion.

For all papers that passed the title and abstract screening, the lead reviewer located the full texts and screened them for eligibility in the review. The second reviewer (AW) screened 20.4% (23/113) of the papers at the full-text screening stage to verify their inclusion or exclusion, and a 100% agreement was achieved between the first and second reviewers. Data were extracted from the final papers that were eligible for inclusion by the first reviewer (DH) using a customized data extraction template that was designed and implemented in Covidence [67] and is shown in Table S5 in Multimedia Appendix 2. The data extraction template was piloted with 3 papers by the lead reviewer and was verified for accuracy by another member of the review team (PR). Minor changes were recommended for the data extraction template, including changing the phrasing of some fields and adding two fields that measured the reproducibility of computational linguistic papers—whether

the authors released their code and data set with the paper. A reviewer (DH) extracted all data from the 35 included articles. The extracted data are available in Table S1 in Multimedia Appendix 2.

To enhance the transparency of the research process, as recommended by the JBI [33], the protocol for this scoping review has been registered with Figshare and is available for public access [69].

Analysis

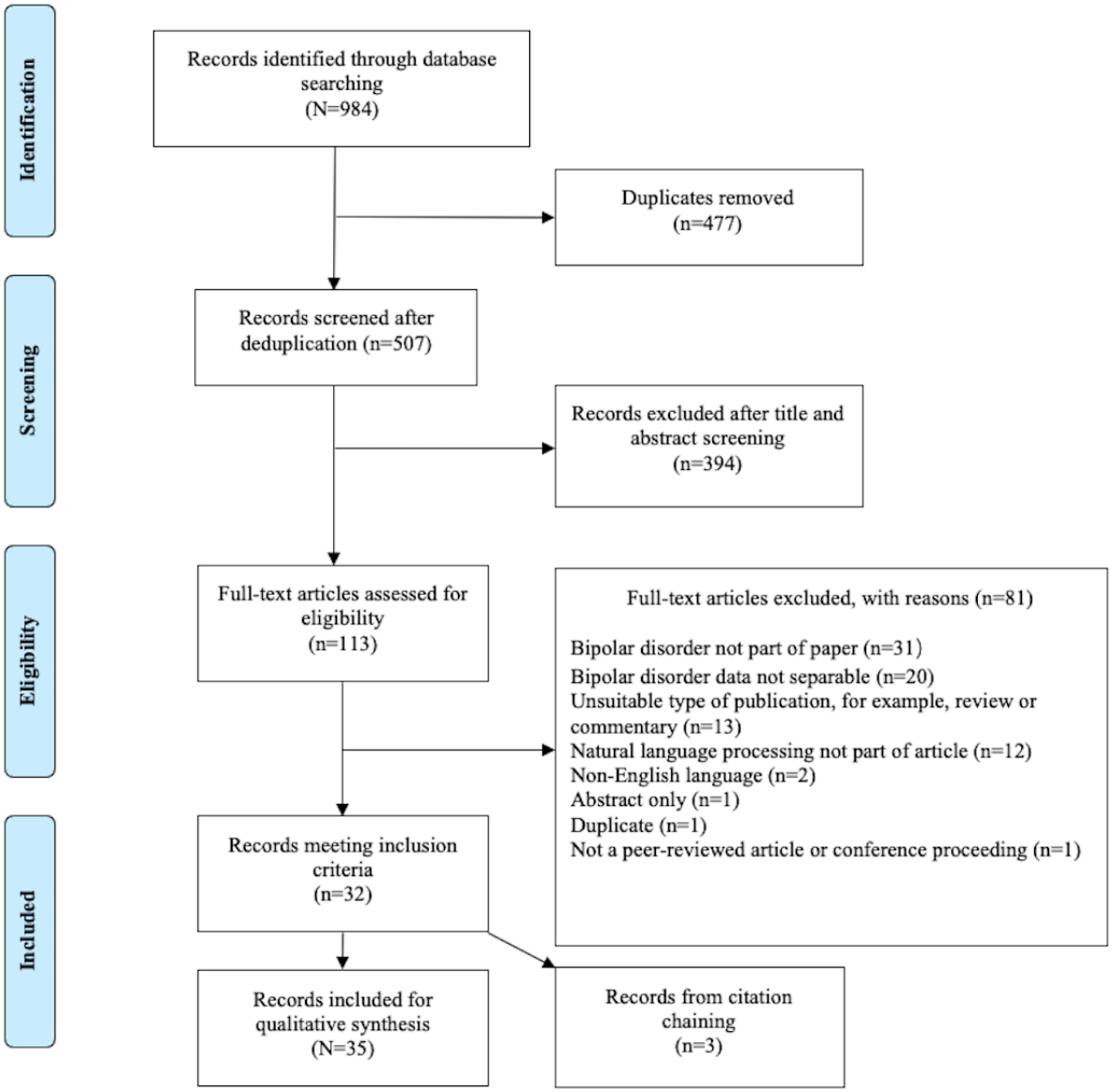
A narrative synthesis [70] of the included studies was undertaken to map the literature as outlined in the research questions. The data were presented using descriptive frequency tables and charts and summarized according to inductively developed objectives.

Results

Overview

The initial search yielded 507 documents after deduplication. Of these, 394 (77.7%) were excluded after title and abstract screening because of ineligibility, leaving 113 (22.2%) for full-text review. After full-text screening, a further 81 (71.6%) articles were excluded. The reasons for the exclusions are shown in Figure 1. After full-text screening, 32 (91%) papers were included in the review, and 3 (9%) additional papers were included after citation screening of these papers, totaling 35 papers for inclusion in the scoping review. The results of the search and the study inclusion process are presented in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart in Figure 1 [71].

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart of search history [71].



Research Question 1: What Trends Can Be Observed in Research Which Uses NLP to Study Bipolar Disorder?

Publication Characteristics

This study identified 35 articles published in 25 different sources, including journals (17/35, 49%), workshops (11/35, 31%), and conference proceedings (7/35, 20%). The publication sources demonstrated the interdisciplinary nature of this type of research, with the studies presenting a crossover among the fields of health care, computational linguistics, and computer science. The most popular source for publication was the *Workshop on Computational Linguistics and Clinical Psychology*, where 6 (17%) of the articles were published. The remaining sources published ≤2 articles each and are detailed in Table S1 in Multimedia Appendix 2. Of the included articles, 97% (34/35) analyzed textual data in English, and 3% (1/35) used Norwegian data.

Table 1 shows the countries of publication represented by the location of the first authors, who were predominantly located in the United States (14/35, 40%), followed by the United Kingdom (7/35, 20%), Taiwan (5/35, 14%), and Australia (3/35, 8%).

In terms of the discipline of the first authors, 88% (31/35) of the articles were first authored by individuals in the fields of NLP, computer science, and bioinformatics (ie, computational fields), whereas only 9% (3/35) of the articles were first authored by individuals with a background in medicine or health care. Of these 3 articles, the disciplines of the first authors included psychiatry (n=2, 67%) and public health (n=1, 33%). There was also an article for which the discipline of the first author could not be confirmed, although the author was based at the Institute of Psychiatry, Psychology and Neuroscience, King's College, London, when the article was published [57].

Table 1. The location of first authors (based on the location of affiliated institution).

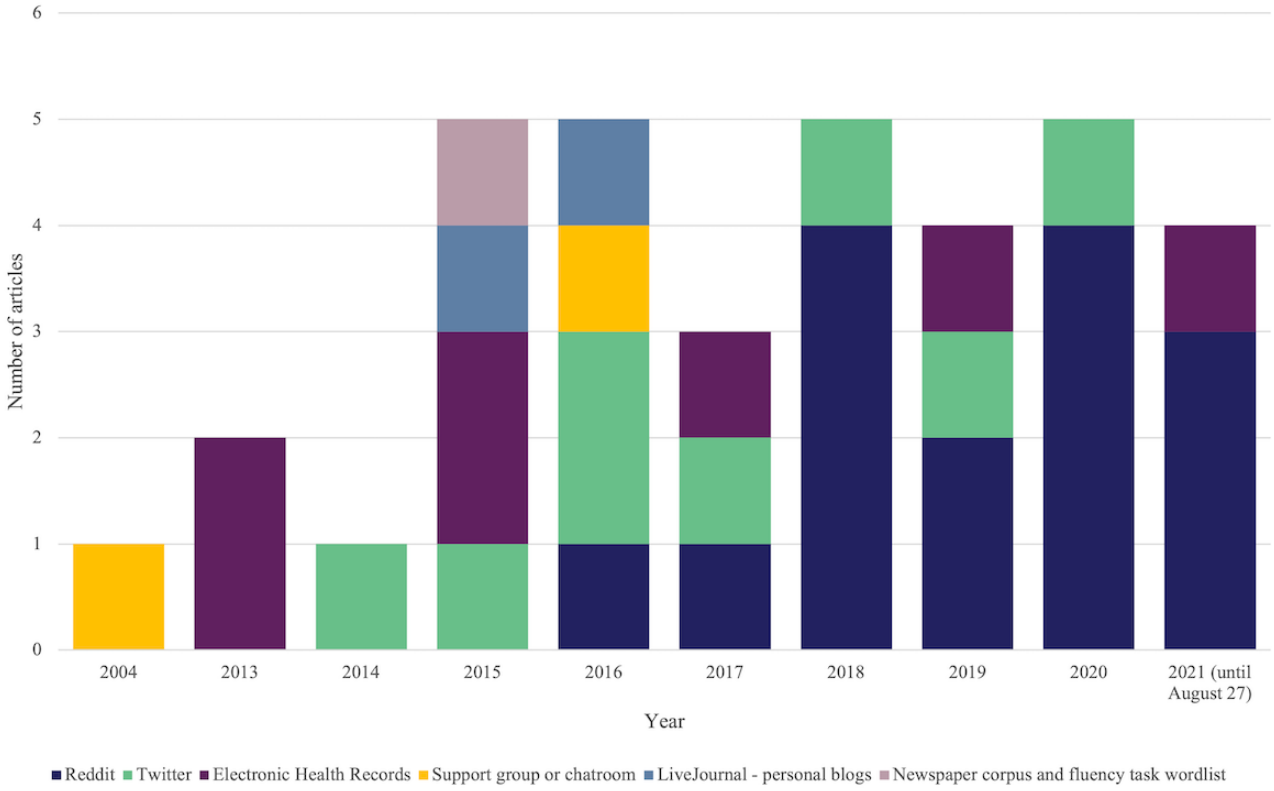
Country (based on registered institution) of first author	Value, n (%)
United States	14 (40)
United Kingdom	7 (20)
Taiwan	5 (14)
Australia	3 (8)
Croatia	1 (3)
United States and Belgium and Germany	1 (3)
Germany	1 (3)
Korea and United States	1 (3)
Brazil	1 (3)
Korea	1 (3)

Data Source

Figure 2 depicts the year-on-year trend in the publication of articles related to NLP and bipolar disorder. It is apparent that there has been an increase in the number of research articles related to this topic, from 1 relevant article in 2004 to 5 relevant articles in 2020. From 2015 onward, interest in this topic has remained fairly constant.

The published articles used a variety of sources for their corpora, including social media (Twitter, Reddit, support groups, chatrooms, and LiveJournal blogs), EHRs, and a newspaper corpus in conjunction with a fluency task wordlist. Figure 2 shows the increased use of social media since 2016, particularly after the publication of the study by Coppersmith et al [38] in 2014, which used Twitter to quantify mental health signals and could be described as a seminal work for this area of research. Since 2017, the only sources of data used in this field of research are Twitter, Reddit, and EHRs. The most commonly used data source for this type of research is Reddit (15/35, 43%), followed by Twitter (8/35, 23%) and EHRs (7/35, 20%). By including blogs, chatrooms, and support groups as a type of social media, 27 (77%) articles relied on data from social media, 7 (20%) used data from EHRs, and 1 (3%) used a newspaper corpus in conjunction with a fluency task wordlist.

Figure 2. Number of studies published yearly by data source.

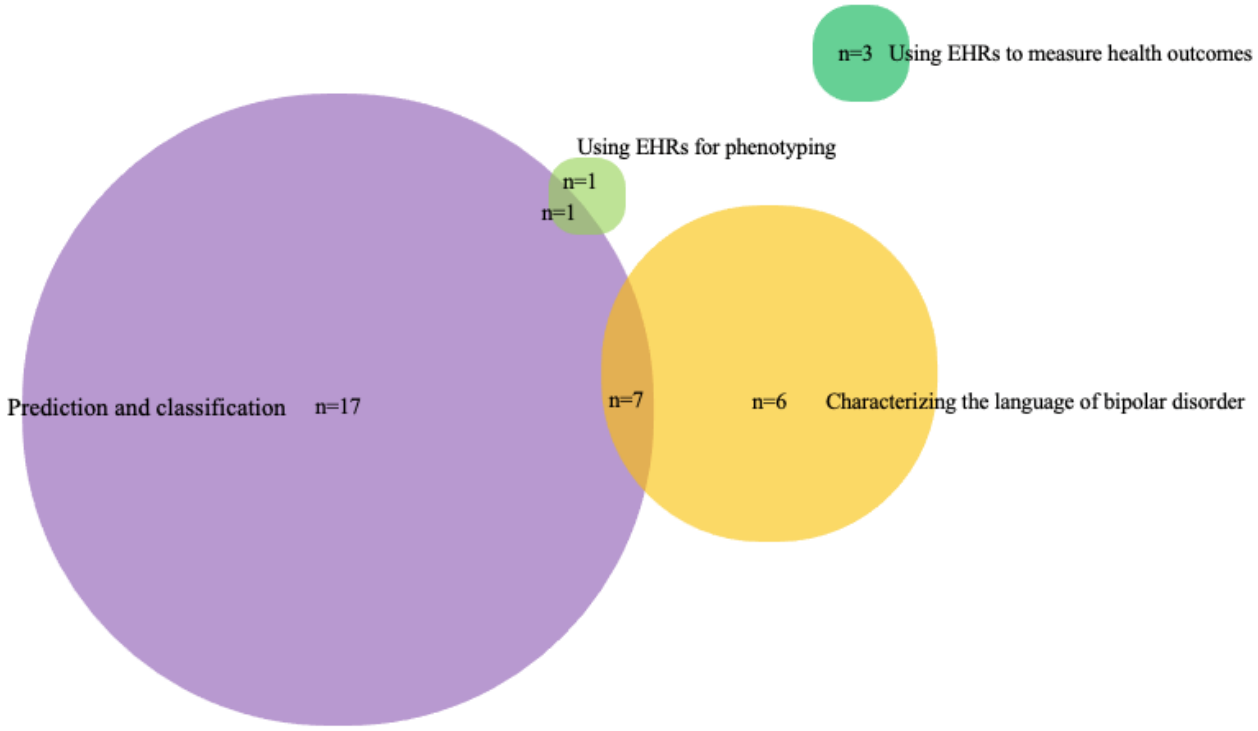


Objectives of the Articles

The primary objectives of the articles were inductively categorized into four broad categories: (1) prediction and classification, (2) characterization of the language of bipolar disorder, (3) use of EHRs to measure health outcomes, and (4) use of EHRs for phenotyping. Figure 3 shows the number of articles that were grouped into each of these objectives and suggests that there is some overlap between these objectives. For example, Low et al [51] used Reddit data to characterize trends in health anxiety and to build a ML classifier that predicted mental health conditions.

Figure 3 suggests that the most prevalent objective was prediction or classification related to bipolar disorder and other mental health conditions, either from social media or using EHRs, and the second most frequent objective was to characterize the language of bipolar disorder and mental health. The 2 least common objectives were to use data from EHRs to measure health outcomes and for phenotyping.

Figure 3. Grouped objectives of the studies. EHR: electronic health record.



Research Question 2: Which NLP Methods Have Been Used in This Research?

Because of the broad variation in the specific aims of each article and the ever-increasing number of NLP tools available to researchers, there is large variation in the tools and methods that were used in the included papers. The following subsections group the articles using the aforementioned 4 objectives, describe the methods identified across the articles, and provide a qualitative summary of the results. Table S1 in Multimedia Appendix 2 provides more fine-grained details of the methods and results reported for each article in this review.

Prediction and Classification

The most frequent objective of the included articles was to use text from social media (n=22) [15-17,34,36-44,48,49,51,54,59-63] or EHRs (n=3) [13,14,64] for prediction or classification purposes; for example, to predict a diagnosis of bipolar disorder based on features in the text. Among the 25 papers categorized into the objective of prediction and classification, 21 (84%) classified posts or users into a bipolar disorder class after comparison with a control group or with other mental health conditions. The aims of the remaining studies included; predicting the emotional tone in a bipolar disorder community, that is, how interactions in a web-based community affect people [15]; predicting the future occurrence of bipolar disorder based on a user's posts in a nonclinical subreddit before joining a bipolar disorder subreddit [63]; performing classification to measure subreddit uniqueness [40]; and using off-the-shelf algorithms to predict the demographic characteristics of people who self-reported a bipolar disorder diagnosis on Reddit [44]. There was large variation in the amount of data collected, with some authors reporting the number of relevant users and posts or comments and some reporting only the number of posts or users. The number of reported users within the bipolar disorder class varied from 50 patients with bipolar disorder listed in EHRs [14] to 19,685 Reddit users [44], and the number of reported posts or documents varied from 1000 blog posts [16] to >21 million Reddit posts [44].

Of the 22 studies that used social media for classification or prediction, 59% (13/22) verified a diagnosis. In the most rigorous cases, diagnoses were verified using detection patterns that incorporated diagnosis keywords collected from the corresponding Diagnostic and Statistical Manual of Mental Disorders (DSM), 5th Edition headings [37]. In other cases, regular expressions were used to pattern-match explicit expressions such as *I was diagnosed with bipolar* or to match mental health keywords used in bios. For the remaining articles (n=9, 41%), the authors used all posts collected from relevant bipolar disorder groups (eg, bipolar disorder portals on Twitter or subreddits related to bipolar disorder) without verifying whether the authors of these posts had received a diagnosis. The 3 studies that used EHRs to predict a diagnosis built their classifiers on a population of individuals within the health records who had received a previous diagnosis of bipolar disorder. The reliability of methods used to establish a diagnosis from social media data should be treated with some caution, because 9 of the articles within this review treated membership in a forum as confirmation of a diagnosis. In reality, forums are likely to include friends, family, and interested observers; therefore, this noisy verification of diagnosis could lead to unreliable data. Even when diagnoses are confirmed through more rigorous pattern matching using regular expressions, there is still a chance that users on the web may not have a genuine diagnosis. However, as described by Coppersmith et al [38], “Given the stigma often associated with mental illness, it seems

unlikely users would tweet that they are diagnosed with a condition they do not have."

Table S3 in Multimedia Appendix 2 shows the variety of ML models that have been applied to the objective of prediction and classification, and Table S4 in Multimedia Appendix 2 shows the different features that have been used as inputs for these models. The pooled data revealed that 19 of the articles used ML methods (see Multimedia Appendix 1 for defining terms) and 13 of the articles also implemented deep learning tools (many of the studies used a combination of both tools to compare the accuracy of different classifiers). Logistic regression was the most commonly used classifier for ML tasks, and convolutional neural networks were popular methods for studies that used deep learning. Several studies that implemented a deep learning methodology (Multimedia Appendix 1) also reported the use of an attention mechanism within their models (n=6). Galassi et al [72] described that the attention mechanism is part of a neural architecture that is able to "dynamically highlight relevant features of the input data, which, in NLP, is typically a sequence of textual elements." The papers in this review that incorporated an attention mechanism described improved performance when compared with baseline methods, because the attention weights were used to demonstrate the most important words or sentences within the text for making classification decisions.

The features used most frequently for classification were derived from Linguistic Inquiry and Word Count (LIWC) [73] (for features relating to emotion and psychological state) and Term Frequency Inverse Document Frequency (TF-IDF) vectors. Pattern of Life (PoL) analytic features were introduced by Coppersmith et al [38], and relay information about the patterns and behavioral tendencies of users measured by social interactions (eg, tweet rate and number of @mentions) and were implemented in 4 of the studies. For studies that relied on deep learning methodologies, a number of different types of word embeddings were used as inputs for the models, including those derived from Bidirectional Encoder Representations [74], Word2vec [75], and global vectors for word representation (GloVe) [76].

In terms of accuracy (reported as overall accuracy, precision, recall, F_1 score, and area under the curve defined in Multimedia Appendix 1) of the studies that aimed to classify a population into a bipolar disorder class (n=21), the following studies reported the highest scores (at 90%/≥0.9). Chang et al [36] reported a precision of 0.96 by using a random forest classifier based on TF-IDF features of Twitter users in single-task learning. Chen et al [17] reported an overall accuracy of 91.9% using the EMOTIVE ontology, LIWC, and Pattern of Life features for Twitter users with a logistic regression classifier in single-task learning. Huang et al [42] reported 95% precision for the female class using a pattern attention mechanism in single-task learning. Jiang et al [48] reported an F_1 score of 0.982 for Reddit users using a Retrieval Augmented Language Model in single-task learning. Kim et al [49] achieved an overall accuracy of 90.2% for Reddit users using a convolutional neural network model with TF-IDF Word2vec vectors for single-task learning. Saravia et al [60] achieved a score of 96% precision for classifying Twitter users as having bipolar disorder using TF-IDF features with a random forest classifier in single-task learning, and Castro et al [13] reported an area under the curve of 0.93 for classifying an individual as having bipolar disorder or not using a logistic regression classifier in single-task learning from EHR data.

There were 4 articles, which used NLP methods for alternative classification purposes. Silveira et al [15] predicted how the emotional states of Reddit users would change after interacting on social media and framed this as a regression task that outperformed the baseline by a score of at least 12.9. Their results showed that general emotional states improved after interacting on the web and that the emotional tone of the final post by the thread author was generally more positive than their initial post. Gkotsis et al [40] used an ML classifier to measure the vocabulary uniqueness between mental health subreddits and demonstrated that there was a shared vocabulary across 3 different bipolar disorder subreddits. Thorstad and Wolff [63] demonstrated that future mental disorders could be predicted with an F_1 score of 0.37 (which, although low, is above chance). Their work described the possibility of building classifiers to identify people at risk for developing mental illnesses. Finally, Jagfeld et al [44] used hybrid models to predict age and gender, which achieved 99% and 97% accuracies on their test set, respectively, as well as an inference model for location that achieved 78.4% test set accuracy.

The literature reports a number of successes that have been achieved in a variety of NLP prediction tasks related to bipolar disorder, and the heterogeneity in the methods of the papers, their data sets, and their individual objectives reflects the wide breadth of the field and the potential for this area of research. The results of each study are provided in Tables S1-S3 in Multimedia Appendix 2 with more details on the methods and tools used.

Characterizing the Language of Bipolar Disorder

A total of 13 papers were grouped into the objective of characterizing the language of bipolar disorder [16,35,37,39,40,42,44, 50,51,53,55,58,66] and used methods to build a more fine-grained picture of the linguistic behaviors of people living with bipolar disorder. Table S1 in Multimedia Appendix 2 provides more information on the focus, method, and main outcomes of each of the papers included in this category. The main patterns that emerged from this synthesis are described here.

LIWC was used by a number of authors to characterize language [16,37,39,40,50,66]. Cohan et al [37] found that bipolar disorder populations were significantly more likely to use first-person singular pronouns than their control group of Reddit users without a self-reported diagnosis, which they have suggested correlates with the LIWC category of authenticity. Gkotsis et al [40] also found a large number of first-person pronouns when comparing one of the bipolar subreddits (r/bipolarSOs) with other mental health groups within their study. The authors reported that this observation has been found in previous research on the language of depression and also touched on the idea of authenticity by suggesting that people with bipolar

disorder may talk about personal issues more sincerely, which may increase their use of personal pronouns.

In all, 2 papers reported that the bipolar disorder community was more likely to talk about topics in the LIWC category of *Health* than a control group of Twitter users [39] and other web-based depression communities on LiveJournal blogs [16]. Yoo et al [66] identified clusters within the bipolar disorder community that were related to emotion and negative feelings, and their LIWC analysis also showed greater use of negative expressions when compared with people who posted on a depression subreddit. Kramer et al [50] described that the use of first-person pronouns was positively correlated with negative emotion words and that the use of *you* was positively correlated with positive emotion words. Huang et al [42] also reported that when using their graph-based algorithm, negative emotions were frequently used by their bipolar disorder group [42] and described that there was a significant difference in the use of tense between men and women in that women tended to use the present tense *I am* and that men preferred the past tense *I was.* Coppersmith et al [39] also described that the use of auxiliary verbs was significant in bipolar disorder users when compared with their control group according to the LIWC analysis.

In terms of the use of web-based social media sites, Jagfeld et al [44] used out of the box NLP models to report that most users who self-reported a diagnosis of bipolar disorder fell into the 30-49 year age range (47.5% of their data set) and were more likely to be classified as female (52.2%). This is in contrast with the demographic information of the general US Reddit adult population, with 64% of population being composed of people between the ages of 18 and 29 years and 67% of the US Reddit users being men [77]. McDonald and Woodward-Kron [53] focused on member role change in web-based communities using corpus methods and showed that users became like veterans the longer they used a web-based forum, dispensing advice using modal declaratives such as *You should consider seeing a professional*. Over the course of time, users preferred to describe themselves as *having bipolar* instead of *being bipolar*, and Kramer et al [50] reported in their study that users wrote more as they spent more time on the site. Park and Conway [55] assessed the readability of posts on Reddit over time and reported that although the posts of people posting on bipolar disorder subreddits were initially significantly more difficult to read than the control group, this improved as members participated more in the community. Rosenstein et al [58] conducted a verbal fluency task to understand how semantic structure is affected by bipolar disorder and discovered that people with bipolar disorder presented lower lexical diversity and semantic coherence than the control group.

Finally, 2 papers observed how external factors can influence the representation of bipolar disorder on the web. Low et al [51] used topic modeling and sentiment analysis to compare health-related anxiety presented on Reddit before the COVID-19 pandemic and during the pandemic. They demonstrated that the bipolar disorder subreddit did not seem to have suffered from induced health anxiety unlike other subreddits that were affected, such as those related to borderline personality disorder and posttraumatic stress disorder. They reported that there was no negative semantic change in the bipolar disorder subreddit by the middle of the pandemic, whereas other subreddits demonstrated significant negative semantic changes at this point. Budenz et al [35] used Twitter to collect tweets from communication spikes caused by external events (eg, the death of mental health advocate Carrie Fisher) to measure the amount of stigma or support presented in the communication. Their results showed that >67,393 (5.3% of the total sample) tweets discussed bipolar disorder, and 64.7% (4709/7281) of the bipolar disorder tweets that displayed stigma or support showed stigmatizing language. This was in contrast with 4.3% (38,336/873,590) of the tweets related to mental health and mental illness more generally that displayed stigmatizing language.

Using EHRs to Measure Health Outcomes

There were 3 articles that were grouped into the category of using EHRs to measure health outcomes [56,57,65]. All 3 studies used the South London and Maudsley Clinical Record Interactive Search (CRIS, 2021) database between 2013 and 2019 to assess different health outcomes of people diagnosed with bipolar disorder.

Wu et al [65] used the database to investigate smoking prevalence and the factors that influence it in populations receiving mental health care. Using open-text fields with General Architecture for Text Engineering (GATE) [78], the authors created a CRIS-IE smoking application using a shallow parsing rule–based approach to keywords. The results of this study demonstrated that patients with schizophrenia and schizoaffective disorder had a higher smoking prevalence than those with bipolar disorder. Patel et al [56] used the CRIS to assess the impact of mood instability on the clinical outcomes of individuals receiving secondary mental health care. TextHunter [79] was used to extract documentation related to mood instability from unstructured free-text fields, and supervised learning was used to develop support vector machine applications that were combined to generate a binary variable of instability. The prevalence of instability within one month of clinical presentation was 22.6% in the bipolar disorder population compared with 12.1% in the overall sample. Finally, Ramu et al [57] extracted descriptions of insight from text fields to determine whether poor insight recorded early after clinical presentation could predict subsequent service use. The authors used TextHunter [79] to create an ML algorithm based on a sample from clinical records to predict good or poor insight or to classify a document as irrelevant. The algorithm identified 61 patients with bipolar disorder who had at least one recording of poor insight, and the authors reported that a higher number of hospitalization episodes, unique antipsychotics, and inpatient days were all significantly correlated with poor insight.

Using EHRs for Phenotyping

The final characteristic used to group papers in this review was the use of EHRs for phenotyping (n=2) [13,52], in which case phenotyping relates to the process of characterizing or determining the observable characteristics of an individual and can refer to anything from a common trait, such as height or hair color, to presence or absence of a disease [80].

Castro et al [13] performed EHR-based phenotyping of bipolar disorder using EHRs to extract diagnostic data and compared the validity of an NLP algorithm with diagnostic interviews conducted by clinicians. The performance of the NLP algorithm for classifying case and control patients was assessed against DSM-IV Structured Clinical Interview for DSM-IV Disorders gold standard interviews, and the algorithm scored a positive predictive value of 0.85. Lyalina et al [52] used EHRs to identify the signature of 3 neuropsychiatric illnesses and to elucidate their phenotypic boundaries. The authors used text mining to annotate notes with concepts from 22 clinically relevant ontologies after preprocessing and negation checking, and enriched concepts were identified by reducing the number of case and control notes to 1000 each. A Fisher exact test was used to measure the enrichment within the sample. Their results demonstrated that the symptoms related to enriched phenotypes of bipolar disorder include migraines, irritable bowel syndrome, sleep disorders, ulcers, and mania and that there is substantial phenotypic overlap between bipolar disorder and schizophrenia. It should be noted that although not eligible for inclusion in this scoping review based on the methodology used, Mota et al [81] presented evidence to suggest that despite often sharing psychotic symptoms such as hallucinations, hyperactivity, and aggressive behavior, schizophrenia and bipolar disorder can successfully be differentiated based on the analysis of dream graphs, but psychometric scales cannot achieve the same result. Their work could provide a framework that uses behavioral biomarkers to drive a more objective, bottom-up search for anatomical and physiological biomarkers [81].

Research Question 3: What Are the Clinical and Practical Applications of the Current Research?

It is important to understand why NLP methods have been applied to the study of mental health conditions and if this type of research is grounded in real-life implementations, particularly when large amounts of potentially sensitive social media data have been used.

The articles used in this review cited various reasons that make this type of study clinically relevant. Many authors have suggested that applying NLP methods to social media data could aid clinicians in their evaluations of bipolar disorder and that improved suicide prevention methods could be designed by combining ML methods and the medical community [34,39]. Sekulić et al [61] stated that the high incidence of suicide in bipolar disorder demonstrates the importance of early detection, and many authors suggested that applying NLP methods to social media data could contribute to the understanding of bipolar disorder and its detection and diagnosis [36,37,48,49,53,54,62,63]. Coppersmith et al [38] also suggested that using social media for large-scale data collection could complement existing methods and potentially make individual and population analyses quicker and cheaper. A number of the authors described that building a varied representation of bipolar disorder (eg, using features such as semantic deficit or attention weights) could provide a better understanding of the user experience, aid in diagnosis [16,42,58,64,66], and generate hypotheses for the clinical settings that may inform the provisioning of appropriate therapeutic resources [51].

Another practical application cited by the authors was the implementation of intervention systems based on flagging social media data for the moderator's attention [15,41,43]. Chen et al [17] and Park and Conway [55] described how different linguistic features can show how mental health conditions fluctuate over time and how these could help to identify worsening mental health. Gkotsis et al [40] suggested that urgency markers could be implemented for targeted interventions. Saha et al [59] reported that NLP methods could be used to screen and monitor health groups, and Saravia et al [60] and Silveira et al [15] suggested that social media data could be used to assist in the potential distribution of treatment to populations that are difficult to reach through traditional approaches. Ethical questions related to invasion of privacy, particularly when referring to populations who may have undetected mental illnesses, are raised by these possible innovations. It must be questioned whether the collection of data from social media platforms from a possibly unsuspecting population is ethical and it is also unclear who would be responsible for such an intervention.

In terms of the relevance of social media itself to people living with bipolar disorder, Kramer et al [50] described the hypothesis that 24 hour access to other people living with the same problem could reduce social isolation, improve coping skills, and improve patient knowledge about their own condition. Jagfeld et al [44] suggested that being aware of the demographics of web-based communities may help clinicians in recommending forums to their clients. Budenz et al [35] also described that social media advocacy can increase the amount of social support for people living with bipolar disorder to minimize the stigmatizing content posted on the web.

Finally, considering the use of NLP and medical records, Castro et al [13] and Dai et al [14] described that specific and predictive diagnostic algorithms could be created to assist with the diagnosis and to improve accuracy, achieving results that are comparable with diagnostic interviews. Other authors demonstrated how data, extracted using NLP, could improve care management and demonstrated the need, for example, to screen for the presence of instability on a routine basis or improve the assessment of smoking behavior [52,56,57,65].

Research Question 4: What Ethical Considerations Are Present in the Literature?

A total of 60% (21/35) of articles used for the review referenced ethical considerations, and 40% (14/35) did not reference any ethical decision-making or design. The ethical considerations that were implemented are shown in Table 2, and Table 3 describes how the authors managed the code and data set release. It is interesting to note that the papers published until 2016 included limited discussion regarding ethical considerations, with only 47% (7/15) of papers published between 2004 and 2016 acknowledging ethical decision-making. In these earlier papers, discussion was generally limited to short statements, such as *all collected data were publicly posted to Twitter between 2008 and 2015* [39] or clarification that ethics approval had been granted.

Ethical considerations became more frequent in papers published from 2017 onward, with 67% (14/21) papers published in

2017-2021 incorporating (generally much more robust) ethics statements. The increased focus on ethics correlates with the drive toward open science, and recent guidelines were implemented by scientific communities, such as the Association for Computational Linguistics, that require authors to upload a checklist for responsible NLP research alongside any paper submission [82] and to include a discussion about positive and negative societal impacts that could stem from the research.

Several articles in this review provided a more detailed discussion of ethical issues. Benton et al [34] suggested that NLP models could be overgeneralized or used to identify specific people, and Cohan et al [37] stated that risks to individuals as a consequence of social media research should always be considered. Various articles [38,55,60,61] described that mental health analyses must be approached sensitively and [55] also described the nature of Reddit and the throwaway accounts that can protect users from social discrimination. The studies by Jagfeld et al [44] and Thorstad and Wolff [63] both described the issue of dual use in which research can be misused to harm the public (eg, by insurance companies) and also suggested that a possible solution to violating user privacy would be to inform people that the casual comments they make on social media may be mined [63]. Finally, Huang et al [42] stated that the practical application of their proposed model would only be used if both health care practitioners and patients agreed to use it.

Table 2. Ethical considerations.[a]

Ethical considerations	Values, n (%)
None	14 (40)
All user information anonymized	10 (29)
Ethical approval granted by relevant institution	9 (26)
Excerpts from data paraphrased or not published	3 (9)
No private tweets or protected user accounts used	2 (6)
Models did not include user features	1 (3)
URLs and usernames containing sensitive information removed	1 (3)
Comply with data usage agreement	1 (3)
Detailed initial psychological evaluations were excluded in the interest of public privacy	1 (3)

[a]Note that n does not equal total sample of 35 papers as some papers appear across multiple rows.

Table 3. Data set and code release.

Code and data set release	Values, n (%)
Data set	
Availability not referenced	15 (43)
Not provided for ethical reasons but potentially available on request	11 (31)
Link to dataset or code to scrape dataset provided	6 (17)
Faulty link provided	2 (6)
Partial access provided	1 (3)
Code	
Not released	27 (77)
Access provided	7 (20)
Available on request	1 (3)

Discussion

Principal Findings

This scoping review highlights the heterogeneity in the existing research that has used NLP methods to study bipolar disorder. The review suggests that the literature has been produced predominantly in the United States and the United Kingdom (21/35, 60%) and that 66% (23/35) of the studies used Twitter or Reddit as a source of data. The studies were predominantly led by authors from the computational and informatics fields (31/35, 88%), with only 3 articles being first authored by a health care expert. The articles were grouped into four inductively developed objectives: (1) prediction and classification, (2) characterization of the language of bipolar disorder, (3) use of EHRs to measure health outcomes, and (4) using EHRs for phenotyping, with most of the articles using NLP methods for prediction and classification purposes. The review suggests that using NLP for the study of mental health and bipolar disorder specifically is a growing field and it seems to have been influenced by the study of Coppersmith et al [38]

when they provided a framework for obtaining quantifiable data in mental health research using Twitter.

The range of technologies that have been applied in the field reflects the ever-increasing possibilities for conducting research on language, with the most recent articles mainly favoring deep learning methodologies and word embeddings. The results from the existing research are varied and promising and indicate the usefulness of NLP methods to aid in diagnosis, predict the emotional impact of web-based interactions, characterize the language used by people living with bipolar disorder, and use phenotypes to better assist in care management. The 13 articles that characterized the language of bipolar disorder provided evidence to suggest that there are some observable linguistic traits that can be identified in a population with bipolar disorder; for example, an increased use of both first-person pronouns and negative emotion expressions, which could be useful in providing a better representation of bipolar disorder and developing early detection or intervention systems.

Future Research

There are 4 areas for further research that are proposed based on the results of this review. First, Sekulić et al [61] referred to the high incidence of suicide in bipolar disorder and suggested that early detection systems could be developed. The use of signposting systems that could flag at-risk users for moderator intervention also has been discussed by several authors included in this review. Considering that bipolar spectrum disorders are associated with significant disinhibition and poor judgment, which can lead to the commission of risky and dangerous behaviors [83], a key area of future research should be to look at how risky behaviors (not just suicide) are discussed. This study would help to better understand how people living with bipolar disorder can be supported by health care providers to facilitate and improve their quality of life. Examples of risk-taking behaviors referenced in the literature on bipolar disorder include binge eating, excessive drinking, gambling, self-injury, and risky spending [83].

Second, several articles described the potential benefits of using social media for people living with bipolar disorder (eg, becoming more informed about bipolar disorder [53], improved emotional state after web-based interactions [15], and improved readability scores over time [55]). Further research could be conducted on how people living with bipolar disorder can best be supported on the web, and specific evaluative frameworks could be implemented for this purpose [50].

Third, although gender was not discussed in detail in this review, there are some contradictory results with regard to the portrayal of gender on social media by people living with bipolar disorder. Although Cohan et al [37] proposed that their Reddit corpus may be gendered toward men (because of the large amount of references to women), Jagfeld et al [44] used predictive algorithms to suggest that more than half of their Reddit corpus comprises women and that feminine-gender–identifying people with a BD diagnosis seem to be more likely to use Reddit and disclose their diagnosis. This area of research was also touched upon by Huang et al [42] who built a set of gender-specific syntactic patterns for bipolar disorder recognition. Further computational linguistic research could be conducted to determine how gender is presented by social media users and whether this correlates with demographic statistics for the diagnosis of bipolar disorder. Future work may demonstrate if there are demographic groups, which are currently undersupported by health care services and are instead seeking help online.

Finally, impaired social and occupational functioning in bipolar disorder has been presented frequently in the wider literature, although a review of the range of functioning in bipolar disorder has demonstrated that 16% of individuals diagnosed with the condition function at a high level and that functioning with bipolar disorder may have been underestimated by some clinical measures [84]. This area has not yet been explored using NLP methods, which presents an opportunity to provide a more balanced perspective on the wide range of ways in which people live with bipolar experiences based on lived experience narratives that are free from the potential ceiling effect of some clinical measures.

Ethics

It is crucial that ethical design underpins research that uses NLP to study bipolar disorder, because there are serious ethical concerns relating to data use and anonymization and the concept of dual use. Although the research community must uphold rigorous ethical standards for data collection and protection, researchers are simultaneously moving toward open science to ensure transparency of research practices and to enable easy access to the data from which important conclusions have been drawn. Therefore, researchers are now faced with a conflict between the objectives of the open science movement and the need to uphold data privacy. Dennis et al [85] described that privacy and open science are on a collision course. A number of ideas have been proposed to manage this conflict, although there is still no clearly defined solution [85-87]. The British Psychological Society stated that internet-mediated research should obtain valid consent when it “cannot be reasonably argued that online data can be considered ‘in the public domain’” and that any data disseminated through the research should maintain the anonymity of the author [88]. They also stated that research should maximize benefits and minimize harm (to the research participants), referring to the fourth main principle of the Code of Human Research Ethics [89]. Friedrich and Zesch [90] described that ethics should be integrated into any NLP project and that NLP researchers should be mindful of the implications of developing any language technology. Benton et al [91] provided guidelines for ethical research using social media data stating that all research should consider the benefits and risks involved from the outset, thus enabling the implementation of strategies to make research as risk averse as possible.

The literature included in this scoping review suggests that there is ambiguity around the best practice for the ethical design of NLP methodologies, with 40% (14/35) of the articles making no reference to ethical decision-making and a wide range of methodologies for the articles that do. Future research that implements NLP methods to study bipolar disorder should be governed by ethical principles, and researchers should be aware that the best intentions could still have potentially harmful

consequences. Although researchers are likely to be governed by the principals of open science, any decisions regarding the collection and sharing of data sets should ultimately be made on a case-by-case basis with consideration for the risk to the data participants and ensuring their privacy.

Limitations

Although this scoping review was conducted according to a scoping review methodology and a previous protocol, there were some limitations that are worth noting.

First, as described throughout the review, there was large variation in the way NLP methods were described and indexed, and so it is possible that some relevant articles were not included in this review through the terms used in the search query.

Second, the data were extracted by only 1 reviewer because of the relatively high number of studies identified in the data extraction phase. An attempt was made to ensure accurate extraction by using a verified and standardized extraction form; however, the data that were extracted and used within the scoping review were predominantly qualitative, so it is likely that there could be researcher bias.

Finally, this review was conducted in an area of research that is constantly growing and developing and therefore only provides a time-stamped representation of the field.

Conclusions

This scoping review provided an overview of 35 papers that applied NLP methods to the study of bipolar disorder. The data indicate that there are increasing opportunities for interaction between the clinical and NLP communities, and existing research shows how the analysis of language can be used to assist with and improve the provision of care for people living with bipolar disorder. There are 4 areas in bipolar disorder research that have been identified that may benefit from NLP methods, including the study of risk-taking behaviors, the research and design of web-based support groups specific to bipolar disorder, the study of social and occupational functioning, and the study of gender representation in bipolar disorder populations on the web.

Acknowledgments

This review was completed as part of an Economic and Social Research Council Collaborative Studentship Competition (CASE) PhD studentship (grant ES/P000665/1). The funder had no role in the study design, collection, analysis, or interpretation of the data, writing of the manuscript, or the decision to submit the paper for publication.

Authors' Contributions

DH designed this study, wrote the protocol, and performed the literature search. Title and abstract screening were performed by DH and AW, and full-text screening was performed primarily by DH, with AW screening 20% of the papers at the full-text stage. PR validated the data extraction template, and DH extracted all data for the review. FL, SJ, and PR provided comments and guidance throughout this study and provided valuable insights for the draft manuscript. All the authors approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Definition of terms.
[DOCX File , 18 KB - mental_v9i4e35928_app1.docx]

Multimedia Appendix 2
Extracted data, data extraction template, and search strategy.
[DOCX File , 302 KB - mental_v9i4e35928_app2.docx]

References

1. Patel V, Saxena S, Lund C, Thornicroft G, Baingana F, Bolton P, et al. The Lancet Commission on global mental health and sustainable development. Lancet 2018 Oct 27;392(10157):1553-1598. [doi: 10.1016/S0140-6736(18)31612-X] [Medline: 30314863]
2. Global Burden of Disease Study 2013 Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. Lancet 2015 Aug 22;386(9995):743-800 [FREE Full text] [doi: 10.1016/S0140-6736(15)60692-4] [Medline: 26063472]
3. GBD Compare | IHME Viz Hub. URL: https://vizhub.healthdata.org/gbd-compare [accessed 2022-04-04]
4. Tatay-Manteiga A, Correa-Ghisays P, Cauli O, Kapczinski FP, Tabarés-Seisdedos R, Balanzá-Martínez V. Staging, neurocognition and social functioning in bipolar disorder. Front Psychiatry 2018 Dec 19;9:709 [FREE Full text] [doi: 10.3389/fpsyt.2018.00709] [Medline: 30618879]

5. Clements C, Jones S, Morriss R, Peters S, Cooper J, While D, et al. Self-harm in bipolar disorder: findings from a prospective clinical database. J Affective Disorders 2015 Mar 01;173:113-119 [FREE Full text] [doi: 10.1016/j.jad.2014.10.012] [Medline: 25462404]
6. Geddes JR, Miklowitz DJ. Treatment of bipolar disorder. Lancet 2013 May 11;381(9878):1672-1682 [FREE Full text] [doi: 10.1016/S0140-6736(13)60857-0] [Medline: 23663953]
7. Bipolar – the facts. Bipolar UK. URL: https://www.bipolaruk.org/faqs/bipolar-the-facts [accessed 2022-04-12]
8. Keramatian K, Pinto JV, Schaffer A, Sharma V, Beaulieu S, Parikh SV, et al. Clinical and demographic factors associated with delayed diagnosis of bipolar disorder: data from Health Outcomes and Patient Evaluations in Bipolar Disorder (HOPE-BD) study. J Affect Disorders 2022 Jan 01;296:506-513 [FREE Full text] [doi: 10.1016/j.jad.2021.09.094] [Medline: 34606817]
9. Mccrone P. Paying the Price The Cost of Mental Health Care in England to 2026. London, United Kingdom: King's Fund; 2008.
10. Mental health action plan 2013-2020. World Health Organization. URL: https://www.who.int/publications/i/item/9789241506021 [accessed 2022-04-04]
11. Conway M, O'Connor D. Social media, big data, and mental health: current advances and ethical implications. Curr Opin Psychol 2016 Jun;9:77-82 [FREE Full text] [doi: 10.1016/j.copsyc.2016.01.004] [Medline: 27042689]
12. Le Glaz A, Haralambous Y, Kim-Dufor D, Lenca P, Billot R, Ryan TC, et al. Machine learning and natural language processing in mental health: systematic review. J Med Internet Res 2021 May 04;23(5):e15708 [FREE Full text] [doi: 10.2196/15708] [Medline: 33944788]
13. Castro VM, Minnier J, Murphy SN, Kohane I, Churchill SE, Gainer V, International Cohort Collection for Bipolar Disorder Consortium. Validation of electronic health record phenotyping of bipolar disorder cases and controls. Am J Psychiatry 2015 Apr;172(4):363-372 [FREE Full text] [doi: 10.1176/appi.ajp.2014.14030423] [Medline: 25827034]
14. Dai HJ, Su CH, Lee YQ, Zhang YC, Wang C, Kuo CJ, et al. Deep learning-based natural language processing for screening psychiatric patients. Front Psychiatry 2020;11:533949 [FREE Full text] [doi: 10.3389/fpsyt.2020.533949] [Medline: 33584354]
15. Silveira B, Silva HS, Murai F, da Silva AP. Predicting user emotional tone in mental disorder online communities. Future Generation Comput Syst 2021 Dec;125:641-651. [doi: 10.1016/j.future.2021.07.014]
16. Nguyen T, O'Dea B, Larsen M, Phung D, Venkatesh S, Christensen H. Differentiating sub-groups of online depression-related communities using textual cues. In: Web Information Systems Engineering – WISE 2015. Cham Switzerland: Springer; 2015.
17. Chen X, Sykora M, Jackson T, Elayan S, Munir F. Tweeting your mental health: an exploration of different classifiers and features with emotional signals in identifying mental health conditions. In: Proceedings of the 51st Hawaii International Conference on System Sciences. 2018 Presented at: 51st Hawaii International Conference on System Sciences; 2018; Hawaii. [doi: 10.24251/hicss.2018.421]
18. Lovejoy C, Buch V, Maruthappu M. Technology and mental health: the role of artificial intelligence. European Psychiatry 2019 Jan;55:1-3. [doi: 10.1016/j.eurpsy.2018.08.004] [Medline: 30384105]
19. Becker D, van Breda W, Funk B, Hoogendoorn M, Ruwaard J, Riper H. Predictive modeling in e-mental health: a common language framework. Internet interventions 2018 Jun;12:57-67 [FREE Full text] [doi: 10.1016/j.invent.2018.03.002] [Medline: 30135769]
20. Carr S. 'AI gone mental': engagement and ethics in data-driven technology for mental health. J Mental Health 2020 Apr 30;29(2):125-130. [doi: 10.1080/09638237.2020.1714011] [Medline: 32000544]
21. Pham MT, Rajić A, Greig JD, Sargeant JM, Papadopoulos A, McEwen SA. A scoping review of scoping reviews: advancing the approach and enhancing the consistency. Res Synth Methods 2014 Dec;5(4):371-385 [FREE Full text] [doi: 10.1002/jrsm.1123] [Medline: 26052958]
22. Daudt H, van Mossel C, Scott S. Enhancing the scoping study methodology: a large, inter-professional team's experience with Arksey and O'Malley's framework. BMC Med Res Methodol 2013 Mar 23;13:48 [FREE Full text] [doi: 10.1186/1471-2288-13-48] [Medline: 23522333]
23. Librenza-Garcia D, Kotzian BJ, Yang J, Mwangi B, Cao B, Pereira Lima LN, et al. The impact of machine learning techniques in the study of bipolar disorder: a systematic review. Neurosci Biobehav Rev 2017 Sep;80:538-554. [doi: 10.1016/j.neubiorev.2017.07.004] [Medline: 28728937]
24. Monteith S, Glenn T, Geddes J, Whybrow P, Bauer M. Big data for bipolar disorder. Int J Bipolar Disord 2016 Dec 11;4(1):10 [FREE Full text] [doi: 10.1186/s40345-016-0051-7] [Medline: 27068058]
25. Alonso S, de la Torre-Díez I, Hamrioui S, López-Coronado M, Barreno D, Nozaleda L, et al. Data mining algorithms and techniques in mental health: a systematic review. J Med Syst 2018 Jul 21;42(9):161 [FREE Full text] [doi: 10.1007/s10916-018-1018-2] [Medline: 30030644]
26. Calvo RA, Milne DN, Hussain MS, Christensen H. Natural language processing in mental health applications using non-clinical texts. Nat Lang Eng 2017 Jan 30;23(5):649-685. [doi: 10.1017/S1351324916000383]

27. Ongsulee P. Artificial intelligence, machine learning and deep learning. In: Proceedings of the 15th International Conference on ICT and Knowledge Engineering (ICT&KE). 2017 Presented at: 15th International Conference on ICT and Knowledge Engineering (ICT&KE); Nov 22-24, 2017; Bangkok, Thailand. [doi: 10.1109/ictke.2017.8259629]
28. Powers DM. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. Int J Mach Learn Technol 2020:37-63 [FREE Full text] [doi: 10.9735/2229-3981]
29. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32 [FREE Full text] [doi: 10.1080/1364557032000119616]
30. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
31. Peters M, Godfrey C, McInerney P, Munn Z, Tricco A, Khalil H. Chapter 11: scoping reviews. In: JBI Manual for Evidence Synthesis. Adelaide, South Australia: JBI Global; 2020.
32. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
33. Peters MD, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, et al. Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth 2020 Oct;18(10):2119-2126. [doi: 10.11124/JBIES-20-00167] [Medline: 33038124]
34. Benton A, Mitchell M, Hovy D. Multi-task learning for mental health using social media text. In: Proceedings of the 15th Conference of the EACL. 2017 Presented at: The 15th Conference of the EACL; Apr 3-7, 2017; Valencia, Spain URL: https://doi.org/10.48550/arXiv.1712.03538
35. Budenz A, Klassen A, Purtle J, Yom Tov E, Yudell M, Massey P. Mental illness and bipolar disorder on Twitter: implications for stigma and social support. J Ment Health 2020 Apr;29(2):191-199. [doi: 10.1080/09638237.2019.1677878] [Medline: 31694433]
36. Chang C, Saravia E, Chen Y. Subconscious Crowdsourcing: a feasible data collection mechanism for mental disorder detection on social media. In: Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). 2016 Presented at: 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM); Aug 18-21, 2016; San Francisco, CA, USA. [doi: 10.1109/asonam.2016.7752261]
37. Cohan A, Desmet B, Yates A, Soldaini L, MacAvaney S, Goharian N. SMHD: A large-scale resource for exploring online language usage for multiple mental health conditions. In: Proceedings of the 27th International Conference on Computational Linguistics. 2018 Presented at: The 27th International Conference on Computational Linguistics; Aug 20-26, 2018; Santa Fe, New Mexico, USA URL: https://doi.org/10.48550/arXiv.1806.05258
38. Coppersmith G, Dredze M, Harman C. Quantifying mental health signals in Twitter. In: Proceedings of the Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2014 Presented at: The Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; Jun, 2014; Baltimore, Maryland, USA. [doi: 10.3115/v1/w14-3207]
39. Coppersmith G, Dredze M, Harman C, Hollingshead K. From ADHD to SAD: analyzing the language of mental health on twitter through self-reported diagnoses. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2015 Presented at: The 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; Jun 5, 2015; Denver, Colorado. [doi: 10.3115/v1/w15-1201]
40. Gkotsis G, Oellrich A, Hubbard T, Dobson R, Liakata M, Velupillai S, et al. The language of mental health problems in social media. In: Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology. 2016 Presented at: The Third Workshop on Computational Linguistics and Clinical Psychology; Jun 16, 2016; San Diego, CA, USA. [doi: 10.18653/v1/w16-0307]
41. Gkotsis G, Oellrich A, Velupillai S, Liakata M, Hubbard T, Dobson R, et al. Characterisation of mental health conditions in social media using Informed Deep Learning. Sci Rep 2017 Mar 22;7:45141 [FREE Full text] [doi: 10.1038/srep45141] [Medline: 28327593]
42. Huang Y, Chen Y, Alvarado F, Lee S, Wu S, Lai Y, et al. Leveraging linguistic characteristics for bipolar disorder recognition with gender differences. In: Proceedings of the 2019 KDD Workshop on Applied Data Science for Healthcare. 2019 Presented at: DSHealth '19: 2019 KDD Workshop on Applied Data Science for Healthcare; 2019; Alaska URL: https://arxiv.org/abs/1907.07366
43. Ive J, Gkotsis G, Dutta R, Stewart R, Velupillai S. Hierarchical neural model with attention mechanisms for the classification of social media text related to mental health. In: Proceedings of the Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic. 2018 Presented at: The Fifth Workshop on Computational Linguistics and Clinical Psychology: From Keyboard to Clinic; Jun 5, 2018; New Orleans, LA. [doi: 10.18653/v1/w18-0607]
44. Jagfeld G, Lobban F, Rayson P, Jones S. Understanding who uses Reddit: profiling individuals with a self-reported bipolar disorder diagnosis. In: Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access. 2021 Presented at: The Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access; Jun, 2021; Online URL: https://doi.org/10.48550/arXiv.2104.11612 [doi: 10.18653/v1/2021.clpsych-1.1]

45. Tigunova A, Mirza P, Yates A, Weikum G. Listening between the lines: learning personal attributes from conversations. In: Proceedings of the WWW '19: The Web Conference. 2019 Presented at: WWW '19: The Web Conference; May 13 - 17, 2019; San Francisco CA USA. [doi: 10.1145/3308558.3313498]
46. Harrigian K. Geocoding without geotags: a text-based approach for reddit. In: Proceedings of the 2018 EMNLP Workshop W-NUT: The 4th Workshop on Noisy User-generated Text. 2018 Presented at: The 2018 EMNLP Workshop W-NUT: The 4th Workshop on Noisy User-generated Text; Nov 1, 2018; Brussels, Belgium. [doi: 10.18653/v1/w18-6103]
47. Wang Z, Jurgens D. It's going to be okay: measuring access to support in online communities. In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing. 2018 Presented at: The 2018 Conference on Empirical Methods in Natural Language Processing; Oct 31 - Nov 4, 2018; Brussels, Belgium. [doi: 10.18653/v1/d18-1004]
48. Jiang Z, Levitan S, Zomick J, Hirschberg J. Detection of mental health from reddit via deep contextualized representations. In: Proceedings of the 11th International Workshop on Health Text Mining and Information Analysis. 2020 Presented at: The 11th International Workshop on Health Text Mining and Information Analysis; Nov 20, 2020; Online. [doi: 10.18653/v1/2020.louhi-1.16]
49. Kim J, Lee J, Park E, Han J. A deep learning model for detecting mental illness from user content on social media. Sci Rep 2020 Jul 16;10(1):11846 [FREE Full text] [doi: 10.1038/s41598-020-68764-y] [Medline: 32678250]
50. Kramer A, Fussell S, Setlock L. Text analysis as a tool for analyzing conversation in online support groups. In: Proceedings of the CHI 2004 Conference on Human Factors in Computing Systems. 2004 Presented at: CHI04: CHI 2004 Conference on Human Factors in Computing Systems; Apr 24 - 29, 2004; Vienna Austria. [doi: 10.1145/985921.986096]
51. Low DM, Rumker L, Talkar T, Torous J, Cecchi G, Ghosh SS. Natural language processing reveals vulnerable mental health support groups and heightened health anxiety on reddit during COVID-19: observational study. J Med Internet Res 2020 Oct 12;22(10):e22635 [FREE Full text] [doi: 10.2196/22635] [Medline: 32936777]
52. Lyalina S, Percha B, LePendu P, Iyer SV, Altman R, Shah N. Identifying phenotypic signatures of neuropsychiatric disorders from electronic medical records. J Am Med Inform Assoc 2013 Dec;20(e2):e297-e305 [FREE Full text] [doi: 10.1136/amiajnl-2013-001933] [Medline: 23956017]
53. McDonald D, Woodward-Kron R. Member roles and identities in online support groups: perspectives from corpus and systemic functional linguistics. Discourse Commun 2016 Jan 25;10(2):157-175. [doi: 10.1177/1750481315615985]
54. Murarka A, Radhakrishnan B, Ravichandran S. Classification of mental illnesses on social media using RoBERTa. In: Proceedings of the 12th International Workshop on Health Text Mining and Information Analysis. 2021 Presented at: The 12th International Workshop on Health Text Mining and Information Analysis; Apr, 2021; Online URL: https://aclanthology.org/2021.louhi-1.7
55. Park A, Conway M. Harnessing reddit to understand the written-communication challenges experienced by individuals with mental health disorders: analysis of texts from mental health communities. J Med Internet Res 2018 Apr 10;20(4):e121 [FREE Full text] [doi: 10.2196/jmir.8219] [Medline: 29636316]
56. Patel R, Lloyd T, Jackson R, Ball M, Shetty H, Broadbent M, et al. Mood instability is a common feature of mental health disorders and is associated with poor clinical outcomes. BMJ Open 2015 May 21;5(5):e007504 [FREE Full text] [doi: 10.1136/bmjopen-2014-007504] [Medline: 25998036]
57. Ramu N, Kolliakou A, Sanyal J, Patel R, Stewart R. Recorded poor insight as a predictor of service use outcomes: cohort study of patients with first-episode psychosis in a large mental healthcare database. BMJ Open 2019 Jun 12;9(6):e028929 [FREE Full text] [doi: 10.1136/bmjopen-2019-028929] [Medline: 31196905]
58. Rosenstein M, Foltz P, Vaskinn A, Elvevåg B. Practical issues in developing semantic frameworks for the analysis of verbal fluency data: a Norwegian data case study. In: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality. 2015 Presented at: Proceedings of the 2nd Workshop on Computational Linguistics and Clinical Psychology: From Linguistic Signal to Clinical Reality; Jun 5, 2015; Denver, Colorado. [doi: 10.3115/v1/w15-1215]
59. Saha B, Nguyen T, Phung D, Venkatesh S. A framework for classifying online mental health-related communities with an interest in depression. IEEE J Biomed Health Inform 2016 Jul;20(4):1008-1015. [doi: 10.1109/JBHI.2016.2543741] [Medline: 27008680]
60. Saravia E, Chang C, De Lorenzo RJ, Chen Y. MIDAS: mental illness detection and analysis via social media. In: Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). 2016 Presented at: 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM); Aug 18-21, 2016; San Francisco, CA, USA. [doi: 10.1109/asonam.2016.7752434]
61. Sekulić I, Gjurković M, Šnajder J. Not just depressed: bipolar disorder prediction on reddit. In: Proceedings of the The 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. 2018 Presented at: The 9th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis; Oct 31, 2018; Brussels, Belgium. [doi: 10.18653/v1/w18-6211]
62. Sekulić I, Strube M. Adapting deep learning methods for mental health prediction on social media. In: Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019). 2019 Presented at: The 5th Workshop on Noisy User-generated Text (W-NUT 2019); Nov 4, 2019; Hong Kong, China. [doi: 10.18653/v1/d19-5542]

63. Thorstad R, Wolff P. Predicting future mental illness from social media: a big-data approach. Behav Res Methods 2019 Aug 29;51(4):1586-1600. [doi: 10.3758/s13428-019-01235-z] [Medline: 31037606]
64. Tran T, Kavuluru R. Predicting mental conditions based on "history of present illness" in psychiatric notes with deep neural networks. J Biomed Inform 2017 Nov;75S:S138-S148 [FREE Full text] [doi: 10.1016/j.jbi.2017.06.010] [Medline: 28606869]
65. Wu C, Chang C, Robson D, Jackson R, Chen S, Hayes R, et al. Evaluation of smoking status identification using electronic health records and open-text information in a large mental health case register. PLoS One 2013;8(9):e74262 [FREE Full text] [doi: 10.1371/journal.pone.0074262] [Medline: 24069288]
66. Yoo M, Lee S, Ha T. Semantic network analysis for understanding user experiences of bipolar and depressive disorders on Reddit. Inf Process Manag 2019 Jul;56(4):1565-1575 [FREE Full text] [doi: 10.1016/j.ipm.2018.10.001]
67. Better systematic review management. Covidence. URL: https://www.covidence.org [accessed 2022-04-04]
68. Mohammad S. Examining citations of natural language processing literature. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020 Presented at: The 58th Annual Meeting of the Association for Computational Linguistics; Jul, 2020; Online. [doi: 10.18653/v1/2020.acl-main.464]
69. How have Natural Language Processing methods been applied to the study of bipolar? Figshare. URL: https://tinyurl.com/mtkwxuyv [accessed 2022-04-04]
70. Rodgers M, Sowden A, Petticrew M, Arai L, Roberts H, Britten N, et al. Testing methodological guidance on the conduct of narrative synthesis in systematic reviews. Evaluation 2009 Jan 01;15(1):49-73. [doi: 10.1177/1356389008097871]
71. Moher D, Liberati A, Tetzlaff J, Altman D, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
72. Galassi A, Lippi M, Torroni P. Attention in natural language processing. IEEE Trans Neural Netw Learn Syst 2021 Oct;32(10):4291-4308. [doi: 10.1109/TNNLS.2020.3019893] [Medline: 32915750]
73. Pennebaker J, Francis M, Booth R. Linguistic inquiry and word count: LIWC 2001. United States: Mahway: Lawrence Erlbaum Associates; 2001.
74. Devlin J, Chang M, Kenton L, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: Proceedings of the The 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). 2019 Presented at: The 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers); Jun 4, 2019; Minneapolis, Minnesota. [doi: 10.18653/v1/N19-1423]
75. Mikolov T, Chen K, Corrado G, Dean J. Efficient Estimation of Word Representations in Vector Space. Arxiv.org 2013:1-12 [FREE Full text] [doi: 10.48550/arXiv.1301.3781]
76. Pennington J, Socher R, Manning C. GloVe: global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014 Presented at: The 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP); Oct 25–29, 2014; Doha, Qatar. [doi: 10.3115/v1/d14-1162]
77. Barthel M, Stocking G, Holcomb J, Mitchell A. Nearly eight-in-ten Reddit users get news on the site. Pew Research Center. 2016. URL: https://www.pewresearch.org/wp-content/uploads/sites/8/2016/02/PJ_2016.02.25_Reddit_FINAL.pdf [accessed 2022-04-05]
78. Cunningham H. GATE: a general architecture for text engineering. In: Proceedings of the 16th conference on Computational linguistics - Volume 2. 1996 Presented at: COLING '96: The 16th conference on Computational linguistics - Volume 2; Aug 5-9, 1996; Copenhagen Denmark. [doi: 10.3115/993268.993365]
79. Jackson MSc RG, Ball M, Patel R, Hayes RD, Dobson RJ, Stewart R. TextHunter--a user friendly tool for extracting generic concepts from free text in clinical research. AMIA Annu Symp Proc 2014;2014:729-738 [FREE Full text] [Medline: 25954379]
80. Phenotype. National Human Genome Research Institute. URL: https://www.genome.gov/genetics-glossary/Phenotype [accessed 2022-04-05]
81. Mota NB, Furtado R, Maia PP, Copelli M, Ribeiro S. Graph analysis of dream reports is especially informative about psychosis. Sci Rep 2014 Jan 15;4:3691 [FREE Full text] [doi: 10.1038/srep03691] [Medline: 24424108]
82. Dodge J. NLP reproducibility checklist. GitHub. URL: https://jessedodge.github.io/NLP_Reproducibility_Checklist_V1.2.pdf [accessed 2022-04-05]
83. Reinharth J, Braga R, Serper M. Characterization of risk-taking in adults with bipolar spectrum disorders. J Nerv Ment Dis 2017;205(7):580-584. [doi: 10.1097/nmd.0000000000000680]
84. Akers N, Lobban F, Hilton C, Panagaki K, Jones S. Measuring social and occupational functioning of people with bipolar disorder: a systematic review. Clin Psychol Rev 2019 Dec;74:101782 [FREE Full text] [doi: 10.1016/j.cpr.2019.101782] [Medline: 31751878]
85. Dennis S, Garrett P, Yim H, Hamm J, Osth A, Sreekumar V, et al. Privacy versus open science. Behav Res Methods 2019 Aug;51(4):1839-1848 [FREE Full text] [doi: 10.3758/s13428-019-01259-5] [Medline: 31152387]
86. Proferes N, Jones N, Gilbert S, Fiesler C, Zimmer M. Studying reddit: a systematic overview of disciplines, approaches, methods, and ethics. Social Media Soc 2021 May 26;7(2):205630512110190 [FREE Full text] [doi: 10.1177/20563051211019004]

87. Walsh CG, Xia W, Li M, Denny JC, Harris PA, Malin BA. Enabling open-science initiatives in clinical psychology and psychiatry without sacrificing patients' privacy: current practices and future challenges. Advances Methods Pract Psychol Sci 2018 Jan 23;1(1):104-114. [doi: 10.1177/2515245917749652]
88. Hewson C, Buchanan T. Ethics guidelines for internet-mediated research. The British Psychology Society. URL: https://www.bps.org.uk/news-and-policy/ethics-guidelines-internet-mediated-research [accessed 2022-04-05]
89. BPS code of human research ethics. The British Pyschological Society. 2021. URL: https://www.bps.org.uk/sites/bps.org.uk/files/Policy/Policy%20-%20Files/BPS%20Code%20of%20Human%20Research%20Ethics.pdf [accessed 2022-04-05]
90. Friedrich A, Zesch T. A crash course on ethics for natural language processing. In: Proceedings of the Fifth Workshop on Teaching NLP. 2021 Presented at: The Fifth Workshop on Teaching NLP; Jun, 2021; Online. [doi: 10.18653/v1/2021.teachingnlp-1.6]
91. Benton A, Coppersmith G, Dredze M. Ethical research protocols for social media health research. In: Proceedings of the First ACL Workshop on Ethics in Natural Language Processing. 2017 Presented at: The First ACL Workshop on Ethics in Natural Language Processing; Apr 4, 2017; Barcelona, Spain. [doi: 10.18653/v1/w17-1612]

Abbreviations

CRIS: Clinical Record Interactive Search
DSM: Diagnostic and Statistical Manual of Mental Disorders
EHR: electronic health record
GloVe: global vectors for word representation
LIWC: Linguistic Inquiry and Word Count
ML: machine learning
NLP: natural language processing PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PoL: Pattern of Life
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
TF-IDF: Term Frequency Inverse Document Frequency

Edited by J Torous; submitted 22.12.21; peer-reviewed by A Le Glaz, DH Kim-Dufor, C Entwistle; comments to author 10.02.22; revised version received 15.03.22; accepted 20.03.22; published 22.04.22.

Please cite as:
Harvey D, Lobban F, Rayson P, Warner A, Jones S
Natural Language Processing Methods and Bipolar Disorder: Scoping Review
JMIR Ment Health 2022;9(4):e35928
URL: https://mental.jmir.org/2022/4/e35928
doi:10.2196/35928
PMID:35451984

Review

Data Visualization for Chronic Neurological and Mental Health Condition Self-management: Systematic Review of User Perspectives

Ashley Polhemus[1,2], MSE[‡]; Jan Novak[3], MS; Shazmin Majid[3,4], MS; Sara Simblett[5], PhD; Daniel Morris[5], BSc; Stuart Bruce[6], BSc; Patrick Burke[6], BA, BSc; Marissa F Dockendorf[7], PhD; Gergely Temesi[8], PhD; Til Wykes[5,9], PhD

[1]Merck Research Labs, Information Technology, Merck, Sharpe, & Dohme, Zurich, Switzerland

[2]Epidemiology, Biostatistics and Prevention Institute, University of Zürich, Zürich, Switzerland

[3]Merck Research Labs, Information Technology, Merck, Sharpe, & Dohme, Prague, Czech Republic

[4]School of Computer Science, University of Nottingham, Nottingham, United Kingdom

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

[6]RADAR-CNS Patient Advisory Board, London, United Kingdom

[7]Global Digital Analytics and Technologies, Merck, Sharpe, & Dohme, Kenilworth, NJ, United States

[8]Merck Research Labs, Information Technology, Merck, Sharpe, & Dohme, Vienna, Austria

[9]National Institute for Health Research, Maudsley Biomedical Research Centre, South London and Maudsley National Health Services Foundation Trust, London, United Kingdom

[‡]the RADAR-CNS consortium

Corresponding Author:
Ashley Polhemus, MSE
Merck Research Labs, Information Technology
Merck, Sharpe, & Dohme
The Circle 66
Zurich, 8058
Switzerland
Phone: 41 764922388
Email: ashley.polhemus@uzh.ch

Abstract

Background: Remote measurement technologies (RMT) such as mobile health devices and apps are increasingly used by those living with chronic neurological and mental health conditions. RMT enables real-world data collection and regular feedback, providing users with insights about their own conditions. Data visualizations are an integral part of RMT, although little is known about visualization design preferences from the perspectives of those living with chronic conditions.

Objective: The aim of this review was to explore the experiences and preferences of individuals with chronic neurological and mental health conditions on data visualizations derived from RMT to manage health.

Methods: In this systematic review, we searched peer-reviewed literature and conference proceedings (PubMed, IEEE Xplore, EMBASE, Web of Science, Association for Computing Machinery Computer-Human Interface proceedings, and the Cochrane Library) for original papers published between January 2007 and September 2021 that reported perspectives on data visualization of people living with chronic neurological and mental health conditions. Two reviewers independently screened each abstract and full-text article, with disagreements resolved through discussion. Studies were critically appraised, and extracted data underwent thematic synthesis.

Results: We identified 35 eligible publications from 31 studies representing 12 conditions. Coded data coalesced into 3 themes: desire for data visualization, impact of visualizations on condition management, and visualization design considerations. Data visualizations were viewed as an integral part of users' experiences with RMT, impacting satisfaction and engagement. However, user preferences were diverse and often conflicting both between and within conditions.

Conclusions: When used effectively, data visualizations are valuable, engaging components of RMT. They can provide structure and insight, allowing individuals to manage their own health more effectively. However, visualizations are not "one-size-fits-all,"

and it is important to engage with potential users during visualization design to understand when, how, and with whom the visualizations will be used to manage health.

(JMIR Ment Health 2022;9(4):e25249) doi:10.2196/25249

KEYWORDS

digital health; remote measurement technology; neurology; mental health; data visualization; user-centered design

Introduction

Despite widespread interest in the use of remote measurement technology (RMT) such as wearable devices and mobile apps in health care settings, the design and practical implementation of these technologies remain challenging. Designing RMT requires a careful balance between (often conflicting) requirements, which address technological, clinical, and regulatory specifications, individual users' needs, and health-management goals. The result must be effective and engaging for users and capable of supplementing health management practices and promoting long-term adherence [1-3]. However, best practices and guidelines for RMT design are sparse and most designs rely heavily on designers' interpretations or inferences of user preferences [4]. Emphasis on user engagement during RMT design has increased in recent years but is not yet a widespread practice [5].

Although determinants of engagement with RMT are diverse and condition-dependent, *access to and interaction with data* frequently emerge as a key factor influencing user motivation and satisfaction with RMT [6,7]. However, data access alone is insufficient to achieve goals like self-management. Data must be organized and presented in ways that meaningfully address questions of self-awareness and self-care [8]. To paraphrase Few [9], a good visualization must clearly indicate relationships, accurately represent the data, enable easy comparisons, clearly show scales and ordering of the data, and encourage people to use the presented information. This is not trivial, as perceptions of whether visualizations are accurate, clear, or easily interpretable are subject to personal mindset and circumstance [8]. Design of good visualizations for health management requires careful consideration of user perspectives and needs throughout the design process, ideally engaging these users through participatory design methods [10,11]. Recent research on health data visualization tends to focus on the needs of health care professionals, typically with regard to electronic medical records and other novel sources of big health data [12-14]. This paradigm is ripe for change. The abundance and accessibility of health data, driven in large part by RMT, provide individuals with unprecedented resources to aid in condition self-management.

Although theories and techniques exist to guide the visualization design process [15-17], these techniques require a fundamental understanding of the purposes for and contexts in which the desired visualizations will be used. Unfortunately, functional guidance describing service-user perspectives and preferences in RMT data visualization is scarce. A recent systematic review explored types of visualizations shown to patients in health research, predominantly in nondigital formats in primary care settings [18]. However, it did not discuss individuals' perspectives on these visualizations. It is still unclear *which* data RMT users wish to access, *how* these data should be visualized, *for what purposes* these visualizations are used, and *which factors moderate* these preferences. The aim of this qualitative systematic review was to identify and synthesize existing studies that report the perspectives of individuals with neurological and mental health conditions on RMT data visualizations. Based on identified themes and gaps in the literature, we suggest both priorities for future research and design considerations for RMT visualizations, which could enrich current service-user engagement practices.

Methods

Study Scope and Research Question

This study adhered to PRISMA-P (Preferred Reporting Items for Systematic review and Meta-Analysis Protocols) guidelines for systematic review conduct and reporting [19,20]. We aimed to address the question, "What are the data visualization preferences and perceptions of people living with chronic neurological and mental health conditions when using RMT to manage health?" The project's patient advisory board, which comprises patient advocates for depression, epilepsy, and multiple sclerosis, was consulted during the study design and data analysis.

Identifying Relevant Studies

This protocol was registered on PROSPERO (International Prospective Register of Systematic Reviews, CRD42019139319) while the review was in its pilot phase [21]. We searched PubMed, IEEE Xplore, EMBASE, Web of Science, proceedings from the Association for Computing Machinery Conference on Human Factors in Computing Systems, and the Cochrane Library for original, peer-reviewed, or gray literature published in English between January 2007 and September 2021. Searches included combinations of terms such as mHealth, along with terms related to data visualization and neurological disease. The search strings are provided in Multimedia Appendix 1. Relevant papers were also identified from manual searches of included studies' reference lists. Studies were screened in 2 stages: abstract screening and full-text review. Eligibility criteria and screening forms were piloted on a set of 50 abstracts and 15 full-text reviews, and criteria were clarified or amended as needed. Two reviewers (AP, JN, SM, or DM) independently screened each abstract. In the case of disagreement, the abstract automatically proceeded to the full-text stage. Two reviewers then independently assessed each full-text paper for eligibility, and disagreements were resolved through discussion. If no consensus could be reached, a third member of the review team reviewed the paper and made a final determination. Agreement between reviewer pairs was determined through Cohen kappa

[22]. Deduplication, record management, and screening were conducted in CADIMA, an open-access systematic review software [23]. Data extraction and coding were conducted using custom forms developed in Microsoft Word (Multimedia Appendix 1).

Eligibility Criteria

We included studies if they met the following criteria:

- All or part of a study population was living with a neurological or mental health condition
- Participants were ≥18 years of age
- RMT for laypeople to track, monitor, or manage their own health was investigated
- Results of any qualitative methods or integrated syntheses of mixed methods were reported
- Patient perspectives on visualizations of health or wellness data were reported

The following studies were ineligible for inclusion:

- Conditions that were not neurological in nature or not associated with mental health
- Perspectives on interface design, intervention design, or any component of RMT design unrelated to data visualization
- Visualizations limited to medication adherence or non–health-related data
- Perspectives of caregivers, health professionals, or others not living with a neurological or mental health condition

We adhered to Davis et al's [24] definition of RMT, which includes "any technology that enables monitoring of a person's health status through a remote interface," which can then either be transmitted to a health care provider or as a tool for self-management of one's health. We purposefully remained broad in our definition of the term "data visualization" because it is understood differently by different people. Therefore, we included any format through which RMT displayed data to service users. We defined chronic neurological or mental health conditions as any long-term, progressive, relapsing, or recurrent conditions related to mental health or dysfunction of the nervous system. Neurodegenerative diseases, depression, anxiety, or bipolar disorder, pain disorders, and sleep disorders, among others were eligible. Mental health conditions were also eligible if they were symptoms or comorbidities of a nonincluded condition. This heterogeneous set of conditions allowed us to identify themes that may be generalizable across conditions and others that are disease-specific.

Data Extraction, Critical Appraisal, and Qualitative Synthesis

Two authors independently reread each included study and extracted quotes related to data visualization preferences. When available, screenshots of data visualizations were also extracted. To ensure that analysis remained grounded in the context of the original studies, data extraction forms included a detailed description of each study's objectives and methods, and annotated PDFs were preserved. Studies were critically appraised with the Mixed Methods Appraisal Tool [25,26]. We then categorized studies as conceptually rich "key papers," "satisfactory papers," which are methodologically acceptable but provide only moderate value to the synthesis, and "fatally flawed papers," which contain major methodological flaws [27,28]. We also noted "minimal impact papers," which provided minimal contribution to the synthesis.

We employed Thomas and Harden's [29] inductive approach to thematic synthesis. Following a reading of the extracted text and its context, 2 authors (AP and JN) independently coded data line by line, producing a draft coding frame. The coding frame was iteratively amended, refined, restructured and the data recoded until no additional codes or disagreements were identified, and categorized into "descriptive themes," which described the structure and content of the codes. Analytical themes, which interpreted the coded data, were developed through iterative rereading and discussion of the codes and thematically organized data (Table 1). An experienced qualitative researcher (SS) oversaw and provided input on this process. Following thematic synthesis, we conducted sensitivity analyses to identify potential differences between conditions. We analyzed 4 subgroups (mental health, neurological conditions, sleep, and pain) that arose from the patient populations and lines of inquiry addressed by included studies. As a form of member-checking, we consulted members of the patient advisory board (authors PB and SB) who reviewed our analysis and interpretations for face validity within the context of their own experiences of RMT and condition self-management.

Table 1. Worked examples of the data synthesis process.

Extracted text	Coding	Descriptive theme	Analytical theme
Participants reported that the mood monitoring surveys and associated graphical feedback were a reason to return to the app and that it increased self-awareness of how their mood fluctuated over time and in relation to use of the intervention content [30].	• Form: Graphical • Self-awareness • See progress	Increased self-awareness	Visualizations enable proactive self-management through improved self-awareness
…It was really nice to look at the circle plot [the pie-chart]. When I've had a hard day, I looked back on the previous day, and saw a big yellow portion [Work and Education] and then it made sense to me why I felt bad today [31].	• Form: Graphical • Identify patterns	Increased self-awareness	Visualizations enable proactive self-management through improved self-awareness

Results

Included studies

Searches returned 2928 unique records. Of these, 177 papers were included in full-text review and 35 were eligible for qualitative synthesis (Figure 1). Reviewer agreement was moderate during abstract screening (weighted κ=0.45) and substantial during full-text review (weighted κ=0.79) [22]. Our relatively low agreement during abstract screening was expected and mitigated by reviewing the full texts of all abstracts, which were judged eligible by at least one reviewer. Multiple papers were identified for 3 research projects: the SPIRIT study (n=3) [32-34], the MoodRhythm app (n=2) [35,36], and the MONARCA (MONitoring, treAtment and pRediCtion of bipolAr Disorder Episodes) project (n=2) [37,38]. For these projects, all identified papers were analyzed as 1 study. Thus, 31 unique studies were included. The characteristics of the included studies are summarized in Table 2 and provided in detail in Multimedia Appendix 2. Most studies (19/31, 61%) addressed mental health conditions. Only 2 studies specifically investigated user perspectives on data visualizations [39,40]. All others reported on data visualization preferences within the broader context of RMT design, evaluation, or usability. We categorized 9 publications as key papers, 16 as satisfactory papers, 6 as minimal impact, and 4 as fatally flawed. We identified 3 themes through content analysis: desire for data visualization, impact of visualizations on condition management, and visualization design considerations (Textbox 1). The final coding frame and illustrative quotes are provided in Multimedia Appendix 2.

Figure 1. PRISMA (Preferred Reporting Items for Systematic review and Meta-Analysis) flow diagram of study screening and selection. RMT: remote measurement technology.

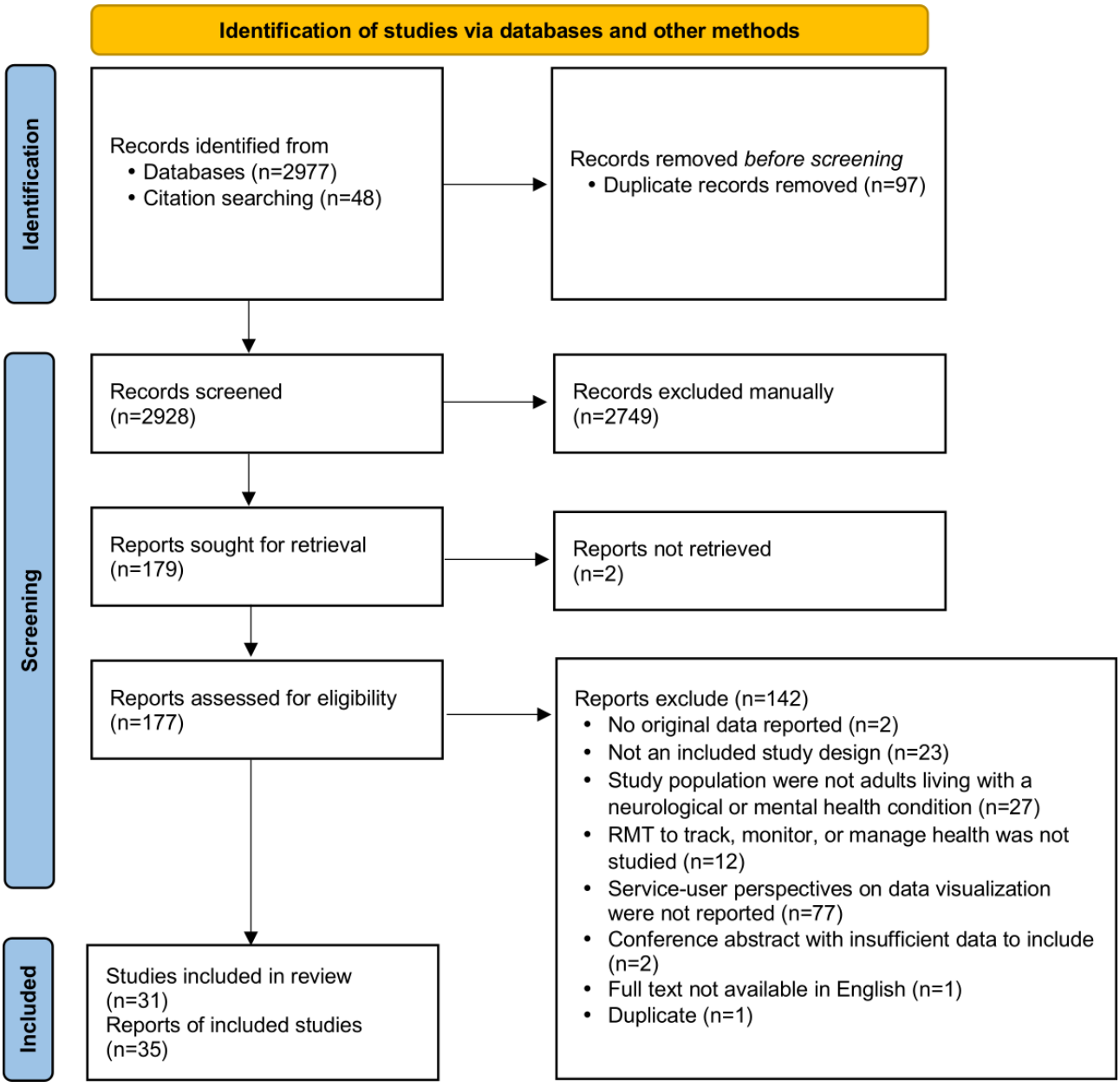


Table 2. Characteristics of the included studies (N=31).

Characteristics	Values, n (%)[a]
Condition	
Bipolar disorder	6 (19)
Depression	6 (19)
Sleep disorders	2 (6)
Schizophrenia	2 (6)
Anxiety	2 (6)
Pain disorders	2 (6)
Posttraumatic stress disorder	2 (6)
Mixed mental health conditions	6 (19)
Parkinson disease	2 (6)
Multiple sclerosis	3 (10)
Motor neuron disease	1 (3)
Epilepsy	1 (3)
Study design	
Qualitative feedback on RMT[b] from a field study	15 (48)
Co-design of an RMT	10 (32)
Qualitative user acceptance testing of an RMT	11 (35)
Exploratory qualitative studies not associated with a specific RMT	10 (32)
Qualitative studies specifically on data visualization preferences	2 (6)
Qualitative methods	
Mixed methods	12 (39)
Interviews	17 (71)
Open-ended surveys	4 (13)
Focus groups	4 (23)
Other	3 (10)

[a]As several projects included multiple conditions, study designs, or qualitative methods, the total exceeds 100%.

[b]RMT: remote measurement technology.

Textbox 1. Identified themes and subthemes.

Theme 1. Desire for data visualization

Theme 2. Impact of visualizations on condition management

- Visualizations enable proactive self-management through improved self-awareness
- Visualizations enable more effective communication with care partners
- Visualizations drive engagement with remote measurement technology

Theme 3. Visualization design considerations

- Format
- Context and Annotation
- Customization
- Moderators of visualization design preferences

Desire for Data Visualization

Users explicitly expressed a desire for data visualizations in 21 of the 31 included studies. Discussed or desired data included mood and disease-specific symptom scores, physical activity, sleep patterns and quality, and rhythm of daily activities. In 6 studies, RMT did not originally include visualizations of users' historic data. Participants expressed their dissatisfaction, prompting authors to add visualizations to their designs [32-35,38,41,42].

Impact of Visualizations on Condition Management

Visualizations Enable Proactive Self-management Through Improved Self-awareness

Participants reported turning to RMT and data visualization tools when they required external structure or organization to manage their conditions. This structure improved users' recall of their own experiences, which were perceived as valuable owing to the difficulty of accurately reflecting on past symptoms [30,31,35,37,41,43]. Objective visualizations of past symptoms also afforded them a sense of validation, making current experiences feel more real or trustworthy [35,41,44]. This was especially important in bipolar disorder or schizophrenia, when symptoms could distort one's perception of reality [35,41].

Graphs and charts enabled users to identify patterns [30,31,34-38,41-55] such as activities that prompted feelings of wellness [31,52,55] or triggered symptoms [31,35,41,46,56]. Thus, visualization of historic data heightened service users' self-awareness, leading them to engage in more proactive self-management [31,35-37,41,46,50,53,55]. Once trends were identified, users could use their newly acquired knowledge to avoid or pre-empt triggers, reducing the severity of oncoming episodes [46]. This awareness gave users a sense of accountability and level of control over their conditions [46,50] and helped them use the management tools at their disposal more effectively [41,50,52]. Some saw timely access to data as essential to proactive self-management, especially when identifying recent triggers or activities that may have affected their current condition [35,36]. Over time and despite fluctuations, visualizations helped users objectively see their own progress, the magnitude of which they may not have fully appreciated otherwise [30,35,41,52,57]. Illustrative quotes are provided in Table 3.

Table 3. Illustrative quotes for visualizations improve self-awareness and enable proactive self-management.

Theme	Illustrative quotes
Provide structure and organization	*…Still, respondents reported turning to external tracking via paper or technology when they need extra support, for instance, when thoughts get "scrambled" or their mind feels too "full" - sensations that are especially common during [bipolar disorder] episodes* [43].
Improve recall of past experiences	*…What I saw [in the trial] is that it helped me keep on track. I try to keep track of the triggers [early warning signs], and my history - and in that way it has helped me enormously. Previously, I went into periods where I encountered random mood swings, up and down, and I did not have any history [data] to relate to, so it kind of surprised me. But now I can actually follow how I'm doing - also back in time - and what caused it. It has really been great, and I think I have been able to keep track of myself* [37] (bipolar disorder).
Validate current experiences	*…I am looking for confirmation that I had similar symptoms in the past, because sometimes due to the nature of bipolar I feel like I can't trust the emotions I have at any given moment (or their possible triggers) and it is a relief to know that these are patterns* [35] (bipolar disorder).
Increase self-awareness, identify patterns	*…I found the most valuable tool to analyze my activities. It provides an understanding of which activities helps me, and which gives problems that I need to be aware of—or completely avoid* [31] (depression).
Enable proactive self-management	*…A majority of respondents described this awareness as an opportunity to become proactive about their condition, helping them make adjustments to preempt mood episode triggers and maintain stability or at least avoid severe episodes. Survey respondents also stated that the feedback provided by tracking keeps them accountable to themselves and that they find visual forms of feedback particularly helpful for identifying personal behavioral and emotional patterns and motivating positive lifestyle choices* [43] (bipolar disorder).
See progress over time	*…I was AMAZED when I scrolled back through the Android weeks to see how much my mood has stabilized since I started [medication]. The weeks themselves weren't as meaningful as the pattern over time* [35] (bipolar disorder).

Visualizations Enable More Effective Communication With Care Partners

Participants also frequently reported using or desiring to use RMT data visualizations to communicate with health care professionals, caregivers, and others, regardless of whether the technology was designed for this purpose [39,41,43-46,50-55,58,59]. Eisner et al [41] described "participants theorized that having access to objective data representing their symptoms, particularly in the graph form, might enable a shared understanding of their experiences, both with the care team…and potentially with the general public."

Participants in several studies reported using visualizations from their self-tracking apps to foster higher-quality dialogue with their care team [39,41,43,45,46,50-52,58,60]. Visualizations were advantageous because they improved recall of past symptoms and relayed clinically relevant patterns that users found difficult to describe [43,46,50,52]. Visual aids also helped patients make the most out of short and infrequent appointments, especially when current health status was not representative of the patient's experiences over previous weeks or months [46,50,55]. Sometimes, these visualizations even served as the basis for defensible positions when broaching difficult topics with their care team [41,46]. For example, 1 participant who experienced difficulties getting a care team to take his concerns

seriously stated, "If you were to answer the questions and go to the doctor and say 'look, these are my results, you can see clearly there's a change, and these are my experiences,' that would be substantial evidence for the doctor to then sit up and take note" [41].

Visualizations Drive Engagement With RMT

Participants also described how visualizations made experiences with RMT more engaging [30,31,34,35,41,42,49,51]. Users with bipolar disorder and insomnia suggested that data access and visualization could entice users to engage more with the RMT, since data and insights could be perceived as rewards [35,49]. However, those with depression, Parkinson disease, and multiple sclerosis warned that inappropriately designed visualizations could be demoralizing [32,34,35,47,57,61], which could lead to disengagement. When visualizations were unavailable, users tended to find RMT unengaging and unmotivating. In 2 studies, the lack of personalized graphs in early prototypes was thought to contribute to study dropout [34,42]. Qualitative feedback in one of these studies, which focused on narcolepsy management, led the designers to conclude, "to keep patients motivated to use the tool over a longer period, a personal visualization of recorded data is required" [42].

Visualization Design Considerations

Participants often reflected on aspects of visualization design, such as format, the need for contextual information, timeliness, and customization (Table 4). Health status, data literacy, and previous experience with RMT appeared to moderate individuals' needs and design preferences.

Table 4. Illustrative quotes on visualization design considerations.

Theme	Illustrative quotes
Visualization design	*…Similar to other personal health systems that use fishes and flowers as metaphors, we were looking for an appropriate metaphor for bipolar disorder. Many attempts were tried, including using metaphors like a scale, an equalizer, a river, a volcano, a dart board, and a radar, but we always had the case that some patients preferred one visualization, and others hated it* [37] (bipolar disorder).
Context and annotation	*…I would like to click [points on the daily mood graph] and see what I did on this day, since I was so well* [31] (depression or anxiety disorder). *…I kinda wish I could put a little note in and be like, "This is why I put this number." I think, yeah. I guess that's actually - it would have been really nice to have some sort of journaly type feature where I could make notes like that* [34] (bipolar disorder).
Timeliness of data access	*Comments mostly differed regarding the frequency of (feedback via data visualizations), again with a great range from daily feedback to one response per month:* *…Yes, one feedback per month. Or maybe every two weeks. In the case of warning signals also more frequently.* *…Yes a short feedback every two days or every day.* *…Yes absolutely. Once a week would be good or every two weeks. Not every day though* [58] (bipolar disorder).
Customization	*…Survey respondents' approaches for tracking multiple indicators vary. Participants were divided between keeping separate journals or tools, each dedicated to chronicling a particular indicator, or tracking all items with a single chart or application. Sometimes, elaborate tracking setups are reported as necessary to accommodate such tracking habits in ways technologies do not currently support* [46] (depression).

Format

Studies described a variety of data visualization formats, including line, bar, and pie charts [45,61], calendar views [31], scales [62], mood clouds [51], traffic lights [63], and overlays on maps [61]. Several studies reported that graphical representations were preferred by most participants [30,51,64], though personal preferences varied [37,40,49]. Some users preferred to aggregate multiple data streams in a single visualization [31,45], while others preferred simple, nongraphical formats such as scales or textual descriptions [61,62]. Participants suggested that images and color were powerful tools to make data visualizations more meaningful and engaging [31,35,45,51,61]. However, these tools could also trigger emotional responses that affected users' self-image [31-33,35,37,41,47,49,61].

Context and Annotation

Users also discussed the importance of contextual information when interpreting visualized data [31,32,34,35,42,49,50,58]. They often wished to annotate their data, thus providing "internal" context in the form of notes alongside a numeric score [31,32,34,42,58]. When used in conjunction with a graphical visualization, these annotations could help users generate insights, which were not evident from numeric data alone. Such internal context was also seen as valuable when using visualizations to communicate with health care providers since it augmented users' memories of past experiences [32,34,42,58]. Often, external context explaining symptom scores or condition-specific concepts were also required to help users interpret visualizations. This was especially important when apps were intended for independent use, outside a health care professional's oversight [32,40,55]. Some participants wanted semipersonalized feedback based on their recent scores [34,35,49,50], which could be as simple as an occasional reassurance that the users' data were normal [50] or as complex as personalized health management advice [34,42,46].

Timeliness

Six studies [35,36,45,55,58,63] discussed the time frame for data access. Participants valued regular data access because it provided them with confidence and insight as they worked to

manage their conditions. Preferences varied, including real-time, daily, or monthly feedback. This variety was observed even within relatively uniform user populations. Several studies concluded that it was best to incorporate flexibility into their designs, enabling each user to choose the time frame and amount of data to visualize.

Customization

Participants often expressed conflicting visualization design preferences [31,38,45,46,48,49,51,61,62], and authors noted this as a design challenge. Unsurprisingly, the ability to customize data visualizations was highly desirable [30,32-34,45,46,51,64], and personalization was seen as key to increase long-term engagement with RMT [37]. This included the ability to select which data to visualize [32,34,37,42,46], how to visualize it [37,46,51], and how often or rapidly to access it [35,36,45,58,63]. Some users even wished to add their own personally relevant data streams to RMT visualizations [42]. Often, users desired flexibility to manipulate and compare data streams, allowing exploration of one's own data and generation of insights [34,45,46].

Moderators of User Needs and Design Preferences

Health Status

Participants indicated that visualizations must be meaningful and sensitive to users' personal experiences of their health conditions and comorbidities. Otherwise, they may unintentionally provide negative feedback. For example, one study on people with Parkinson disease [61] visualized walking activity as progress across a map of France. However, slow progress was interpreted as negative feedback within the context of the users' mobility impairments [61]. Participants with multiple sclerosis recommended that visualizations should be designed to emphasize an individual's progress against their own goals rather than an uncontextualized or absolute score [47]. When disease progression is tracked over the long term, it should be visualized in a positive light and contextualized within the resources, management strategies, and self-care activities that individuals can continue to use to manage their conditions [57]. Use of bright color was frequently seen as desirable and engaging in day-to-day health management [31,35,51], but muted colors were perceived as preferable when someone was acutely depressed or struggling to maintain a routine, since harsh colors (ie, red) emphasized lack of progress [32,33,35,51,61]. For conditions that caused vision impairment, such as multiple sclerosis, legibility and color contrast were considered important for visualizations [52,57]. Visualizations must also account for possible changes in health status due to disease progression or episodic fluctuation, which may alter user needs or functional abilities [41,42,46,47,62].

Data Literacy and "Data People"

Participants often referred to themselves in terms of whether they were "data people" when describing their desired level of complexity in visualizations. Data people tended to want more control over their visualizations to allow self-exploration of their symptoms and trends [45,46]. However, those who did not identify as data people preferred simpler visualizations. These individuals also tended to question the accuracy of purely numeric information [40,41]. This further highlights the need for flexibility, allowing users to choose how and how much data are displayed on a single visualization [64].

Experience With Self-monitoring

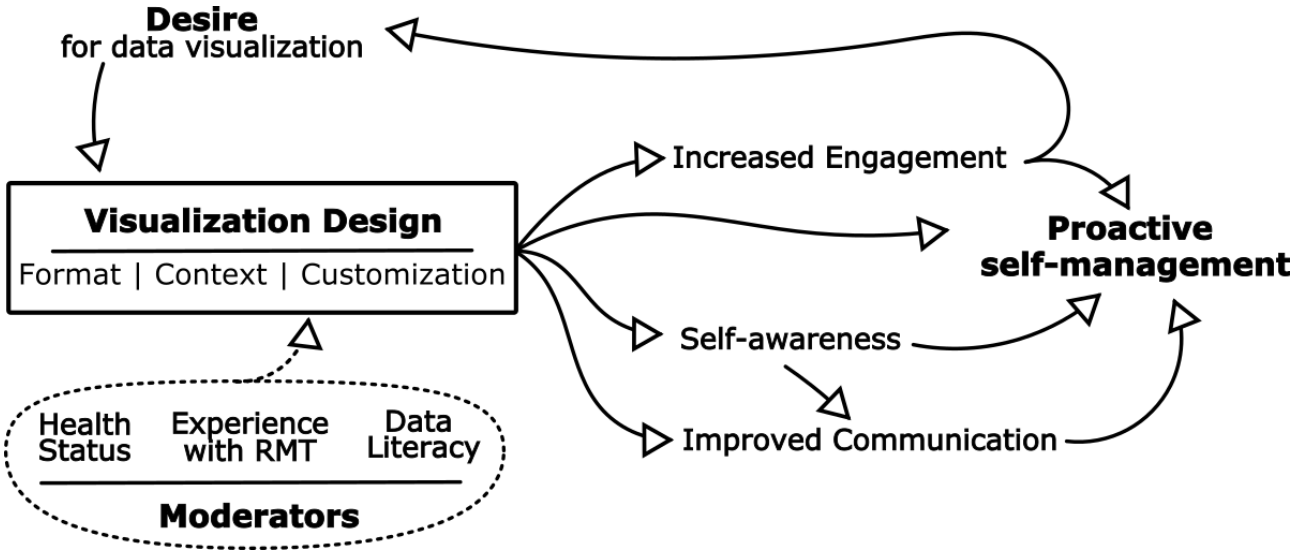
Some users described that their visualization preferences could evolve when self-monitoring was practiced regularly [34,35,40,46]. Users described developing their skills with just a few variables at first and then moving on to complex customized visualizations with many data streams [34,35,40,46]. Once they were accustomed to tracking many variables, self-monitoring became a mental task that no longer required RMT or visualizations. However, users reported returning to external tracking methods during acute phases or when something disrupted their routines. Those who had extensive experience with RMT also tended to request more customizable features such as selecting data streams to display or requesting personalized advice [34,46]. If customization is not available, there is a risk of disengagement with the RMT as users' management practices evolve over time [46].

Discussion

Principal Findings

When designed appropriately, data visualizations are perceived as valuable, engaging components of RMT. They can be validating and empowering, allowing users to become more active, responsible participants in their own care. Individual experiences of health conditions were perceived as highly personal, leading to requests for contextual aids, flexibility, and personalization. Factors such as health status, data literacy, and past experiences affected user needs, thereby moderating design preferences. Figure 2 describes the themes and relationships identified in this study.

Figure 2. Data visualization as an integral personal component of health management. RMT: remote measurement technology.



Ample literature discusses scientifically derived best practices for app interface design, engagement strategies, and intervention design in various medical conditions and contexts [65-67]. Many of these modern practices emphasize user engagement, co-design, and human-centric design methods with an emphasis on scientific rigor [5,66,68]. In fact, many included studies reported on app development processes that adhered to these principles. Unfortunately, specific literature on visualization design within the context of RMT-enabled health management is sparse. This review aimed to address this gap. Based on our synthesis alone, we cannot suggest concrete examples of good design or universal design preferences. On the contrary, this review highlighted the diversity of preferences both across and within conditions and study samples. However, with this in mind, we pose several recommendations for RMT visualization design.

First, visualizations of personal data were consistently considered valuable, engaging, and even necessary components of RMT. When designed appropriately, they can be an integral part of condition self-management. Therefore, incorporation of data visualizations should be considered when designing RMT. However, designs should be sensitive to the experiences and health statuses of their target audiences. Participants were regularly concerned that visualizations that inadvertently emphasized slow progress, poor health, or unattained goals could prompt disengagement with RMT during periods of relapse or perceptible disease progression [32-34,42]. These warnings are consistent with the findings of Lee et al [69] who suggest that negative emotions and subsequent reductions in health self-efficacy reduce health information–seeking behavior [69].

Designs should be informed by a deep understanding of the service users' needs and experiences. For example, study participants with rapid-cycling bipolar disorder suggested that tracking apps and visualizations were not useful to them since most apps allowed only 1 data entry per day [42,46]. Such mismatches between user needs and design features are avoidable if sufficient time is taken *early in the design process* to truly empathize with service users [5]. Designers should work with members of the target audience to understand their lived experiences, needs, and challenges. It is critical to engage with potential users during the design phase to understand *when, how, and with whom* data visualizations will be used. Fit-for-purpose visualizations can then be designed with these factors in mind.

Included studies repeatedly suggested that visualizations are not "one-size-fits-all." Symptoms, triggers, contextual factors, and tracking needs differ from person to person, context to context, and timepoint to timepoint. As such, individual data visualization preferences are highly variable and often dynamic. This is consistent with Paterson's [70] "shifting perspectives model of chronic illness," which describes how individuals' perceptions of their conditions, symptoms, and health fluctuate between predominant feelings of wellness and illness over time. Sometimes, individuals may predominantly self-identify as well. In a health-tracking context, this may lead service users to track symptoms less frequently or focus on wellness data (eg, physical activity, sleep) [31]. During such times, reminders of feelings of illness prompted by symptom tracking may even be detrimental to the service user's self-image. However, when feelings of control over health status is threatened, such as during a relapse, individuals tend to shift their perspectives and more readily identify as ill [70]. During these times, individuals may be more inclined to self-track their symptoms to identify trends and regain control [46,50]. Owing to this interpersonal and intrapersonal variability, flexible design is highly desirable when it is feasible to implement. Designers should remain mindful of sources of heterogeneity in their target audience's needs and subsequently identify RMT features that are most likely to require flexibility or personalization.

We also identified areas where tensions may emerge between user preferences and those of other stakeholders. Although this study focused on RMT designed for individual use, participants, especially those with mental health conditions, perceived data visualizations as tools to help communicate experiences to their care partners. However, clinicians often find it difficult to interpret and manage the visualizations derived from RMT [71], especially if visualizations are not standardized across the

various apps used by patients. If visualizations are to be used as communication tools, they must be interpretable, meaningful, and actionable by those involved in patient care [72]. Otherwise, confusion or even miscommunication may ensue, potentially wasting valuable face-to-face time during short appointments with health care professionals [73]. Therefore, engagement with other possible users such as caregivers or health care professionals may also be valuable during visualization design.

Limitations

As with all qualitative metasyntheses, the themes identified here are influenced by the topics, context, and limitations of the included studies. All results reported here must be interpreted as third-order constructs [74] or interpretations of interpretations of an individual's original feedback. Although this is the aim of a meta-synthesis, it does pose limitations to remaining grounded in the authors' and participants' original meanings and contexts. As reviewers, we attempted to minimize our assumptions during coding and synthesis by continually questioning each other's interpretations of the data. However, we acknowledge that is impossible to fully isolate our synthesis from our own experiences and worldviews. Study methodology was heterogeneous and the authors' place in the data was not always clear. Most comments on visualizations were brief and conducted as part of a larger studies, yielding relatively shallow data. Few studies compared preferences between multiple visualization designs, limiting our ability to compare the merits or relative preferences between design options. Most studies used user experience and co-design methods, which generally did not meet traditional expectations of qualitative research rigor. Although we opted for broad search strategies and sensitive review methods to identify all relevant studies, data visualization was almost always a secondary topic in the included papers and was not always clearly discussed. It is therefore a possibility that our strategies did not identify all eligible papers. Those papers that were included predominantly reflected the perspectives of individuals with mental health conditions, and it is unclear how these results may translate to other populations.

Future Work

Few studies directly addressed research questions related to data visualization preferences. Rather, data visualization was usually approached as peripheral piece of a larger project. Therefore, this review should serve as a starting point for targeted work on health data visualization design. Additional qualitative research on service-user data visualization preferences is warranted, especially within the context of neurological condition self-management. This research should involve individuals directly through interviews, focus groups, or other qualitative methods and should focus on generating "thick" qualitative data that describe more specific uses and contexts. This work should verify and build upon the concepts developed here. Ideally, this work should span multiple conditions and generate diverse perspectives. These perspectives should then be used to generate more detailed design considerations and recommendations for RMT visualizations.

Conclusions

When designed appropriately, data visualizations are valuable engaging components of RMT. They can provide structure and insight, allowing individuals, especially those with mental health conditions, to manage their own health more effectively. However, visualizations are not "one-size-fits-all," and it is important to engage with potential users during visualization design to understand when, how, and with whom visualizations will be used to manage health. The considerations presented here should serve as the basis for future designs and discussions to ensure that visualizations address users' needs and preferences.

Acknowledgments

This review was conducted as part of the Remote Assessment of Disease and Relapse-Central Nervous System (RADAR-CNS) research program. The RADAR-CNS project has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement 115902 [75]. This joint undertaking receives support from the European Union's Horizon 2020 research and innovation program and European Federation of Pharmaceutical Industries and Associations (EFPIA). This communication reflects the views of the RADAR-CNS consortium, and neither Innovative Medicines Initiative nor the European Union and EFPIA are liable for any use that may be made of the information contained herein. This paper represents independent research part funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley National Health Service (NHS) Foundation Trust and King's College London. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care. TW is a National Institute of Health Research senior investigator.

Conflicts of Interest

AP, JN, SM, MFD, and GT are employees of Merck, Sharp, and Dohme, Inc, and may own stock or stock options.

Multimedia Appendix 1
Search strategy and data extraction form.
[PDF File (Adobe PDF File), 100 KB - mental_v9i4e25249_app1.pdf]

Multimedia Appendix 2

Characteristics of the included studies, critical appraisal, coding, and illustrative quotes.
[PDF File (Adobe PDF File), 258 KB - mental_v9i4e25249_app2.pdf]

References

1. Gurupur VP, Wan TTH. Challenges in implementing mHealth interventions: a technical perspective. Mhealth 2017;3:32 [FREE Full text] [doi: 10.21037/mhealth.2017.07.05] [Medline: 28894742]
2. Mohammadzadeh N, Safdari R. Patient monitoring in mobile health: opportunities and challenges. Med Arch 2014;68(1):57-60 [FREE Full text] [doi: 10.5455/medarh.2014.68.57-60] [Medline: 24783916]
3. Nouri R, R Niakan Kalhori S, Ghazisaeedi M, Marchand G, Yasini M. Criteria for assessing the quality of mHealth apps: a systematic review. J Am Med Inform Assoc 2018 Aug 01;25(8):1089-1098 [FREE Full text] [doi: 10.1093/jamia/ocy050] [Medline: 29788283]
4. Orlowski S, Matthews B, Bidargaddi N, Jones G, Lawn S, Venning A, et al. Mental Health Technologies: Designing With Consumers. JMIR Hum Factors 2016 Jan 28;3(1):e4 [FREE Full text] [doi: 10.2196/humanfactors.4336] [Medline: 27026210]
5. Altman M, Huang TTK, Breland JY. Design Thinking in Health Care. Prev Chronic Dis 2018 Sep 27;15:E117 [FREE Full text] [doi: 10.5888/pcd15.180128] [Medline: 30264690]
6. Palacholla RS, Fischer N, Coleman A, Agboola S, Kirley K, Felsted J, et al. Provider- and Patient-Related Barriers to and Facilitators of Digital Health Technology Adoption for Hypertension Management: Scoping Review. JMIR Cardio 2019 Mar 26;3(1):e11951 [FREE Full text] [doi: 10.2196/11951] [Medline: 31758771]
7. Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, et al. Barriers to and Facilitators of Engagement With Remote Measurement Technology for Managing Health: Systematic Review and Content Analysis of Findings. J Med Internet Res 2018 Jul 12;20(7):e10480 [FREE Full text] [doi: 10.2196/10480] [Medline: 30001997]
8. Marcengo A, Rapp A. Visualization of human behavior data: the quantified self. In: Huang ML, Huang W, editors. Innovative Approaches of Data Visualization and Visual Analytics. Hershey, PA: IGI Global; 2014:236-235.
9. Few S. Data visualization for human perception. In: The Encyclopedia of Human-Computer Interaction, 2nd Ed. Aarhus, Denmark: The Interaction Design Foundation; 2010.
10. Woods L, Cummings E, Duff J, Walker K. Design Thinking for mHealth Application Co-Design to Support Heart Failure Self-Management. Stud Health Technol Inform 2017;241:97-102. [Medline: 28809190]
11. Bate P, Robert G. Experience-based design: from redesigning the system around the patient to co-designing services with the patient. Qual Saf Health Care 2006 Oct;15(5):307-310 [FREE Full text] [doi: 10.1136/qshc.2005.016527] [Medline: 17074863]
12. Schmidt JM, Irvine JM, Miller S. Data visualization. In: De Georgia M, Loparo K, editors. Neurocritical Care Informatics. Berlin, Heidelberg: Springer; 2019:135-147.
13. Schlee W, Hall DA, Edvall NK, Langguth B, Canlon B, Cederroth CR. Visualization of Global Disease Burden for the Optimization of Patient Management and Treatment. Front Med (Lausanne) 2017;4:86 [FREE Full text] [doi: 10.3389/fmed.2017.00086] [Medline: 28674694]
14. Park S, Bekemeier B, Flaxman A, Schultz M. Impact of data visualization on decision-making and its implications for public health practice: a systematic literature review. Inform Health Soc Care 2021 Sep 28:1-19. [doi: 10.1080/17538157.2021.1982949] [Medline: 34582297]
15. Chittaro L. Visualizing Information on Mobile Devices. Computer 2006 Mar;39(3):40-45. [doi: 10.1109/mc.2006.109]
16. Ware C. In: Kaufmann M, editor. Visual Thinking for Design. Amsterdam: Elsevier; 2008.
17. Few S. Data visualization past, present and future. Perceptual Edge. 2009. URL: http://mail.perceptualedge.com/articles/Whitepapers/Data_Visualization.pdf [accessed 2020-06-29]
18. Turchioe MR, Myers A, Isaac S, Baik D, Grossman LV, Ancker JS, et al. A Systematic Review of Patient-Facing Visualizations of Personal Health Data. Appl Clin Inform 2019 Aug;10(4):751-770 [FREE Full text] [doi: 10.1055/s-0039-1697592] [Medline: 31597182]
19. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. BMJ 2015 Jan 02;350:g7647 [FREE Full text] [doi: 10.1136/bmj.g7647] [Medline: 25555855]
20. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
21. Polhemus A, Novak J, Majid S, Simblett S, Wykes T, Dockendorf M, et al. Service user preferences on data visualization by remote measurement technologies for chronic central nervous system disease management. PROSPERO: International Prospective Register of Systematic Reviews. 2019. URL: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019139319 [accessed 2020-06-29]
22. Cohen J. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 2016 Jul 02;20(1):37-46. [doi: 10.1177/001316446002000104]

23. Kohl C, McIntosh EJ, Unger S, Haddaway NR, Kecke S, Schiemann J, et al. Online tools supporting the conduct and reporting of systematic reviews and systematic maps: a case study on CADIMA and review of existing tools. Environ Evid 2018 Feb 1;7(1):1-17. [doi: 10.1186/s13750-018-0115-5]
24. Davis MM, Freeman M, Kaye J, Vuckovic N, Buckley DI. A systematic review of clinician and staff views on the acceptability of incorporating remote monitoring technology into primary care. Telemed J E Health 2014 May;20(5):428-438 [FREE Full text] [doi: 10.1089/tmj.2013.0166] [Medline: 24731239]
25. Pace R, Pluye P, Bartlett G, Macaulay AC, Salsberg J, Jagosh J, et al. Testing the reliability and efficiency of the pilot Mixed Methods Appraisal Tool (MMAT) for systematic mixed studies review. Int J Nurs Stud 2012 Jan;49(1):47-53. [doi: 10.1016/j.ijnurstu.2011.07.002] [Medline: 21835406]
26. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. EFI 2018 Dec 18;34(4):285-291. [doi: 10.3233/EFI-180221]
27. Malpass A, Shaw A, Sharp D, Walter F, Feder G, Ridd M, et al. "Medication career" or "moral career"? The two sides of managing antidepressants: a meta-ethnography of patients' experience of antidepressants. Soc Sci Med 2009 Jan;68(1):154-168. [doi: 10.1016/j.socscimed.2008.09.068] [Medline: 19013702]
28. Dixon-Woods M, Booth A, Sutton AJ. Synthesizing qualitative research: a review of published reports. Qualitative Research 2016 Aug 15;7(3):375-422. [doi: 10.1177/1468794107078517]
29. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008 Jul 10;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
30. Fuller-Tyszkiewicz M, Richardson B, Klein B, Skouteris H, Christensen H, Austin D, et al. A Mobile App-Based Intervention for Depression: End-User and Expert Usability Testing Study. JMIR Ment Health 2018 Aug 23;5(3):e54 [FREE Full text] [doi: 10.2196/mental.9445] [Medline: 30139722]
31. Rohani D, Tuxen N, Lopategui A, Faurholt-Jepsen M, Kessing L, Bardram J. Personalizing mental health: a feasibility study of a mobile behavioral activation tool for depressed patients. : Association for Computing Machinery; 2019 Presented at: 3th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 20-23; Trento, Italy. [doi: 10.1145/3329189.3329214]
32. Bauer AM, Hodsdon S, Bechtel JM, Fortney JC. Applying the Principles for Digital Development: Case Study of a Smartphone App to Support Collaborative Care for Rural Patients With Posttraumatic Stress Disorder or Bipolar Disorder. J Med Internet Res 2018 Jun 06;20(6):e10048 [FREE Full text] [doi: 10.2196/10048] [Medline: 29875085]
33. Bauer AM, Hodsdon S, Hunter S, Choi Y, Bechtel J, Fortney JC. Lessons from the Deployment of the SPIRIT App to Support Collaborative Care for Rural Patients with Complex Psychiatric Conditions. Proc ACM Int Conf Ubiquitous Comput 2017 Sep;2017:772-780 [FREE Full text] [doi: 10.1145/3123024.3125610] [Medline: 29075683]
34. Bauer AM, Iles-Shih M, Ghomi RH, Rue T, Grover T, Kincler N, et al. Acceptability of mHealth augmentation of Collaborative Care: A mixed methods pilot study. Gen Hosp Psychiatry 2018;51:22-29 [FREE Full text] [doi: 10.1016/j.genhosppsych.2017.11.010] [Medline: 29272712]
35. Matthews M, Voida S, Abdullah S, Doherty G, Choudhury T, Im S, et al. In situ design for mental illness: considering the pathology of bipolar disorder in mHealth design. : Association for Computing Machinery; 2015 Presented at: 17th International Conference on Human-Computer Interaction with Mobile Devices and Services; August 24-27; Copenhagen, Denmark. [doi: 10.1145/2785830.2785866]
36. Matthews M, Abdullah S, Murnane E, Voida S, Choudhury T, Gay G, et al. Development and Evaluation of a Smartphone-Based Measure of Social Rhythms for Bipolar Disorder. Assessment 2016 Aug;23(4):472-483 [FREE Full text] [doi: 10.1177/1073191116656794] [Medline: 27358214]
37. Bardram J, Frost M, Szántó K, Faurholt-Jepsen M, Vinberg M, Kessing L. Designing mobile health technology for bipolar disorder: a field trial of the MONARCA system. : Association for Computing Machinery; 2013 Presented at: SIGCHI Conference on Human Factors in Computing Systems; April 27-May 3; Paris, France. [doi: 10.1145/2470654.2481364]
38. Bardram J, Frost M, Szántó K, Marcu G. The MONARCA self-assessment system: a persuasive personal monitoring system for bipolar patients. : Association for Computing Machinery; 2012 Presented at: 2nd ACM SIGHIT International Health Informatics Symposium; January 28-30; Miami, Florida, USA. [doi: 10.1145/2110363.2110370]
39. Jonassaint CR, Rao N, Sciuto A, Switzer GE, De Castro L, Kato GJ, et al. Abstract Animations for the Communication and Assessment of Pain in Adults: Cross-Sectional Feasibility Study. J Med Internet Res 2018 Aug 03;20(8):e10056 [FREE Full text] [doi: 10.2196/10056] [Medline: 30076127]
40. Adams P, Murnane E, Elfenbein M, Wethington E, Gay G. Supporting the self-management of chronic pain conditions with tailored momentary self-assessments. 2017 Presented at: 2017 CHI Conference on Human Factors in Computing Systems; May 6-11; Denver, Colorado, USA. [doi: 10.1145/3025453.3025832]
41. Eisner E, Drake RJ, Berry N, Barrowclough C, Emsley R, Machin M, et al. Development and Long-Term Acceptability of ExPRESS, a Mobile Phone App to Monitor Basic Symptoms and Early Signs of Psychosis Relapse. JMIR Mhealth Uhealth 2019 Mar 29;7(3):e11568 [FREE Full text] [doi: 10.2196/11568] [Medline: 30924789]
42. Quaedackers L, De Wit J, Pillen S, Van Gilst M, Batalas N, Lammers GJ, et al. A Mobile App for Longterm Monitoring of Narcolepsy Symptoms: Design, Development, and Evaluation. JMIR Mhealth Uhealth 2020 Jan 07;8(1):e14939 [FREE Full text] [doi: 10.2196/14939] [Medline: 31909723]

43. Hoffman L, Benedetto E, Huang H, Grossman E, Kaluma D, Mann Z, et al. Augmenting Mental Health in Primary Care: A 1-Year Study of Deploying Smartphone Apps in a Multi-site Primary Care/Behavioral Health Integration Program. Front Psychiatry 2019;10:94 [FREE Full text] [doi: 10.3389/fpsyt.2019.00094] [Medline: 30873053]
44. Ando H, Ashcroft-Kelso H, Halhead R, Chakrabarti B, Young CA, Cousins R, et al. Experience of telehealth in people with motor neurone disease using noninvasive ventilation. Disabil Rehabil Assist Technol 2021 Jul;16(5):490-496. [doi: 10.1080/17483107.2019.1659864] [Medline: 31512539]
45. Anderson K, Burford O, Emmerton L. Mobile Health Apps to Facilitate Self-Care: A Qualitative Study of User Experiences. PLoS One 2016;11(5):e0156164 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]
46. Murnane EL, Cosley D, Chang P, Guha S, Frank E, Gay G, et al. Self-monitoring practices, attitudes, and needs of individuals with bipolar disorder: implications for the design of technologies to manage mental health. J Am Med Inform Assoc 2016 May;23(3):477-484. [doi: 10.1093/jamia/ocv165] [Medline: 26911822]
47. Giunti G, Kool J, Rivera Romero O, Dorronzoro Zubiete E. Exploring the Specific Needs of Persons with Multiple Sclerosis for mHealth Solutions for Physical Activity: Mixed-Methods Study. JMIR Mhealth Uhealth 2018 Feb 09;6(2):e37 [FREE Full text] [doi: 10.2196/mhealth.8996] [Medline: 29426814]
48. Forchuk C, Reiss JP, O'Regan T, Ethridge P, Donelle L, Rudnick A. Client perceptions of the mental health engagement network: a qualitative analysis of an electronic personal health record. BMC Psychiatry 2015 Oct 14;15:250 [FREE Full text] [doi: 10.1186/s12888-015-0614-7] [Medline: 26467210]
49. Erten-Uyumaz B, Ahuja M, Vacaretu T, Rama M, Overeem S, Visser T, et al. Design and evaluation of a negotiation-based sleep scheduler app for insomnia treatment. : Association for Computing Machinery; 2019 Presented at: 13th EAI International Conference on Pervasive Computing Technologies for Healthcare; May 20-23; Trento, Italy. [doi: 10.1145/3329189.3329218]
50. Saunders KEA, Bilderbeck AC, Panchal P, Atkinson LZ, Geddes JR, Goodwin GM. Experiences of remote mood and activity monitoring in bipolar disorder: A qualitative study. Eur Psychiatry 2017 Mar;41:115-121 [FREE Full text] [doi: 10.1016/j.eurpsy.2016.11.005] [Medline: 28135594]
51. McClelland GT, Fitzgerald M. A participatory mobile application (app) development project with mental health service users and clinicians. Health Education Journal 2018 Jun 05;77(7):815-827. [doi: 10.1177/0017896918773790]
52. Wannheden C, Revenäs Å. How People with Parkinson's Disease and Health Care Professionals Wish to Partner in Care Using eHealth: Co-Design Study. J Med Internet Res 2020 Sep 21;22(9):e19195 [FREE Full text] [doi: 10.2196/19195] [Medline: 32955448]
53. Van Tiem J, Moeckli J, Suiter N, Fuhrmeister L, Pham K, Dindo L, et al. "A link to the outside:" Patient perspectives on a mobile texting program to improve depression self-management. Patient Educ Couns 2021 Sep;104(9):2154-2158. [doi: 10.1016/j.pec.2021.02.018] [Medline: 33627234]
54. Elsborg M, Bruun A, Jensen R. Supporting anxiety patients? self-reflection through visualization of physiological data. : Association for Computing Machinery; 2020 Presented at: 32nd Australian Conference on Human-Computer Interaction; December 2-4; Sydney, Australia. [doi: 10.1145/3441000.3441037]
55. Ruzic L, Sanford JA. Needs Assessment—mHealth Applications for People Aging with Multiple Sclerosis. J Healthc Inform Res 2018 May 10;2(1-2):71-98. [doi: 10.1007/s41666-018-0023-z]
56. Herbst E, McCaslin SE, Hassanbeigi Daryani S, Laird KT, Hopkins LB, Pennington D, et al. A Qualitative Examination of Stay Quit Coach, A Mobile Application for Veteran Smokers With Posttraumatic Stress Disorder. Nicotine Tob Res 2020 Apr 17;22(4):560-569. [doi: 10.1093/ntr/ntz037] [Medline: 30874289]
57. Schleimer E, Pearce J, Barnecut A, Rowles W, Lizee A, Klein A, et al. A Precision Medicine Tool for Patients With Multiple Sclerosis (the Open MS BioScreen): Human-Centered Design and Development. J Med Internet Res 2020 Jul 06;22(7):e15605 [FREE Full text] [doi: 10.2196/15605] [Medline: 32628124]
58. Daus H, Kislicyn N, Heuer S, Backenstrass M. Disease management apps and technical assistance systems for bipolar disorder: Investigating the patients´ point of view. J Affect Disord 2018 Mar 15;229:351-357. [doi: 10.1016/j.jad.2017.12.059] [Medline: 29331693]
59. Yoo S, Lim K, Baek H, Jang S, Hwang G, Kim H, et al. Developing a mobile epilepsy management application integrated with an electronic health record for effective seizure management. Int J Med Inform 2020 Feb;134:104051. [doi: 10.1016/j.ijmedinf.2019.104051] [Medline: 31837500]
60. Ando H, Ashcroft-Kelso H, Halhead R, Chakrabarti B, Young CA, Cousins R, et al. Experience of telehealth in people with motor neurone disease using noninvasive ventilation. Disabil Rehabil Assist Technol 2021 Jul;16(5):490-496. [doi: 10.1080/17483107.2019.1659864] [Medline: 31512539]
61. Garzo A, Silva PA, Garay-Vitoria N, Hernandez E, Cullen S, Cochen De Cock V, et al. Design and development of a gait training system for Parkinson's disease. PLoS One 2018;13(11):e0207136 [FREE Full text] [doi: 10.1371/journal.pone.0207136] [Medline: 30418993]
62. Buitenweg DC, Bongers IL, van de Mheen D, van Oers HA, van Nieuwenhuizen C. Cocreative Development of the QoL-ME: A Visual and Personalized Quality of Life Assessment App for People With Severe Mental Health Problems. JMIR Ment Health 2019 Mar 28;6(3):e12378 [FREE Full text] [doi: 10.2196/12378] [Medline: 30920381]

63. Ospina-Pinillos L, Davenport T, Mendoza Diaz A, Navarro-Mancilla A, Scott EM, Hickie IB. Using Participatory Design Methodologies to Co-Design and Culturally Adapt the Spanish Version of the Mental Health eClinic: Qualitative Study. J Med Internet Res 2019 Aug 02;21(8):e14127 [FREE Full text] [doi: 10.2196/14127] [Medline: 31376271]
64. Proudfoot J, Parker G, Hadzi Pavlovic D, Manicavasagar V, Adler E, Whitton A. Community attitudes to the appropriation of mobile phones for monitoring and managing depression, anxiety, and stress. J Med Internet Res 2010 Dec 19;12(5):e64 [FREE Full text] [doi: 10.2196/jmir.1475] [Medline: 21169174]
65. Farao J, Malila B, Conrad N, Mutsvangwa T, Rangaka MX, Douglas TS. A user-centred design framework for mHealth. PLoS One 2020;15(8):e0237910 [FREE Full text] [doi: 10.1371/journal.pone.0237910] [Medline: 32813711]
66. Mummah SA, Robinson TN, King AC, Gardner CD, Sutton S. IDEAS (Integrate, Design, Assess, and Share): A Framework and Toolkit of Strategies for the Development of More Effective Digital Interventions to Change Health Behavior. J Med Internet Res 2016 Dec 16;18(12):e317 [FREE Full text] [doi: 10.2196/jmir.5927] [Medline: 27986647]
67. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
68. Polhemus AM, Novák J, Ferrao J, Simblett S, Radaelli M, Locatelli P, et al. Human-Centered Design Strategies for Device Selection in mHealth Programs: Development of a Novel Framework and Case Study. JMIR Mhealth Uhealth 2020 May 07;8(5):e16043 [FREE Full text] [doi: 10.2196/16043] [Medline: 32379055]
69. Lee SY, Hwang H, Hawkins R, Pingree S. Interplay of Negative Emotion and Health Self-Efficacy on the Use of Health Information and Its Outcomes. Communication Research 2008 Apr 14;35(3):358-381. [doi: 10.1177/0093650208315962]
70. Paterson BL. The shifting perspectives model of chronic illness. J Nurs Scholarsh 2001;33(1):21-26. [doi: 10.1111/j.1547-5069.2001.00021.x] [Medline: 11253576]
71. Zhu H, Colgan J, Reddy M, Choe EK. Sharing Patient-Generated Data in Clinical Practices: An Interview Study. AMIA Annu Symp Proc 2016;2016:1303-1312 [FREE Full text] [Medline: 28269928]
72. Andrews JA, Craven MP, Jamnadas-Khoda J, Lang AR, Morriss R, Hollis C, RADAR-CNS Consortium. Health Care Professionals' Views on Using Remote Measurement Technology in Managing Central Nervous System Disorders: Qualitative Interview Study. J Med Internet Res 2020 Jul 24;22(7):e17414 [FREE Full text] [doi: 10.2196/17414] [Medline: 32706664]
73. Caprani N, Doyle J, Komaba Y, Inomata A. Exploring healthcare professionals' preferences for visualising sensor data. : Association for Computing Machinery; 2015 Presented at: 2015 British HCI Conference; July 13; Lincoln, Lincolnshire, United Kingdom. [doi: 10.1145/2783446.2783564]
74. Noblit GW, Hare RD. Meta-Ethnography: Synthesizing Qualitative Studies. Newbury Park, California, USA: SAGE Publications, Inc; 1988.
75. Innovative Medicines Initiative. URL: https://www.imi.europa.eu/ [accessed 2022-01-25]

Abbreviations

EFPIA: European Federation of Pharmaceutical Industries and Associations
MONARCA: MONitoring, treAtment and pRediCtion of bipolAr Disorder Episodes
NHS: National Health Service
NIHR: National Institute for Health Research
PRISMA-P: Preferred Reporting Items for Systematic review and Meta-Analysis Protocols
PROSPERO: International Prospective Register of Systematic Reviews
RADAR-CNS: Remote Assessment of Disease and Relapse-Central Nervous System
RMT: remote measurement technology

Edited by J Torous; submitted 25.10.20; peer-reviewed by A Alasmari, I Shubina, H Daus, M Ghazisaeedi, CY Lin; comments to author 30.12.20; revised version received 29.03.21; accepted 20.10.21; published 28.04.22.

Please cite as:
Polhemus A, Novak J, Majid S, Simblett S, Morris D, Bruce S, Burke P, Dockendorf MF, Temesi G, Wykes T
Data Visualization for Chronic Neurological and Mental Health Condition Self-management: Systematic Review of User Perspectives
JMIR Ment Health 2022;9(4):e25249
URL: https://mental.jmir.org/2022/4/e25249
doi:10.2196/25249
PMID:35482368

Original Paper

Comparing the Ratio of Therapist Support to Internet Sessions in a Blended Therapy Delivered to Trauma-Exposed Veterans: Quasi-experimental Comparison Study

Marylene Cloitre[1], PhD; Amber Bush Amspoker[2], PhD; Terri L Fletcher[2], PhD; Julianna B Hogan[2], PhD; Christie Jackson[1], PhD; Adam Jacobs[1], MS; Rayan Shammet[1], MA; Sarah Speicher[1], MS; Miryam Wassef[2], LCSW; Jan Lindsay[2], PhD

[1]National Center for PTSD Dissemination and Training Division, Veterans Administration Palo Alto Health Care System, Palo Alto, CA, United States

[2]Veterans Administration Health Services Research & Development Center for Innovations in Quality, Effectiveness and Safety, Michael E DeBakey Veterans Affairs Medical Center, Baylor Medical College, Houston, TX, United States

Corresponding Author:
Marylene Cloitre, PhD
National Center for PTSD Dissemination and Training Division, Veterans Administration Palo Alto Health Care System
795 Willow Road
Palo Alto, CA, 94025
United States
Phone: 1 415 603 8093
Email: marylene.cloitre@gmail.com

Abstract

Background: Blended models of therapy, which incorporate elements of both internet and face-to-face methods, have been shown to be effective, but therapists and patients have expressed concerns that fewer face-to-face therapy sessions than self-guided internet sessions may be associated with lower therapeutic alliance, lower program completion rates, and poorer outcomes.

Objective: A multisite quasi-experimental comparison study with a noninferiority design implemented in routine clinical care was used to assess webSTAIR, a 10-module blended therapy derived from STAIR (skills training in affective and interpersonal regulation) for trauma-exposed individuals delivered with 10 weekly therapist sessions (termed Coach10) compared to 5 biweekly sessions (Coach5). It was hypothesized that Coach5 would be as good as Coach10 in a range of outcomes.

Methods: A total of 202 veterans were enrolled in the study with 101 assigned to Coach5 and 101 to Coach10. Posttraumatic stress disorder (PTSD) symptoms, depression, emotion regulation, interpersonal problems, and social functioning measures were collected pre-, mid-, and posttreatment, and at a 3-month follow-up. Noninferiority analyses were conducted on symptom outcome measures. Comparisons were made of continuous and categorical measures regarding participant and therapist activities.

Results: Participants reported moderate to severe levels of baseline PTSD, depression, or both. Significant reductions were obtained in all symptom measures posttreatment and at the 3-month follow up. Coach5 was not inferior to Coach10 in any outcome. Therapeutic alliance was at an equivalently high level across the 2 treatment conditions; completion rates and web usage were similar. Total session time was substantially less for the Coach5 therapists than the Coach10 therapists. Both programs were associated with a low, but equal number of therapist activities related to scheduling and crisis or motivational sessions.

Conclusions: A blended model delivered with 5 sessions of therapist support was noninferior to 10 sessions in individuals with moderate to severe symptoms. Future studies identifying patient characteristics as moderators of outcomes with high versus low doses of therapist support will help create flexible, technology-based intervention programming.

(JMIR Ment Health 2022;9(4):e33080) doi:10.2196/33080

KEYWORDS

PTSD; depression; veterans; blended therapy; iCBT; web-based; webSTAIR; noninferiority; mental health; digital health

Introduction

Background

Meta-analyses and reviews have found that trauma-exposed individuals receive moderate benefits for posttraumatic stress disorder (PTSD) and depression symptom reduction from internet-based interventions; larger effects were found in studies that included therapist support as compared to completely self-guided approaches [1-3]. Nevertheless, internet interventions, even those with some therapist support, do not provide full recovery for everyone [1,3], particularly over the long term [2], indicating the value of further exploration and improvement in technology-supported treatments. An emerging alternative approach is the "blended intervention," which integrates face-to-face therapy with internet approaches. Blended therapy is characterized by continued therapist input alongside internet self-help to allow greater flexibility and personalization within the overall therapy process [4,5]. Blended therapies provide an alternative model of care which may improve treatment outcomes as well as increase engagement with mental health services among those who prefer more intensive therapist guidance.

Blended therapy approaches have been found to be positively viewed by both therapists and patients and preferred over stand-alone internet programs [4-7]. To date, several case reports and open trials of individuals with anxiety or depression have reported that blended interventions that focus on web-based interventions but also provide substantial therapist support are feasible and highly acceptable to clients [8-11]. Three randomized controlled trials (RCTs) have supported their efficacy. An early study comparing subjects who received an intervention to those who were placed on a waitlist found moderate to large between-group effect sizes for clients with social phobia who received 9 weeks of internet-based cognitive behavioral therapy (CBT) integrated with two 3-hour exposure sessions and email support from a therapist [12]. An RCT of problem-solving therapy (PST) for individuals with anxiety or depression found that receiving web-based PST with support was superior to being placed on a waitlist, while 3 other forms of delivery—web alone, with support as requested, or with weekly emails—were not [13]. Last, a 4-arm RCT, which included patients with anxiety or depression, found that the blended approach was superior not only to the no-treatment condition but also to a face-to-face–only condition and internet-only condition [14]. The findings from the latter two studies suggest the possibility of synergistic effects in combining these two intervention approaches.

Based on the above successes, investigation of blended therapies for trauma-exposed populations is justified. A recent open trial study assessed patient satisfaction and outcomes in a blended model delivered to rural trauma-exposed veterans. The program was entirely virtual; the patients completed the web-based program concurrent with a face-to-face coaching session via video conferencing [15]. The program, webSTAIR, is a 10-session, transdiagnostic, trauma-informed program derived from STAIR (skills training in affective and interpersonal regulation), a CBT approach with empirical support [16]. Significant improvements in PTSD, depression, and social functioning were obtained with moderate to large effect sizes at posttreatment and at a 3-month follow up. Analyses of posttreatment interviews revealed themes regarding the value and importance of the therapists, particularly regarding their ability to provide support, accountability, and effective tailoring of the interventions, activities that have been described as key therapist functions in technology-based interventions [17].

Objectives

One important long-term goal is to identify the optimal amount of therapist support relative to self-guided work to maximize outcomes. Both patients and therapists tend to take a "more is better" perspective when considering the presence of therapists in blended treatments. Results from a Delphi survey study found that therapists preferred that 75% of sessions be face-to-face, while most patients preferred 50% to 60% [18]. Moreover, both patients and therapists express concern that less therapist involvement will be associated with lower therapeutic alliance, lower completion rates, and reduced effectiveness [19,20]. To our knowledge, however, no studies have been conducted that consider the impact of the amount of therapist support (ie, number of sessions) on outcomes.

The purpose of this study was to systematically assess and compare the impact of 2 different ratios of therapist sessions to self-guided work on therapeutic alliance, completion rates, and symptom reduction among trauma-exposed veterans. The specific goals of the project were (1) to replicate the results of the first webSTAIR study by using a 1:1 ratio of therapist sessions to self-guided web-based sessions and (2) to assess outcomes where therapist support was reduced by 50%, (ie, the ratio was 1:2), with 1 therapist session for every 2 self-guided web-based sessions. The 1:1 ratio was adopted as the anchoring reference point for this investigation based on the high patient satisfaction ratings and the large effect sizes reported in the first webSTAIR study, which used this ratio [15], and because this ratio provided the maximum number of therapist sessions in a blended therapy program in which the intervention of interest was the web-based intervention. The selection of the 1:2 ratio was based on a pilot study that delivered webSTAIR with this ratio, yielding significant symptom reduction with a large effect size [21]. That study also included posttreatment interviews that identified an association between patient satisfaction and the number of therapist sessions; veterans reported in the interviews that self-guided work fostered independence and mastery.

We hypothesized that providing 1 therapist session for every 2 self-guided sessions in this 10-module treatment (ie, the webSTAIR Coach5 condition) would not be inferior to providing 1 therapist session for each self-guided session (ie, the webSTAIR Coach10 condition) in regard to PTSD, depression, emotion regulation, interpersonal problems, or social functioning outcomes. We also compared therapeutic alliance, completion rates, and web usage (measured as time in minutes) among the participants. The amount of time therapists spent in session and the amount of activity related to rescheduling appointments and providing additional sessions as needed for crisis management and motivational support was also compared.

Methods

Study Design

The study was funded by the Office of Rural Health and dedicated to rural veterans, with a focus on rural women veterans, who have been identified as under-represented in mental health services relative to both urban male and female veterans [22]. This was a naturalistic evaluation study, in which delivery of the program was conducted as part of routine care in mental health outpatient clinics within the Veterans Health Administration. The study used a quasi-experimental comparison design, in which 9 service sites were assigned to either the Coach5 or the Coach10 condition, matched on 3 characteristics: percentage of rural veterans enrolled in the mental health service, projected number of veterans expected to be enrolled in the study per month, and job description of the therapists trained to deliver the intervention (eg, psychologist, social worker, or mental health technician).

Procedures

Candidates for the program were referred by clinical therapists. Individuals were eligible for inclusion if they reported a history of trauma exposure and were currently experiencing symptoms of PTSD, depression, or both, as indicated by a positive screen on the Primary Care PTSD Screen [23] or the 2-item Patient Health Questionnaire (PHQ-2) [24]. Additional inclusion criteria were an expressed willingness to complete assessment and treatment procedures and an interest in working on improving emotion regulation skills and interpersonal relationships. Exclusion criteria were active suicidal or homicidal ideation, psychosis, mania, cognitive impairment, inability to attend regular telemental health appointments, primary substance or alcohol use difficulties, current interpersonal violence, lack of a private place to connect for sessions, engagement in concurrent trauma-focused treatment for PTSD, and receipt of inpatient or residential PTSD care in the past year. Although the target population for enrollment into the study was rural male veterans and rural women veterans, any veteran who satisfied the above criteria and could not easily access in-clinic care (because of, for example, health concerns, time constraints, or elder or childcare responsibilities) was accepted into the program.

Outcomes

Assessment and Symptom Measures

All assessments were conducted by the study coordinator via telephone at pretreatment, the midpoint (session 5), posttreatment, and at the 3-month follow up. The initial assessment included an inquiry about frequency of traumatic events using an adapted version of the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5)-derived Life Events Checklist (LEC-5) [25]. PTSD symptoms were measured using the PTSD Checklist for DSM-5 (PCL-5) [26], depression was measured with the Patient Health Questionnaire (PHQ)-9 [27], emotion regulation with the Difficulties in Emotion Regulation Scale (DERS)-36 [28], interpersonal problems with the Inventory of Interpersonal Problems (IIP)-32 [29], and social functioning by the brief, 21-item version of the World Health Organisation Disability Assessment Schedule (WHODAS)-2 [15,30], designated as the WHODAS-21. Probable diagnoses of PTSD (PCL-5 score>33) and depressive disorder (PHQ score>10) were calculated at baseline. Therapeutic alliance was measured with the patient version of the Working Alliance Inventory (WAI) [31].

Participant Web Usage and Therapist Time and Activities

Usage data, such as total number of minutes each user engaged in webSTAIR, was a built-in analytic feature of the program. The therapists used an online survey not integrated with the webSTAIR site to log the number of minutes they were engaged with the veterans after each session as well as the number of times they engaged in additional interaction with the veteran since their last session. Interactions included phone calls, instant messages, and emails to reschedule appointments as well as crisis or motivational sessions deemed clinically relevant by the therapist. The therapists were allowed to spend up to 20 minutes for crisis or motivational sessions.

WebSTAIR Intervention

The webSTAIR program consists of 10 web-based modules adapted from STAIR [32]. The first 5 modules review emotional awareness, emotion management, and distress tolerance while the final 5 modules raise awareness about relationship patterns and provide interpersonal skills training regarding effective assertiveness, interpersonal flexibility, and compassion for the self and others. Modules include text, video, and audio delivery of psychoeducation, as well as interactive exercises and worksheets to aid the patient in learning and practicing the material.

Therapist Sessions and Adherence

The therapists followed a manual that provided instructions for each session. The sessions were 45-50 minutes long and were organized with the same general tasks and goals. The therapists had 4 tasks: to clarify key concepts presented in the modules, to reinforce engagement and enthusiasm for the material, to help the participant tailor the skills to their own life experiences and concerns, and to support the participant in completing the modules on a weekly basis. The overall goal of the sessions was to help the participants make the most of the webSTAIR material by reinforcing work they did independently. The Coach5 condition required that the therapists review materials from 2 modules in a session while the Coach10 condition involved review of materials from only 1 module. Topics and content covered by the instruction manuals did not differ between the 2 treatment conditions. The therapists completed self-reported adherence ratings after each session. Self-reported adherence ratings have been found to be reliable and were chosen as being appropriate for this resource-limited intervention approach [33].

Therapists

All webSTAIR therapists were licensed mental health staff working in Veterans Health Administration clinics. Five were licensed clinical psychologists, 3 were licensed social workers, and 1 was a licensed professional mental health counselor. All the therapists were women. Training for Coach5 and Coach10 was implemented separately, and each group was provided their own manual describing the topics and content for each of their sessions. Training also covered unique issues related to the use

of web-based technology in comparison to traditional face-to-face psychotherapy. The Coach5 and Coach10 therapists each received weekly group phone supervision sessions with an experienced clinical psychologist and certified STAIR trainer. The therapists also attended weekly implementation meetings to address questions and concerns related to implementing a web-based telemental health intervention.

Statistical Analysis Plan

Study noncompleters were defined as those who did not complete either the posttreatment or 3-month follow-up assessments. Pretreatment demographic variables and clinical characteristics were compared in the 2 treatment groups (Coach5 and Coach10) with independent samples and a 2-tailed *t* test for continuous variables and the chi-square test for categorical variables. The pretreatment demographic variables and clinical characteristics of study completers and noncompleters were compared with similar analyses.

We then examined linear changes in outcomes over time separately within each treatment group. For the Coach5 and Coach10 groups, an unconditional linear growth curve model using Proc Mixed in SAS (version 9.4; SAS Institute, Inc) was employed to examine linear changes over time for each of the 5 outcomes. In each model, time was the sole predictor and was coded such that 0 equaled baseline, 1 equaled the midpoint assessment, 2 equaled the posttreatment assessment, and 3 equaled the 3-month follow-up assessment. Both the intercept and time were included as random effects in all models. For each condition, within-group effect sizes from pre- to posttreatment and from pretreatment to the 3-month follow-up were calculated. Unlike superiority trials, where an intent-to-treat (ITT) analysis is more conservative as it makes 2 groups more similar, for a noninferiority trial, an observed or per-protocol analysis yields a more conservative estimate as it exaggerates the differences between treatment groups [33]. Therefore, a series of ANCOVA models was first calculated using the observed data to examine the treatment conditions (ie, Coach5 and Coach10) as predictors of each outcome at the posttreatment assessment. Treatment-group comparisons of posttreatment assessment outcomes were then repeated with another set of ITT analyses that used the multiple imputation procedures Proc MI and Mianalyze in SAS. Respective baseline scores and characteristics that differed between the treatment conditions were included as covariates in all models. Similar observed and ITT analyses were then used to examine treatment group differences in the 3-month follow-up assessment.

Examination of between-group changes relative to noninferiority margins was accomplished with 4 steps. First, noninferiority margins (*F*) for the posttreatment and follow-up assessments were calculated for each outcome using data from the initial webSTAIR study [19]. For each outcome, we calculated the noninferiority margin, which was 80% of the change in Coach10 from baseline to the posttreatment assessment, and then calculated the margin from baseline to the 3-month follow-up assessment. Second, we examined the actual difference between Coach5 and Coach10 in the change from baseline to posttreatment and from baseline to the 3-month follow up using the differences in differences (ie, the change score) approach. Baseline was always coded 0, and the follow-up assessment (posttreatment or 3-month follow-up) was coded 1. For each of the 5 outcomes, Proc Mixed in SAS was employed with time, treatment group, and the time by treatment group interaction as predictors. The intercept was included as a random effect. A significant time by treatment group interaction indicated a significant difference between groups in the change in outcome. Of primary interest was the point estimate of the difference in differences. Third, we calculated the 95% CI for each difference in differences. Finally, we plotted the posttreatment estimate of the difference in differences and its 95% CI relative to the noninferiority margin.

If we consider the difference in effect between Coach5 and Coach10, superiority of Coach5 versus Coach10 implies that the lower bound of the 95% CI is greater than 0. Noninferiority of Coach5 allows the lower bound of the 95% CI to extend below 0, so long as it remains above the noninferiority margin and so long as the upper bound of the 95% CI is greater than 0. In the case that the lower bound of the 95% CI falls below and the upper bound of the 95% CI falls above the noninferiority margin, the study result is indeterminate. Finally, inferiority of Coach5 versus Coach10 is implied if the upper bound of the 95% CI falls below the noninferiority margin [34]. All analyses involved 2-sided significance testing and were conducted in SAS. Effect sizes were also calculated.

A series of 2-tailed independent sample *t* tests were conducted to examine mean differences between Coach5 and Coach10 in (1) participant therapeutic alliance score, (2) number of modules completed by participants, (3) average amount of time spent (in minutes) per module and in total across all modules, (4) total therapist time spent (in minutes) providing treatment, (5) number of times the therapist made a phone call, texted, or emailed to reschedule an appointment, and (6) the number of times the therapist provided an additional session to address a crisis or enhance motivation. The chi-square test was used to examine the difference between Coach5 and Coach10 in program completion rate.

Ethics Approval

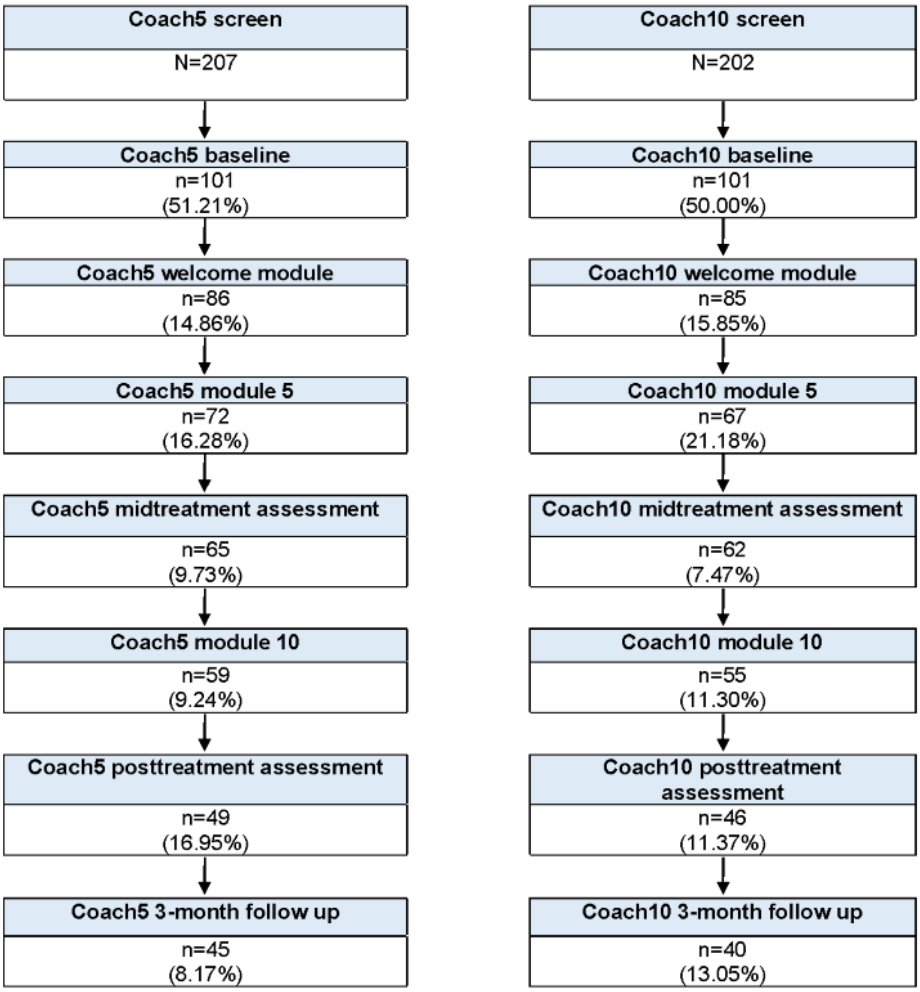
This project was funded by the Department of Veterans Affairs Office of Rural Health. All procedures involved in the evaluation were reviewed and exempted by the local academic institutional review board (APP03H08).

Results

Participants

Given the quasi-experimental design, we wished to assess the similarity of the 2 conditions in enrollment rates and participant characteristics. As indicated by the Consolidated Standards of Reporting Trials (CONSORT) chart (Figure 1), there were no differences between the 2 treatment conditions in enrollment numbers, attrition rate from screening to baseline assessment, or from baseline assessment to program enrollment. The percentages in each box in the figure represent the amount of attrition that occurred relative to the previous step in the study.

Figure 1. CONSORT chart for Coach5 and Coach10 conditions.



There were no differences in sociodemographic characteristics between the 2 treatment conditions in age, gender, ethnicity or minority status, education, or employment status. The baseline symptom profiles were also similar (Table 1). The average age of the participants was 44.11 years (SD 11.73) with a range of 22-77 years. Participants who identified as women were the largest gender represented, followed by men and then transgender individuals: 60.4% (122/202), 38.6% (78/202), and 0.5% (1/202), respectively. The majority of the participants 63.9% (129/202) had some college education or a high school diploma and 45.5% (92/202) were working full- or part-time. A total of 39.6% (80/202) of participants identified as a member of a racial or ethnic minority group. Treatment symptom measures did not differ between the 2 conditions with the exception of the WHODAS-21 results (Table 1). The participants who received Coach5 had higher WHODAS-21 scores (mean 49.38, SD 16.34) compared to those who received Coach10 (mean 44.97, SD 15.01; t_{200}=2.00, *P*=.047). Baseline WHODAS-21 scores were included as a covariate in subsequent outcome models that compared treatment groups.

Table 1. Comparison of sociodemographic and baseline clinical characteristics of Coach5 and Coach10 participants.

Variable	Overall, N=202	Coach5, n=101	Coach10, n =101	*P* value (*t* test or chi-square test)
Completed study, n (%)	100 (49.5)	54 (52.5)	47 (46.5)	.40
Age, mean (SD)	44.10 (11.73)	44.96 (11.54)	43.06 (11.90)	.25
Gender, n (%)				.08
Male	78 (38.6)	33 (32.7)	45 (44.6)	
Female	122 (60.4)	67 (66.3)	55 (54.5)	
Transgender	1 (0.5)	0 (0)	1 (1)	
Race, n (%)				.12
White	122 (60.4)	59 (58.4)	63 (62.4)	
Black or African American	37 (18.3)	19 (18.8)	18 (17.8)	
Asian	2 (1)	1 (1)	1 (2)	
Hispanic, Latino(a)	11 (5.6)	6 (5.9)	5 (4.6)	
Multiracial	17 (8.4)	13 (12.9)	4 (4)	
Other[a]	13 (6.4)	3 (3)	10 (9.9)	
Education level, n (%)				.47
Some high school	1 (0.5)	1 (1)	0 (0)	
Earned high school degree	25 (12.4)	12 (11.9)	13 (12.9)	
Some college/2-year degree	104 (51.5)	56 (55.6)	48 (47.5)	
Earned 4-year degree	51 (25.3)	24 (23.8)	27 (26.7)	
Postgraduate degree	21 (10.4)	8 (7.9)	13 (12.9)	
Employment status, n (%)				.26
Full-time	73 (36.1)	31 (30.7)	42 (41.6)	
Part-time	19 (9.4)	11 (10.9)	8 (7.9)	
Not currently working	62 (30.7)	31 (30.7)	31 (30.7)	
Retired	48 (23.8)	28 (27.7)	20 (19.8)	
Relationship status, n (%)				.12
Married/partnered	115 (56.9)	52 (51.5)	62 (62.4)	
Single	38 (18.8)	22 (21.8)	16 (15.8)	
Divorced	47 (23.3)	25 (24.8)	22 (21.8)	
Widowed	2 (1)	2 (2)	0 (0)	
Baseline outcomes				
PCL[b] total score, mean (SD) (range 2-78)	50.66 (15.52)	50.85 (16.09)	50.47 (15.01)	.86
PHQ[c] total score, mean (SD) (range 2-27)	15.64 (5.40)	15.96 (5.38)	15.32 (5.44)	.40
DERS[d] total score, mean (SD) (range 49-169)	108.74 (25.19)	108.72 (26.11)	108.75 (24.36)	.99
IIP-32[e] score mean (SD) (range 0.22-3.31)	1.86 (0.53)	1.90 (0.53)	1.83 (0.53)	.38
WHODAS-2[f] total score (SD) (range 0-81)	47.17 (15.81)	49.38 (16.34)	44.97 (15.01)	.05

[a]Other included American Native Indian, Alaskan Native, Pacific Islander, Middle Eastern, and North African.

[b]PCL-5: posttraumatic stress disorder checklist for DSM-5

[c]PHQ-9: Patient Health Questionaire-9

[d]DERS: Difficulties in Emotion Regulation Scale

[e]IIP-32: Inventory in Interpersonal Problems-32

[f]WHODAS-21: brief version of World Health Organization Disability Assessment Schedule-2.

A total of 86.1% of participants (174/202) had a probable diagnosis of PTSD regardless of depression status and 86.6% (175/202) had a probable diagnosis of depression regardless of PTSD status. A total of 80.2% (162/202) had both disorders and 12.4% (25/202) had either one or the other, yielding a total of 92.6% (187/202) of the participants with at least one probable disorder. The study was completed by 49.5% (100/202) of participants. Compared to noncompleters, the completers were older (mean age 46.11 years vs 41.95 years; *P*=.01), more likely to rent versus own their home (97/100, 97% compared to 89/102, 87.3%; *P*=.01), more likely to be white (69/100, 69% vs 53/102, 52%; *P*=.01), and more likely to live in rural or highly rural locations (53/100, 67% vs 53/102, 52%; *P*=.03) Study completers and noncompleters were similar in all other demographic and pretreatment variables (*P*>.05 for all values, data not presented).

Symptom Outcomes

Table 2 presents the observed mean for each outcome over time as well as within-group changes from baseline to posttreatment and from baseline to the 3-month follow up, organized by treatment group. All outcomes for both the Coach5 and Coach10 groups were significantly improved at the posttreatment assessment and the 3-month follow up relative to baseline. Furthermore, a series of ANCOVAs, controlling for baseline differences in WHODAS-21 scores, revealed no difference between the Coach5 and Coach10 groups in any of the 5 outcomes at the posttreatment assessment or at the 3-month follow up (see Table 3). ITT analyses revealed identical findings.

Table 2. Repeated measures tests and within-group effect sizes for the Coach5 and Coach10 groups.

Outcome measure	Baseline: Coach5, n=101; Coach10, n=101	Posttreatment: Coach5, n=49; Coach10, n=46	Three-month follow up: Coach5, n=45; Coach 10, n=40	Within-group baseline to posttreatment		Within-group baseline to 3-month follow up	
				P value	Cohen *d* (95% CI)	*P* value	Cohen *d* (95% CI)
PCL-5[a]score, mean (SD)							
Coach5	50.85 (16.09)	40.04 (19.50)	39.96 (19.84)	<.001	0.66 (0.38-0.94)	<.001	0.67 (0.37-0.97)
Coach10	50.47 (15.01)	34.59 (19.66)	35.20 (19.42)	<.001	1.04 (0.73-1.35)	<.001	1.00 (0.67-10.33)
PHQ-9[b]score, mean (SD)							
Coach5	15.96 (5.38)	12.39 (7.13)	12.58 (6.26)	<.001	0.65 (0.32-0.98)	<.001	0.62 (0.28-0.96)
Coach10	15.32 (5.44)	9.91 (6.77)	10.35 (6.76)	<.001	0.65 (0.26-1.03)	0.003	0.90 (0.55-1.24)
DERS-36[c]score, mean (SD)							
Coach5	108.72 (26.11)	90.51 (29.71)	88.13 (26.80)	<.001	0.69 (0.37-1.00)	<.001	0.69 (0.35-1.02)
Coach10	108.75 (24.36)	88.07 (26.07)	85.85 (25.95)	<.001	0.83 (0.51-1.16)	<.001	0.92 (0.56-1.28)
IIP-32[d]score, mean (SD)							
Coach5	1.90 (0.53)	1.67 (0.62)	1.58 (0.65)	.002	0.43 (0.15-0.70)	<.001	0.59 (0.33-0.85)
Coach10	1.83 (0.53)	1.47 (0.65)	1.54 (0.65)	<.001	0.67 (0.36-0.72)	.001	0.54 (0.26-0.81)
WHODAS-21[e]score, mean (SD)							
Coach5	49.38 (16.34)	44.14 (17.70)	42.46 (18.83)	.001	0.32 (0.06-0.57)	<.001	0.42 (0.17-0.66)
Coach10	44.97 (15.01)	35.43 (19.33)	35.03 (18.62)	.04	0.62 (0.30-0.95)	.04	0.65 (0.35-0.95)

[a]PCL-5: posttraumatic stress disorder checklist for DSM-5

[b]PHQ-9: Patient Health Questionaire-9

[c]DERS: Difficulties in Emotion Regulation Scale

[d]IIP-32: Inventory in Interpersonal Problems-32

[e]WHODAS-21: brief version of World Health Organization Disability Assessment Schedule-2

Table 3. Between-group effect sizes at posttreatment assessment and 3-month follow up.

Outcome measure	Coach 5 vs Coach 10 at posttreatment assessment			Coach 5 vs Coach 10 at 3-month assessment		
	F test (*df*)	Treatment group *P* value	Cohen *d* (adjusted)	*F* test (*df*)	Treatment group *P* value	Cohen *d* (adjusted)
PCL-5[a] total score	0.06 (1,91)	.81	0.05	0.01 (1,81)	.93	0.02
PHQ-9[b] total score	0.07 (1,91)	.80	0.05	0.05 (1,81)	.82	0.05
DERS-36[c] total score	0.05 (1,91)	.83	0.05	0.69 (1,81)	.41	0.18
IIP-32[d] mean score	0.03 (1,91)	.87	0.03	3.00 (1,81)	.09	0.38
WHODAS-21[e] total score	0.27 (1,92)	.60	0.11	0.31 (1,82)	.58	0.12

[a]PCL-5: posttraumatic stress disorder checklist for DSM-5

[b]PHQ-9: Patient Health Questionaire-9

[c]DERS: Difficulties in Emotion Regulation Scale

[d]IIP-32: Inventory in Interpersonal Problems-32

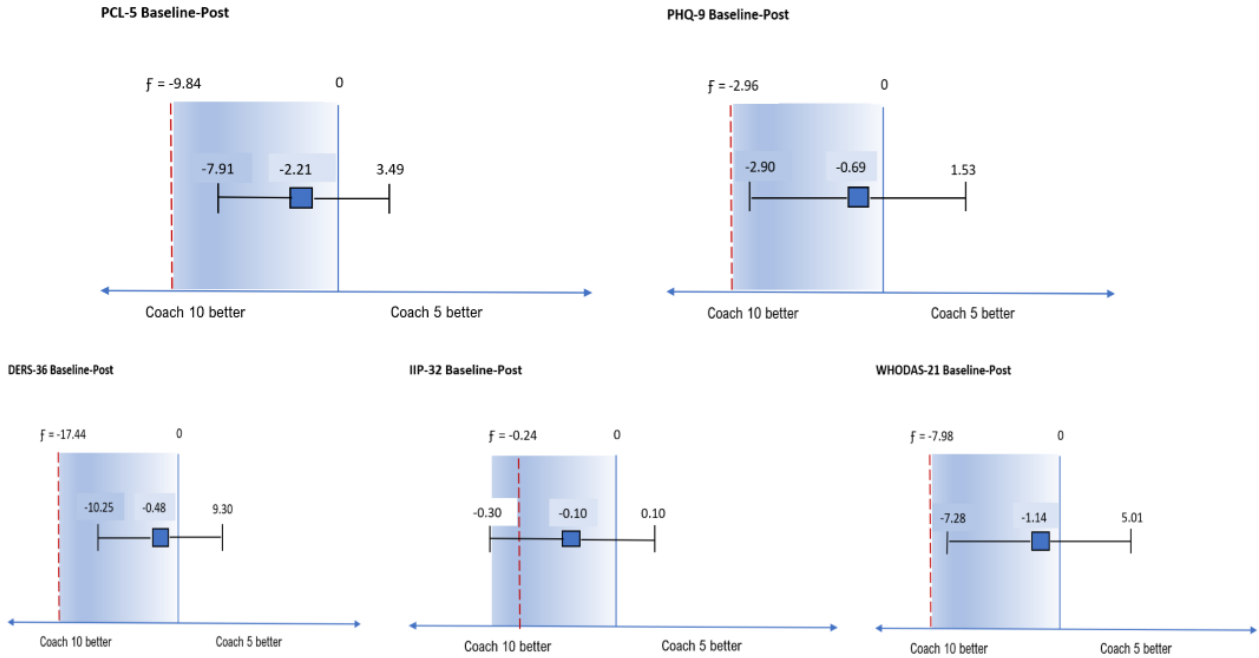
[e]WHODAS-21: brief version of World Health Organization Disability Assessment Schedule-2

Tests of Noninferiority

Figure 2 shows the CIs for the mean difference in scores for the 5 outcome measures between conditions for the completer sample at posttreatment assessment. If Coach10 is superior to Coach5, the difference in the change score is negative. The value of *F* (shown as the dotted vertical line) is the predetermined minimum clinically significant difference (ie, the margin of noninferiority). For 4 of 5 outcomes, the lower bound of the 95% CI was less than 0 and yet also greater than the margin of noninferiority (ie, it is in the shaded area) and the higher bound of the 95% CI was greater than 0, which indicates noninferiority of Coach5 with respect to Coach10 in improvement from baseline to posttreatment assessment. For the IIP, the lower bound of the 95% CI was less than 0 but also less than the margin of noninferiority (ie, it crossed the margin of inferiority), which indicates inconclusive results.

When examining changes from baseline to the 3-month follow up, the lower bound of the 95% CI was within the shaded area for PCL-5, DERS-36, IIP-32, and WHODAS-21, indicating the noninferiority of Coach5 with respect to Coach10 in improvement from baseline to the 3-month follow up for these 4 outcomes. Conversely, for the PHQ-2 the lower bound of the 95% CI was less than the margin of noninferiority (ie, it crossed the margin of noninferiority), which indicates inconclusive results.

Figure 2. Noninferiority figures for 5 outcomes. PCL-5: posttraumatic stress disorder checklist for DSM-5; PHQ-9: Patient Health Questionaire-9; DERS: Difficulties in Emotion Regulation Scale; IIP-32: Inventory in Interpersonal Problems-32; WHODAS-21: brief version of World Health Organization Disability Assessment Schedule-2.



Therapeutic Alliance

The average WAI score in the Coach5 and Coach10 conditions was not significantly different, with a mean of 6.66 (SD 0.49) and 6.54 (SD 0.50) respectively (t_{128}=1.34, *P*=.18). The score for WAI items ranged from 1 to 7, with a higher score indicating a better relationship. The results of this assessment indicate that participants' perception of the working alliance with their therapists was very high, and equally so, across the 2 conditions.

Completion Rates

The total number of modules completed by the Coach5 and Coach10 groups did not differ (6.88, SD 4.19 vs 6.66, SD 4.11, respectively; t_{200}=0.37, *P*=.71). The percentage of participants who completed half of the program (up to and including the fifth module) did not differ between the Coach5 (72/101, 71.3%) and Coach10 (67/101, 66.4%) groups. The total percentage of participants who completed all 10 modules also did not differ between the Coach5 (59/101, 58.4%) and Coach10 (52/101, 51.5%) groups (χ^2_1=0.93, *P*=.32).

Web Usage

Coach5 participants spent an average of 31.07 minutes (SD 8.30) per module with time per module ranging from 43.66 to 18.80 minutes. Coach10 participants spent an average of 33.06 minutes (SD 11.05) per module with a range of 38.98 to 14.36 minutes. The average amount of time participants who completed the program spent on the website was 341.83 minutes (SD 160.46) for Coach5 and 363.63 (SD 160.46) for Coach10. The difference in total time spent on the program between the 2 conditions was not significant (t_{20} =2.02, *P*=.16).

Therapist Session Time and Activities

As expected, the Coach5 therapists spent significantly less time than the Coach10 therapists in session time (t_{28}=28.44, *P*<.001), with the Coach5 therapists spending an average of 332.67 minutes (SD 49.10) across the 5 coaching sessions and the Coach 10 therapists spending an average of 505.09 minutes (SD 111.79) across all 10 sessions. The number of contacts (phone calls, instant messages, or emails) did not differ between the Coach5 and the Coach10 therapists, with values of 2.83 (SD 3.28) and 2.02 (SD 2.85), respectively. The number of phone calls for interventional purposes (crisis management or motivational interventions) was low and did not differ between the Coach5 (M 1.10, SD 1.65) and Coach10 therapists (M 0.62, SD 1.06).

Therapist Adherence Ratings

The adherence rating indicates the percentage of intervention items that were completed by the therapist in each session. Overall, the average adherence ratings were high and were statistically similar across both the Coach5 (M 0.90, SD 0.14) and Coach10 conditions (M 0.97, SD 0.06).

Discussion

Five sessions of therapist support for a transdiagnostic trauma-informed intervention, delivered to veterans with moderate to severe symptoms of PTSD, depression, or both was found to be noninferior to a 10-session delivery approach across several outcomes. Participants obtained significant benefits from webSTAIR with both conditions. In addition, therapeutic alliance was strong and did not differ between 5-session and 10-session delivery. The time participants spent on the program was equivalent, as was the completion rate. This study demonstrates the feasibility and effectiveness of blended models in routine clinical care for patients with moderate to severe symptoms. It also contributes to the exploration of the therapist role in technology-based intervention and provides parameters regarding the amount of therapist support that is associated with good outcomes for patients.

Symptom reduction was significant for PTSD, depression, emotion regulation, interpersonal problems, and overall functioning at both posttreatment and a 3-month follow up for both treatment conditions and was associated with predominantly moderate to large effect sizes. Most of the measures demonstrated noninferiority at both posttreatment and follow up. Two measures, interpersonal problems and depression, were associated with some variability, indicating either noninferiority or an inconclusive result depending on the timepoint. However, overall, all 5 outcomes for the Coach5 group were not inferior to the Coach 10 group at any time point.

On average, participants spent approximately 30 minutes per module and completed about 6 to 7 modules. The percentage of participants who completed the entire program (ie, all 10 modules) was 58% (59/101) for the Coach5 group and 51% (51/101) for the Coach10 group. The level of success that this represents is unknown. Because the project was an evaluation program, not a research study, there were no special efforts made to retain participants in the project (eg, ensure before enrollment that a participant had no plans for long travel or hospitalization that might disrupt completion), as would occur in a research study. RCTs of internet-based interventions for PTSD populations have reported drop-out rates of 25%, and by this measure, the completion rates in this study are inferior. On the other hand, in studies of naturalistic use of technology-based interventions, retention rates appear to fall to 20% by the fifth session [35], and by this measure, our completion rate is superior. Additional studies are needed to determine the completion rates associated with blended therapies delivered in a clinical service context.

It is notable that the large majority of patients enrolled in the program reported severe symptoms of PTSD, depression, or both, with the overwhelming majority (187/202, 92.6%) meeting criteria for a probable diagnosis of one or the other disorder and over 80% (162/202) meeting a probable diagnosis for both. The results indicate that either of these blended approaches is effective and safe for highly symptomatic patients. It is often thought that technology-based therapies are most appropriate primarily for individuals with low to moderate symptoms. Technology-based programs have typically been introduced into mental health services as part of "stepped care" delivered to enrolling patients whose symptoms are relatively mild or as part of maintenance care once face-to-face treatment has been completed. Our findings indicate that technology-based care with therapist support is effective even for severely affected patients. It is interesting to speculate whether referral of relatively severe patients to this program by clinical therapists

was motivated by knowledge of high level of therapist involvement. The blended therapy model may introduce a new way of thinking about mental health care that provides high quality care to patients but reduces therapist burden in terms of time and effort.

The overall noninferiority of Coach5 compared to Coach10 provides therapists and mental health services reassurance that a reduction in therapist time does not lead to a reduction in good outcomes for patients or in the relationship with the patient. Therapists have been concerned that reduction in therapist time might lead to increased need to manage crisis events [19,20], representing a reduction in quality of care for patients and their well-being as well as a "hidden cost" in regard to therapist time and effort. However, the Coach5 and Coach10 programs were associated with a relatively low need for additional support and intervention, with therapists in both conditions reporting on average 1 phone session per patient and 3 contacts related to rescheduling of sessions in addition to the technology-based program.

Finally, the finding of noninferiority may be viewed as counterintuitive if one assumes a "dose-response" effect in psychotherapy treatment, where more therapist intervention is better. We anchored the study to a 1:1 ratio of therapist session to self-guided work. This ratio represents the maximum dose of therapist support in a program in which the technology intervention is primary. Its selection was supported by evidence from previous studies indicating the success of this ratio [15,36]. The absence of worse outcomes with a reduction in the number of therapist sessions can be interpreted in several ways. However, one possibility is that there are mechanisms at work other than an additive dose-response effect. For example, it is possible that self-guided work and the therapist sessions made unique and complementary contributions to the program outcomes and may ultimately have provided greater benefits than therapist-only or internet-only approaches [18]. Evidence for this view is supported by qualitative analyses from interviews completed in one of our earlier studies [21], which found that the self-guided work facilitated a sense of autonomy and mastery, while the therapist sessions provided emotional and practical support as well as clarity in tailoring the tools to specific problems and life experience. The potential presence of a dynamic, reciprocal relationship between these 2 treatment components deserves further investigation via assessment of cross-lagged effects and identification of underlying mechanisms of change.

The strengths of this study include the delivery of the treatment in a usual care context, provided by clinic staff and delivered to trauma-exposed patients with relatively severe and diverse symptoms and problems. Two limitations of the study are the absence of randomization, which could have resulted in unidentified patient factors influencing outcomes, and the potential lack of generalizability of the findings from veterans to other trauma-exposed populations. Future studies are warranted evaluating other combinations of therapist-supported and self-guided work and delivery to different trauma populations. In addition, studies identifying patient characteristics as moderators of outcomes in high versus low doses of therapist support will help create flexible technology-based intervention programming that facilitates engagement of a greater number of individuals and tailoring of therapist time and attention relative to client need.

Conflicts of Interest

None declared.

References

1. Kuester A, Niemeyer H, Knaevelsrud C. Internet-based interventions for posttraumatic stress: A meta-analysis of randomized controlled trials. Clin Psychol Rev 2016 Feb;43:1-16. [doi: 10.1016/j.cpr.2015.11.004] [Medline: 26655959]
2. Lewis C, Roberts NP, Simon N, Bethell A, Bisson JI. Internet-delivered cognitive behavioural therapy for post-traumatic stress disorder: systematic review and meta-analysis. Acta Psychiatr Scand 2019 Dec;140(6):508-521. [doi: 10.1111/acps.13079] [Medline: 31359407]
3. Sijbrandij M, Kunovski I, Cuijpers P. Effectiveness of Internet-Delivered Cognitive Behavioral Therapy for Posttraumatic Stress Disorder: A Systematic Review and Meta-Analysis. Depress Anxiety 2016 Sep;33(9):783-791. [doi: 10.1002/da.22533] [Medline: 27322710]
4. Erbe D, Eichert H, Riper H, Ebert DD. Blending Face-to-Face and Internet-Based Interventions for the Treatment of Mental Disorders in Adults: Systematic Review. J Med Internet Res 2017 Sep 15;19(9):e306 [FREE Full text] [doi: 10.2196/jmir.6588] [Medline: 28916506]
5. Andersson G, Titov N, Dear BF, Rozental A, Carlbring P. Internet-delivered psychological treatments: from innovation to implementation. World Psychiatry 2019 Feb;18(1):20-28 [FREE Full text] [doi: 10.1002/wps.20610] [Medline: 30600624]
6. Wentzel J, van DVR, Bohlmeijer ET, van GJEWC. Mixing Online and Face-to-Face Therapy: How to Benefit From Blended Care in Mental Health Care. JMIR Ment Health 2016;3(1):e9 [FREE Full text] [doi: 10.2196/mental.4534] [Medline: 26860537]
7. Overholser J. Adapting computerized treatments into traditional psychotherapy for depression. Stud Health Technol Inform 2012;181:32-36. [Medline: 22954823]
8. Hantsoo L, Epperson CN, Thase ME, Kim DR. Antepartum depression: treatment with computer-assisted cognitive-behavioral therapy. Am J Psychiatry 2013 Aug;170(8):929-930 [FREE Full text] [doi: 10.1176/appi.ajp.2013.13020254] [Medline: 23903340]

9. Jacmon J, Malouff J, Taylor N. Treatment of Major Depression: Effectiveness of Cognitive Behavior Therapy with an Internet Course as a Central Component. EJAP 2010 Jan 03;5(2):1-8. [doi: 10.7790/ejap.v5i2.153]
10. Nordmo M, Sinding AI, Carlbring P, Andersson G, Havik OE, Nordgreen T. Internet-delivered cognitive behavioural therapy with and without an initial face-to-face psychoeducation session for social anxiety disorder: A pilot randomized controlled trial. Internet Interventions 2015 Nov;2(4):429-436. [doi: 10.1016/j.invent.2015.10.003]
11. Wilhelmsen M, Lillevoll K, Risør MB, Høifødt R, Johansen M, Waterloo K, et al. Motivation to persist with internet-based cognitive behavioural treatment using blended care: a qualitative study. BMC Psychiatry 2013;13:296 [FREE Full text] [doi: 10.1186/1471-244X-13-296] [Medline: 24199672]
12. Andersson G, Carlbring P, Holmström A, Sparthan E, Furmark T, Nilsson-Ihrfelt E, et al. Internet-based self-help with therapist feedback and in vivo group exposure for social phobia: a randomized controlled trial. J Consult Clin Psychol 2006 Aug;74(4):677-686. [doi: 10.1037/0022-006X.74.4.677] [Medline: 16881775]
13. Kleiboer A, Donker T, Seekles W, van SA, Riper H, Cuijpers P. A randomized controlled trial on the role of support in Internet-based problem solving therapy for depression and anxiety. Behav Res Ther 2015 Sep;72:63-71. [doi: 10.1016/j.brat.2015.06.013] [Medline: 26188373]
14. Sethi S, Campbell AJ, Ellis LA. The Use of Computerized Self-Help Packages to Treat Adolescent Depression and Anxiety. Journal of Technology in Human Services 2010 Aug 31;28(3):144-160. [doi: 10.1080/15228835.2010.508317]
15. Bauer A, Amspoker AB, Fletcher TL, Jackson C, Jacobs A, Hogan J, et al. A Resource Building Virtual Care Programme: improving symptoms and social functioning among female and male rural veterans. Eur J Psychotraumatol 2021 Feb 04;12(1):1860357 [FREE Full text] [doi: 10.1080/20008198.2020.1860357] [Medline: 34025913]
16. Jain S, Ortigo K, Gimeno J, Baldor DA, Weiss BJ, Cloitre M. A Randomized Controlled Trial of Brief Skills Training in Affective and Interpersonal Regulation (STAIR) for Veterans in Primary Care. J Trauma Stress 2020 Aug;33(4):401-409. [doi: 10.1002/jts.22523] [Medline: 32506563]
17. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
18. van DVR, Witting M, Riper H, Kooistra L, Bohlmeijer ET, van GLJ. Blending online therapy into regular face-to-face therapy for depression: content, ratio and preconditions according to patients and therapists using a Delphi study. BMC Psychiatry 2014;14(1):355 [FREE Full text] [doi: 10.1186/s12888-014-0355-z] [Medline: 25496393]
19. Becker EM, Jensen-Doss A. Computer-assisted therapies: examination of therapist-level barriers to their use. Behav Ther 2013 Dec;44(4):614-624. [doi: 10.1016/j.beth.2013.05.002] [Medline: 24094787]
20. Cook JE, Doyle C. Working alliance in online therapy as compared to face-to-face therapy: preliminary results. Cyberpsychol Behav 2002 Apr;5(2):95-105. [Medline: 12025884]
21. Fletcher TL, Amspoker AB, Wassef M, Hogan JB, Helm A, Jackson C, et al. Increasing access to care for trauma‐exposed rural veterans: A mixed methods outcome evaluation of a web‐based skills training program with telehealth‐delivered coaching. The Journal of Rural Health 2021 Oct 14:1-8. [doi: 10.1111/jrh.12628]
22. Brooks E, Dailey N, Bair B, Shore J. Rural Women Veterans Demographic Report: Defining VA Users' Health and Health Care Access in Rural Areas. The Journal of Rural Health 2013 Jul 19;30(2):146-152. [doi: 10.1111/jrh.12037]
23. Prins A, Ouimette P, Kimerling R, Camerond RP, Hugelshofer DS, Shaw-Hegwer J, et al. The primary care PTSD screen (PC–PTSD): development and operating characteristics. Prim Care Psych 2004 Jan 01;9(1):9-14. [doi: 10.1185/135525703125002360]
24. Kroenke K, Spitzer RL, Williams JBW. The Patient Health Questionnaire-2: validity of a two-item depression screener. Med Care 2003 Nov;41(11):1284-1292. [doi: 10.1097/01.MLR.0000093487.78664.3C] [Medline: 14583691]
25. Weathers FW, Blake DD, Schnurr PP, Kaloupek DG, Marx BP, Keane TM. The Life Events Checklist for DSM-5 (LEC-5). 2013. URL: https://www.ptsd.va.gov/professional/assessment/te-measures/life_events_checklist.asp [accessed 2022-04-12]
26. Blevins CA, Weathers FW, Davis MT, Witte TK, Domino JL. The Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5): Development and Initial Psychometric Evaluation. J Trauma Stress 2015 Dec;28(6):489-498. [doi: 10.1002/jts.22059] [Medline: 26606250]
27. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [Medline: 11556941]
28. Gratz KL, Roemer L. Multidimensional Assessment of Emotion Regulation and Dysregulation: Development, Factor Structure, and Initial Validation of the Difficulties in Emotion Regulation Scale. Journal of Psychopathology and Behavioral Assessment 2004 Mar;26(1):41-54. [doi: 10.1023/B:JOBA.0000007455.08539.94]
29. Barkham M, Hardy GE, Startup M. The IIP-32: a short version of the Inventory of Interpersonal Problems. Br J Clin Psychol 1996 Feb;35(1):21-35. [doi: 10.1111/j.2044-8260.1996.tb01159.x] [Medline: 8673033]
30. Üstün T, Chatterji S, Kostanjsek N, Rehm J, Kennedy C, Epping-Jordan J, et al. Developing the World Health Organization Disability Assessment Schedule 2.0. In: Bull World Health Organ. Malta: World Health Organization; May 20, 2010:815-823.
31. Horvath AO, Greenberg LS. Development and validation of the Working Alliance Inventory. Journal of Counseling Psychology 1989;36(2):223-233. [doi: 10.1037/0022-0167.36.2.223]
32. Fung H. Treating survivors of childhood abuse and interpersonal trauma: STAIR Narrative Therapy (2nd ed.). In: China Journal of Social Work. New York, Ny: Guilford Press; Mar 13, 2022:1-3.

33. Serfaty M, Shafran R, Vickerstaff V, Aspden T. A pragmatic approach to measuring adherence in treatment delivery in psychotherapy. Cogn Behav Ther 2020 Sep;49(5):347-360. [doi: 10.1080/16506073.2020.1717594] [Medline: 32114905]
34. Oczkowski SJW. A clinician's guide to the assessment and interpretation of noninferiority trials for novel therapies. Open Med 2014;8(2):e67-e72 [FREE Full text] [Medline: 25009686]
35. Eysenbach G. The Law of Attrition. J Med Internet Res 2005 Mar 31;7(1):e11. [doi: 10.2196/jmir.7.1.e11]
36. Craske MG, Stein MB, Sullivan G, Sherbourne C, Bystritsky A, Rose RD, et al. Disorder-Specific Impact of Coordinated Anxiety Learning and Management Treatment for Anxiety Disorders in Primary Care. Arch Gen Psychiatry 2011 Apr 04;68(4):378-388. [doi: 10.1001/archgenpsychiatry.2011.25]

Abbreviations

CBT: cognitive behavioral therapy
DERS: Difficulties in Emotion Regulation Scale
DSM-5: Diagnostic and Statistical Manual, Fifth Edition
IIP: Inventory of Interpersonal Problems
ITT: intent-to-treat
PCL-5: PTSD Checklist for DSM-5
PHQ: Patient Health Questionnaire
PST: problem-solving therapy
PTSD: posttraumatic stress disorder
RCT: randomized controlled trials
WAI: Working Alliance Inventory
WHODAS: World Health Organisation Disability Assessment Schedule

Edited by J Torous; submitted 24.08.21; peer-reviewed by A Pencer, I Choi; comments to author 31.12.21; accepted 18.02.22; published 27.04.22.

Please cite as:
Cloitre M, Amspoker AB, Fletcher TL, Hogan JB, Jackson C, Jacobs A, Shammet R, Speicher S, Wassef M, Lindsay J
Comparing the Ratio of Therapist Support to Internet Sessions in a Blended Therapy Delivered to Trauma-Exposed Veterans: Quasi-experimental Comparison Study
JMIR Ment Health 2022;9(4):e33080
URL: https://mental.jmir.org/2022/4/e33080
doi:10.2196/33080
PMID:35475777

Original Paper

Detecting Mental Health Behaviors Using Mobile Interactions: Exploratory Study Focusing on Binge Eating

Julio Vega[1], PhD; Beth T Bell[2], PhD; Caitlin Taylor[3], MRes; Jue Xie[4], PhD; Heidi Ng[4], MDiet; Mahsa Honary[5], PhD; Roisin McNaney[4], PhD

[1]Department of Medicine, University of Pittsburgh, Pittsburgh, PA, United States
[2]University of York, York, United Kingdom
[3]York St John University, York, United Kingdom
[4]Department of Human Centred Computing, Monash University, Clayton, Australia
[5]Lancaster University, Lancaster, United Kingdom

Corresponding Author:
Roisin McNaney, PhD
Department of Human Centred Computing
Monash University
Wellington Rd
Clayton, VIC 3800
Australia
Phone: 61 3 9902 6000
Email: roisin.mcnaney@monash.edu

Abstract

Background: Binge eating is a subjective loss of control while eating, which leads to the consumption of large amounts of food. It can cause significant emotional distress and is often accompanied by purging behaviors (eg, meal skipping, overexercising, or vomiting).

Objective: The aim of this study was to explore the potential of mobile sensing to detect indicators of binge-eating episodes, with a view toward informing the design of future context-aware mobile interventions.

Methods: This study was conducted in 2 stages. The first involved the development of the DeMMI (Detecting Mental health behaviors using Mobile Interactions) app. As part of this, we conducted a consultation session to explore whether the types of sensor data we were proposing to capture were useful and appropriate, as well as to gather feedback on some specific app features relating to self-reporting. The second stage involved conducting a 6-week period of data collection with 10 participants experiencing binge eating (logging both their mood and episodes of binge eating) and 10 comparison participants (logging only mood). An optional interview was conducted after the study, which discussed their experience using the app, and 8 participants (n=3, 38% binge eating and n=5, 63% comparisons) consented.

Results: The findings showed unique differences in the types of sensor data that were triangulated with the individuals' episodes (with nearby Bluetooth devices, screen and app use features, mobility features, and mood scores showing relevance). Participants had a largely positive opinion about the app, its unobtrusive role, and its ease of use. Interacting with the app increased participants' awareness of and reflection on their mood and phone usage patterns. Moreover, they expressed no privacy concerns as these were alleviated by the study information sheet.

Conclusions: This study contributes a series of recommendations for future studies wishing to scale our approach and for the design of bespoke mobile interventions to support this population.

(JMIR Ment Health 2022;9(4):e32146) doi:10.2196/32146

KEYWORDS

eating disorder; binge eating; mental health; mobile sensing; context-aware computing; NAP; EMA; mobile phone

Introduction

Background

Binge eating is classified as a distinct period during which an individual experiences a subjective loss of control over eating, eats notably more or differently than usual, and feels unable to stop eating or limit the type or amount of food eaten [1]. It is thought to affect approximately 5% of women and 4% of men worldwide in some form over *their life course* [2]. Binge eating is a precursor for, and is symptomatic of, clinical eating disorders, including bulimia nervosa and binge eating disorder [3]. In bulimia nervosa, binge eating is typically followed by the purging of calories from the body in an attempt to counteract a binge eating episode—for example, through vomiting, use of laxatives or diuretics, extreme dieting, or excessive exercise—whereas individuals with binge eating disorder do not engage in purging practices [4]. Binge eating is often a hidden behavior conducted in secret [5] and can lead to extreme feelings of shame, worthlessness, and lack of control, which can have major impacts on an individual's mental health and emotional well-being [6]. Research has shown that clinical binges (ie, from those with an eating disorder diagnosis) yield the same subjective experiences before and after the binge as those reported from a nonclinical binge eating episode [7].

Anxiety and depression are prevalent in people with binge-eating behaviors [8-10], and research indicates that binge eating is used as a strategy for regulating negative affect (often used as an umbrella term to refer to emotive experiences such as mood, emotion, impulses, and stress response [11]). In addition, as people with binge-eating behaviors are often within the normal to obese BMI category range [12], people in this population report feelings of being unworthy of mental health support and not identifying themselves as having an eating disorder. All this, paired with the secretive nature of binge-eating behaviors, means that the identification and treatment of binge eating are particularly challenging [5].

A growing body of literature has identified opportunities for mobile sensing (ie, the collection and use of data collected from sensors embedded within mobile devices such as smartphones) to detect mental health behaviors such as schizophrenia [13], bipolar disorder [14], and depression [15]. With the exception of the study by Juarascio et al [16], who looked at heart rate variability as a risk predictor for emotional eating episodes, no research has explicitly investigated the potential for mobile sensing in relation to disordered eating behaviors. This is an underexplored area of research, which has significant opportunities for (1) identifying the contextual and situational factors associated with binge-eating episodes and (2) providing improved access to support and behavior prevention through context-aware interventions [17,18].

Our research aimed to contextualize the experiences of people engaging in self-identified episodes of binge eating, with or without subsequent purging activities, by understanding the types of activities that a person might be engaging with in and around the occurrence of an episode. As this was a first-of-its-kind exploratory work, we used a broad range of mobile sensors that already exist in mobile phones (eg, location sensors can indicate whether someone is spending a lot of time at home; movement sensors can give us an idea of how much activity a person has been engaging in; app usage sensors can indicate how much time someone is spending on social media or healthy eating and fitness apps). We asked participants to provide daily self-reports of their mood (collected in both the morning and evening) and to self-report any episodes of binge eating (logged through a button press to capture the time of the episode, with the option of providing further information in the form of free-flowing text). This provided us with a measure of the differences in behavioral features extracted from smartphone data on days with and without incidents of binge eating in an attempt to inform future context-aware mobile interventions to support this population.

We describe the development of the DeMMI (Detecting Mental health behaviors through Mobile Interactions) app, which was refined in consultation with service users. We then describe a 6-week remote study of 20 participants (10 with experiences of binge eating and 10 without any mental health issues, who reported twice daily mood logs and acted as a comparison group). Our contributions from this paper are three-fold: (1) first, we provide a set of reflections around the challenges of conducting work with the binge-eating population and the benefits of remote, anonymous engagement; (2) second, we provide unique insights into the successes and challenges surrounding our mobile sensing approach (from a pilot study perspective) and how this might be better scaled in the future for larger-scale studies over longer periods; and (3) finally, we provide a set of recommendations for the design of future context-aware interventions aimed at supporting people experiencing binge-eating behaviors.

Use of Ecological Momentary Assessment for Monitoring Mood and Binge Eating

Cross-sectional and longitudinal studies are useful for understanding the long-term risk factors of poor mental health, including those that contribute to binge eating (see the study by Burton and Abbott [19] for review). However, they are much less useful for understanding the more immediate contextual and situational factors that directly contribute to fluctuations in mood that accompany binge-eating behavior. Ecological Momentary Assessment (EMA) techniques are better able to provide information on these contextual factors. EMA involves the recording of problematic moods, thoughts, and behaviors, as well as the events that immediately precede them, to identify predictive patterns. Although early EMA involved the use of paper-based diaries, more recent research has used digital tools (eg, mobile phones and web-based apps) to gather real-time self-reported data. Data collected through digitally enhanced EMA not only have the potential to enhance the understanding of binge eating in research settings but can also be used in therapeutic settings, as well as by individuals, to better understand and monitor individual patterns related to poor mental health. Indeed, a large proportion of smartphone apps specifically for binge eating [20-23], as well as for mental health [24-30], typically involves self-reporting and repeatedly prompting participants over time [31]. Such mobile monitoring apps are typically well-received by young people [32].

EMA relies on the self-report of affective states; however, there is much heterogeneity in the way these affective states are measured. The self-report questions used in smartphone-based EMA are often literal translations of clinical tools [24,26,33,34]. For example, The Positive and Negative Affect Scale (PANAS) involves participants indicating the extent to which they are experiencing 10 types of positive (eg, *alert*) and negative (eg, *upset*) affect using 5-point Likert scales [35]. Although the original PANAS is generally considered too long to be applied with high frequency in EMA [36], subscales and shortened forms have been delivered using smartphone EMA with good response rates, even when used multiple times daily [33]. Furthermore, other approaches have attempted to reduce the burden on participants through the use of visual scales. For example, the Self-Assessment Manikin uses 3 icon-based scales to measure pleasure, arousal, and dominance [37]. In their original form, the icons are abstract outlines of a human-like figure; however, other implementations have also used more realistic representations [26]. Studies have found preferences for these briefer visual scales when compared with more repetitive traditional EMA [26,38].

Mobile Sensing Approaches for Supporting Mental Health and Binge Eating

The ubiquity and sensing capabilities of smartphones make them attractive tools for the passive collection of multimodal sensor data 24/7. Compared with EMA, they are objective, less burdensome, have a higher temporal resolution, and provide rich data streams to infer aspects of users' social context and behavior in naturalistic conditions [39,40]. Research has shown the potential of these data to monitor and support mental health conditions, including depression [15,38,41-44], schizophrenia [13,45-48], bipolar disorder [14,49-53], stress [54-58] and anxiety [59-62]. Typically, behavioral features (metrics quantifying aspects of individuals' routines and activities) are computed from smartphone and wearable data, and their role in tracking, classifying, or predicting events of interest (eg, depressive states and hospital readmission) is explored via analytical methods. Clinical scales, medical records, and patient-reported outcomes are often used as the ground truth to validate models built on top of behavioral features. However, to the best of our knowledge, only Juarascio et al [16] explored mobile sensing in the context of eating episodes associated with negative emotions. They monitored 21 people with clinically significant emotional eating behaviors for 4 weeks using a wrist-worn device. The results showed that time and frequency domain features of heart rate variability could be used to classify 30-minute periods with and without emotional eating episodes better than chance. Although they showcased the importance of wearable data in supporting binge-eating monitoring, patient perspectives on mobile sensing and the role of smartphone phone data remain unexplored.

For eating disorders more broadly, research has shown that smartphone apps could increase patients' access to treatment [18] because of the anonymity they afford when considering the barriers people face in seeking clinical help (eg, shame and fear of stigma) [5]. Furthermore, in light of near-ubiquitous smartphone use in modern society, these devices are uniquely positioned to support access to resources by promoting help-seeking and self-management behaviors [18,63]. Smartphones can enable personalized monitoring, which can aid in the identification of high-risk situations derived from behavioral and situational contexts extracted from multimodal data. Alongside their capabilities for digital intervention provision, they offer a powerful platform for delivering support at optimal times [17,18]. Currently, there are a number of apps designed primarily for people with disordered eating behaviors that have been studied in the literature, including *Recovery Record* and *RiseUp* [18,64]. Both apps use self-monitoring techniques and provide users with a set of coping strategies to try. In particular, *Recovery Record* uses EMA to facilitate self-monitoring [65,66] and has some features that are similar to the DeMMI app (eg, the ability to track mood and episodes of binge eating). However, this requires a significant amount of active tracking from the user (ie, daily diaries, logging of meals, and the feelings surrounding them), which can be a laborious task. Our study is interested in how passive approaches to monitoring can be leveraged, allowing us to potentially automatically detect contextual or situational triggers for episodes of binge eating. This would ultimately remove some of the tracking burdens from the user and provide indications of where digital interventions might be best positioned to help them. As a first step toward this goal, our study explores the individual differences in smartphone behavioral features between days with and without binge-eating episodes, framed around the experiences and needs of our users, as well as the current and future challenges faced by the mobile sensing community in detecting binge-eating episodes and delivering digital interventions.

Methods

Overview

This study was conducted in 2 stages. The first involved the development of the DeMMI app. As part of this, we conducted a consultation session to explore whether the types of sensor data we were proposing to capture were seen as useful and appropriate, as well as to gather feedback on some specific app features relating to self-report. The second stage involved conducting a 6-week period of data collection with 10 participants experiencing binge eating (logging both their mood and episodes of binge eating) and 10 comparison participants (logging only mood).

First, we present the ethical considerations of this study. We then present the 2 stages of research separately, first describing the development of the DeMMI app before moving on to discuss our fieldwork study methods and findings.

Ethical Considerations

Ethical approval for this work was obtained from York St John, United Kingdom, University Ethics Committee (RECPSY00012), and the study adhered to the British Psychological Society ethical guidelines. The activities for the stage 1 consultation session were constructed collaboratively within the research team, which was made up of a clinical mental health professional and multiple highly experienced researchers with expertise in engaging people with a range of mental health issues in qualitative workshops and interviews.

The session was led by one of these experts and supported by an experienced postgraduate student working in the field of disordered eating and self-harm. Both facilitators were careful in creating an open and nonjudgmental space during the sessions. In stage 2, the research team created a safeguarding protocol before the commencement of remote participant recruitment. Participants were fully informed that their data were not being actively monitored and that the research team members were not mental health professionals. However, we conducted weekly well-being checks via WhatsApp or SMS text messaging (depending on participant preference) during the study. These asked participants how they were managing with the study and offered them an opportunity to reach out for support if required. Although we did not have any requests for support during the study, we were prepared to point the participants to local services.

Phase 1: DeMMI App Development

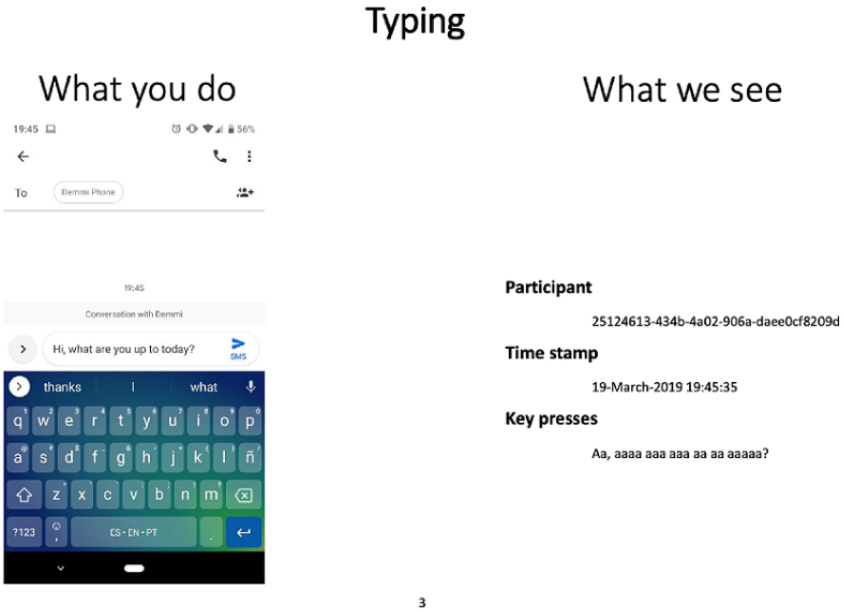
Overview

Given the sensitivity of the data we wanted to collect, as well as an acknowledgment that reporting on disordered eating behaviors might, in itself, be considered a trigger, we first conducted a consultation to (1) gain an understanding of early perceptions toward the data we intended to collect with the app; (2) gather ideas on the best ways of collecting self-reported data in a sensitive way; and (3) understand any specific opinions potential users might have around the rate of data capture, anonymity, and offers of support. We report the main insights drawn from this consultation to provide context for our design decisions before moving on to describe the DeMMI app itself.

Consultation Activity

We engaged service users (n=2) from York, United Kingdom, in a consultation session to develop and iterate key design decisions related to the DeMMI app. One of the service users had lived experience of binge eating, and the other had lived experience of self-harm (as a parallel project aimed to explore the feasibility of DeMMI app use among individuals who self-harm). The consultation explored the participants' perceptions of mobile data collection through both passive and active means. In terms of passive data capture, we explored participants' general perceptions of mobile sensed data collection and sought specific feedback on the intended data collection proposed within the study—specifically, how they felt about us collecting potentially invasive sensor data. To facilitate this activity, we used a bespoke *What You Do/What We See* resource (Figure 1; the full resource can be viewed in Multimedia Appendix 1), which showed participants the output of every sensor. We discussed each sensor in turn and responded transparently to any questions. In terms of active data capture, we explored participants' general perceptions of using apps to log mood and behavior, as well as their specific feedback on our proposed app logging mechanisms in the DeMMI app. There were 2 specific features for which we were interested in gaining feedback. First, we were interested in understanding service users' perspectives on the use of the PANAS (which is widely used in EMA research with the general population [33-35]) to assess mood multiple times each day. Second, we were interested in the use of a 1-click logging mechanism of binge-eating episodes based on the *oops* button developed by Tulu et al [67]. During the session, data were audio recorded to allow the team to listen back to the session; however, as we only had 2 participants, we did not thematically analyze the data. Instead, we took notes during the session and cross-checked all key findings with the participants.

Figure 1. Example of one of our What You Do/What We See slides used to explain to participants what smartphone data we wanted to capture (ie, the participants' interaction) and exactly what we would see from this interaction. Slides were created to represent the following: app notifications, app use, typing, battery, calls, SMS text messages, data sent or received, screen locks or unlocks, time zone, Wi-Fi nearby devices, Bluetooth nearby devices, ambient light, weather, location, ambient noise, and activity recognition data.



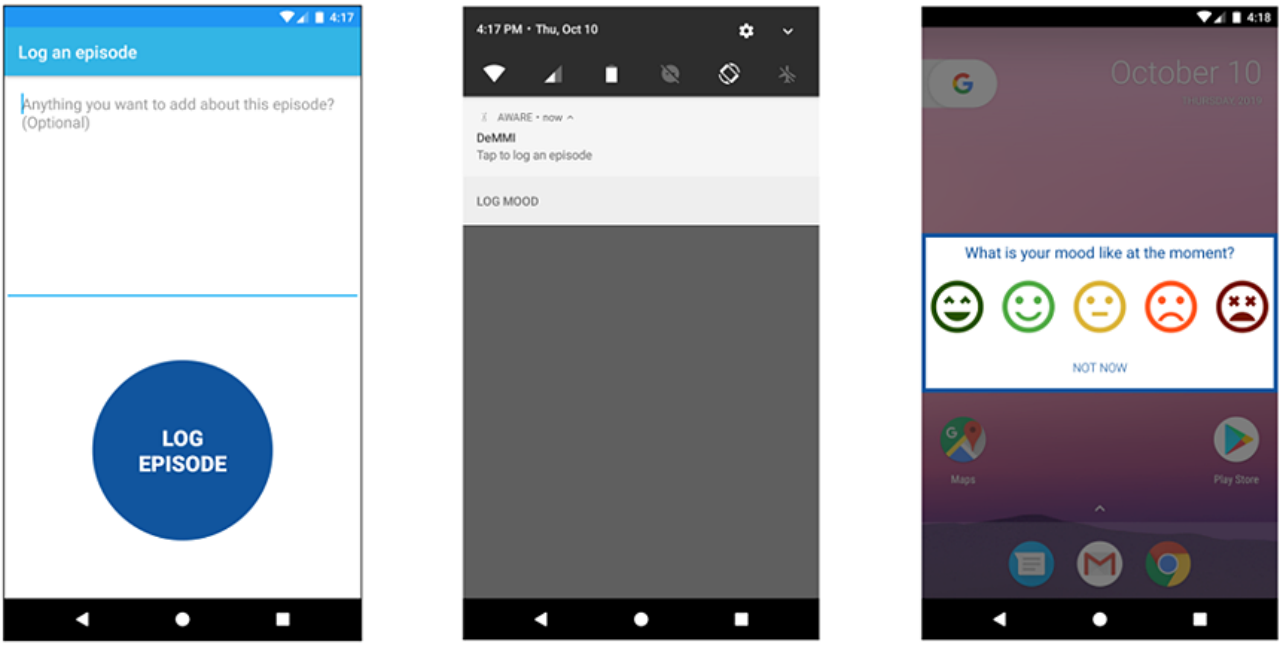
The key findings from the consultation may be summarized as follows:

1. Participants were generally positive about the use of mobile sensing data for research and intervention purposes, identifying multiple beneficial uses, including trigger identification, improved (ie, more accurate and less effortful) behavioral logging, crisis management to prompt positive behaviors, and supported self-reflection. Any concerns they had regarding passively sensed mobile data (eg, *that's creepy*) were overcome once data privacy had been reassured using the *What You Do/What We See* activity.
2. Discussions of the *oops* feature were mixed and led to modifications. Participants liked the idea of an easily identifiable, user-friendly, and low-effort button to quickly and efficiently record any incidences of problem behavior but felt that labeling this as *Oops* was condescending, whereas *Log* was found to be more amenable. Participants also described how this method of behavior logging could be improved by allowing the optional entry of descriptive text—either at the time of making the log or afterward—so that they could add additional contextual information.
3. Finally, participants expressed concerns regarding the use of the PANAS, both in terms of the content of the scale and the frequency of its proposed use in the study (eg, If I had to do this three times a day, it would be a bit of a trigger for me...Just the word, ashamed, like if I had to go over that three times a day, I probably wouldn't do it, to be honest"). A simple way of logging mood (eg, using smiley faces) was suggested as a preferable alternative.

DeMMI App Development

The DeMMI app is based on the AWARE [68] client for Android 7.0 or newer, which we used to collect smartphone sensor data 24/7 from 14 sensors. The AWARE framework is an open-source mobile sensing platform used in context-aware mobile computing research. We collected data on accelerometer readings, app use, notifications, battery, Bluetooth, calls, keyboard events, ambient light, locations, SMS text messages, physical activity, screen power events, screen touch events, and Wi-Fi data. On the basis of our consultation, we modified the client to allow participants to log binge-eating episodes using a *Log episode* button placed below a textbox for open-ended feedback related to the episode (see the left screenshot of Figure 2); both were shown after tapping on the main body of a persistent notification labeled "DeMMI. Tap to log an episode." Participants were also able to log their mood using a scale with emoji faces, ranging in expression and color, to visually represent affective states on a 5-point scale ranging from very positive (represented by a very happy, smiling face) to very negative (represented by a very sad face; see the right screenshot of Figure 2). This scale was automatically shown on screen every day at 9 AM and 9 PM, and as soon as a participant tapped a face, the app logged the choice and hid the instrument. We provided a *Not Now* button so that participants could ignore the prompt and allowed them to report their mood outside the scheduled times by tapping the bottom area of the persistent notification labeled as "LOG MOOD."

Figure 2. Screenshots of our button to self-report episodes (left), the persistent notification that allowed participants to open the episode and mood reporting screens (middle), and the mood reporting strings with 5 face emojis (right).



Phase 2: Fieldwork

The COVID-19 Context

This study was conducted with participants located in England, United Kingdom, between June 1 and August 14, 2020, as the COVID-19 lockdown restrictions were gradually eased. Before this point, individuals had not been permitted to leave home except for limited purposes (eg, shopping for essential items, exercise, and medical care), and all nonessential shops, libraries, places of worship, and playgrounds were closed. During the time of our study, changing restrictions permitted individuals from ≥1 household to meet outdoors in groups of 6 while maintaining 2 m social distance (June 1), whereas individuals living alone could form a *support bubble* with 1 other household (June 13). From July 4, the premises reopened with strict social distancing and hygiene measures in place, and gatherings of up to 30 people were allowed both inside and outside private dwellings. Those facing disordered eating behaviors are thought to be particularly vulnerable during the pandemic. A study by Branley-Bell and Talbot [69] found that the COVID-19 pandemic had a profound negative impact on people with eating disorders, whereas Schelgl et al [70] found that 49% of patients reported a deterioration in eating disorder symptoms because of the COVID-19 pandemic, and 47% of binge-eating group patients reported an increase in binge-eating symptoms.

Participants

Participants were recruited via several channels, including social media, the regional branch of a UK-based mental health charity (York Mind), and York St John University research participant recruitment forums. The study was advertised as a mobile sensing study for mental health, recruiting participants aged ≥18 years who currently reside in the United Kingdom and have an Android phone as their primary device. Participants were recruited into two categories: (1) those with experiences of binge eating, defined as "Eating an amount of food that you consider excessive, usually very quickly during a single session, eating until you feel uncomfortably full, eating when you're not hungry, eating alone, eating secretly, and feeling depressed, guilt, ashamed or disgusted after eating. This is often, but not always, accompanied by behaviours to counter the binge-eating (e.g. skipping meals, vomiting, over-exercising)" and (2) those with no current mental health difficulties (comparisons). Participants with experiences of binge eating were not required to have a clinical diagnosis to take part in the study. Participants registered their interest by either directly emailing a designated member of the research team or completing a web-based form. Prospective participants were emailed the study information sheet and consent form and could ask any questions via email or phone call. All participation in the study was conducted remotely. A total of 20 participants took part in the study (n=10 [50%] with experiences of binge eating and 10 [50%] with no history of mental health issues). All participants were aged between 18 and 36 (mean 25) years and were almost exclusively female (with 1 male in each group).

Study Methods

Once recruited into the study, the participants were sent an onboarding pack via email. This included a link to the DeMMI app Android Application Package (an Android package file format allowing the app to be downloaded via the link) and a set of step-by-step instructions for downloading, opening, optimizing battery life, and setting permissions on the app (Multimedia Appendix 2). Instructions on how to join the study with a unique identifier were also provided, as were instructions on how to log mood, log episodes, and uninstall the app.

Participants were asked to run DeMMI for a total of 6 weeks but were instructed that they could uninstall the app at any time, and following removal, all collected data would be deleted from their phone. All participants were asked to log their mood twice a day (at 9 AM and 9 PM), and the binge-eating group was asked to log any episodes of binge eating using the *Log episode* button, with the option to provide free-text information regarding the episode. To enhance engagement in the study and ensure that safeguarding protocols were being followed, participants were contacted once a week via SMS text messages to flag any potential problems. Following the study, participants were given the opportunity to take part in an optional interview to discuss their experiences during the study and provide any feedback regarding how we might improve the app functionality in the future. The interviews were conducted via telephone or video call (depending on participants' preferences).

Data Analysis

Quantitative

Smartphone behavioral feature analysis was conducted on the data collected from the binge-eating group, with comparisons across days that an episode had been reported and days that an episode had not been reported. We used the Reproducible Analysis Pipeline for Data Streams (RAPIDS) [71,72] to preprocess, clean, and extract behavioral features from the smartphone data collected with the AWARE framework. RAPIDS is a reproducible pipeline that allows for the processing of mobile sensing data. According to our protocol, every participant had to be monitored for 42 days (ie, 6 weeks); however, in practice, their smartphones could run out of battery, our sensing app could crash, or it could have issues synchronizing the data. Therefore, we expected some of these days to be missing all, most, or some of the data. We measured the quality of our smartphone data through the concept of valid sensed days; we labeled a sensed day as valid if we had 8 hours of data with at least 30 sensed minutes each. A sensed minute is a 60-second window with at least one row of data from any smartphone sensor.

Once the data were processed, we used the Nonoverlap of All Pairs (NAP) index [73] to measure the probability that a behavioral feature value drawn at random from any episode day would exceed that of a feature value drawn at random from any nonepisode day. The NAP analysis provided an indication of effect size and offered directions for future work that might use such sensor-driven approaches in the context of binge eating.

Qualitative

All free-text episode logs collected during the study were collated and subjected to content analysis [74] to explore any recurrent themes of discussion across participants that might provide future directions for focus. Our analysis involved

assigning codes to the lines of the data and grouping them into themes. The interviews were all audio recorded and transcribed verbatim. Interview transcripts were thematically analyzed using a deductive approach that saw codes created at the paragraph and sentence level [75].

Results

Overview

In this section, we describe and summarize the smartphone data, mood survey notifications, mood scores, and collected binge-eating episode data. To explore the relationship between smartphone data and binge-eating episodes, we computed 12 behavioral features related to social interaction and physical activity across location, Bluetooth, physical activity, and screen sensor data and analyzed the difference in their values between days with and without self-reported binge-eating episodes. We have shared the code of our mobile app and analysis pipeline to enable the reproduction of our methods [76,77].

We present our analysis in 3 parts. We first look at the entire cohort of data for all 20 participants (both the comparison group and the binge-eating group) and provide an outline of the valid sensed days that we were able to collect (providing an indication of how long participants in both groups engaged in the study before deleting the app and how well the software functioned in relation to collecting the sensor data). Within this cohort level data, we also provide a comparison of mood scores across all 20 participants to explore whether there was any difference between mood reporting across the 2 groups. In the second part, we look specifically at the binge-eating data and the differences in sensor data collected on days when an episode was reported versus days when an episode was not reported. We first provide a content analysis of binge-eating episodes and then discuss the NAP analysis conducted. Finally, we report the interview data across both groups, focusing on the usability and acceptability of our approach and the app itself.

Part 1: Entire Cohort Data

Valid Sensed Days Across All Participants

Participants were asked to keep the app running on their phones for 6 weeks (approximately 42 days). In relation to study engagement, the 2 groups were relatively similar; the binge-eating group had a total of 395 total days with the app, with 20% (2/10) deleting the app before the end of the study (P02BE after 15 days; P06BE after 36 days). The comparison group had a total of 397 days with the app, with 10% (1/10) of participants deleting the app before the end of the study (P13C after 5 days). There were some differences in the valid sensed days; however, the binge-eating group had 72.4% (286/395) of valid sensed days compared with the comparison group, which had 61.7% (245/397) of valid sensed days. See Table 1 for a full breakdown of the data.

Table 1. Participants in the binge eating (N=395 days) and comparison groups (N=397 days) along with the number of days they were monitored and the number of valid sensed days.

Participant ID	Group	Days with the app, N	Valid sensed days, n (%)
P01BE	Binge eating	45	0 (0)
P02BE	Binge eating	15	12 (80)
P03BE	Binge eating	44	42 (95)
P04BE	Binge eating	43	31 (72)
P05BE	Binge eating	41	38 (93)
P06BE	Binge eating	36	6 (17)
P07BE	Binge eating	43	38 (88)
P08BE	Binge eating	43	39 (91)
P09BE	Binge eating	43	41 (95)
P10BE	Binge eating	43	40 (93)
P11C	Comparison	44	31 (70)
P12C	Comparison	44	35 (80)
P13C	Comparison	5	0 (0)
P14C	Comparison	43	40 (93)
P15C	Comparison	44	9 (20)
P16C	Comparison	44	20 (45)
P17C	Comparison	43	42 (98)
P18C	Comparison	43	2 (5)
P19C	Comparison	44	27 (61)
P20C	Comparison	43	39 (91)

Mood Logs Across All Participants

Mood logs were collected twice daily at 9 AM and 9 PM. Overall, the binge-eating group logged their mood when prompted by our app notifications with a log rate of 87.1% (548/629 times). The comparison group logged their mood at a higher rate of 90.4% (638/706 times).

However, as seen in Table 2, the binge-eating group had lower overall positive mood scores and was more likely to score their mood as negative (64/596, 10.7% compared with 17/671, 2.5% in the comparison group) or very negative (19/596, 3.2% compared with 0.4%, 3/671 in the comparison group). The binge-eating group was also more likely to report a lower mood log in the mornings (34/298, 11.4%) and evenings (39/298, 13.1%), reporting negative or very negative scores compared with comparisons (6/330, 1.82%, and 14/330, 4.2%).

Table 2. Count of mood score categories from 0 (very positive) to 4 (very negative) for the comparison and binge-eating groups.

Mood Score	Binge-Eating Group	Comparison Group
0 Very Positive	34	52
1 Positive	335	453
2 Neutral	144	146
3 Negative	64	17
4 Very Negative	19	3

We computed the rate between the number of mood surveys displayed on the screen and the number of surveys that should have been triggered (2 per sensed day). Our participants had a mean trigger rate of 92.9% (986/1062; SD 10, minimum 59.4, maximum 100), which shows our surveys were delivered reliably. We also calculated the rate between the number of surveys that were answered and the surveys displayed on the screen. We obtained an average answer rate of 88.4 (SD 6.71, minimum 67.5, maximum 96.2), suggesting that our mood survey instrument was easy to use, which was further confirmed by the fact that the number of ignored surveys was on average 1.3 per week throughout the study's 6 weeks. We also gave participants the opportunity to report their moods outside the scheduled time. This functionality was rarely used, with a total of 95 self-initiated mood reports and an average of 5 responses (SD 5.9, minimum 0, maximum 18) across all participants. However, we highlight that 26% (25/95) of these surveys were used to correct a previous self-reported mood score (defined as any new report made within 2 minutes of the previous one).

Part 2: Binge-Eating Data

Episode Logs

We had 98 episodes reported across our 10 participants in the binge-eating group (Table 3). Most logs occurred in the evening between 6 PM and midnight (41/98, 42%) or in the afternoon between noon and 6 PM (35/98, 36%), with a smaller number occurring in the morning between 6 AM and noon (14/98, 14%) and a very small number occurring late at night between midnight and 6 AM (8/98, 8%).

Table 3. The total number of episode logs (N=98) and episodes logged with additional text for all binge-eating group participants.

Participant ID	Episode logs, n (%)	Episodes logged with additional text (n=62), n (%)
P01BE	6 (6)	0 (0)
P02BE	3 (3)	0 (0)
P03BE	9 (9)	3 (5)
P04BE	6 (6)	2 (3)
P05BE	8 (8)	7 (11)
P06BE	23 (23)	23 (37)
P07BE	19 (19)	18 (29)
P08BE	16 (16)	2 (3)
P09BE	6 (6)	5 (8)
P10BE	2 (2)	2 (3)

Of the 98 episode logs, 36 (37%) were recorded without any additional text. We conducted a content analysis of the remaining 63% (62/98) of self-reported episodes with additional text to explore any related themes of reporting that cut across participants. It should be noted that some episodes involved multiple themes. There were six overarching themes identified from our content analysis: (1) perceived lack of control, (2) emotions, (3) invasive negative thoughts, (4) disordered eating behavior explicitly reported, (5) situational influence, and (6) behavior avoided.

We had 36 explicit reports of binge-eating behavior across participants:

> *I was watching movies and binged snacks even though i'm not hungry* [P05BE]
>
> *Made coffee muffins when partner was out. Ate 6 in a row.* [P06BE]
>
> *massively over-snacked after not eating much lunch at 12pm* [P08BE]

There were 17% (6/36) of reports that displayed a perceived lack of control:

> *tried really hard to not binge, but couldn't stop myself* [P03BE]
>
> *I'm on my own and I can't sleep, and I can't stop snacking* [P05BE]

Invasive negative thoughts were also explicitly reported 33% (12/36) times: "Self-critical thoughts. Strong emotions" (P04BE); "feel like failure" (P06BE); and "I'm having really bad intrusive thoughts :(" (P09BE). Moreover, participants reported on their feelings and emotions after a binge-eating episode: "feeling quite ashamed now" (P03BE), "feel useless and angry," and "Feel disappointed and dumb" (P06BE). Participants also reported 15 situational influences, which led them to disordered eating behaviors:

> *Had to do a job interview today but didn't go to plan* [P03BE]
>
> *jeans I had ordered arrived today. They didn't fit and all I wanted to do was sit and cry and eat but I couldn't because I wasn't alone* [P05BE]
>
> *every time I see an ad for that massive popcorn chicken from KFC I wanna binge eat the entire thing* [P06BE]

Finally, there were 17% (6/36) of remaining reports related to disordered eating behaviors that were avoided, despite the thought processes around binge eating being there, because of factors such as not being alone or because of the acknowledgment of the negative effects of the disordered eating behavior:

> *want to make myself sick in the bathrooms but it's empty and I'm scared my partner will hear.* [P06BE]
>
> *didn't realise how much I was eating till I finished[...] but I am trying to stop purging as it is starting to have negative physical effects* [P03BE]

In the case of P06BE, there were several reporting incidents that were framed positively (5/36, 14%):

> *Feel good didn't have cake etc at cafe as I would 100% scoff and eat too much and feel ill like i usually do* [P06BE]

This may have been perceived as a restraint in relation to the avoidance of a binge-eating episode; however, it is worth noting that the deliberate restriction of food can also be problematic and a prelude to binging.

Smartphone Data on Episode Versus Nonepisode Days

P01BE had only 2% (1/45) valid days, P02BE had 93% (14/15) of valid days, and P06BE had only 17% (6/36) valid days; as such, they were excluded from our behavioral feature analysis. This left us with the remaining 70% (7/10) of participants. We explored the difference in magnitude between days when our participants did and did not report an episode in relation to two self-reported mood scores (those logged in the morning between 6 AM and noon and those logged in the evening between 6 PM and midnight), as well as 12 behavioral smartphone features extracted in daily segments that span midnight to midnight. We included the following smartphone features: geographical location variance, total distance traveled, radius of gyration, time at home, stationary (to moving) ratio, the total stationary time, number of distinct Bluetooth devices sensed around the phone, number of screen unlocks, total screen time, time of first screen use, total number of mobile apps used, and the entropy of all mobile apps used (a wider variety of apps produce a higher entropy value). The sensors that provide these features, a layman's description, and the implementation of these features can be found in the RAPIDS documentation [71].

We framed our problem as 7 n-of-1, or single-case, experiments with alternating AB phases. A phase is formed by consecutive days without binge-eating episodes, whereas a B phase is formed by consecutive days with binge-eating episodes. Overall, 82% (36/44) of the B phases across all participants were 1 day long (mean 1.3, range 1-4 days). The A phase across all participants (N=48) was, on average, 5 days long (range 1-28). Table 4 shows the breakdown of the average phase length and range for each participant.

Table 4. Average length (days) of phase A (no episode; N=48) and phase B (episode; N=44).

Participant ID	A phases, n (%)	A phases length, mean (range)	B phases, n (%)	B phases length, mean (range)
P03BE	7 (15)	5.1 (1-10)	6 (14)	1.3 (1-2)
P04BE	6 (13)	6.3 (1-21)	5 (11)	1 (1)
P05BE	6 (13)	5.5 (1-13)	6 (14)	1.3 (1-3)
P07BE	9 (19)	3.1 (1-9)	9 (20)	1.7 (1-4)
P08BE	11 (23)	2.6 (1-7)	11 (25)	1.3 (1-3)
P09BE	6 (13)	6.3 (1-14)	5 (11)	1 (1)
P10BE	3 (6)	13.7 (1-28)	2 (5)	1 (1)

We used the NAP index [73] to measure the probability that a behavioral feature value drawn at random from any phase B will exceed that of a feature value drawn at random from any phase A. We used the R implementation provided by

Pustejovsky et al [78], concatenating all (1–N)A phases into a single A phase and all (1–M)B phases into a single B phase and setting a confidence threshold of 0.95 and an expected direction of improvement given by the sign of the standardized mean difference of A and B values ($[mean_B - mean_A]/SD_{[A,B]}$).

In Table 5, we share the behavioral features of each participant that had a NAP medium effect (an index between 0.66 and 0.92) or a NAP strong effect (an index between 0.93 and 1) [73]. Only the strong effects belonging to P10BE were statistically significant after adjusting for multiple tests within participants using the Benjamini and Hochberg method [79]. P04BE, P05BE, P07BE, and P10BE self-reported feeling worse on days when they binge ate; P04BE had morning mood at a medium effect; P05BE and P07BE had evening mood at a medium effect; and P10BE had evening mood at a strong effect. P05BE, P08BE, and P09BE had screen features with a medium effect (P08BE and P09BE used or unlocked their phones less on days when they binge ate). P03BE had app entropy with a medium effect (he or she used a wider variety of apps on days when they binge ate). P04BE and P10BE had time at home with a medium effect (they spent more time at home when they binge ate). P04BE, P05BE, and P10BE had a medium and strong effect for total distance (P04BE traveled more and the others 2 less in days when they binge ate). P09BE and P10BE had a stationary time and ratio with a medium effect (they moved around less when they binge ate). Finally, P10BE had location variance and radius of gyration with a strong effect (he or she traveled less when they binge ate) and unique Bluetooth devices with a medium effect (he or she was arguably around fewer people or public places when they binge ate).

Table 5. Smartphone features that showed a medium or high NAP[a] effect between phases A and B of participants that binge eat.

Participant ID	Feature	Standardized mean difference	NAP (SE)	NAP, 95% CI	NAP effect	Adjusted *P* value
P03BE	App entropy	0.67	0.69 (0.1)	0.46-0.84	Medium	.45
P04BE	Time at home	0.64	0.73 (0.9)	0.46-0.89	Medium	.45
P04BE	Mood morning	0.62	0.68 (0.1)	0.41-0.86	Medium	.45
P04BE	Total distance	0.38	0.67 (0.16)	0.40-0.85	Medium	.45
P05BE	Total distance	−0.58	0.73 (0.11)	0.51-0.87	Medium	.29
P05BE	Mood evening	0.79	0.69 (0.13)	0.47-0.85	Medium	.29
P05BE	Screen unlocks	0.36	0.67 (0.08)	0.45-0.83	Medium	.33
P07BE	Mood evening	0.70	0.68 (0.08)	0.50-0.81	Medium	.25
P08BE	Screen unlocks	−0.63	0.69 (0.08)	0.51-0.83	Medium	.29
P09BE	Stationary time	0.79	0.68 (0.16)	0.41-0.86	Medium	.44
P09BE	Screen time	−0.81	0.67 (0.16)	0.41-0.85	Medium	.44
P10BE	Mood evening	1.67	0.99 (0.02)	0.58-1.00	Strong	.046
P10BE	Radius of gyration	−0.46	0.98 (0.02)	0.57-1.00	Strong	.046
P10BE	Location variance	−0.32	0.98 (0.02)	0.57-1.0	Strong	.046
P10BE	Total distance	−0.68	0.98 (0.02)	0.57-1.0	Strong	.046
P10BE	Stationary ratio	1.47	0.91 (0.09)	0.5-0.99	Medium	.07
P10BE	Time at home	1.01	0.88 (0.06)	0.46-0.98	Medium	.09
P10BE	Bluetooth devices	−0.88	0.83 (0.07)	0.42-0.97	Medium	.13

[a]NAP: Nonoverlap of All Pairs.

Part 3: Poststudy Interviews

Poststudy interviews were optional. Among the 10 binge-eating group participants, 3 (30%) consented to the poststudy interview; among the 10 comparison participants, 5 (50%) provided consent. Of the 20 participants, this provided a total of 8 (40%) interviews for analysis. Interviews lasted between 14 and 42 minutes each, were conducted via telephone, and were audio transcribed for later deductive thematic analysis. A total of 45 codes were initially created and then further grouped into themes. There were four overarching themes identified: (1) positive and negative impacts of lockdown, (2) phone habits, (3) mood and episode logging, and (4) the usability of the DeMMI app.

Positive and Negative Impact of Lockdown

In the interviews, participants were asked how their moods might have changed during the lockdown. Most participants discussed their mood negatively, with both groups reporting fluctuations in their mood during the lockdown and possibly triggering episodes for those in the binge-eating group:

> *I was doing a lot worse when lockdown got really bad.* [P09BE]

> *it was very much up and down...* [P13C]

> *I definitely think it's been a lot more up and down...as soon as something tiny goes wrong...I would have like a full-on breakdown. Then I'll comfort eat, have a binge or start like picking at myself* [P07BE]

Several recurring themes were noticed in the interviews, which were related to the negative impacts of the lockdown on one's mood. Some participants highlighted feelings of hopelessness and that the uncertainty of the situation was causing stress and anxiety:

> *It can feel quite hopeless as we don't know when it will change* [P10BE]

> *I think like everyone, it's had a negative impact...there was that sort of novelty factor and we weren't quite sure how long it was gonna last.* [P11C]

> *it was definitely a bit more scary, and because we didn't know how long it was gonna go on for. A bit unsettling...that was a bit stressful* [P15C]

However, there were also some positives reported by both groups. Approximately 25% (2/8) of participants mentioned how the lockdown had encouraged them to have better time management, as the time usually spent getting ready and commuting to work could be used to spend more time with people around them and participating in activities that they enjoyed, such as exercising and volunteering:

> *there's been a lot of benefits of lockdown...I managed to get into a nice little rhythm once my routine kinda reset* [P15C]

> *it's [lockdown] given everyone a focus which in some ways has helped my social anxiety, I've been volunteering which I wasn't before* [P10BE]

Furthermore, although some participants highlighted that being able to exercise "makes [them] feel good and it kinda like refreshes and resets [them]" (P1C), for members of the binge-eating group "not being able to go and do exercise as much, just feeling very tired all day" (P07BE) was often seen as a cause of stress that could trigger a vicious cycle of binge-eating episodes and dietary restrictions. Despite this, the same binge-eating group participant reported how having regular face-to-face social interactions during and after the lockdown led to a positive impact on their mood:

> *if I go into placement. I work well, it's a good day, but if I'm at home it's not a good day* [P07BE]

However, with lockdown limiting face-to-face contact, participants in this group could be at an even higher risk of triggering a binge-eating episode.

Phone Habits

Long periods of staying at home during a pandemic, with limited knowledge of the virus, were seen to take a toll on participants' mental health, and with restrictions on socializing, participants discussed beginning to form new habits, especially with their phone and social media use. During the lockdown, most of the participants reported a noticeable increase in their phone use:

> *I've been on my phone a lot since lockdown started.* [P09BE]

> *I definitely use my phone a lot more than I used to...I wouldn't have done as much if I didn't have as much free time as I do now.* [P15C]

> *I would say that I was probably spending a bit more screen time in lockdown...Not having that structure.* [P14C]

A participant clarified that this was not because of taking part in the study:

> *I think it's more lockdown that's increased my phone behaviours not really the app.* [P07BE]

The major causes for this change in phone use behaviors noted by many participants were procrastination and boredom, particularly as the lockdown period was during the summer break for university students and/or some participants were forced to take a break from work:

> *A lot more procrastination, just not really using it for anything useful.* [P07BE]

> *I'm into the habit when I'm not doing anything, I'm much more procrastinating on my phone than I used to be.* [P09BE]

> *I was using that a lot more, and maybe even out of boredom.* [P13C]

> *maybe out of boredom a little bit and maybe just out of habit as well.* [P14C]

A few participants also discussed the negative impact of social media and phone use on one's mental health:

> *the increase in my phone usage probably contributed to the decline in my mental health.* [P10BE]

> *when I'm on social media and I see negative things. So, when I get news alerts it's always negative...I think Twitter was the worst for me...when everything was going on with coronavirus and like the Black Lives Matter movements...everything really made me feel down and depressed.* [P17C]

Moreover, 60% (3/5) of the comparison group participants discussed the positive impact of using the DeMMI app:

> *it made me think about my use of social media or being on my phone and if I thought that correlated with my mood...I was thinking about how work affected my mood, rather than my usage of my phone.* [P14C]

> *I did think about phone usage as well because when I was thinking about what the app is sort of looking at, I would reflect and think that actually I probably use my phone a lot more than I thought I did.* [P13C]

> *I noticed how often I was using my phone, and I was quite conscious of it in the first couple of days, I did notice patterns in my mood too.* [P15C]

This seemingly led some participants to reflect on how their phone use correlated with their mood.

Mood and Episode Logging

The binge-eating group participants were asked to expand on what they felt triggered them to experience an episode and how

they classified an episode during the study (it should be noted that the comparison group was only instructed to log their mood). For one participant, the impact of social relationships on the occurrence of episodes was noted, and the passing of a close family member had affected their mood:

> *in terms of the binge eating...Cause, my boyfriend at the time...we had some sort of issues* [P07BE]

> *probably about 6 weeks ago there was a big trigger...my uncle had been ill for a while...And then he died about 6 weeks ago* [P07BE]

Other binge-eating group participants noted that "intrusive thoughts" (P09BE) and "the increase in my phone usage probably contributed to the decline in my mental health" (P10BE), which might trigger episodes of binge eating. In the interviews, participants were also asked whether they noticed any patterns in their behavior. Approximately 67% (2/3) of the binge-eating group participants also noticed that their mood had improved with the return of social interactions and having more occupied time:

> *So now it's a lot less time for overthinking and worrying about things, and more time to actually just be doing stuff.* [P07BE]

None of the participants in the comparison group reported any concerns with mood logging. On the other hand, one of the binge-eating group participants found it invasive at times because of logging in a public environment:

> *Sometimes it was annoying because I wanted to look at my phone for something else. Sometimes if I was with someone, I didn't want them to see what I was clicking so I didn't do it.* [P10BE]

However, another binge-eating group participant noted that the app provided them with a level of accountability over binge-eating episodes:

> *having to log it and recording my episodes has actually been really helpful as like a deterrent, especially for my binge eating. Because it was giving me some accountability for doing it and if I do binge eat then I have to record it. I actually think that it made my like threshold for having an episode a little bit higher.* [P07BE]

Usability of the DeMMI App

When discussing the usability of the DeMMI app, participants mostly provided positive feedback for the app. Many participants commented that it was "easy to use" (P13C, P14C, P15C, P17C, and P10BE). Some of the participants elaborated on the user-friendly aspects of the app:

> *I liked [how] you could just click on the notification bar and then just tap to add episode* [P07BE]

> *I liked how it popped up with the pop up telling you to log your mood now. Because otherwise 100% I would have forgotten.* [P09BE]

The main issue reported by the participants was related to the app crashing. Approximately 25% (2/8) of the participants commented the following:

> *I had a lot of issues with the app force closing and not knowing if it was still running, I don't think in the end I could enable all the features of it.* [P07BE]

> *The only thing that did happen, was occasionally when I was browsing the web, so not really when I was using any apps, it would occasionally say the app wasn't responding.* [P15C]

Participants provided several specific examples of feedback that could be used to improve the app in the future. 50% (4/8) of participants proposed an option for additional mood logs; a participant suggested the following:

> *Maybe one around lunch time and one around evening maybe?* [P09BE]

Another elaborated with a similar suggestion, saying the following:

> *Because there was lots of days where in the morning and evening I would put a neutral or a happy face, then in the middle I would dip to close to having an episode but not quite have an episode.* [P07BE]

P10BE commented that they would like the ability to adjust the time of the mood logs:

> *I am often only just up by 9am or still asleep, so checking in later for a mood score would have been better...sometimes I needed to urgently look at something on my phone, so I clicked off it quickly without thinking and missed logging it.*

All of the binge-eating group participants could see the benefits of a future app version to identify or predict a potential episode, as well as the addition of mindfulness and relaxation exercises, safe practices, or positive affirmations:

> *even it's just recognising you're about to have an episode...I think that might in itself even be a bit useful. Like giving the accountability before and after.* [P07BE]

> *if I pre-put in some like songs or something that I like and then it would identify it and then prompt you to play the song, rather than needing to go and think about oh I'm going to put on some relaxing music. Or maybe some exercise which uses your mind a bit to distract you.* [P07BE]

> *Maybe if you could tailor them to what you find useful. I wouldn't like an app to tell me what to do. Maybe if you'd set up a reminder to do mindfulness or something.* [P10BE]

> *like positive affirmations...Like, a flowchart on what to do.* [P09BE]

Finally, the participants were asked about privacy concerns regarding automated data collection by the app. Most participants voiced no privacy concerns:

> *I kinda trusted the study* [P13C]

Moreover, 25% (2/8) of participants explained as follows:

> *but if this was just another app on the app store for sure I would be having privacy concerns.* [P09BE]

> *no [concerns] because I was told you could see what apps and stuff were open, but you couldn't see any messages. I was worried initially when I was signing up to the study, but after I read everything and emailed you it made me feel a bit better about it.* [P17C]

Further open questions were asked surrounding the clarity of the information sheet that was distributed to the participants. Most participants were satisfied with the level of information that was provided to them to understand the purpose of the study and how their data would be collected and handled:

> *I didn't have any concerns or anything about that...I think it was very clearly set out, I don't think anything stood out particularly.* [P14C]

> *No [concerns], I was really happy with it...in terms of privacy there was nothing that really stood out in the information sheet that I was, ooh I'm not sure about that or I don't understand that...it was all pretty straight forward.* [P14C]

Several participants further highlighted the importance of explicitly discussing data privacy and security within the study:

> *the information sheet where they were pointing out the privacy settings, I think that stood out the most to me. Because that was the information that I wanted to know the most before I started the study.* [P17C]

> *I did at first, just in terms of there's so much data that you can really get passively from someone's phone...But I knew what was being done with the data, so I wasn't too concerned about it.* [P15C]

Discussion

Principal Findings

Our preliminary quantitative results suggest that every participant had various smartphone features that were meaningfully different between days with and without binge-eating episodes. This, in turn, could encourage researchers to investigate fully data-driven approaches to find *hidden* links between smartphone behavioral features and these episodes either via interpretable or black box predictive approaches. However, our qualitative work paints a more nuanced image of the research needed to deliver effective, safe, and ethical digital interventions.

Perceived Episode Indicators

In the feedback attached to the self-reported episodes, participants described a variety of affective states, comorbid mental health disorders, social interactions, and daily life experiences that either preceded or happened during their binge-eating episode. These findings correspond with the existing literature surrounding binge-eating episodes, which has also identified these factors as potential antecedents of binge-eating episodes [80]. We refer to these as indicators because of their predictive potential and are keen to emphasize that indicators are not necessarily causal contributors to binge-eating episodes. It may be that they are provoked by the latent trigger itself; for example, calling a relative for support may co-occur with binge eating but might not necessarily cause a binge eating episode. Keeping in mind that the basic premise behind a digital intervention based on mobile data is to find behaviors that can be measured to deliver treatment before, during, or after a binge-eating episode, it is crucial to understand and catalog these indicators independently of their causal role.

The idiosyncratic nature of the indicators identified in this study warrants further examination in the form of a longitudinal observational study, which can help clarify the relevance, scalability, and focus of quantifying indicators using mobile data (in our case, screen unlock time, app entropy, time at home, stationary-to-location ratio, total distance, stationary time, screen unlocks, evening mood score, and time of the first screen unlock). For example, a scenario where most of these indicators are related to affective states or mental health disorders such as depression would support the idea of leveraging previous works on general or personalized mobile monitoring of these constructs [81]. Under a different scenario, if most of these indicators are related to situational influences such as social interactions with certain people, work activities, or leisure activities, then it is likely that models to detect these events will have to be highly personalized, given the differences in people's routines. For example, researchers might have to monitor digital communication between participants and specific relatives or friends or adapt to people's work, school, or leisure settings (monitoring sleep patterns, drinking patterns, and physical activity). In practice, there might not be clear-cut lines among affect, mental health, and situational influences, as the latter is likely to affect the former, and a participant could report episodes around both types of indicators; however, some might be easier to quantify using smartphone or wearable data (see the *Computation Amenability of Indicators* section).

It is also worth noting that the idiosyncratic nature of the indicators we identified might reflect the idiosyncratic nature of binge eating more broadly. In this study, we defined binge eating in inclusive terms, did not state that it must form part of a specific diagnosis (eg, bulimia nervosa and binge-eating disorder), did not exclude binge eating that is comorbid with other mental health conditions (eg, anxiety), and did not specify whether binge eating must occur in the presence or absence of purging. Previous EMA research suggests that different factors may be more or less important to different clusters of participants. For example, a review by Dingemans et al [82] found differences in the affective dynamics associated with binge eating with purging compared with binge eating without purging. Thus, the previously described work on binge-eating indicators would benefit from collecting this information to understand whether there are commonalities among certain participant clusters.

Frequency and Life Span of Episode Indicators

We need to understand how often and for how long these indicators happen such that researchers focus their efforts on the most common ones for a participant or a cluster of participants. It is possible that the time a person is exposed to or experiences an indicator will vary and that their relevance fluctuates over time. The former means that the timescale at which mobile data are analyzed will depend on the indicator

(eg, should we look at anxiety levels in days or hours in the past?), and the latter implies that models will have to adapt over time to changes in people's routines and personal circumstances (eg, if being alone triggers binge-eating episodes, what happens when a young adult moves to live on their own after sharing a house with other people during university?). Further compacting this issue, research examining binge-eating episodes has suggested that they may not necessarily be discrete events. More specifically, Wilson and Sysko [83] discuss how binge eating may be best thought of in terms of binge-eating days rather than binge-eating episodes. This is especially important when binge eating occurs without purge behavior, as purge behavior is often considered a clear indicator that an episode has finished [84].

Severity and Impact of the Episodes Around Certain Indicators

Paying particular attention to the situational influence indicators (eg, being alone or having an argument with a loved one), it is unclear (1) how often they are a proxy for binge-eating episodes (and their likelihood of triggering a false positive intervention); (2) whether the severity of the episodes they pinpoint is similar, as measured by objective means (such as the amount of consumed food, its nutritional value, or the duration of the binge) or subjective means (such as the extent to which they experience a lack of control); and (3) whether the psychological and physical impact of such episodes is similar across episodes (eg, people might not always engage in purging behaviors after an episode). Researchers and patients might prefer to investigate, quantify, and monitor the indicators that pinpoint episodes with the most negative effects. Our lightweight approach to episode logging was well-received by the participants, who provided additional text in 63% (62/98) of the cases. Leveraging natural language processing approaches to gain a better understanding of specific indicator types and their severity of impact on the person is a promising direction for future work.

Computation Amenability of Indicators

We expect to be able to quantify and detect each indicator to a different degree using smartphone or wearable data. In our study, we computed smartphone features that measured constructs we considered to be roughly related to the indicators reported by our participants. However, once researchers have a better idea of the breadth and depth of indicators in a population, they can decide which behavioral features might be more relevant and need to be extracted. For example, it might be very difficult to measure or anticipate the effect of daily activities such as shopping on body image (P03BE); however, if it turns out that a considerable number of binge-eating episodes occur around body image issues, then we could focus on measuring the affective state induced by them. Alternatively, research has indicated that some individuals may engage with certain apps (eg, calorie-tracking apps) in different ways around binge-eating episodes [1]. Similarly, certain episodes could happen around indicators related to physical activity, sleep disorders, or communication patterns that previous research has had positive results quantifying using smartphones and wearables [47,59,85-87], which are a more direct measurement of the observed phenomena compared with affect and psychological constructs.

Intervention Candidates for Episode Indicators

Once a relevant indicator is identified and can be reliably measured, the next question is how to intervene to try to prevent binge-eating episodes from occurring and/or provide support for its duration. There is a pressing need for effective binge-eating interventions. The currently available treatments are only effective for up to 50% of individuals with binge-eating disorders and 30% of individuals with bulimia nervosa [88,89]. Of these treatments, the most commonly prescribed is cognitive behavioral therapy (CBT), which involves restructuring an individual's thoughts, feelings, and behaviors to support more productive outcomes and reduce binge-eating occurrences [90]. Although CBT is typically delivered in therapeutic settings, research has begun to consider its potential when delivered remotely through smartphones (akin to a self-help tool) [17]. Initial research in this space suggests that smartphone-enhanced CBT can be as effective as therapist-led CBT and may be even more effective in attaining some outcomes, such as meal adherence [22]. Such smartphone-based approaches may also be useful at the subclinical level, where individuals binge eat and experience significant distress without a diagnosis.

There are several types of interventions that may be appropriate for in situ delivery in response to the mobile sensed occurrence of an indicator. These interventions can be loosely categorized as (1) prompting and (2) self-management, both of which may be compatible with the CBT framework. Prompting interventions refer to those that are aimed at nudging participants to change their behavior, prompting self-reflection that helps put things in perspective, or even making them consciously aware that an indicator that is typically associated with their binge-eating behavior has been detected. Self-management refers to interventions that support participants by providing access to web-based tools that foster positive mental or physical health. These web-based tools could take many forms, including the automatic recommendation of activities aimed at de-escalating the situation (such as distraction activities available in apps like *Calm Harm* or *Recovery Record*); support messages that are meaningful for the individual; or open communication channels to family, friends, or health care providers that the participant agrees to. Indeed, our participants provided similar recommendations for activities that might support them in future app iterations during the poststudy interviews. Our empirical program suggests that such interventions would need to be personalized to the participant, consistent with previous work on the design of smartphone-based interventions to support mental health [91]. This personalization could be achieved with the support of a therapist (ie, a therapist-mediated intervention) or be self-led (ie, a self-help approach).

The Risks, Cost, and Effectiveness of an Intervention

Researchers must systematically consider the cost-effectiveness of prompting, self-management, and other types of interventions. This type of analysis has been conducted for HIV [92], physical activity [93], smoking [94], alcohol consumption [95], and CBT-guided self-help interventions for binge eating [96] and should take into account the time and expertise that these kinds of digital interventions would demand from participants and their health care providers [17]. In addition, researchers need

to be aware of the risks of delivering an intervention when it is not needed, failing to deliver an intervention when previous deliveries have been successful and participants rely on them, delivering an intervention aimed at disrupting a particular indicator that in turn puts the participant at risk of engaging in situations that could still trigger an episode (eg, suggesting someone to avoid texting a relative without knowing that this could make them anxious and in turn provoke an episode), and the long-term side effects of following certain interventions (eg, spending less time outside). Designing future solutions that actively consider responsible innovation and the possible negative consequences of certain app features will help avoid unintended consequences [1].

In the end, it is fair to assume that binge-eating monitoring and digital interventions will need to account for the frequency, life span, and computational amenability of the episodes' indicators and find a trade-off between the severity and impact of such episodes and the goal, cost, risk, and the effectiveness of candidate interventions. This sensor-informed context could support people and their therapists in identifying the triggers of their binge-eating episodes or simply augment the nondigital strategies they already use. As with other forms of retrospective self-reporting [97-99], the consequences of showing participants historical contextual information, the validity of these reports, and cognitive biases that might come into play should be studied. To our knowledge, this is an open problem in binge-eating research.

Usability and Acceptance

Although we wanted to know how to improve the functionality of the DeMMI app in our interviews, we did not set out to follow a usability engineering method for evaluating usability. Usability discussions were organically gathered from the interviews and identified as one of the themes in our qualitative analysis. At this early stage of our research, when we were simply collecting data as unobtrusively as possible, we wanted participants to have minimal tasks to complete when interacting with the app; they were only expected to log episodes and/or mood. Poor system usability has been highlighted as one of the factors affecting patient acceptance of health technologies [100]. As such, we envision the need to conduct usability evaluations such as heuristic evaluation and end user testing when the DeMMI app is further developed to include intervention-based functionality. End user testing examines how users conduct certain tasks or follow processes and is mainly focused on user experiences within the system [101]. Heuristics evaluation is conducted by usability experts and is concerned with the assessment of the system against a set of heuristic guidelines [102]. This type of usability engineering method would be an essential step when end users are expected to navigate through the app and engage with intervention-based content to support them in managing their disordered eating behaviors. Future work should adapt usability evaluation approaches taken, for example, by Honary et al [103] and Or and Tao [101], which are recommended for digital health solutions.

Reflections on Approaches to Participant Engagement

Participant insights played a pivotal role in uncovering the link between mobile data and binge-eating occurrence. That said, participant engagement was a significant challenge that we experienced in this study. Our most successful recruitment medium was an Instagram campaign in which 177 people responded via email. However, only 15 (8.5%) people consented to participate after receiving the study information, and of those 15, only 10 (67%) installed our app. This low recruitment rate could be caused by concerns surrounding the sharing of mobile data that has the potential to expose web browsing and communication habits. However, as noted in both our consultation and poststudy interviews, and echoing the findings from the study by Honary et al [91], participants were generally satisfied with the information we had provided relating to how we would capture and use their sensor data through our *What You Do/What We See* resource, which we have made available for reuse (Multimedia Appendix 1). The clarity and transparency regarding what could be considered invasive data capture were enough to alleviate participants' initial concerns, and we greatly suggest that future researchers use a similar approach in their research to increase participants' literacy surrounding mobile sensed data. Furthermore, only 30% (3/10) of our binge-eating group participants agreed to be interviewed at the end of the study. This could be an exception, given our low number of participants but could also be linked to the shame and fear of stigma reported by those who binge eat [5]. Exit interviews are an important part of the research process; therefore, moving forward, we aim to explore the potential of questionnaire-based exit interviews that may be perceived by participants as more confidential.

Limitations

This was an exploratory study aiming to explore the type of mobile sensing data that might be relevant for detecting episodes of binge eating. We acknowledge the limitations of our small sample size (20 participants), which makes it difficult to conclude any definitive findings, particularly given the individual differences between the participants. However, this preliminary study was conducted to inform the design of future larger-scale trials in this space. Given what we now know about participants' engagement with the study, their acceptance of the methods, and their willingness to provide self-reported data, we are confident that a larger-scale study would be feasible. In addition, our study monitored a UK-only cohort of people with binge-eating behaviors, limiting our results' generalizability to other contexts. During the study, our participants' general behavior might not be representative of their usual routines because of mobility limitations during the COVID-19 lockdown. Our smartphone monitoring app was only compatible with Android and stopped collecting sensor data on 20% (4/20) of our participants' phones, likely because of a software bug related to data synchronization, which, as it is, could limit the deployment of future studies using the same app. The exploratory nature of our study calls for the collection of data from multiple smartphone sensors. However, this might have influenced our low initial consent rate compared with the number of people initially interested in participating. Clearer study information materials provided early in the recruitment process and a more constrained sensing approach might alleviate this limitation.

Conclusions

We conducted a preliminary analysis of the differences in smartphone-based behavioral features between days with and without binge-eating episodes to explore the feasibility of using mobile sensing to detect these events. We contextualized the experiences of people who binge eat and reflected on the challenges and opportunities for working with this population. In addition, we discussed the need to understand participants' personal and social contexts preceding and accompanying their binge-eating episodes to be able to weigh the benefits, constraints, and risks of monitoring them using smartphones, as well as the implications of leveraging the insights extracted from these data sources to plan for safer and more effective digital interventions.

Acknowledgments

This study was funded through a UK Engineering and Physical Sciences Research Council National Institute for Health Research NewMinds Network plus grant. The authors thank Sarah Clinch (University of Manchester, United Kingdom), Aaron Sefi (Kooth, United Kingdom), and Leo Kroll (clinical psychologist) for their early input in the study design.

Authors' Contributions

All the authors contributed to the preparation of this manuscript. The analysis of quantitative data was conducted by JV and JX; the analysis of qualitative data was conducted by HN and RM; and the study was conducted by BTB and CT.

Conflicts of Interest

None declared.

Multimedia Appendix 1
What You Do/What We See resource.
[PDF File (Adobe PDF File), 1609 KB - mental_v9i4e32146_app1.pdf]

Multimedia Appendix 2
Installation instructions provided to participants.
[PDF File (Adobe PDF File), 737 KB - mental_v9i4e32146_app2.pdf]

References

1. Honary M, Bell BT, Clinch S, Wild SE, McNaney R. Understanding the role of healthy eating and fitness mobile apps in the formation of maladaptive eating and exercise behaviors in young people. JMIR mHealth uHealth 2019 Jun 18;7(6):e14239 [FREE Full text] [doi: 10.2196/14239] [Medline: 31215514]
2. Wick MR, Fitzgerald EH, Keel PK. Epidemiology of binge eating. In: Frank GK, Berner LA, editors. Binge Eating. Cham: Springer; Jun 02, 2020.
3. Diagnostic and Statistical Manual of Mental Disorders, 3rd ed. Washington, DC: American Psychiatric Association; 1980.
4. Bulimia nervosa - Symptoms and causes. Mayo Clinic. URL: https://www.mayoclinic.org/diseases-conditions/bulimia/symptoms-causes/syc-20353615 [accessed 2022-03-18]
5. Hepworth N, Paxton SJ. Pathways to help-seeking in bulimia nervosa and binge eating problems: a concept mapping approach. Int J Eat Disord 2007 Sep;40(6):493-504. [doi: 10.1002/eat.20402] [Medline: 17573682]
6. Swan S, Andrews B. The relationship between shame, eating disorders and disclosure in treatment. Br J Clin Psychol 2003 Nov;42(Pt 4):367-378. [doi: 10.1348/014466503322528919] [Medline: 14633413]
7. Jansen A, van den Hout M, Griez E. Clinical and non-clinical binges. Behav Res Ther 1990;28(5):439-444. [doi: 10.1016/0005-7967(90)90164-e]
8. Ivezaj V, Saules KK, Hoodin F, Alschuler K, Angelella NE, Collings AS, et al. The relationship between binge eating and weight status on depression, anxiety, and body image among a diverse college sample: a focus on Bi/Multiracial women. Eat Behav 2010 Jan;11(1):18-24. [doi: 10.1016/j.eatbeh.2009.08.003] [Medline: 19962116]
9. Abebe DS, Lien L, Torgersen L, von Soest T. Binge eating, purging and non-purging compensatory behaviours decrease from adolescence to adulthood: a population-based, longitudinal study. BMC Public Health 2012 Jan 13;12:32 [FREE Full text] [doi: 10.1186/1471-2458-12-32] [Medline: 22244266]
10. Schulz S, Laessle RG. Associations of negative affect and eating behaviour in obese women with and without binge eating disorder. Eat Weight Disord 2010 Dec;15(4):287-293. [doi: 10.1007/BF03325311] [Medline: 21406953]
11. Gross JJ, Uusberg H, Uusberg A. Mental illness and well-being: an affect regulation perspective. World Psychiatry 2019 Jun;18(2):130-139 [FREE Full text] [doi: 10.1002/wps.20618] [Medline: 31059626]
12. Dingemans AE, van Furth EF. Binge Eating Disorder psychopathology in normal weight and obese individuals. Int J Eat Disord 2012 Jan;45(1):135-138 [FREE Full text] [doi: 10.1002/eat.20905] [Medline: 22170025]

13. Barnett I, Torous J, Staples P, Sandoval L, Keshavan M, Onnela JP. Relapse prediction in schizophrenia through digital phenotyping: a pilot study. Neuropsychopharmacology 2018 Jul;43(8):1660-1666 [FREE Full text] [doi: 10.1038/s41386-018-0030-z] [Medline: 29511333]
14. Matthews M, Abdullah S, Murnane E, Voida S, Choudhury T, Gay G, et al. Development and evaluation of a smartphone-based measure of social rhythms for bipolar disorder. Assessment 2016 Aug;23(4):472-483 [FREE Full text] [doi: 10.1177/1073191116656794] [Medline: 27358214]
15. Musolesi M, Canzian L. Trajectories of depression: unobtrusive monitoring of depressive states by means of smartphone mobility traces analysis. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2015 Sep 07 Presented at: UbiComp '15: The 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 7-11, 2015; Osaka, Japan p. 1293-1304. [doi: 10.1145/2750858.2805845]
16. Juarascio AS, Crochiere RJ, Tapera TM, Palermo M, Zhang F. Momentary changes in heart rate variability can detect risk for emotional eating episodes. Appetite 2020 Sep 01;152:104698. [doi: 10.1016/j.appet.2020.104698] [Medline: 32278643]
17. Juarascio AS, Parker MN, Lagacey MA, Godfrey KM. Just-in-time adaptive interventions: a novel approach for enhancing skill utilization and acquisition in cognitive behavioral therapy for eating disorders. Int J Eat Disord 2018 Aug;51(8):826-830 [FREE Full text] [doi: 10.1002/eat.22924] [Medline: 30051495]
18. Juarascio AS, Manasse SM, Goldstein SP, Forman EM, Butryn ML. Review of smartphone applications for the treatment of eating disorders. Eur Eat Disord Rev 2015 Jan;23(1):1-11. [doi: 10.1002/erv.2327] [Medline: 25303148]
19. Burton AL, Abbott MJ. Conceptualising binge eating: a review of the theoretical and empirical literature. Behav Change 2017 Aug 14;34(3):168-198. [doi: 10.1017/bec.2017.12]
20. Ambwani S, Roche MJ, Minnick AM, Pincus AL. Negative affect, interpersonal perception, and binge eating behavior: an experience sampling study. Int J Eat Disord 2015 Sep;48(6):715-726. [doi: 10.1002/eat.22410] [Medline: 25946681]
21. Goldschmidt AB, Crosby RD, Engel SG, Crow SJ, Cao L, Peterson CB, et al. Affect and eating behavior in obese adults with and without elevated depression symptoms. Int J Eat Disord 2014 Apr;47(3):281-286 [FREE Full text] [doi: 10.1002/eat.22188] [Medline: 24014067]
22. Hilbert A, Tuschen-Caffier B. Maintenance of binge eating through negative mood: a naturalistic comparison of binge eating disorder and bulimia nervosa. Int J Eat Disord 2007 Sep;40(6):521-530. [doi: 10.1002/eat.20401] [Medline: 17573697]
23. Smyth JM, Wonderlich SA, Heron KE, Sliwinski MJ, Crosby RD, Mitchell JE, et al. Daily and momentary mood and stress are associated with binge eating and vomiting in bulimia nervosa patients in the natural environment. J Consult Clin Psychol 2007 Aug;75(4):629-638. [doi: 10.1037/0022-006X.75.4.629] [Medline: 17663616]
24. Andrewes HE, Hulbert C, Cotton SM, Betts J, Chanen AM. An ecological momentary assessment investigation of complex and conflicting emotions in youth with borderline personality disorder. Psychiatry Res 2017 Jun;252:102-110. [doi: 10.1016/j.psychres.2017.01.100] [Medline: 28259033]
25. Houben M, Claes L, Vansteelandt K, Berens A, Sleuwaegen E, Kuppens P. The emotion regulation function of nonsuicidal self-injury: a momentary assessment study in inpatients with borderline personality disorder features. J Abnorm Psychol 2017 Jan;126(1):89-95. [doi: 10.1037/abn0000229] [Medline: 27808541]
26. Khue LM, Ouh EL, Jarzabek S. Mood self-assessment on smartphones. In: Proceedings of the Conference on Wireless Health. 2015 Oct 14 Presented at: WH '15: Wireless Health 2015 Conference; October 14 - 16, 2015; Bethesda, Maryland p. 1-8. [doi: 10.1145/2811780.2811921]
27. Lathia N, Sandstrom GM, Mascolo C, Rentfrow PJ. Happier people live more active lives: using smartphones to link happiness and physical activity. PLoS One 2017 Jan 04;12(1):e0160589 [FREE Full text] [doi: 10.1371/journal.pone.0160589] [Medline: 28052069]
28. Panza E. Minority stress and eating behavior among overweight and obese sexual minority women. Ph. D. Thesis - Rutgers University-School of Graduate Studies, New Brunswick, New Jersey. 2018 Oct. URL: https://rucore.libraries.rutgers.edu/rutgers-lib/59107/ [accessed 2022-03-29]
29. Santangelo PS, Koenig J, Funke V, Parzer P, Resch F, Ebner-Priemer UW, et al. Ecological momentary assessment of affective and interpersonal instability in adolescent non-suicidal self-injury. J Abnorm Child Psychol 2017 Oct;45(7):1429-1438. [doi: 10.1007/s10802-016-0249-2] [Medline: 27995358]
30. Schwartz S, Schultz S, Reider A, Saunders EF. Daily mood monitoring of symptoms using smartphones in bipolar disorder: a pilot study assessing the feasibility of ecological momentary assessment. J Affect Disord 2016 Feb;191:88-93 [FREE Full text] [doi: 10.1016/j.jad.2015.11.013] [Medline: 26655117]
31. Pejovic V, Lathia N, Mascolo C, Musolesi M. Mobile-based experience sampling for behaviour research. In: Emotions and Personality in Personalized Services. Cham: Springer; Jul 14, 2016:141-161.
32. Dubad M, Winsper C, Meyer C, Livanou M, Marwaha S. A systematic review of the psychometric properties, usability and clinical impacts of mobile mood-monitoring applications in young people. Psychol Med 2018 Jan;48(2):208-228. [doi: 10.1017/S0033291717001659] [Medline: 28641609]
33. Kanjo E, Kuss D, Ang C. NotiMind: utilizing responses to smart phone notifications as affective sensors. IEEE Access 2017;5:22023-22035. [doi: 10.1109/ACCESS.2017.2755661]
34. Stieger S, Lewetz D. A week without using social media: results from an ecological momentary intervention study using smartphones. Cyberpsychol Behav Soc Netw 2018 Oct;21(10):618-624. [doi: 10.1089/cyber.2018.0070] [Medline: 30334650]

35. Watson D, Clark LA, Tellegen A. Development and validation of brief measures of positive and negative affect: the PANAS scales. J Pers Soc Psychol 1988 Jun;54(6):1063-1070. [doi: 10.1037//0022-3514.54.6.1063] [Medline: 3397865]
36. Pollak JP, Adams P, Gay G. PAM: a photographic affect meter for frequent, in situ measurement of affect. In: Proceedings of the International Conference on Human Factors in Computing Systems. 2011 Presented at: CHI 2011 International Conference on Human Factors in Computing Systems; May 7-12, 2011; Vancouver, BC, Canada p. 725-734. [doi: 10.1145/1978942.1979047]
37. Bradley MM, Lang PJ. Measuring emotion: the self-assessment manikin and the semantic differential. J Behav Ther Exp Psychiatry 1994 Mar;25(1):49-59. [doi: 10.1016/0005-7916(94)90063-9] [Medline: 7962581]
38. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, et al. StudentLife: Using Smartphones to Assess Mental Health and Academic Performance of College Students. Cham: Springer; 2017:7-33.
39. Griffin B, Saunders KE. Smartphones and wearables as a method for understanding symptom mechanisms. Front Psychiatry 2020 Jan 17;10:949 [FREE Full text] [doi: 10.3389/fpsyt.2019.00949] [Medline: 32009990]
40. Onnela JP, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. Neuropsychopharmacology 2016 Jun;41(7):1691-1696 [FREE Full text] [doi: 10.1038/npp.2016.7] [Medline: 26818126]
41. Aledavood T, Torous J, Hoyos AM, Naslund JA, Onnela JP, Keshavan M. Smartphone-based tracking of sleep in depression, anxiety, and psychotic disorders. Curr Psychiatry Rep 2019 Jun 04;21(7):49 [FREE Full text] [doi: 10.1007/s11920-019-1043-y] [Medline: 31161412]
42. Masud MT, Mamun MA, Thapa K, Lee DH, Griffiths MD, Yang SH. Unobtrusive monitoring of behavior and movement patterns to detect clinical depression severity level via smartphone. J Biomed Inform 2020 Mar;103:103371 [FREE Full text] [doi: 10.1016/j.jbi.2019.103371] [Medline: 31935462]
43. Mehrotra A, Hendley R, Musolesi M. Towards multi-modal anticipatory monitoring of depressive states through the analysis of human-smartphone interaction. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct. New York, USA: Association for Computing Machinery; 2016 Sep Presented at: UbiComp '16: The 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 12 - 16, 2016; Heidelberg Germany p. 1132-1138. [doi: 10.1145/2968219]
44. Asare KO, Terhorst Y, Vega J, Peltonen E, Lagerspetz E, Ferreira D. Predicting depression from smartphone behavioral markers using machine learning methods, hyperparameter optimization, and feature importance analysis: exploratory study. JMIR Mhealth Uhealth 2021 Jul 12;9(7):e26540 [FREE Full text] [doi: 10.2196/26540] [Medline: 34255713]
45. Ben-Zeev D, Wang R, Abdullah S, Brian R, Scherer EA, Mistler LA, et al. Mobile behavioral sensing for outpatients and inpatients with schizophrenia. Psychiatr Serv 2016 May 01;67(5):558-561 [FREE Full text] [doi: 10.1176/appi.ps.201500130] [Medline: 26695497]
46. Buck B, Scherer E, Brian R, Wang R, Wang W, Campbell A, et al. Relationships between smartphone social behavior and relapse in schizophrenia: a preliminary report. Schizophr Res 2019 Jun;208:167-172 [FREE Full text] [doi: 10.1016/j.schres.2019.03.014] [Medline: 30940400]
47. Staples P, Torous J, Barnett I, Carlson K, Sandoval L, Keshavan M, et al. A comparison of passive and active estimates of sleep in a cohort with schizophrenia. NPJ Schizophr 2017 Oct 16;3(1):37 [FREE Full text] [doi: 10.1038/s41537-017-0038-0] [Medline: 29038553]
48. Wang R, Wang W, Aung MS, Ben-Zeev D, Brian R, Campbell AT, et al. Predicting symptom trajectories of schizophrenia using mobile sensing. Proc ACM Interact Mob Wearable Ubiquitous Technol 2017 Sep 11;1(3):1-24. [doi: 10.1145/3130976]
49. Alvarez-Lozano J, Osmani V, Mayora O, Frost M, Bardram J, Faurholt-Jepsen M, et al. Tell me your apps and I will tell you your mood: correlation of apps usage with bipolar disorder state. In: Proceedings of the 7th International Conference on PErvasive Technologies Related to Assistive Environments. New York, USA: Association for Computing Machinery; 2014 May Presented at: PETRA '14: The 7th International Conference on PErvasive Technologies Related to Assistive Environments; May 27 - 30, 2014; Rhodes Greece p. 1-7. [doi: 10.1145/2674396.2674408]
50. Beiwinkel T, Kindermann S, Maier A, Kerl C, Moock J, Barbian G, et al. Using smartphones to monitor bipolar disorder symptoms: a pilot study. JMIR Ment Health 2016 Jan 06;3(1):e2 [FREE Full text] [doi: 10.2196/mental.4560] [Medline: 26740354]
51. Carr O, Andreotti F, Saunders K, Palmius N, Goodwin G, De Vos M. Monitoring depression in bipolar disorder using circadian measures from smartphone accelerometers. arXiv. Preprint posted online on July 4, 2020. [FREE Full text]
52. Faurholt-Jepsen M, Bauer M, Kessing LV. Smartphone-based objective monitoring in bipolar disorder: status and considerations. Int J Bipolar Disord 2018 Jan 23;6(1):6 [FREE Full text] [doi: 10.1186/s40345-017-0110-8] [Medline: 29359252]
53. Puiatti A, Mudda S, Giordano S, Mayora O. Smartphone-centred wearable sensors network for monitoring patients with bipolar disorder. Annu Int Conf IEEE Eng Med Biol Soc 2011;2011:3644-3647. [doi: 10.1109/IEMBS.2011.6090613] [Medline: 22255129]
54. Bauer G, Lukowicz P. Can smartphones detect stress-related changes in the behaviour of individuals? In: Proceedings of the IEEE International Conference on Pervasive Computing and Communications Workshops. 2012 Presented at: IEEE International Conference on Pervasive Computing and Communications Workshops; March 19-23, 2012; Lugano, Switzerland p. 423-426. [doi: 10.1109/PerComW.2012.6197525]

55. Ciman M, Wac K, Gaggi O. iSensestress: assessing stress through human-smartphone interaction analysis. In: Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare. Brussels, Belgium: ICST; 2020 May 20 Presented at: PervasiveHealth '15: 9th International Conference on Pervasive Computing Technologies for Healthcare; May 20-23, 2015; Istanbul, Turkey p. 84-91. [doi: 10.4108/icst.pervasivehealth.2015.259280]
56. Ferdous R, Osmani V, Mayora O. Smartphone app usage as a predictor of perceived stress levels at workplace. In: Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare. Istanbul, Turkey: ICST; 2015 Aug 03 Presented at: PervasiveHealth '15: 9th International Conference on Pervasive Computing Technologies for Healthcare; May 20–23, 2015; Istanbul, Turkey p. 225-228. [doi: 10.4108/icst.pervasivehealth.2015.260192]
57. Garcia-Ceja E, Osmani V, Mayora O. Automatic stress detection in working environments from smartphones' accelerometer data: a first step. IEEE J Biomed Health Inform 2016 Jul;20(4):1053-1060. [doi: 10.1109/JBHI.2015.2446195] [Medline: 26087509]
58. Mayya S, Jilla V, Tiwari VN, Nayak MM, Narayanan R. Continuous monitoring of stress on smartphone using heart rate variability. In: Proceedings of the IEEE 15th International Conference on Bioinformatics and Bioengineering (BIBE). 2015 Presented at: IEEE 15th International Conference on Bioinformatics and Bioengineering (BIBE); November 2-4, 2015; Belgrade, Serbia p. 1-5. [doi: 10.1109/BIBE.2015.7367627]
59. Boukhechba M, Huang Y, Chow P, Fua K, Teachman BA, Barnes LE. Monitoring social anxiety from mobility and communication patterns. In: Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and 2017 ACM International Symposium on Wearable Computers. New York, USA: Association for Computing Machinery; 2017 Presented at: UbiComp '17: The 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing; September 11-15, 2017; Maui, Hawaii p. 749-753. [doi: 10.1145/3123024.3125607]
60. Fukazawa Y, Ito T, Okimura T, Yamashita Y, Maeda T, Ota J. Predicting anxiety state using smartphone-based passive sensing. J Biomed Inform 2019 May;93:103151 [FREE Full text] [doi: 10.1016/j.jbi.2019.103151] [Medline: 30880254]
61. Gong J, Huang Y, Chow PI, Fua K, Gerber MS, Teachman BA, et al. Understanding behavioral dynamics of social anxiety among college students through smartphone sensors. Information Fusion 2019 Sep;49:57-68. [doi: 10.1016/j.inffus.2018.09.002]
62. Rashid H, Mendu S, Daniel KE, Beltzer ML, Teachman BA, Boukhechba M, et al. Predicting subjective measures of social anxiety from sparsely collected mobile sensor data. Proceedings of the ACM On Interactive Mobile, Wearable, and Ubiquitous Technologies 2020 Sep 04;4(3):1-24. [doi: 10.1145/3411823]
63. Berry N, Bucci S, Lobban F. Use of the internet and mobile phones for self-management of severe mental health problems: qualitative study of staff views. JMIR Ment Health 2017 Nov 01;4(4):e52 [FREE Full text] [doi: 10.2196/mental.8311] [Medline: 29092809]
64. Fairburn CG, Rothwell ER. Apps and eating disorders: a systematic clinical appraisal. Int J Eat Disord 2015 Nov;48(7):1038-1046 [FREE Full text] [doi: 10.1002/eat.22398] [Medline: 25728705]
65. Lindgreen P, Lomborg K, Clausen L. Patient experiences using a self-monitoring app in eating disorder treatment: qualitative study. JMIR mHealth and uHealth 2018 Jun 22;6(6):e10253 [FREE Full text] [doi: 10.2196/10253] [Medline: 29934285]
66. Technology enabled best practice for eating disorder treatment. Recovery Record. URL: https://www.recoveryrecord.com/ [accessed 2022-03-18]
67. Tulu B, Ruiz C, Allard J, Acheson J, Busch A, Roskuski A, et al. SlipBuddy: a mobile health intervention to prevent overeating. In: Proceedings of the 50th Hawaii International Conference on System Sciences. 2017 Presented at: 50th Hawaii International Conference on System Sciences; January 4-7, 2017; Waikōloa Beach. [doi: 10.24251/HICSS.2017.436]
68. Ferreira D, Kostakos V, Dey AK. AWARE: Mobile context instrumentation framework. Frontiers in ICT. 2015 Apr 20. URL: https://www.frontiersin.org/articles/10.3389/fict.2015.00006/full [accessed 2022-03-18]
69. Branley-Bell D, Talbot CV. Exploring the impact of the COVID-19 pandemic and UK lockdown on individuals with experience of eating disorders. J Eat Disord 2020 Aug 31;8:44 [FREE Full text] [doi: 10.1186/s40337-020-00319-y] [Medline: 32874585]
70. Schlegl S, Maier J, Meule A, Voderholzer U. Eating disorders in times of the COVID-19 pandemic: results from an online survey of patients with anorexia nervosa. Int J Eat Disord 2020 Aug 25;53(11):1791-1800. [doi: 10.1002/eat.23374]
71. Vega J. Reproducible analysis pipeline for data streams. RAPIDS 1.8. URL: https://www.rapids.science/latest/ [accessed 2022-03-18]
72. Vega J, Li M, Aguillera K, Goel N, Joshi E, Khandekar K, et al. Reproducible analysis pipeline for data streams: open-source software to process data collected with mobile devices. Front Digit Health 2021 Nov 21;3:769823 [FREE Full text] [doi: 10.3389/fdgth.2021.769823] [Medline: 34870271]
73. Parker RI, Vannest K. An improved effect size for single-case research: nonoverlap of all pairs. Behav Ther 2009 Dec;40(4):357-367. [doi: 10.1016/j.beth.2008.10.006] [Medline: 19892081]
74. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Apr;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
75. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]
76. JulioV / demmi_aware. GitHub. URL: https://github.com/JulioV/demmi_aware [accessed 2022-03-18]

77. JulioV/demmi-analysis. GitHub. URL: https://github.com/JulioV/demmi-analysis [accessed 2022-03-18]
78. Pustejovsky JE, Swan DM. SingleCaseES: a calculator for single-case effect sizes. Cran R Projecft. 2021 Sep 30. URL: https://CRAN.R-project.org/package=SingleCaseES [accessed 2022-03-18]
79. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Society: Ser B (Methodol) 2018 Dec 05;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
80. Haedt-Matt AA, Keel PK. Revisiting the affect regulation model of binge eating: a meta-analysis of studies using ecological momentary assessment. Psychol Bull 2011 Jul;137(4):660-681 [FREE Full text] [doi: 10.1037/a0023660] [Medline: 21574678]
81. Jacobson NC, Chung YJ. Passive sensing of prediction of moment-to-moment depressed mood among undergraduates with clinical levels of depression sample using smartphones. Sensors (Basel) 2020 Jun 24;20(12):3572 [FREE Full text] [doi: 10.3390/s20123572] [Medline: 32599801]
82. Dingemans A, Danner U, Parks M. Emotion regulation in binge eating disorder: a review. Nutrients 2017 Nov 22;9(11):1274 [FREE Full text] [doi: 10.3390/nu9111274] [Medline: 29165348]
83. Wilson GT, Sysko R. Frequency of binge eating episodes in bulimia nervosa and binge eating disorder: diagnostic considerations. Int J Eat Disord 2009 Nov;42(7):603-610 [FREE Full text] [doi: 10.1002/eat.20726] [Medline: 19610014]
84. Rossiter EM, Agras WS, Telch CF, Bruce B. The eating patterns of non-purging bulimic subjects. Int J Eat Disord 1992 Mar;11(2):111-120. [doi: 10.1002/1098-108x(199203)11:2<111::aid-eat2260110203>3.0.co;2-j]
85. Fellger A, Sprint G, Weeks D, Crooks E, Cook DJ. Wearable device-independent next day activity and next night sleep prediction for rehabilitation populations. IEEE J Transl Eng Health Med 2020 Aug 05;8:2700509 [FREE Full text] [doi: 10.1109/JTEHM.2020.3014564] [Medline: 32802598]
86. Case MA, Burwick HA, Volpp KG, Patel MS. Accuracy of smartphone applications and wearable devices for tracking physical activity data. J Am Med Assoc 2015 Feb 10;313(6):625-626. [doi: 10.1001/jama.2014.17841] [Medline: 25668268]
87. Gal R, May AM, van Overmeeren EJ, Simons M, Monninkhof EM. The effect of physical activity interventions comprising wearables and smartphone applications on physical activity: a systematic review and meta-analysis. Sports Med Open 2018 Sep 03;4(1):42 [FREE Full text] [doi: 10.1186/s40798-018-0157-9] [Medline: 30178072]
88. Hilbert A, Petroff D, Herpertz S, Pietrowsky R, Tuschen-Caffier B, Vocks S, et al. Meta-analysis of the efficacy of psychological and medical treatments for binge-eating disorder. J Consult Clin Psychol 2019 Jan;87(1):91-105. [doi: 10.1037/ccp0000358] [Medline: 30570304]
89. Linardon J, Wade TD. How many individuals achieve symptom abstinence following psychological treatments for bulimia nervosa? A meta-analytic review. Int J Eat Disord 2018 Apr;51(4):287-294. [doi: 10.1002/eat.22838] [Medline: 29417609]
90. Linardon J, Wade TD, de la Piedad Garcia X, Brennan L. The efficacy of cognitive-behavioral therapy for eating disorders: a systematic review and meta-analysis. J Consult Clin Psychol 2017 Nov;85(11):1080-1094. [doi: 10.1037/ccp0000245] [Medline: 29083223]
91. Honary M, Bell B, Clinch S, Vega J, Kroll L, Sefi A, et al. Shaping the design of smartphone-based interventions for self-harm. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. New York, USA: Association for Computing Machinery; 2020 Apr 23 Presented at: CHI '20: CHI Conference on Human Factors in Computing Systems; April 25 - 30, 2020; Honolulu HI USA p. 1-14. [doi: 10.1145/3313831]
92. Cohen DA, Wu SY, Farley TA. Comparing the cost-effectiveness of HIV prevention interventions. J Acquir Immune Defic Syndr 2004 Nov 01;37(3):1404-1414. [doi: 10.1097/01.qai.0000123271.76723.96] [Medline: 15483470]
93. Cobiac LJ, Vos T, Barendregt JJ. Cost-effectiveness of interventions to promote physical activity: a modelling study. PLoS Med 2009 Jul 14;6(7):e1000110 [FREE Full text] [doi: 10.1371/journal.pmed.1000110] [Medline: 19597537]
94. Tsevat J. Impact and cost-effectiveness of smoking interventions. Am J Med 1992 Jul 15;93(1A):43-47. [doi: 10.1016/0002-9343(92)90627-n] [Medline: 1497003]
95. Wutzke SE, Shiell A, Gomel MK, Conigrave KM. Cost effectiveness of brief interventions for reducing alcohol consumption. Soc Sci Med 2001 Mar;52(6):863-870. [doi: 10.1016/s0277-9536(00)00189-1] [Medline: 11234861]
96. Lynch FL, Striegel-Moore RF, Dickerson JH, Perrin N, Debar L, Wilson GT, et al. Cost-effectiveness of guided self-help treatment for recurrent binge eating. J Consult Clin Psychol 2010 Jun;78(3):322-333 [FREE Full text] [doi: 10.1037/a0018982] [Medline: 20515208]
97. Reychav I, Beeri R, Balapour A, Raban DR, Sabherwal R, Azuri J. How reliable are self-assessments using mobile technology in healthcare? The effects of technology identity and self-efficacy. Comput Hum Behav 2019 Feb;91:52-61. [doi: 10.1016/j.chb.2018.09.024]
98. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient compliance with paper and electronic diaries. Control Clin Trials 2003 Apr;24(2):182-199. [doi: 10.1016/s0197-2456(02)00320-3] [Medline: 12689739]
99. Swan M. The quantified self: fundamental disruption in big data science and biological discovery. Big Data 2013 Jun;1(2):85-99. [doi: 10.1089/big.2012.0002] [Medline: 27442063]
100. Or CK, Karsh BT. A systematic review of patient acceptance of consumer health information technology. J Am Med Inform Assoc 2009 Jul;16(4):550-560 [FREE Full text] [doi: 10.1197/jamia.M2888] [Medline: 19390112]
101. Or C, Tao D. Usability study of a computer-based self-management system for older adults with chronic diseases. JMIR Res Protoc 2012 Nov 08;1(2):e13 [FREE Full text] [doi: 10.2196/resprot.2184] [Medline: 23612015]

102. Nielsen J. Usability Engineering. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc; 1993.
103. Honary M, Fisher NR, McNaney R, Lobban F. A web-based intervention for relatives of people experiencing psychosis or bipolar disorder: design study using a user-centered approach. JMIR Ment Health 2018 Dec 07;5(4):e11473 [FREE Full text] [doi: 10.2196/11473] [Medline: 30530457]

Abbreviations

CBT: cognitive behavioral therapy
DeMMI app: Detection of mental health behaviors using Mobile Interactions
EMA: Ecological Momentary Assessment
NAP: Nonoverlap of All Pairs
PANAS: Positive and Negative Affect Scale
RAPIDS: Reproducible Analysis Pipeline for Data Streams

Edited by J Torous; submitted 16.07.21; peer-reviewed by I Barnett, C Or; comments to author 18.09.21; revised version received 16.01.22; accepted 17.01.22; published 25.04.22.

Please cite as:
Vega J, Bell BT, Taylor C, Xie J, Ng H, Honary M, McNaney R
Detecting Mental Health Behaviors Using Mobile Interactions: Exploratory Study Focusing on Binge Eating
JMIR Ment Health 2022;9(4):e32146
URL: https://mental.jmir.org/2022/4/e32146
doi:10.2196/32146
PMID:35086064

Original Paper

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

Michael Mullarkey[1], PhD; Mallory Dobias[1], MA; Jenna Sung[1], MA; Isaac Ahuvia[1], BA; Jason Shumake[1], PhD; Christopher Beevers[1], PhD; Jessica Schleider[1], 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

(JMIR Ment Health 2022;9(4):e33473) doi:10.2196/33473

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.

Question and hypotheses	Sampling plan (eg, power analysis)[a]	Analysis plan[a]	Interpretation given to different outcomes
Does the Contain COVID Anxiety SSI[b] decrease generalized anxiety symptoms from baseline to 2 weeks later more than the placebo SSI Remain COVID Free?			
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.	H1: n=400 for 95% power	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.	Support for H1: If the *P* value for condition is <.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 >.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.
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.	H0: n=150 for 95% power	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.	Support for H0: If the *P* value for condition is <.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 Remain COVID Free SSI or the *P* value or intervention order is >.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 <.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 >.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.
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 <.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 >.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.

Question and hypotheses	Sampling plan (eg, power analysis)[a]	Analysis plan[a]	Interpretation given to different outcomes
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.	H0: n=150 for 95% power	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*=–0.63 to *d*=0.21.	Support for H0: If the *P* value for condition is <.0167 in the H1 linear model and the 95% CI for the difference in perceived control over anxiety is negative and does not include 0 when participants are randomized to the Remain COVID Free SSI or the *P* value or intervention order is >.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 <.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 described in the "Support for H0" section is >.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.
Does completing the Contain COVID Anxiety SSI have an association with social distancing intentions statistically equivalent to 0?			
H1: Social distancing intentions do not increase or decrease pre-Contain COVID Anxiety SSIs to immediate post-Contain COVID Anxiety SSI.	H1: n=154 for 95% power	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.	Support for H1: The *P* value for the paired equivalence test described in the analysis plan column is <.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 <.0167).
H0: Social distancing intentions either increase or decrease pre- to immediate post-Contain COVID Anxiety SSI	H0: n=156 for 95% power	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.	Support for H0: The *P* value for the paired equivalence test in the analysis plan column is >.0167, and the *P* value for the paired *t* test in the analysis plan column is <.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).

[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.

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 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.

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-probability sample was stratified on age, sex, and ethnicity. To help further maximize the generalizability of our findings, there were no prescreening inclusion nor 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 their beliefs about the intentions of others to engage in social distancing behaviors like avoiding public spaces and private gatherings.

The participants were then randomized to and spent approximately 8 minutes completing one of the SSIs described in the following sections (which one depended on the number generated by the random sequence from the Qualtrics randomizer described in the previous paragraphs), immediately followed by approximately 2 minutes completing the ACQ-EC scale, the Hand Washing Intentions scale, several questions about social distancing from a standardized Centers for Disease Control and Prevention (CDC) item bank, and the comprehension questions. They were then sent to a Prolific link for compensation (with an incentive of US $2.17 for the 20-minute survey, or US $6.50 per hour) and reminded of the follow-up survey 2 weeks later. Participants also received a reminder through the Prolific platform 2 weeks later to participate in the 2-minute follow-up survey containing a measure of generalized anxiety symptoms, perceived control over anxiety, and an anchor-based question (see [60]) to help determine the smallest subjectively noticeable difference in generalized anxiety symptoms over 2 weeks. They were then debriefed, provided with mental health resources, and sent to a Prolific link for compensation (US $0.22 for the 2-minute survey, or US $6.50 per hour). We focus here on describing 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

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 chosen in step 2. 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 on10 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. 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.

Demographics	Treatment received	
	Active Contain COVID Anxiety (n=261)	Placebo Remain COVID Free (n=261)
Age (years), mean (SD)	46.02 (15.65)	46.19 (15.71)
Race/ethnicity, n (%)		
American Indian and/or Alaska Native	0 (0)	1 (0.4)
Asian	19 (7.3)	16 (6.1)
African American	31 (12.0)	39 (15.0)
Hispanic or Latino/a	16 (6.1)	11 (4.2)
Native Hawaiian or Pacific Islander	1 (0.4)	0 (0)
White, non-Hispanic	187 (72.0)	190 (73.0)
More than one race	4 (1.5)	3 (1.1)
Other	3 (1.1)	1 (0.4)
Gender, n (%)		
Agender	2 (0.8)	2 (0.8)
Genderqueer or gender fluid	3 (1.1)	0 (0)
Man	127 (49.0)	129 (49.0)
Trans man	3 (1.1)	1 (0.4)
Woman	125 (48.0)	129 (49.0)
Other	1 (0.4)	0 (0)
Sexual orientation, n (%)		
Asexual	6 (2.3)	5 (1.9)
Bisexual	20 (7.7)	18 (6.9)
Gay	3 (1.1)	7 (2.7)
Heterosexual	218 (84.0)	220 (85.0)
Lesbian	2 (0.8)	2 (0.8)
Pansexual	3 (1.1)	6 (2.3)
Queer	5 (1.9)	2 (0.8)
Questioning or unsure	2 (0.8)	0 (0)
Other	2 (0.8)	0 (0)
Unknown	0 (0)	1 (0.4)
Education, n (%)		
Less than high school degree	1 (0.4)	2 (0.8)
High school degree	27 (10.0)	29 (11.0)
Some college, no degree	74 (28.0)	67 (26.0)
Associate degree	26 (10.0)	31 (12.0)
Bachelor’s degree	77 (30.0)	99 (38.0)
Master’s degree	46 (18.0)	24 (9.2)
Professional degree	4 (1.5)	4 (1.5)
Doctorate	6 (2.3)2	5 (1.9)
Annual income (US $), n (%)		
Less than 10,000	17 (6.5)	14 (5.4)
10,000-19,999	24 (9.2)	22 (8.4)
20,000-29,999	29 (11.0)	28 (11.0)

Demographics	Treatment received	
	Active Contain COVID Anxiety (n=261)	Placebo Remain COVID Free (n=261)
30,000-39,999	23 (8.8)	24 (9.2)
40,000-49,999	21 (8.0)	33 (13.0)
50,000-59,999	26 (10.0)	25 (9.6)
60,000-69,999	15 (5.7)	24 (9.2)
70,000-79,999	28 (11.0)	20 (7.7)
80,000-89,999	15 (5.7)	10 (3.8)
90,000-99,999	15 (5.7)	12 (4.6)
100,000-149,999	27 (10.0)	30 (11.0)
150,000 or more	21 (8.0)	19 (7.3)
Relationship status, n (%)		
No current relationship	101 (39.0)	118 (45.0)
Relationship, not living together	25 (9.6)	20 (7.7)
Relationship, living together	24 (9.2)	22 (8.4)
Engaged	3 (1.1)	3 (1.1)
Married	108 (41.0)	98 (38.0)
Has children, n (%)	117 (45.0)	127 (49.0)
Health insurance covers mental health, n (%)	195 (75.0)	196 (75.0)
Received psychotherapy in the past year, n (%)	56 (21.0)	54 (21.0)
Received medication for mental health in the past year, n (%)	56 (21.0)	54 (21.0)
Perceived need for mental health treatment in the past year, n (%)	87 (33.0)	89 (34.0)
Baseline IDAS[a]-Dysphoria (1-5), mean (SD)	2.00 (0.90)	2.09 (0.92)
Baseline GAD-7[b] (0-3), mean (SD)	0.75 (0.74)	0.77 (0.75)
Baseline self-hate (1-7), mean (SD)	2.07 (1.65)	2.20 (1.59)
Baseline perceived control over anxiety (0-5), mean (SD)	2.88 (1.34)	2.85 (1.29)
Baseline Hand Washing Intentions (1-7), mean (SD)	5.08 (1.59)	5.01 (1.61)
Baseline social distancing intentions of others (1-4), mean (SD)	2.27 (1.12)	2.32 (1.12)

[a]IDAS: Inventory of Depression and Anxiety Symptoms.

[b]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_{520}=–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_{520}=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 for only participants who answered both comprehension questions correctly (n=387). Within-group effect sizes indicated small but nonzero increases in generalized anxiety in both the Contain COVID Anxiety (t_{260}=2.00; d_z=0.12, 95% CI 0.002 to 0.25) and Remain COVID Free (t_{260}=2.41; d_z=0.15, 95% CI 0.03 to 0.27) groups.

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 of –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_{520}=–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_{520}=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_z=0.06, 95% CI –0.06 to 0.19) and Remain COVID Free (t_{260}=1.63; d_z=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.
[DOCX File , 56 KB - mental_v9i4e33473_app1.docx]

Multimedia Appendix 2
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 2699 KB - mental_v9i4e33473_app2.pdf]

Multimedia Appendix 3
COVID Anxiety SSI RR preprocessing and analyses.
[DOCX File , 3560 KB - mental_v9i4e33473_app3.docx]

References

1. Balz D, Guskin E. Poll finds universal lifestyle changes, rising stress and growing fears about catching coronavirus. The Washington Post. 2020 Mar 27. URL: https://tinyurl.com/4uwxh38k [accessed 2022-03-20]
2. Dai Y, Hu G, Xiong H, Qiu H, Yuan X. Psychological impact of the coronavirus disease 2019 (COVID-19) outbreak on healthcare workers in China. medRxiv. 2020 Mar 06. URL: https://www.medrxiv.org/content/10.1101/2020.03.03.20030874v1 [accessed 2022-03-20]
3. Ying Y, Ruan L, Kong F, Zhu B, Ji Y, Lou Z. Mental health status among family members of health care workers in Ningbo, China, during the coronavirus disease 2019 (COVID-19) outbreak: a cross-sectional study. BMC Psychiatry 2020 Jul 17;20(1):379 [FREE Full text] [doi: 10.1186/s12888-020-02784-w] [Medline: 32680478]
4. Qian M, Wu Q, Wu P, Hou Z, Liang Y, Cowling BJ, et al. Psychological responses, behavioral changes and public perceptions during the early phase of the COVID-19 outbreak in China: a population based cross-sectional survey. medRxiv. 2020 Feb 20. URL: https://www.medrxiv.org/content/10.1101/2020.02.18.20024448v1 [accessed 2022-03-20]
5. Phillips MR, Zhang J, Shi Q, Song Z, Ding Z, Pang S, et al. Prevalence, treatment, and associated disability of mental disorders in four provinces in China during 2001–05: an epidemiological survey. The Lancet 2009 Jun;373(9680):2041-2053. [doi: 10.1016/s0140-6736(09)60660-7]
6. Gruber J, Prinstein MJ, Clark LA, Rottenberg J, Abramowitz JS, Albano AM, et al. Mental health and clinical psychological science in the time of COVID-19: Challenges, opportunities, and a call to action. Am Psychol 2021 Apr;76(3):409-426 [FREE Full text] [doi: 10.1037/amp0000707] [Medline: 32772538]
7. Kirzinger A, Kearney A, Hamel L, Brodie M. KFF Health Tracking Poll - Early April 2020: The Impact Of Coronavirus On Life In America. Kaiser Family Foundation. 2020 Apr 02. URL: https://www.kff.org/health-reform/report/kff-health-tracking-poll-early-april-2020/ [accessed 2022-03-20]
8. Liu C, Yang YZ, Zhang XM, Xu X, Dou QL, Zhang WW, et al. The prevalence and influencing factors in anxiety in medical workers fighting COVID-19 in China: a cross-sectional survey. Epidemiol Infect 2020 May 20;148:e98 [FREE Full text] [doi: 10.1017/S0950268820001107] [Medline: 32430088]
9. Huang Y, Zhao N. Generalized anxiety disorder, depressive symptoms and sleep quality during COVID-19 outbreak in China: a web-based cross-sectional survey. Psychiatry Res 2020 Jun;288:112954 [FREE Full text] [doi: 10.1016/j.psychres.2020.112954] [Medline: 32325383]
10. Deacon B, Lickel J, Abramowitz JS. Medical utilization across the anxiety disorders. J Anxiety Disord 2008;22(2):344-350. [doi: 10.1016/j.janxdis.2007.03.004] [Medline: 17420113]
11. Coley KC, Saul MI, Seybert AL. Economic burden of not recognizing panic disorder in the emergency department. J Emerg Med 2009 Jan;36(1):3-7. [doi: 10.1016/j.jemermed.2007.06.002] [Medline: 17933481]
12. Frazier PA. Perceived control and distress following sexual assault: a longitudinal test of a new model. J Pers Soc Psychol 2003 Jun;84(6):1257-1269. [doi: 10.1037/0022-3514.84.6.1257] [Medline: 12793588]
13. Bárez M, Blasco T, Fernández-Castro J, Viladrich C. Perceived control and psychological distress in women with breast cancer: a longitudinal study. J Behav Med 2009 Apr 25;32(2):187-196. [doi: 10.1007/s10865-008-9180-5] [Medline: 18815877]
14. Moser DK, Dracup K. Psychosocial recovery from a cardiac event: The influence of perceived control. Heart & Lung 1995 Jul;24(4):273-280. [doi: 10.1016/s0147-9563(05)80070-6]
15. Otto MW. Anxiety and its disorders: the nature and treatment of anxiety and panic. Psychosomatics 1989 Aug;30(3):349-350. [doi: 10.1016/s0033-3182(89)72285-4]
16. Barlow DH. Anxiety and Its Disorders: The Nature and Treatment of Anxiety and Panic, 2nd edition. New York, NY: Guilford Publications; 2004.
17. Mineka S, Zinbarg R. Conditioning and ethological models of anxiety disorders: stress-in-dynamic-context anxiety models. Nebr Symp Motiv 1996;43:135-210. [Medline: 8912309]
18. Zvolensky MJ, Eifert GH, Lejuez C, Hopko DR, Forsyth JP. Assessing the perceived predictability of anxiety-related events: a report on the perceived predictability index. Journal of Behavior Therapy and Experimental Psychiatry 2000 Sep;31(3-4):201-218. [doi: 10.1016/s0005-7916(01)00006-4]
19. Brown TA, White KS, Forsyth JP, Barlow DH. The structure of perceived emotional control: Psychometric properties of a revised anxiety control questionnaire. Behavior Therapy 2004 Feb;35(1):75-99. [doi: 10.1016/s0005-7894(04)80005-4]

20. Rapee RM, Craske MG, Brown TA, Barlow DH. Measurement of perceived control over anxiety-related events. Behavior Therapy 1996 Jan;27(2):279-293. [doi: 10.1016/s0005-7894(96)80018-9]
21. Rapee RM. Perceived threat and perceived control as predictors of the degree of fear in physical and social situations. Journal of Anxiety Disorders 1997 Sep;11(5):455-461. [doi: 10.1016/s0887-6185(97)00022-4]
22. Brown TA, Naragon-Gainey K. Evaluation of the unique and specific contributions of dimensions of the triple vulnerability model to the prediction of DSM-IV anxiety and mood disorder constructs. Behav Ther 2013 Jun;44(2):277-292 [FREE Full text] [doi: 10.1016/j.beth.2012.11.002] [Medline: 23611077]
23. Gallagher MW, Bentley KH, Barlow DH. Perceived control and vulnerability to anxiety disorders: a meta-analytic review. Cogn Ther Res 2014 Jun 13;38(6):571-584. [doi: 10.1007/s10608-014-9624-x]
24. Mineka S, Thomas C. Mechanisms of change in exposure therapy for anxiety disorders. In: Dalgleish T, Power MJ, editors. Handbook of Cognition and Emotion. Hoboken, NJ: John Wiley & Sons, Ltd; 1999:747-764.
25. Sanderson WC, Rapee RM, Barlow DH. The influence of an illusion of control on panic attacks induced via inhalation of 5.5% carbon dioxide-enriched air. Arch Gen Psychiatry 1989 Feb 01;46(2):157-162. [doi: 10.1001/archpsyc.1989.01810020059010] [Medline: 2492423]
26. Zvolensky MJ, Eifert GH, Lejuez CW. Offset control during recurrent 20% carbon dioxide-enriched air induction: relation to individual difference variables. Emotion 2001 Jun;1(2):148-165. [doi: 10.1037/1528-3542.1.2.148] [Medline: 12899194]
27. Wadsworth L, Hayes-Skelton S. Exploring perceived control, a low-control task, and a brief acceptance intervention in a low and high transdiagnostic anxiety sample. Neurology, Psychiatry and Brain Research 2020 Feb;35:1-9. [doi: 10.1016/j.npbr.2019.11.001]
28. Schmidt NB, Lerew DR. Prospective evaluation of perceived control, predictability, and anxiety sensitivity in the pathogenesis of panic. Journal of Psychopathology and Behavioral Assessment 2002;24(4):207-214. [doi: 10.1023/A:1020795114296]
29. Wadsworth L, Wessman I, Beard C, Bjorgvinsson T. Levels of perceived control and treatment response in a brief partial hospital setting. Neurology, Psychiatry and Brain Research 2019 Dec;34:1-8. [doi: 10.1016/j.npbr.2019.08.001]
30. Mahon JN, Rohan KJ, Nillni YI, Zvolensky MJ. The role of perceived control over anxiety in prospective symptom reports across the menstrual cycle. Arch Womens Ment Health 2015 Apr 1;18(2):239-246. [doi: 10.1007/s00737-014-0456-1] [Medline: 25269759]
31. Treanor M, Erisman SM, Salters-Pedneault K, Roemer L, Orsillo SM. Acceptance-based behavioral therapy for GAD: effects on outcomes from three theoretical models. Depress Anxiety 2011 Feb 13;28(2):127-136 [FREE Full text] [doi: 10.1002/da.20766] [Medline: 21284065]
32. Schleider JL, Weisz JR. Reducing risk for anxiety and depression in adolescents: Effects of a single-session intervention teaching that personality can change. Behav Res Ther 2016 Dec;87:170-181 [FREE Full text] [doi: 10.1016/j.brat.2016.09.011] [Medline: 27697671]
33. Schleider JL, Abel MR, Weisz JR. Do immediate gains predict long-term symptom change? Findings from a randomized trial of a single-session intervention for youth anxiety and depression. Child Psychiatry Hum Dev 2019 Oct 16;50(5):868-881 [FREE Full text] [doi: 10.1007/s10578-019-00889-2] [Medline: 30993499]
34. Schleider JL, Weisz JR. Little treatments, promising effects? Meta-analysis of single-session interventions for youth psychiatric problems. J Am Acad Child Adolesc Psychiatry 2017 Feb;56(2):107-115. [doi: 10.1016/j.jaac.2016.11.007] [Medline: 28117056]
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]
36. Aafjes-van Doorn K, Sweeney K. The effectiveness of initial therapy contact: A systematic review. Clin Psychol Rev 2019 Dec;74:101786. [doi: 10.1016/j.cpr.2019.101786] [Medline: 31733564]
37. Tanner-Smith EE, Lipsey MW. Brief alcohol interventions for adolescents and young adults: a systematic review and meta-analysis. J Subst Abuse Treat 2015 Apr;51:1-18 [FREE Full text] [doi: 10.1016/j.jsat.2014.09.001] [Medline: 25300577]
38. Loughman A, Haslam N. Neuroscientific explanations and the stigma of mental disorder: a meta-analytic study. Cogn Res Princ Implic 2018 Nov 14;3(1):43 [FREE Full text] [doi: 10.1186/s41235-018-0136-1] [Medline: 30426319]
39. Schleider JL, Dobias ML, Sung JY, Mullarkey MC. Future directions in single-session youth mental health interventions. J Clin Child Adolesc Psychol 2020 Dec 04;49(2):264-278 [FREE Full text] [doi: 10.1080/15374416.2019.1683852] [Medline: 31799863]
40. Cuijpers P, Kleiboer A, Karyotaki E, Riper H. Internet and mobile interventions for depression: Opportunities and challenges. Depress Anxiety 2017 Jul 04;34(7):596-602. [doi: 10.1002/da.22641] [Medline: 28471479]
41. Schleider JL, Burnette JL, Widman L, Hoyt C, Prinstein MJ. Randomized trial of a single-session growth mind-set intervention for rural adolescents' internalizing and externalizing problems. J Clin Child Adolesc Psychol 2020 Jun 20;49(5):660-672 [FREE Full text] [doi: 10.1080/15374416.2019.1622123] [Medline: 31219698]
42. Schleider J, Weisz J. A single-session growth mindset intervention for adolescent anxiety and depression: 9-month outcomes of a randomized trial. J Child Psychol Psychiatry 2018 Feb;59(2):160-170. [doi: 10.1111/jcpp.12811] [Medline: 28921523]
43. Eskreis-Winkler L, Milkman KL, Gromet DM, Duckworth AL. A large-scale field experiment shows giving advice improves academic outcomes for the advisor. Proc Natl Acad Sci U S A 2019 Jul 23;116(30):14808-14810 [FREE Full text] [doi: 10.1073/pnas.1908779116] [Medline: 31285348]

44. Piff PK, Wiwad D, Robinson AR, Aknin LB, Mercier B, Shariff A. Shifting attributions for poverty motivates opposition to inequality and enhances egalitarianism. Nat Hum Behav 2020 May 16;4(5):496-505. [doi: 10.1038/s41562-020-0835-8] [Medline: 32203322]
45. Kazdin AE, Blase SL. Rebooting psychotherapy research and practice to reduce the burden of mental illness. Perspect Psychol Sci 2011 Jan 03;6(1):21-37 [FREE Full text] [doi: 10.1177/1745691610393527] [Medline: 26162113]
46. Kazdin AE. Annual Research Review: Expanding mental health services through novel models of intervention delivery. J Child Psychol Psychiatry 2019 Apr 13;60(4):455-472. [doi: 10.1111/jcpp.12937] [Medline: 29900543]
47. Cuijpers P, Reijnders M, Huibers MJ. The role of common factors in psychotherapy outcomes. Annu Rev Clin Psychol 2019 May 07;15(1):207-231. [doi: 10.1146/annurev-clinpsy-050718-095424] [Medline: 30550721]
48. Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. Trials 2015 Nov 03;16(1):495 [FREE Full text] [doi: 10.1186/s13063-015-1023-4] [Medline: 26530985]
49. Zweben A, Barrett D, Berger L, Murray KT. Recruiting and retaining participants in a combined behavioral and pharmacological clinical trial. J Stud Alcohol Suppl 2005 Jul(15):72-81; discussion 65. [doi: 10.15288/jsas.2005.s15.72] [Medline: 16223059]
50. Geldsetzer P. Use of rapid online surveys to assess people's perceptions during infectious disease outbreaks: a cross-sectional survey on COVID-19. J Med Internet Res 2020 Apr 02;22(4):e18790 [FREE Full text] [doi: 10.2196/18790] [Medline: 32240094]
51. Schleider JL, Dobias M, Sung J, Mumper E, Mullarkey MC. Acceptability and utility of an open-access, online single-session intervention platform for adolescent mental health. JMIR Ment Health 2020 Jun 30;7(6):e20513 [FREE Full text] [doi: 10.2196/20513] [Medline: 32602846]
52. A scalable, single session intervention for perceived control over anxiety during COVID-19. Center for Open Science. 2021. URL: https://osf.io/haers/ [accessed 2022-03-20]
53. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD. How will country-based mitigation measures influence the course of the COVID-19 epidemic? The Lancet 2020 Mar;395(10228):931-934 [FREE Full text] [doi: 10.1016/s0140-6736(20)30567-5]
54. Webb TL, Sheeran P. Does changing behavioral intentions engender behavior change? A meta-analysis of the experimental evidence. Psychol Bull 2006 Mar;132(2):249-268. [doi: 10.1037/0033-2909.132.2.249] [Medline: 16536643]
55. Chandler J, Shapiro D. Conducting clinical research using crowdsourced convenience samples. Annu Rev Clin Psychol 2016 Mar 28;12(1):53-81. [doi: 10.1146/annurev-clinpsy-021815-093623] [Medline: 26772208]
56. Palan S, Schitter C. Prolific.ac—A subject pool for online experiments. Journal of Behavioral and Experimental Finance 2018 Mar;17:22-27. [doi: 10.1016/j.jbef.2017.12.004]
57. Schueller SM, Steakley-Freeman DM, Mohr DC, Yom-Tov E. Understanding perceived barriers to treatment from web browsing behavior. J Affect Disord 2020 Apr 15;267:63-66. [doi: 10.1016/j.jad.2020.01.131] [Medline: 32063574]
58. Perski O, Lumsden J, Garnett C, Blandford A, West R, Michie S. Assessing the psychometric properties of the digital behavior change intervention engagement scale in users of an app for reducing alcohol consumption: evaluation study. J Med Internet Res 2019 Nov 20;21(11):e16197 [FREE Full text] [doi: 10.2196/16197] [Medline: 31746771]
59. Honaker J, King G, Blackwell M. II: A program for missing data. J. Stat. Soft 2011;45(7):1. [doi: 10.18637/jss.v045.i07]
60. Anvari F, Lakens D. Using anchor-based methods to determine the smallest effect size of interest. Journal of Experimental Social Psychology 2021 Sep;96:104159. [doi: 10.1016/j.jesp.2021.104159]
61. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care 2008 Mar;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]
62. Clark DM, Layard R, Smithies R, Richards DA, Suckling R, Wright B. Improving access to psychological therapy: Initial evaluation of two UK demonstration sites. Behav Res Ther 2009 Nov;47(11):910-920 [FREE Full text] [doi: 10.1016/j.brat.2009.07.010] [Medline: 19647230]
63. Form: CDC COVID-19 Community Survey Question Bank (DRAFT). National Library of Medicine. URL: https://cde.nlm.nih.gov/formView?tinyId=Kcceysolt [accessed 2022-03-20]
64. Curran PG, Hauser KA. I'm paid biweekly, just not by leprechauns: Evaluating valid-but-incorrect response rates to attention check items. Journal of Research in Personality 2019 Oct;82:103849. [doi: 10.1016/j.jrp.2019.103849]
65. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York, NY: John Wiley & Sons; 1987. URL: https://onlinelibrary.wiley.com/doi/pdf/10.1002/9780470316696.fmatter [accessed 2022-03-20]
66. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. Prev Sci 2007 Sep;8(3):206-213. [doi: 10.1007/s11121-007-0070-9] [Medline: 17549635]
67. Wickham H. Easily Install and Load the "Tidyverse" R package. tidyverse. URL: https://tidyverse.tidyverse.org/ [accessed 2022-03-20]
68. Buchanan EM, Gillenwaters AM, Scofield JE, Valentine KD. MOTE: Effect Size and Confidence Interval Calculato. rdrr. URL: https://rdrr.io/cran/MOTE/ [accessed 2022-03-20]

69. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
70. Larsen ME, Huckvale K, Nicholas J, Torous J, Birrell L, Li E, et al. Using science to sell apps: Evaluation of mental health app store quality claims. NPJ Digit Med 2019 Mar 22;2(1):18 [FREE Full text] [doi: 10.1038/s41746-019-0093-1] [Medline: 31304366]
71. Wasil AR, Gillespie S, Shingleton R, Wilks CR, Weisz JR. Examining the reach of smartphone apps for depression and anxiety. Am J Psychiatry 2020 May 01;177(5):464-465. [doi: 10.1176/appi.ajp.2019.19090905] [Medline: 32354266]
72. Schleider JL, Weisz JR. Parent expectancies and preferences for mental health treatment: the roles of emotion mind-sets and views of failure. J Clin Child Adolesc Psychol 2018 Jan 24;47(sup1):S480-S496. [doi: 10.1080/15374416.2017.1405353] [Medline: 29364729]
73. Sung JY, Mumper E, Schleider JL. Empowering anxious parents to manage child avoidance behaviors: randomized control trial of a single-session intervention for parental accommodation. JMIR Ment Health 2021 Jul 06;8(7):e29538 [FREE Full text] [doi: 10.2196/29538] [Medline: 34255718]
74. Martinez-Martin N, Dasgupta I, Carter A, Chandler JA, Kellmeyer P, Kreitmair K, et al. Ethics of digital mental health during COVID-19: crisis and opportunities. JMIR Ment Health 2020 Dec 22;7(12):e23776 [FREE Full text] [doi: 10.2196/23776] [Medline: 33156811]
75. Luedtke A, Sadikova E, Kessler RC. Sample size requirements for multivariate models to predict between-patient differences in best treatments of major depressive disorder. Clinical Psychological Science 2019 Jan 18;7(3):445-461. [doi: 10.1177/2167702618815466]
76. Cuijpers P, Berking M, Andersson G, Quigley L, Kleiboer A, Dobson KS. A meta-analysis of cognitive-behavioural therapy for adult depression, alone and in comparison with other treatments. Can J Psychiatry 2013 Jul;58(7):376-385. [doi: 10.1177/070674371305800702] [Medline: 23870719]
77. Schumann G, Binder EB, Holte A, de Kloet ER, Oedegaard KJ, Robbins TW, et al. Stratified medicine for mental disorders. Eur Neuropsychopharmacol 2014 Jan;24(1):5-50 [FREE Full text] [doi: 10.1016/j.euroneuro.2013.09.010] [Medline: 24176673]

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

Edited by J Torous; submitted 08.09.21; peer-reviewed by S Schueller, B Nievas Soriano, B Kaveladze; comments to author 29.09.21; revised version received 08.02.22; accepted 22.02.22; published 12.04.22.

Please cite as:
Mullarkey M, Dobias M, Sung J, Ahuvia I, Shumake J, Beevers C, Schleider J
Web-Based Single Session Intervention for Perceived Control Over Anxiety During COVID-19: Randomized Controlled Trial
JMIR Ment Health 2022;9(4):e33473
URL: https://mental.jmir.org/2022/4/e33473
doi:10.2196/33473
PMID:35230962

Original Paper

Improving Web-Based Treatment Intake for Multiple Mental and Substance Use Disorders by Text Mining and Machine Learning: Algorithm Development and Validation

Sytske Wiegersma[1], MSc; Maurice Hidajat[2], BSc; Bart Schrieken[2], MSc; Bernard Veldkamp[1], Prof Dr; Miranda Olff[3,4], Prof Dr

[1]Department of Research Methodology, Measurement and Data Analysis, University of Twente, Enschede, Netherlands

[2]Interapy PLC, Amsterdam, Netherlands

[3]Department of Psychiatry, Amsterdam University Medical Centres, location Academic Medical Centre, Amsterdam, Netherlands

[4]ARQ National Psychotrauma Centre, Diemen, Netherlands

Corresponding Author:
Sytske Wiegersma, MSc
Department of Research Methodology, Measurement and Data Analysis
University of Twente
PO Box 217
Enschede, 7500 AE
Netherlands
Phone: 31 53 489 3349
Email: sytske@wiegersma.nl

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, underdetection 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.

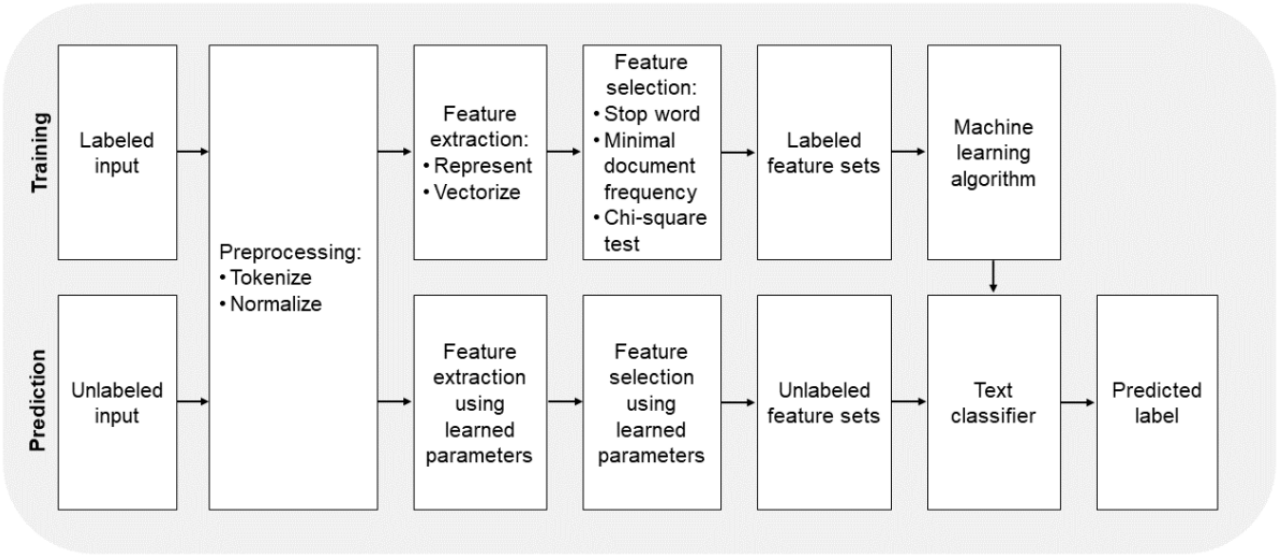
Automated Text-Screening Model

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 γ [42], which controls model flexibility [43], 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 of the 7 classes (the positive class) to the remaining 6 (together forming the negative class). As described by James et al [44], this results in 7 separate binary classification models, each with their own parameters $\beta_{0k},\beta_{1k},\ldots,\beta_{pk}$, with k denoting the k^{th} class and p the number of learned parameters. Each new, unlabeled input text x was provided with the class label for which the confidence score $\beta_{0k}+\beta_{1k}x_1+\beta_{2k}x_2+\cdots+\beta_{pk}x_p$ was the largest. This showed that there was a high level of confidence that the input text belonged to this class and not to one of the other 6 classes.

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*.

The number of true positives for class *A* (TP_A) were the number of times a document was labeled with *A* and the true label was indeed *A*. The false positives for class *A* (FP_A) were the instances that were incorrectly labeled by the classifier as *A*, whereas the true label was not *A*. This was calculated for class *A* by using the following formula:

$$E_{B,A}+E_{C,A}+E_{D,A}+E_{E,A}+E_{F,A}+E_{G,A}$$

The false negatives for class *A* (FN_A) were the instances with true label *A* for which the classifier predicted a different label. This was calculated for class *A* by using the following formula:

$$E_{A,B}+E_{A,C}+E_{A,D}+E_{A,E}+E_{A,F}+E_{A,G}$$

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

True label	Predicted label						
	$Class_A$	$Class_B$	$Class_C$	$Class_D$	$Class_E$	$Class_F$	$Class_G$
$Class_A$	*TP* $_A$ [a,b]	$E_{A,B}$	$E_{A,C}$	$E_{A,D}$	$E_{A,E}$	$E_{A,F}$	$E_{A,G}$
$Class_B$	$E_{B,A}$	*TP* $_B$ [b]	$E_{B,C}$	$E_{B,D}$	$E_{B,E}$	$E_{B,F}$	$E_{B,G}$
$Class_C$	$E_{C,A}$	$E_{C,B}$	*TP* $_C$ [b]	$E_{C,D}$	$E_{C,E}$	$E_{C,F}$	$E_{C,G}$
$Class_D$	$E_{D,A}$	$E_{D,B}$	$E_{D,C}$	*TP* $_D$ [b]	$E_{D,E}$	$E_{D,F}$	$E_{D,G}$
$Class_E$	$E_{E,A}$	$E_{E,B}$	$E_{E,C}$	$E_{E,D}$	*TP* $_E$ [b]	$E_{E,F}$	$E_{E,G}$
$Class_F$	$E_{F,A}$	$E_{F,B}$	$E_{F,C}$	$E_{F,D}$	$E_{F,E}$	*TP* $_F$ [b]	$E_{F,G}$
$Class_G$	$E_{G,A}$	$E_{G,B}$	$E_{G,C}$	$E_{G,D}$	$E_{G,E}$	$E_{G,F}$	*TP* $_G$ [b]

[a]TP: true positive.

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

Performance Metrics

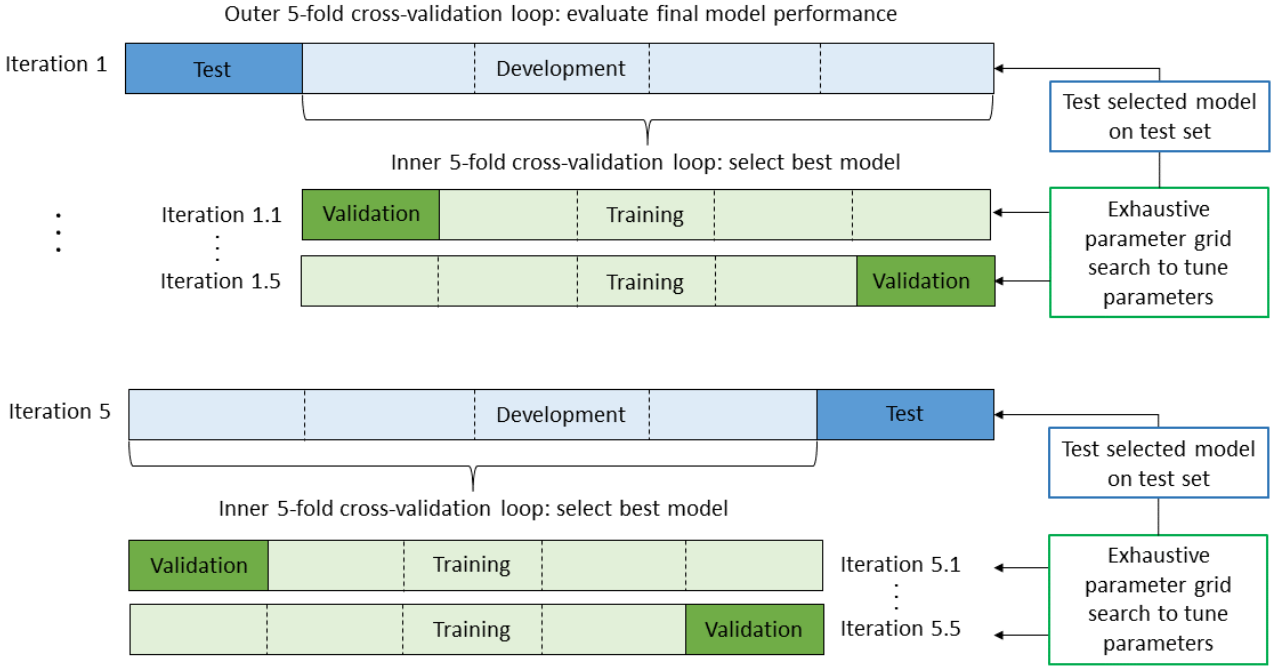
The correct predictions (TPs and TNs) and errors (FPs and FNs) were then used to calculate performance metrics for each class. Bird et al [31] define several metrics, the simplest of which is accuracy, a measure for the proportion of correctly labeled input texts in the test set. The recall, also called 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

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 × (*Precision*×*Recall*)/(*Precision*+*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).

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

[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 3. Best parameters selected by exhaustive grid search.

Parameter	Best value
Remove stop words	Yes
Minimal *x*[a] documents	1
Representation scheme	Unigrams
Term weight	Term frequency
Select *k*[b] best features	470
Regularization parameter *C*	1

[a]*x*: number of documents a feature should be present in.

[b]*k*: 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.

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

English keyword (Dutch stem)	Chi-square (*df*)	*P* value	Addiction[a]	Anxiety[a]	Eating[a]	Mood[a]	PTSD[a,b]	Panic[a]	Somatic[a]
eating (eten)	20.2 (1)	.003	0	0	*5*	0	0	0	0
heavy-headed (lustelos)	20.0 (1)	.003	9	40	6	*105*	27	23	74
control (control)	19.6 (1)	.003	9	53	49	45	36	*102*	38
ate (geget)	19.3 (1)	.004	0	0	*7*	1	0	0	0
underweight (ondergewicht)	18.9 (1)	.004	0	0	*6*	1	0	0	0
nutrition (voeding)	18.9 (1)	.004	0	2	*9*	1	1	0	0
gloomy (somber)	18.6 (1)	.005	3	32	6	*112*	32	40	38
normal (normal)	18.4 (1)	.005	8	58	55	44	83	*105*	63
addictive (verslav)	17.9 (1)	.007	*10*	4	4	3	2	1	4

[a]Occurrence frequencies for each feature in each class (disorder).

[b]PTSD: posttraumatic stress disorder.

[c]The frequency for the class in which it occurs the most is presented in italics.

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).

Disorder	Patients in test set, n (%)	Precision	Recall	F_1 score	Overall accuracy[a]
Addiction	40 (3.41)	0.25	0.33	0.28	—[b]
Anxiety	220 (18.76)	0.44	0.35	0.39	—
Eating	50 (4.26)	0.75	0.82	0.78	—
Mood	220 (18.76)	0.44	0.50	0.47	—
PTSD[c]	203 (17.31)	0.57	0.52	0.54	—
Panic	220 (18.76)	0.57	0.55	0.56	—
Somatic	220 (18.76)	0.46	0.50	0.48	—
Weighted average	N/A[d]	0.50	0.49	0.49	0.49

[a]Accuracy is the overall accuracy of the classifier averaged over all classes.

[b]Data not available for separate classes.

[c]PTSD: posttraumatic stress disorder.

[d]N/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.

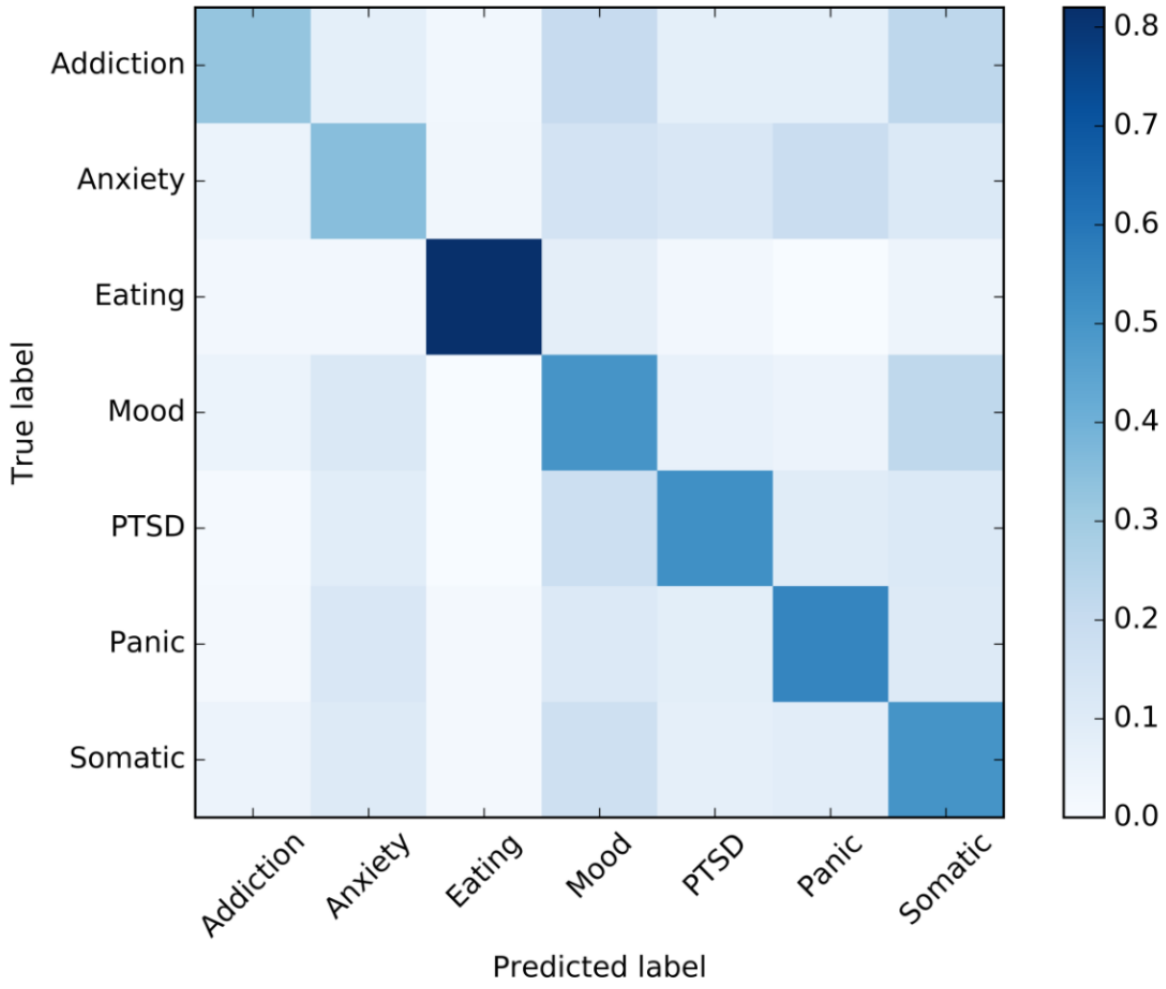
True disorder	Predicted disorder						
	Addiction	Anxiety	Eating	Mood	PTSD[a]	Panic	Somatic
Addiction (N=40), n (%)	*13 (32.5)* [b]	3 (7.5)	1 (2.5)	8 (20)	3 (7.5)	3 (7.5)	9 (22.5)
Anxiety (N=220), n (%)	11 (5)	*77 (35)*	6 (2.7)	33 (15)	27 (12.3)	41 (18.6)	25 (11.4)
Eating (N=50), n (%)	1 (2)	1 (2)	*41 (82)*	4 (8)	1 (2)	0 (0)	2 (4)
Mood (N=220), n (%)	11 (5)	26 (11.8)	0 (0)	*110 (50)*	14 (6.4)	10 (4.5)	49 (22.3)
PTSD (N=203), n (%)	2 (1)	18 (8.9)	0 (0)	36 (17.7)	*105 (51.7)*	19 (9.4)	23 (11.3)
Panic (N=220), n (%)	4 (1.8)	27 (12.3)	3 (1.4)	24 (10.9)	18 (8.2)	*121 (55)*	23 (10.4)
Somatic (N=220), n (%)	10 (4.5)	23 (10.5)	4 (1.8)	37 (16.8)	16 (7.3)	19 (8.6)	*111 (50.5)*

[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 (eg, 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

Disease Study 2015. Lancet 2016 Oct 08;388(10053):1545-1602 [FREE Full text] [doi: 10.1016/S0140-6736(16)31678-6] [Medline: 27733282]
2. Global Burden of Disease Study 2016 (GBD 2016) data resources. Global Health Data Exchange. URL: http://ghdx.healthdata.org/gbd-2016 [accessed 2022-01-24]
3. The European Mental Health Action Plan 2013–2020. Geneva: World Health Organization; 2015.
4. Collins PY, Patel V, Joestl SS, March D, Insel TR, Daar AS, et al. Grand challenges in global mental health. Nature 2011 Jul 06;475(7354):27-30 [FREE Full text] [doi: 10.1038/475027a] [Medline: 21734685]
5. Linden M, Schermuly-Haupt M. Definition, assessment and rate of psychotherapy side effects. World Psychiatry 2014 Oct;13(3):306-309 [FREE Full text] [doi: 10.1002/wps.20153] [Medline: 25273304]
6. Collins K, Westra H, Dozois D, Burns D. Gaps in accessing treatment for anxiety and depression: challenges for the delivery of care. Clin Psychol Rev 2004 Sep;24(5):583-616. [doi: 10.1016/j.cpr.2004.06.001] [Medline: 15325746]
7. Kohn R, Saxena S, Levav I, Saraceno B. The treatment gap in mental health care. Bull World Health Organ 2004 Nov;82(11):858-866 [FREE Full text] [Medline: 15640922]
8. Magruder KM, McLaughlin KA, Elmore Borbon DL. Trauma is a public health issue. Eur J Psychotraumatol 2017;8(1):1375338 [FREE Full text] [doi: 10.1080/20008198.2017.1375338] [Medline: 29435198]
9. Bourla A, Mouchabac S, El Hage W, Ferreri F. e-PTSD: an overview on how new technologies can improve prediction and assessment of Posttraumatic Stress Disorder (PTSD). Eur J Psychotraumatol 2018;9(sup1):1424448 [FREE Full text] [doi: 10.1080/20008198.2018.1424448] [Medline: 29441154]
10. Olff M. Mobile mental health: a challenging research agenda. Eur J Psychotraumatol 2015;6:27882 [FREE Full text] [doi: 10.3402/ejpt.v6.27882] [Medline: 25994025]
11. Frewen P, Schmahl C, Olff M. Interdisciplinary approaches to understand traumatic stress as a public health problem. Eur J Psychotraumatol 2017;8(Suppl 5):1441582 [FREE Full text] [doi: 10.1080/20008198.2018.1441582] [Medline: 33178404]
12. Berger T, Caspar F, Richardson R, Kneubühler B, Sutter D, Andersson G. Internet-based treatment of social phobia: a randomized controlled trial comparing unguided with two types of guided self-help. Behav Res Ther 2011 Mar;49(3):158-169. [doi: 10.1016/j.brat.2010.12.007] [Medline: 21255767]
13. Berger T, Hämmerli K, Gubser N, Andersson G, Caspar F. Internet-based treatment of depression: a randomized controlled trial comparing guided with unguided self-help. Cogn Behav Ther 2011;40(4):251-266. [doi: 10.1080/16506073.2011.616531] [Medline: 22060248]
14. Ruwaard J, Lange A, Schrieken B, Dolan CV, Emmelkamp P. The effectiveness of online cognitive behavioral treatment in routine clinical practice. PLoS One 2012;7(7):e40089 [FREE Full text] [doi: 10.1371/journal.pone.0040089] [Medline: 22792217]
15. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]
16. Cuijpers P, Donker T, van Straten A, Li J, Andersson G. Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. Psychol Med 2010 Dec;40(12):1943-1957. [doi: 10.1017/S0033291710000772] [Medline: 20406528]
17. Emmelkamp PM. Technological innovations in clinical assessment and psychotherapy. Psychother Psychosom 2005;74(6):336-343. [doi: 10.1159/000087780] [Medline: 16244509]
18. Kaltenthaler E, Brazier J, De Nigris E, Tumur I, Ferriter M, Beverley C, et al. Computerised cognitive behaviour therapy for depression and anxiety update: a systematic review and economic evaluation. Health Technol Assess 2006 Sep;10(33):iii, xi-ixiv, 1 [FREE Full text] [doi: 10.3310/hta10330] [Medline: 16959169]
19. Feldman R, Sanger J. The Text Mining Handbook Advanced Approaches in Analyzing Unstructured Data. Cambridge, UK: Cambridge University Press; 2006.
20. Abbe A, Grouin C, Zweigenbaum P, Falissard B. Text mining applications in psychiatry: a systematic literature review. Int J Methods Psychiatr Res 2016 Jun;25(2):86-100 [FREE Full text] [doi: 10.1002/mpr.1481] [Medline: 26184780]
21. Wongkoblap A, Vadillo MA, Curcin V. Researching mental health disorders in the era of social media: systematic review. J Med Internet Res 2017 Jun 29;19(6):e228 [FREE Full text] [doi: 10.2196/jmir.7215] [Medline: 28663166]
22. He Q, Veldkamp BP, Glas CA, de Vries T. Automated assessment of patients' self-narratives for posttraumatic stress disorder screening using natural language processing and text mining. Assessment 2017 Mar;24(2):157-172. [doi: 10.1177/1073191115602551] [Medline: 26358713]
23. Neuman Y, Cohen Y, Assaf D, Kedma G. Proactive screening for depression through metaphorical and automatic text analysis. Artif Intell Med 2012 Sep;56(1):19-25. [doi: 10.1016/j.artmed.2012.06.001] [Medline: 22771201]
24. Digitale Indicatiehulp Psychische Problemen: inhoudelijke dossier. Digital Indication Aid for Mental Health Problems: content file. Unpublished report. Diemen, The Netherlands: Interapy; 2015.
25. van Bebber J, Wigman J, Wunderink L, Tendeiro J, Wichers M, Broeksteeg J, et al. Identifying levels of general distress in first line mental health services: can GP- and eHealth clients' scores be meaningfully compared? BMC Psychiatry 2017 Dec 01;17(1):382 [FREE Full text] [doi: 10.1186/s12888-017-1552-3] [Medline: 29191173]

26. Terluin B. De vierdimensionale klachtenlijst (4DKL): een vragenlijst voor het meten van distress, depressie, angst en somatisatie. Huisarts en Wetenschap 1996 Nov;39(12):538-547 [FREE Full text]
27. Terluin B, Rhenen WV, Schaufeli WB, De Haan M. The four-dimensional symptom questionnaire (4DSQ): measuring distress and other mental health problems in a working population. Work Stress 2004 Jul;18(3):187-207. [doi: 10.1080/0267837042000297535]
28. Terluin B, van Marwijk HW, Adèr HJ, de Vet HC, Penninx BW, Hermens ML, et al. The Four-Dimensional Symptom Questionnaire (4DSQ): a validation study of a multidimensional self-report questionnaire to assess distress, depression, anxiety and somatization. BMC Psychiatry 2006 Aug 22;6:34 [FREE Full text] [doi: 10.1186/1471-244X-6-34] [Medline: 16925825]
29. Perkins J. Python 3 Text Processing with NLTK 3 Cookbook. Birmingham, United Kingdom: Packt Publishing Ltd; 2014.
30. Snowball homepage. Snowball. URL: http://snowballstem.org/ [accessed 2022-01-25]
31. Bird S, Klein E, Loper E. Natural Language Processing With Python. Sebastopol: O'reilly Media, Inc; 2009.
32. Shen D, Sun J, Yang Q, Chen Z. Text classification improved through multigram models. In: Proceedings of the 15th ACM international conference on Information and knowledge management. 2006 Presented at: Proceedings of the 15th ACM international conference on Information and knowledge management; Nov 6 - 11, 2006; Arlington Virginia USA. [doi: 10.1145/1183614.1183710]
33. Forman G. An extensive empirical study of feature selection metrics for text classification. J Mach Learn Res 2003 Jan 3;3:1289-1305 [FREE Full text] [doi: 10.5555/944919.944974]
34. Jurafsky D, Martin JH. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Upper Saddle River, N.J: Prentice Hall; 2009.
35. Campbell RS, Pennebaker JW. The secret life of pronouns: flexibility in writing style and physical health. Psychol Sci 2003 Jan;14(1):60-65. [doi: 10.1111/1467-9280.01419] [Medline: 12564755]
36. Joachims T. Text categorization with support vector machines: learning with many relevant features. In: Machine Learning: ECML-98. Berlin, Heidelberg: Springer; 1998.
37. Vapnik V. The Nature of Statistical Learning. New York, NY: Springer; 1995.
38. Galar M, Fernández A, Barrenechea E, Bustince H, Herrera F. An overview of ensemble methods for binary classifiers in multi-class problems: experimental study on one-vs-one and one-vs-all schemes. Pattern Recognit 2011 Aug;44(8):1761-1776. [doi: 10.1016/j.patcog.2011.01.017]
39. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning. New York, NY: Springer Science & Business Media; 2009.
40. King G, Zeng L. Logistic regression in rare events data. Polit Anal 2017 Jan 04;9(2):137-163. [doi: 10.1093/oxfordjournals.pan.a004868]
41. Fan R, Chang K, Hsieh C, Wang X, Lin C. LIBLINEAR: a library for large linear classification. J Mach Learn Res 2008 Jan 6;9:1871-1874 [FREE Full text] [doi: 10.5555/1390681.1442794]
42. Duan K, Keerthi S, Poo AN. Evaluation of simple performance measures for tuning SVM hyperparameters. Neurocomputing 2003 Apr;51:41-59. [doi: 10.1016/s0925-2312(02)00601-x]
43. Ben-Hur A, Weston J. A user's guide to support vector machines. In: Data Mining Techniques for the Life Sciences. Totowa, New Jersey, United States: Humana Press; 2010.
44. James G, Witten D, Hastie T, Tibshirani R. An Introduction to Statistical Learning With Applications in R. New York: Springer; 2013.
45. Yang Y. An evaluation of statistical approaches to text categorization. Inf Retrieval 1999;1:69-90. [doi: 10.1023/A:1009982220290]
46. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2. 1995 Presented at: Proceedings of the 14th international joint conference on Artificial intelligence - Volume 2; Aug 20 - 25, 1995; Montreal Quebec Canada URL: http://dl.acm.org
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]
48. van Loo HM, Romeijn J. Psychiatric comorbidity: fact or artifact? Theor Med Bioeth 2015;36(1):41-60 [FREE Full text] [doi: 10.1007/s11017-015-9321-0] [Medline: 25636962]
49. Kessler RC, Sonnega A, Bromet E, Hughes M, Nelson CB. Posttraumatic stress disorder in the National Comorbidity Survey. Arch Gen Psychiatry 1995 Dec;52(12):1048-1060. [doi: 10.1001/archpsyc.1995.03950240066012] [Medline: 7492257]
50. Smith M. Psychiatric disorders. In: Essential Human Disease for Dentists. London, United Kingdom: Churchill Livingstone; 2006.
51. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). Washington, DC: American Psychiatric Publishing; 2013.
52. World Health Organization. Task Shifting: Rational Redistribution of Tasks Among Health Workforce Teams : Global Recommendations and Guidelines. Geneva: World Health Organization; 2008.

53. Regier DA, Farmer ME, Rae DS, Locke BZ, Keith SJ, Judd LL, et al. Comorbidity of mental disorders with alcohol and other drug abuse. Results from the Epidemiologic Catchment Area (ECA) study. JAMA 1990 Nov 21;264(19):2511. [doi: 10.1001/jama.1990.03450190043026]
54. Torrens M, Gilchrist G, Domingo-Salvany A. Psychiatric comorbidity in illicit drug users: substance-induced versus independent disorders. Drug Alcohol Depend 2011 Jan 15;113(2-3):147-156. [doi: 10.1016/j.drugalcdep.2010.07.013] [Medline: 20801586]
55. Conway K, Compton W, Stinson F, Grant B. Lifetime comorbidity of DSM-IV mood and anxiety disorders and specific drug use disorders: results from the National Epidemiologic Survey on Alcohol and Related Conditions. J Clin Psychiatry 2006 Feb;67(2):247-257. [doi: 10.4088/jcp.v67n0211] [Medline: 16566620]
56. Bennasar M, Setchi R, Hicks Y, Bayer A. Cascade classification for diagnosing dementia. In: Proceedings of the 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC). 2014 Presented at: 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC); Oct 5-8, 2014; San Diego, CA, USA. [doi: 10.1109/smc.2014.6974308]
57. Hashem A, Rasmy ME, Wahba KM, Shaker OG. Single stage and multistage classification models for the prediction of liver fibrosis degree in patients with chronic hepatitis C infection. Comput Methods Programs Biomed 2012 Mar;105(3):194-209. [doi: 10.1016/j.cmpb.2011.10.005] [Medline: 22070853]
58. Wallace B, Trikalinos T, Laws M, Wilson I, Charniak E. A generative joint, additive, sequential model of topics and speech acts in patient-doctor communication. In: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. 2013 Presented at: Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing; Oct 18-21, 2013; Seattle, Washington, USA URL: https://aclweb.org/
59. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Use of sentiment analysis for capturing patient experience from free-text comments posted online. J Med Internet Res 2013 Nov 01;15(11):e239 [FREE Full text] [doi: 10.2196/jmir.2721] [Medline: 24184993]
60. Keski-Rahkonen A, Tozzi F. The process of recovery in eating disorder sufferers' own words: an internet-based study. Int J Eat Disord 2005;37 Suppl:S80-6; discussion S87. [doi: 10.1002/eat.20123] [Medline: 15852327]

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

Edited by R Kukafka, G Eysenbach; submitted 05.06.20; peer-reviewed by L Ospina-Pinillos, J Wang; comments to author 24.07.20; revised version received 01.11.20; accepted 28.09.21; published 11.04.22.

Please cite as:
Wiegersma S, Hidajat M, Schrieken B, Veldkamp B, Olff M
Improving Web-Based Treatment Intake for Multiple Mental and Substance Use Disorders by Text Mining and Machine Learning: Algorithm Development and Validation
JMIR Ment Health 2022;9(4):e21111
URL: https://mental.jmir.org/2022/4/e21111
doi:10.2196/21111
PMID:35404261

Original Paper

A Serious Game for Young People With First Episode Psychosis (OnTrack>The Game): Qualitative Findings of a Randomized Controlled Trial

Samantha Jankowski[1*], MA; Kathleen Ferreira[2*], PhD; Franco Mascayano[1*], PhD; Effy Donovan[2*], BA; Reanne Rahim[1], MA; Michael L Birnbaum[3], MD; Sabrina Yum-Chan[3], BS; Deborah Medoff[4], PhD; Bethany Marcogliese[2], PhD; Lijuan Fang[4], PhD; Terriann Nicholson[5], MD; Lisa Dixon[1,5], MPH, MD

[1]Department of Psychiatry, New York State Psychiatric Institute, New York, NY, United States

[2]Department of Research and Evaluation, C4 Innovations, Needham, MA, United States

[3]Department of Psychiatry, The Zucker Hillside Hospital, Glen Oaks, NY, United States

[4]Department of Psychiatry, University of Maryland School of Medicine, Baltimore, MD, United States

[5]Department 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.

Engagement

- 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?

Recovery

- 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?

Empowerment

- Did playing the game or using the videos site help you think about ways to speak up for yourself?
- Did you think more about how you can participate in and make decisions about your treatment?

Hope

- 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?

Understanding of psychosis

- How was the content relevant to your own life and experiences with your symptoms and treatment?
- Did you learn anything about your symptoms or treatment?

Self-stigma

- After playing the game or using the videos site, did you feel differently about yourself?

Public stigma

- How do you think the game or videos site portrayed individuals living with mental illness?

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.

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.

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

Characteristics	Treatment group		Comparison		
	OnTrack>The Game group (n=11)	Recovery videos group (n=5)	Chi-square test (*df*)	*t* test[a] (*df*)	*P* value
Sex, n (%)			0.04 (1)	N/A[b]	.84
Men	6 (55)	3 (60)			
Women	5 (45)	2 (40)			
Age (years), mean (SD)	21.55 (4.01)	22.40 (1.95)	N/A	0.57 (13.79)	.58
Race, n (%)			1.9 (3)	N/A	.60
White	3 (27)	2 (40)			
Black or African American	5 (45)	1 (20)			
Asian	1 (9)	0 (0)			
Unknown	2 (18)	2 (40)			
Ethnicity, n (%)			0.04 (1)	N/A	.84
Yes, Hispanic or Latino	5 (45)	2 (40)			
Not Hispanic or Latino	6 (55)	3 (60)			

[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.

Textbox 2. Key themes and subthemes identified.

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 [my]self 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 RVs, 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 access to computers), 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 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

1. GBD 2017 Disease Injury Incidence Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis

for the Global Burden of Disease Study 2017. Lancet 2018 Nov 10;392(10159):1789-1858 [FREE Full text] [doi: 10.1016/S0140-6736(18)32279-7] [Medline: 30496104]
2. Kane JM, Robinson DG, Schooler NR, Mueser KT, Penn DL, Rosenheck RA, et al. Comprehensive versus usual community care for first-episode psychosis: 2-year outcomes from the NIMH raise early treatment program. Am J Psychiatry 2016 Apr 01;173(4):362-372 [FREE Full text] [doi: 10.1176/appi.ajp.2015.15050632] [Medline: 26481174]
3. Nordentoft M, Jeppesen P, Abel M, Kassow P, Petersen L, Thorup A, et al. OPUS study: suicidal behaviour, suicidal ideation and hopelessness among patients with first-episode psychosis. One-year follow-up of a randomised controlled trial. Br J Psychiatry Suppl 2002 Sep 02;43(S43):s98-106. [doi: 10.1192/bjp.181.43.s98] [Medline: 12271808]
4. Srihari VH, Tek C, Kucukgoncu S, Phutane VH, Breitborde NJ, Pollard J, et al. First-episode services for psychotic disorders in the U.S. Public sector: a pragmatic randomized controlled trial. Psychiatr Serv 2015 Jul;66(7):705-712 [FREE Full text] [doi: 10.1176/appi.ps.201400236] [Medline: 25639994]
5. Lal S, Malla A. Service engagement in first-episode psychosis: current issues and future directions. Can J Psychiatry 2015 Aug;60(8):341-345 [FREE Full text] [doi: 10.1177/070674371506000802] [Medline: 26454555]
6. Casey D, Brown L, Gajwani R, Islam Z, Jasani R, Parsons H, et al. Predictors of engagement in first-episode psychosis. Schizophr Res 2016 Aug;175(1-3):204-208 [FREE Full text] [doi: 10.1016/j.schres.2016.04.030] [Medline: 27132495]
7. Cabassa LJ, Piscitelli S, Haselden M, Lee RJ, Essock SM, Dixon LB. Understanding pathways to care of individuals entering a specialized early intervention service for first-episode psychosis. Psychiatr Serv 2018 Jun 01;69(6):648-656 [FREE Full text] [doi: 10.1176/appi.ps.201700018] [Medline: 29493414]
8. Fitzgerald M, Ratcliffe G. Serious games, gamification, and serious mental illness: a scoping review. Psychiatr Serv 2020 Feb 01;71(2):170-183. [doi: 10.1176/appi.ps.201800567] [Medline: 31640521]
9. Shandley K, Austin D, Klein B, Kyrios M. An evaluation of 'Reach Out Central': an online gaming program for supporting the mental health of young people. Health Educ Res 2010 Aug;25(4):563-574. [doi: 10.1093/her/cyq002] [Medline: 20150530]
10. Ferchaud A, Seibert J, Sellers N, Escobar Salazar N. Reducing mental health stigma through identification with video game avatars with mental illness. Front Psychol 2020 Sep 9;11:2240 [FREE Full text] [doi: 10.3389/fpsyg.2020.02240] [Medline: 33013585]
11. Fleming TM, Bavin L, Stasiak K, Hermansson-Webb E, Merry SN, Cheek C, et al. Serious games and gamification for mental health: current status and promising directions. Front Psychiatry 2016;7:215 [FREE Full text] [doi: 10.3389/fpsyt.2016.00215] [Medline: 28119636]
12. Olivet J, Haselden M, Piscitelli S, Kenney R, Shulman A, Medoff D, et al. Results from a pilot study of a computer-based role-playing game for young people with psychosis. Early Interv Psychiatry 2019 Aug 15;13(4):767-772. [doi: 10.1111/eip.12556] [Medline: 29542863]
13. Amsalem D, Yang L, Jankowski S, Lieff S, Markowitz J, Dixon L. Reducing stigma toward individuals with schizophrenia using a brief video: a randomized controlled trial of young adults. Schizophr Bull 2021 Jan 23;47(1):7-14 [FREE Full text] [doi: 10.1093/schbul/sbaa114] [Medline: 33484269]
14. Amsalem D, Markowitz JC, Jankowski SE, Yang LH, Valeri L, Lieff SA, et al. Sustained effect of a brief video in reducing public stigma toward individuals with psychosis: a randomized controlled trial of young adults. Am J Psychiatry 2021 Jul 26;178(7):635-642. [doi: 10.1176/appi.ajp.2020.20091293] [Medline: 33900809]
15. Thornicroft G, Mehta N, Clement S, Evans-Lacko S, Doherty M, Rose D, et al. Evidence for effective interventions to reduce mental-health-related stigma and discrimination. Lancet 2016 Mar;387(10023):1123-1132. [doi: 10.1016/s0140-6736(15)00298-6]
16. Pescosolido BA, Martin JK. The stigma complex. Annu Rev Sociol 2015 Aug 14;41(1):87-116 [FREE Full text] [doi: 10.1146/annurev-soc-071312-145702] [Medline: 26855471]
17. Neal JW, Neal ZP, VanDyke E, Kornbluh M. Expediting the analysis of qualitative data in evaluation. Am J Eval 2014 May 27;36(1):118-132. [doi: 10.1177/1098214014536601]
18. Norman RM, Malla AK, Manchanda R, Harricharan R, Takhar J, Northcott S. Social support and three-year symptom and admission outcomes for first episode psychosis. Schizophr Res 2005 Dec 15;80(2-3):227-234. [doi: 10.1016/j.schres.2005.05.006] [Medline: 15964175]
19. Sündermann O, Onwumere J, Kane F, Morgan C, Kuipers E. Social networks and support in first-episode psychosis: exploring the role of loneliness and anxiety. Soc Psychiatry Psychiatr Epidemiol 2014 Mar 17;49(3):359-366 [FREE Full text] [doi: 10.1007/s00127-013-0754-3] [Medline: 23955376]
20. van der Lubbe L, Gerritsen C, Klein M, Hindriks K. Empowering vulnerable target groups with serious games and gamification. Entertainment Comput 2021 May;38:100402. [doi: 10.1016/j.entcom.2020.100402]
21. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
22. Wisniewski H, Gorrindo T, Rauseo-Ricupero N, Hilty D, Torous J. The role of digital navigators in promoting clinical care and technology integration into practice. Digit Biomark 2020 Nov 26;4(Suppl 1):119-135 [FREE Full text] [doi: 10.1159/000510144] [Medline: 33442585]

Abbreviations

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

Edited by J Torous; submitted 13.09.21; peer-reviewed by J D'Arcey, P Whelan; comments to author 30.10.21; revised version received 29.12.21; accepted 10.01.22; published 06.04.22.

Please cite as:
Jankowski S, Ferreira K, Mascayano F, Donovan E, Rahim R, Birnbaum ML, Yum-Chan S, Medoff D, Marcogliese B, Fang L, Nicholson T, Dixon L
A Serious Game for Young People With First Episode Psychosis (OnTrack>The Game): Qualitative Findings of a Randomized Controlled Trial
JMIR Ment Health 2022;9(4):e33526
URL: https://mental.jmir.org/2022/4/e33526
doi:10.2196/33526
PMID:35384847

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 Kock[1,2], MA, PhD, CPsychol; Helen Ann Latham[3], BSc, MBChB; Richard G Cowden[4], PhD; Breda Cullen[5], DClinPsy, PhD; Katia Narzisi[1], PhD; Shaun Jerdan[1], BSc, MRes, PhD; Sarah-Anne Munoz[1], MA, MSc, PhD; Stephen J Leslie[6], BSc, MBChB, PhD; Andreas Stamatis[7], PhD; Jude Eze[8], BSc, MSc, PhD

[1]Division of Rural Health and Wellbeing, Institute of Health Research and Innovation, School of Health, University of the Highlands and Islands, Inverness, United Kingdom

[2]Department of Clinical Psychology, New Craigs Psychiatric Hospital, Inverness, United Kingdom

[3]Nairn Town and Country Hospital, NHS Highland, United Kingdom

[4]Human Flourishing Program, Harvard University, Cambridge, MA, United States

[5]Institute of Health and Wellbeing, University of Glasgow, Glasgow, United Kingdom

[6]Cardiac Unit, Raigmore Hospital, NHS Highland, Inverness, United Kingdom

[7]Exercise and Nutrition Sciences, State University of New York, Plattsburgh, NY, United States

[8]Epidemiology 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

d=–0.15, 95% CI –0.41 to 0.10). 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 to October 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 codesigned by the University of the Highlands and Islands (UHI), the NHSH, and the software and technical team that supports the MPS app. 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 adduced 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 lsmean 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 (χ^2_2=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).

Figure 1. Details of enrolment into the trial.

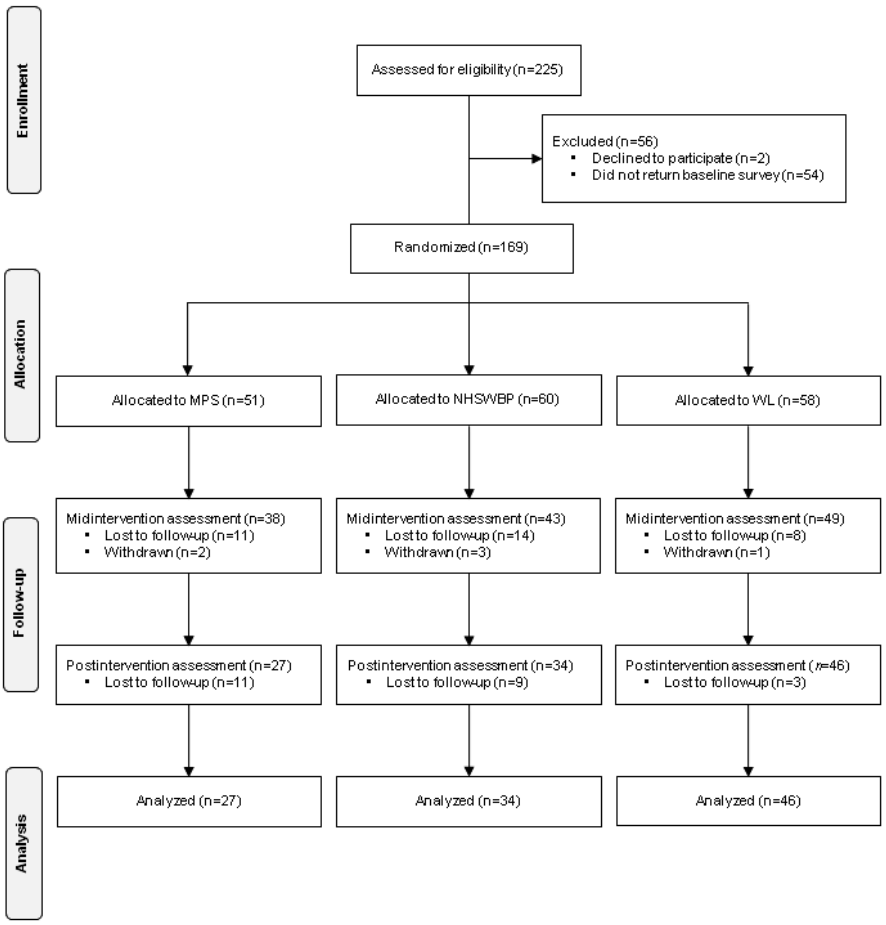


Table 1. Distribution of participant characteristics at baseline and postintervention.

Characteristic	Baseline (N=169), n (%)	Postintervention (N=107), n (%)	Attrition rate
Gender			
Female	149 (88.2)	94 (87.9)	0.37
Male	20 (11.8)	13 (12.1)	0.35
Age (years)			
18-25	4 (2.4)	4 (3.7)	N/A[a]
26-30	10 (5.9)	4 (3.7)	0.60
31-40	31 (18.3)	16 (15.0)	0.48
>40	124 (73.4)	83 (77.6)	0.33
Education level			
Undergraduate or lower	65 (38.5)	40 (37.4)	0.38
Postgraduate or higher	104 (61.5)	67 (62.6)	0.36
Type of employment			
Nurse	48 (28.4)	30 (28.0)	0.38
Doctor	39 (23.1)	21 (19.6)	0.46
Allied health professional	21 (12.4)	12 (11.2)	0.43
Administrative	16 (9.5)	13 (12.1)	0.19
Carer	6 (3.6)	4 (3.7)	0.33
Health care assistant	8 (4.7)	4 (3.7)	0.50
Other	31 (18.3)	23 (21.5)	0.26
Years of employment experience			
<2	14 (8.3)	11 (10.5)	0.21
2-5	13 (7.8)	8 (7.6)	0.38
5-10	21 (12.6)	10 (9.5)	0.52
>10	119 (71.3)	76 (72.4)	0.36
Workplace			
Community, GP[b], and PC[c]	73 (43.7)	41 (39.0)	0.44
Hospital	74 (44.3)	51 (48.6)	0.31
Other	20 (12.0)	14 (13.3)	0.30
N/A	N/A	1 (1.0)	N/A
Hours worked/week			
<20	8 (4.7)	3 (2.8)	0.62
20-30	31 (18.3)	18 (16.8)	0.42
30-40	100 (59.2)	61 (57.0)	0.39
>40	30 (17.8)	25 (23.4)	0.17
Work with COVID-19 patients			
No	129 (77.2)	84 (80.0)	0.35
Yes	38 (22.8)	21 (20.0)	0.45
Level of disruption			
No disruption	3 (1.8)	1 (0.9)	0.67
Minor	15 (8.9)	10 (9.3)	0.33
Moderate	65 (38.0)	41 (38.3)	0.37

Characteristic	Baseline (N=169), n (%)	Postintervention (N=107), n (%)	Attrition rate
Major	66 (39.0)	40 (37.4)	0.39
Severe	20 (12.0)	15 (14.0)	0.25
Shielding			
No	145 (85.8)	93 (86.9)	0.36
Yes	7 (4.1)	5 (4.7)	0.29
Family member is shielding	17 (10.1)	9 (8.4)	0.47
Psychiatric disorder			
No	131 (77.5)	78 (72.9)	0.40
Yes	38 (22.5)	29 (27.1)	0.23

[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.

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

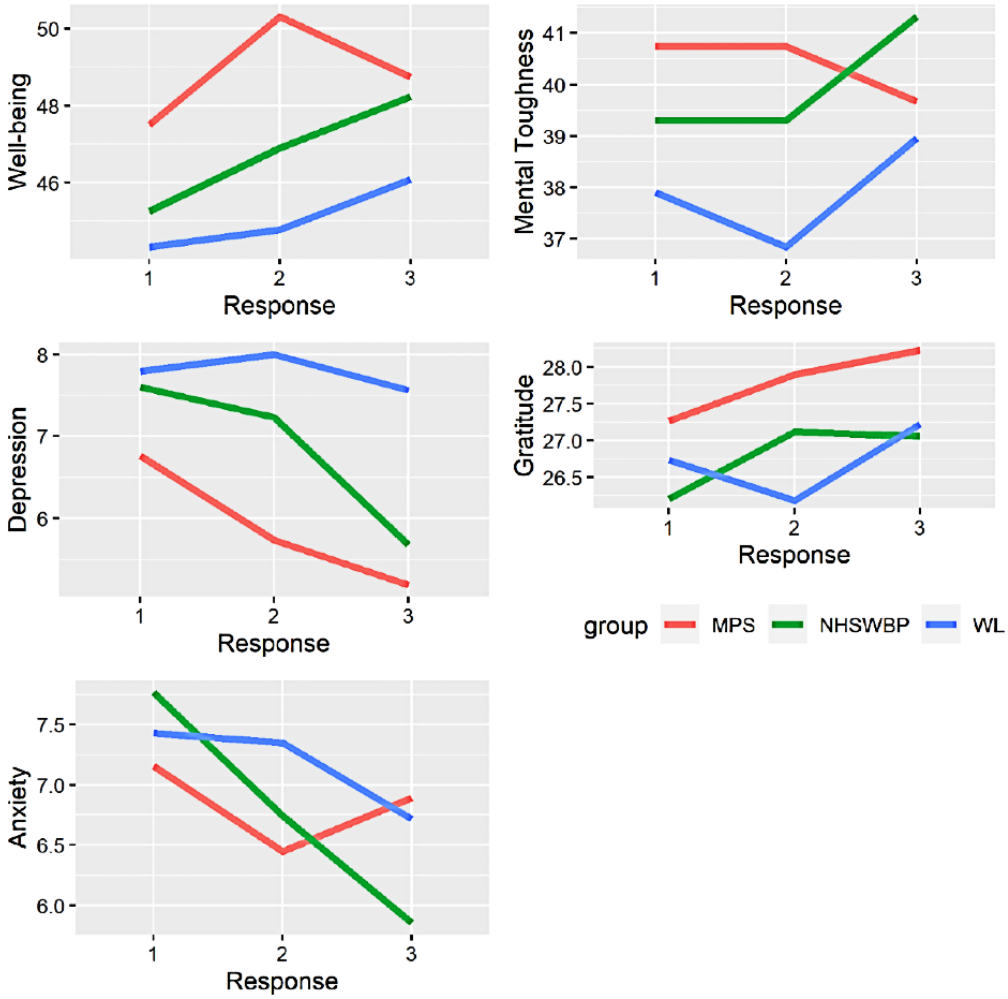
[a]MPS: My Possible Self.

[b]NHSWBP: National Health Service Highland Staff Wellbeing Project.

[c]WL: waitlist.

[d]MT: mental toughness.

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.



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.

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 0.07-0.66; 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*=0.13-0.27).

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

Between-group effects	Anxiety, Cohen *d* (95% CI)	Depression, Cohen *d* (95% CI)	Mental well-being, Cohen *d* (95% CI)	MT[a], Cohen *d* (95% CI)	Gratitude, Cohen *d* (95% CI)
NHSWBP[b] vs MPS[c]	0.01 (–0.26 to 0.28)	0.19 (–0.12 to 0.50)	–0.20 (–0.48 to 0.08)	–0.07 (–0.41 to 0.27)	–0.26 (–0.57 to 0.06)
WL[d] vs MPS	0.07 (–0.20 to 0.33)	0.37 (0.07-0.66)	–0.04 (–0.62 to –0.08)	–0.31 (–0.64 to 0.02)	–0.28 (–0.58 to 0.02)
WL vs NHSWBP	0.06 (–0.19 to 0.31)	0.18 (–0.11 to 0.46)	–0.15 (–0.41 to 0.10)	–0.24 (–0.55 to 0.07)	–0.02(–0.32 to 0.27)

[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.

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

[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.

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

[a]MPS: My Possible Self.

[b]NHSWBP: National Health Service Highland Staff Wellbeing Project.

[c]WL: waitlist.

[d]MT: 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 NHSWPB intervention. All groups experienced an increase 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).
[PDF File (Adobe PDF File), 1563 KB - mental_v9i4e34002_app1.pdf]

References

1. Daly M, Robinson E. Longitudinal changes in psychological distress in the UK from 2019 to September 2020 during the COVID-19 pandemic: Evidence from a large nationally representative study. Psychiatry Res 2021 Jun;300:113920 [FREE Full text] [doi: 10.1016/j.psychres.2021.113920] [Medline: 33882397]
2. Lamb D, Greenberg N, Stevelink SAM, Wessely S. Mixed signals about the mental health of the NHS workforce. Lancet Psychiatry 2020 Dec;7(12):1009-1011. [doi: 10.1016/s2215-0366(20)30379-5]
3. Petrie K, Crawford J, Baker STE, Dean K, Robinson J, Veness BG, et al. Interventions to reduce symptoms of common mental disorders and suicidal ideation in physicians: a systematic review and meta-analysis. Lancet Psychiatry 2019 Mar;6(3):225-234. [doi: 10.1016/s2215-0366(18)30509-1]
4. Gold J. Covid-19: adverse mental health outcomes for healthcare workers. BMJ 2020 May 05;369:m1815. [doi: 10.1136/bmj.m1815] [Medline: 32371465]
5. Tawfik DS, Scheid A, Profit J, Shanafelt T, Trockel M, Adair KC, et al. Evidence relating health care provider burnout and quality of care. Ann Intern Med 2019 Oct 08;171(8):555. [doi: 10.7326/m19-1152]
6. Kisely S, Warren N, McMahon L, Dalais C, Henry I, Siskind D. Occurrence, prevention, and management of the psychological effects of emerging virus outbreaks on healthcare workers: rapid review and meta-analysis. BMJ 2020 May 05;369:m1642 [FREE Full text] [doi: 10.1136/bmj.m1642] [Medline: 32371466]
7. Brooks S, Dunn R, Amlôt R, Rubin G, Greenberg N. A systematic, thematic review of social and occupational factors associated with psychological outcomes in healthcare employees during an infectious disease outbreak. J Occup Environ Med 2018;60(3):248-257. [doi: 10.1097/jom.0000000000001235]
8. De Kock JH, Latham HA, Leslie SJ, Grindle M, Munoz S, Ellis L, et al. A rapid review of the impact of COVID-19 on the mental health of healthcare workers: implications for supporting psychological well-being. BMC Public Health 2021 Jan 09;21(1):104 [FREE Full text] [doi: 10.1186/s12889-020-10070-3] [Medline: 33422039]
9. The Mental Health and Wellbeing of the Medical Workforce: Now and Beyond COVID-19. URL: https://www.bma.org.uk/media/2475/bma-covid-19-and-nhs-staff-mental-health-wellbeing-report-may-2020.pdf [accessed 2022-03-14]
10. Thornicroft G. Stigma and discrimination limit access to mental health care. Epidemiol Psichiatr Soc 2008 May 18;17(1):14-19. [doi: 10.1017/s1121189x00002621] [Medline: 18444452]
11. Knaak S, Mantler E, Szeto A. Mental illness-related stigma in healthcare: barriers to access and care and evidence-based solutions. Healthc Manage Forum 2017 Mar 16;30(2):111-116 [FREE Full text] [doi: 10.1177/0840470416679413] [Medline: 28929889]
12. Henderson C, Noblett J, Parke H, Clement S, Caffrey A, Gale-Grant O, et al. Mental health-related stigma in health care and mental health-care settings. Lancet Psychiatry 2014 Nov;1(6):467-482. [doi: 10.1016/s2215-0366(14)00023-6]
13. Galbraith N, Boyda D, McFeeters D, Hassan T. The mental health of doctors during the COVID-19 pandemic. BJPsych Bull 2021 Apr 28;45(2):93-97 [FREE Full text] [doi: 10.1192/bjb.2020.44] [Medline: 32340645]
14. Sorkin DH, Janio EA, Eikey EV, Schneider M, Davis K, Schueller SM, et al. Rise in use of digital mental health tools and technologies in the United States during the COVID-19 pandemic: survey study. J Med Internet Res 2021 Apr 16;23(4):e26994 [FREE Full text] [doi: 10.2196/26994] [Medline: 33822737]
15. COVID-19: Digital Health Trends & 2021 Opportunities Report: Analysis of Digital Health Usage from January 2020 to January 2021. URL: https://orchahealth.com/covid-19-digital-health-trends-2021-opportunities-report/ [accessed 2022-03-14]
16. Cuijpers P, Donker T, van Straten A, Li J, Andersson G. Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. Psychol Med 2010 Apr 21;40(12):1943-1957. [doi: 10.1017/s0033291710000772]

17. Rauschenberg C, Schick A, Hirjak D, Seidler A, Paetzold I, Apfelbacher C, et al. Evidence synthesis of digital interventions to mitigate the negative impact of the COVID-19 pandemic on public mental health: rapid meta-review. J Med Internet Res 2021 Mar 10;23(3):e23365 [FREE Full text] [doi: 10.2196/23365] [Medline: 33606657]
18. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. Lancet Psychiatry 2020 Jun;7(6):547-560. [doi: 10.1016/s2215-0366(20)30168-1]
19. Proudfoot J, Clarke J, Birch M, Whitton AE, Parker G, Manicavasagar V, et al. Impact of a mobile phone and web program on symptom and functional outcomes for people with mild-to-moderate depression, anxiety and stress: a randomised controlled trial. BMC Psychiatry 2013 Nov 18;13(1):312. [doi: 10.1186/1471-244x-13-312]
20. Gucciardi DF, Hanton S, Gordon S, Mallett CJ, Temby P. The concept of mental toughness: tests of dimensionality, nomological network, and traitness. J Pers 2015 Feb 27;83(1):26-44. [doi: 10.1111/jopy.12079] [Medline: 24428736]
21. Li D, Zhang W, Li X, Li N, Ye B. Gratitude and suicidal ideation and suicide attempts among Chinese adolescents: direct, mediated, and moderated effects. J Adolesc 2012 Feb;35(1):55-66. [doi: 10.1016/j.adolescence.2011.06.005] [Medline: 21774977]
22. NHS Apps Library 2021. URL: https://www.nhs.uk/apps-library/my-possible-self/ [accessed 2022-03-14]
23. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
24. Lamb D, Gnanapragasam S, Greenberg N, Bhundia R, Carr E, Hotopf M. The psychosocial impact of the COVID-19 pandemic on 4,378 UK healthcare workers and ancillary staff: initial baseline data from a cohort study collected during the first wave of the pandemic. medRxiv 2021:0887. [doi: 10.1101/2021.01.21.20240887]
25. Luo M, Guo L, Yu M, Jiang W, Wang H. The psychological and mental impact of coronavirus disease 2019 (COVID-19) on medical staff and general public: a systematic review and meta-analysis. Psychiatry Res 2020 Sep;291:113190 [FREE Full text] [doi: 10.1016/j.psychres.2020.113190] [Medline: 32563745]
26. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
27. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes 2007 Nov 27;5(1):63 [FREE Full text] [doi: 10.1186/1477-7525-5-63] [Medline: 18042300]
28. Ahmed MZ, Ahmed O, Aibao Z, Hanbin S, Siyu L, Ahmad A. Epidemic of COVID-19 in China and associated psychological problems. Asian J Psychiatr 2020 Jun;51:102092 [FREE Full text] [doi: 10.1016/j.ajp.2020.102092] [Medline: 32315963]
29. Greenberg N, Weston D, Hall C, Caulfield T, Williamson V, Fong K. Mental health of staff working in intensive care during Covid-19. Occup Med (Lond) 2021 Apr 09;71(2):62-67 [FREE Full text] [doi: 10.1093/occmed/kqaa220] [Medline: 33434920]
30. Cowden RG, Worthington EL, Griffin BJ, Davis DE, Grubbs JB. Trait tendencies to forgive, punish, and exonerate oneself: a multi-study investigation. J Res Pers 2020 Jun;86:103934. [doi: 10.1016/j.jrp.2020.103934]
31. Boniol M, McIsaac M, Xu L, Wuliji T, Diallo K, Campbell J. Gender Equity in the Health Workforce: Analysis of 104 Countries. URL: https://apps.who.int/iris/bitstream/handle/10665/311314/WHO-HIS-HWF-Gender-WP1-2019.1-eng.pdf?sequence=1&isAllowed=y [accessed 2022-03-14]
32. Stamatis A, Morgan GB, Papadakis Z, Mougios V, Bogdanis G, Spinou A. Cross-cultural invariance of the mental toughness index among American and Greek athletes. Curr Psychol 2019 Nov 09;40(12):5793-5800. [doi: 10.1007/s12144-019-00532-2]
33. Cowden RG. Mental Toughness Inventory: factorial validity and ethnic group measurement equivalence in competitive tennis. Curr Psychol 2018 Feb 2;39(2):736-741. [doi: 10.1007/s12144-018-9798-6]
34. McCullough ME, Emmons RA, Tsang J. The grateful disposition: a conceptual and empirical topography. J Pers Soc Psychol 2002;82(1):112-127. [doi: 10.1037/0022-3514.82.1.112]
35. Kong F, You X, Zhao J. Evaluation of the Gratitude Questionnaire in a Chinese sample of adults: factorial validity, criterion-related validity, and measurement invariance across sex. Front Psychol 2017 Sep 01;8:1498 [FREE Full text] [doi: 10.3389/fpsyg.2017.01498] [Medline: 28919873]
36. Langer, Ulloa VG, Aguilar-Parra JM, Araya-Véliz C, Brito G. Validation of a Spanish translation of the Gratitude Questionnaire (GQ-6) with a Chilean sample of adults and high schoolers. Health Qual Life Outcomes 2016 Mar 31;14(1):53 [FREE Full text] [doi: 10.1186/s12955-016-0450-6] [Medline: 27036956]
37. R Core Team. The R Project for Statistical Computing. URL: https://www.r-project.org/ [accessed 2022-03-14]
38. Bates D, Maechler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. J Stat Softw 2015;67(1):1-48.
39. Tukey JW. Comparing individual means in the analysis of variance. Biometrics 1949 Jun;5(2):99. [doi: 10.2307/3001913]
40. Kenward MG, Roger JH. Small sample inference for fixed effects from restricted maximum likelihood. Biometrics 1997 Sep;53(3):983. [doi: 10.2307/2533558]
41. Hedges LV. Effect sizes in cluster-randomized designs. J Educ Behav Stat 2016 Oct 26;32(4):341-370. [doi: 10.3102/1076998606298043]

42. Westfall J, Kenny DA, Judd CM. Statistical power and optimal design in experiments in which samples of participants respond to samples of stimuli. J Exp Psychol Gen 2014 Oct;143(5):2020-2045. [doi: 10.1037/xge0000014] [Medline: 25111580]
43. Lenth RV. Using lsmeans. J Stat Softw 2017;69:1-33.
44. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, PAFS Consensus Group. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ 2016 Oct 24;355:i5239 [FREE Full text] [doi: 10.1136/bmj.i5239] [Medline: 27777223]
45. Zhang Y, Hedo R, Rivera A, Rull R, Richardson S, Tu XM. Post hoc power analysis: is it an informative and meaningful analysis? Gen Psychiatr 2019 Aug 08;32(4):e100069 [FREE Full text] [doi: 10.1136/gpsych-2019-100069] [Medline: 31552383]
46. Watkins PC, McCurrach D. Progress in the science of gratitude. In: Snyder CR, Lopez SJ, Edwards LM, Marques SC, editors. The Oxford Handbook of Positive Psychology. UK: Oxford; 2021:570-585.
47. Haghighi M, Gerber M. Does mental toughness buffer the relationship between perceived stress, depression, burnout, anxiety, and sleep? Int J Stress Manag 2019 Aug;26(3):297-305. [doi: 10.1037/str0000106]
48. Stamatis A, Grandjean P, Morgan G, Padgett RN, Cowden R, Koutakis P. Developing and training mental toughness in sport: a systematic review and meta-analysis of observational studies and pre-test and post-test experiments. BMJ Open Sport Exerc Med 2020;6(1):e000747 [FREE Full text] [doi: 10.1136/bmjsem-2020-000747] [Medline: 32577300]
49. Seligman MEP, Csikszentmihalyi M. Positive psychology: an introduction. Am Psychol 2000;55(1):5-14. [doi: 10.1037/0003-066x.55.1.5]
50. De Kock JH, Pillay B. South Africa's rural mental health human resource crisis: a situation analysis and call for innovative task-shifting. Fam Med Prim Care Rev 2018;20(2):124-130 [FREE Full text] [doi: 10.5114/fmpcr.2018.76454]
51. NHS Highland. Welcome to NHS Careers. URL: https://www.nhshighland.scot.nhs.uk/CAREERS/Pages/Home.aspx [accessed 2022-03-14]
52. Cowden RG, Captari LE, Chen ZJ, De Kock JH, Houghtaling A. Effectiveness of an intensive experiential group therapy program in promoting mental health and well-being among mass shooting survivors: a practice-based pilot study. Prof Psychol Res Pract 2022 Jan 31:Advance online publication. [doi: 10.1037/pro0000443]

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:
De Kock JH, Latham HA, Cowden RG, Cullen B, Narzisi K, Jerdan S, Munoz SA, Leslie SJ, Stamatis A, Eze J
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

Original Paper

Predicting Uptake of the COVID Coach App Among US Military Veterans: Funnel Analysis Using a Probability-Based Panel

Beth K Jaworski[1], PhD; Katherine Taylor[1], MPH, PsyD; Kelly M Ramsey[1]; Adrienne J Heinz[1], PhD; Sarah Steinmetz[1], PhD; Jason E Owen[1], MPH, PhD; Jack Tsai[2,3], PhD; Robert H Pietrzak[4,5], MPH, PhD

[1]National Center for Posttraumatic Stress Disorder, Dissemination & Training Division, US Department of Veterans Affairs, Menlo Park, CA, United States

[2]National Center on Homelessness Among Veterans, US Department of Veterans Affairs, Tampa, FL, United States

[3]School of Public Health, San Antonio Campus, University of Texas Health Science Center at Houston, San Antonio, TX, United States

[4]National Center for Posttraumatic Stress Disorder, VA Connecticut Healthcare System, US Department of Veterans Affairs, New Haven, CT, United States

[5]Department 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 peripandemic 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.

(JMIR Ment Health 2022;9(4):e36217) doi:10.2196/36217

KEYWORDS

COVID-19; coronavirus; mobile app; mHealth; digital health; mental health; public mental health; stress; coping; older adults; veterans

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

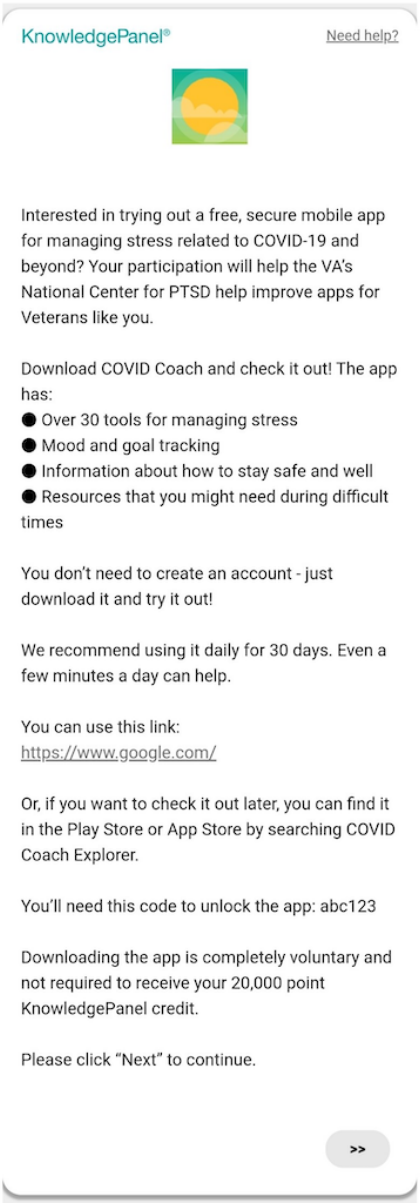
Data were analyzed from the 2019-2020 National Health and Resilience in Veterans Study (NHRVS), which surveyed a nationally representative, prospective cohort of US military veterans. The NHRVS sample was drawn from KnowledgePanel, a research panel of more than 50,000 households maintained by Ipsos, Inc. KnowledgePanel is a probability-based, online, nonvolunteer access survey panel of a nationally representative sample of US adults that covers approximately 98% 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 (χ^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.

Characteristics	Measures
Sociodemographic characteristics	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
Military characteristics	Combat veteran status (combat exposure, not), military branch (Army, Navy, Air Force, Marine Corps, other); combat veteran status (yes, no), years of military service
Prepandemic psychosocial risk factors	
Adverse childhood experiences	Adverse Childhood Experiences Questionnaire (ACEQ) score [45]
Total traumas	Items endorsed on Life Events Checklist for DSM-5[a] (LEC-5) [46]
Lifetime MDD[b] or PTSD[c]	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[e] or DUD[f]	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
Number of medical conditions	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
Personality characteristics	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	COVID-19 infection status (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?").

[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 χ^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 χ^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 χ^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. χ^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 χ^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 $P>.05$).

Table 2. Characteristics of US veterans by device type used for survey completion.

Characteristic	Completed survey on a desktop or personal computer (n=2264; weighted 65.8%)	Completed survey on a mobile device (n=814; weighted 34.2%)	*t* (*df*=3070) or χ^2	*P* value
Age, weighted mean (SD)	64.8 (14.6)	55.5 (15.3)	16.53	<.001
Female gender, n (weighted %)	183 (6.0)	161 (17.0)	92.28 (*df*=1)	<.001
Race/ethnicity, n (weighted %)			71.11 (*df*=3)	<.001
White, non-Hispanic	1932 (83.5)	609 (70.5)		
Black, non-Hispanic	131 (8.7%)	81 (13.7%)		
Hispanic	130 (4.3)	86 (9.6)		
Bi/Multiracial or Other	71 (3.6)	38 (6.3)		
College graduate or higher education, n (weighted %)	1072 (33.9)	336 (28.4)	9.47 (*df*=1)	.002
Married/partnered, n (weighted %)	1634 (73.8)	590 (75.0)	0.54 (*df*=1)	.46
Currently employed, n (weighted %)	787 (45.0)	408 (61.2)	72.55 (*df*=1)	<.001
Annual household income US $60,000 or above, n (weighted %)	1328 (60.3)	525 (64.4)	4.96 (*df*=1)	.03
Region of country, n (weighted %)			31.60 (*df*=3)	<.001
South	835 (38.9)	353 (47.6)		
West	597 (25.2)	174 (17.9)		
Midwest	528 (20.8)	187 (21.8)		
Northeast	304 (15.1)	100 (12.7)		
Non-metro residence	331 (15.8)	143 (17.5)	1.41	.24
Military branch, n (weighted %)			10.93 (*df*=4)	.03
Army	900 (47.9)	298 (45.3)		
Navy	467 (19.5)	206 (21.2)		
Air Force	557 (19.5)	176 (17.8)		
Marine Corps	128 (4.9)	61 (7.5)		
Other	212 (8.2)	73 (8.2)		
Combat veteran, n (weighted %)	781 (33.3)	271 (37.5)	5.60 (*df*=1)	.02
10+ years of military service, n (weighted %)	809 (35.0)	323 (39.5)	6.23 (*df*=1)	.01
Health variables, n (weighted %)				
Lifetime MDD[a] and/or PTSD[b]	356 (16.7)	215 (27.8)	51.83 (*df*=1)	<.001
Lifetime AUD[c] and/or DUD[d]	924 (41.0)	343 (48.2)	14.16 (*df*=1)	<.001
Number of medical conditions, weighted mean (SD)	2.9 (2.2)	2.7 (2.0)	3.02	.001
Own any type of cell phone, n (weighted %)	2,103 (91.9%)	793 (97.2%)	32.87 (*df*=1)	<.001
Own a smartphone, n (weighted %)	1,698 (76.6%)	745 (91.1%)	96.55 (*df*=1)	<.001

[a]MDD: major depressive disorder.

[b]PTSD: posttraumatic stress disorder.

[c]AUD: alcohol use disorder.

[d]DUD: 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.

Characteristic	*Did not* download COVID Coach (n=780, weighted 65.8%)	Downloaded COVID Coach (n=34, weighted 3.3%)	*t* (*df*=1049) or χ^2	*P* value
Age (years), weighted mean (SD)	55.6 (15.3)	51.8 (15.5)	1.44	.08
Female gender, n (weighted %)	153 (16.8)	8 (22.9)	0.87 (*df*=1)	.35
Nonwhite race/ethnicity, n (weighted %)	196 (30.6)	9 (32.4)	0.05 (*df*=1)	.83
College graduate or higher education, n (weighted %)	313 (27.5)	23 (55.9)	13.02 (*df*=1)	<.001
Married/partnered, n (weighted %)	561 (74.7)	29 (82.4)	1.02 (*df*=1)	.31
Currently employed, n (weighted %)	389 (61.2)	19 (61.8)	0.01 (*df*=1)	.94
Annual household income US $60,000 or higher, n (weighted %)	503 (64.6)	22 (61.8)	0.12 (*df*=1)	.73
Region of country, n (weighted %)			3.85 (*df*=3)	.28
South	338 (47.3)	15 (58.8)		
West	167 (18.2)	7 (5.9)		
Midwest	180 (21.8)	7 (20.6)		
Northeast	95 (12.7)	5 (14.7)		
Nonmetro residence, n (weighted %)	643 (17.9)	28 (8.8)	1.87 (*df*=1)	.17
Military branch, n (weighted %)			2.05 (*df*=4)	.73
Army	285 (45.5)	13 (41.2)		
Navy	200 (21.3)	6 (17.6)		
Air Force	168 (17.6)	8 (23.5)		
Marine Corps	57 (7.4)	4 (11.8)		
Other	70 (8.3)	3 (5.9)		
Combat veteran, n (weighted %)	258 (37.4)	13 (41.2)	0.20 (*df*=1)	.66
10+ years of military service, n (weighted %)	310 (39.6)	13 (35.3)	0.26 (*df*=1)	.61
Prepandemic variables				
Adverse childhood experiences, weighted mean (SD)	1.8 (2.2)	2.6 (3.0)	2.14	.02
Number of traumas, weighted mean (SD)	9.4 (9.2)	12.7 (10.2)	2.05	.02
Lifetime MDD[a] and/or PTSD[b], n (weighted %)	209 (31.6)	15 (35.3)	0.21 (*df*=1)	.65
Lifetime AUD[c] and/or DUD[d], n (weighted %)	327 (48.3)	16 (44.1)	0.23 (*df*=1)	.63
Extraversion, weighted mean (SD)	3.7 (1.6)	4.9 (1.3)	4.56	<.001
Agreeableness, weighted mean (SD)	4.9 (1.4)	5.0 (1.2)	0.26	.40
Conscientiousness, weighted mean (SD)	5.8 (1.1)	5.7 (1.4)	0.35	.36
Emotional stability, weighted mean (SD)	5.1 (1.4)	5.1 (1.5)	0.07	.47
Openness to experiences, weighted mean (SD)	4.8 (1.2)	5.0 (1.1)	0.86	.20
Pandemic-related variables				
COVID-19 infection to self, n (weighted %)	70 (8.5)	6 (17.6)	3.41 (*df*=1)	.07

Characteristic	*Did not* download COVID Coach (n=780, weighted 65.8%)	Downloaded COVID Coach (n=34, weighted 3.3%)	*t* (*df*=1049) or χ^2	*P* value
COVID-19 infection to household member, n (weighted %)	61 (9.0)	4 (15.2)	1.43 (*df*=1)	.23
COVID-19 infection to nonhousehold member, n (weighted %)	360 (45.9%)	18 (55.9%)	1.32 (*df*=1)	.25
Know someone who died from COVID-19, n (weighted %)	55 (6.2)	3 (5.7)	0.01 (*df*=1)	.91
Pandemic-related media exposure, weighted mean (SD)	1.6 (2.0)	2.4 (2.6)	2.30	.01
Pandemic-related worries, weighted mean (SD)	–0.1 (1.0)	0.2 (1.1)	1.39	.08
Pandemic-related social restriction stress, weighted mean (SD)	0.0 (1.1)	0.3 (1.1)	1.80	.04
Pandemic-related financial stress, weighted mean (SD)	0.1 (1.1)	0.2 (1.4)	0.70	.24
Pandemic-related worsening of relationships, weighted mean (SD)	0.0 (1.1)	0.5 (1.1)	2.21	.01
Pandemic-related PTSD symptoms, n (weighted %)	113 (13.7)	12 (29.4)	6.65 (*df*=1)	.01

[a]MDD: major depressive disorder.

[b]PTSD: posttraumatic stress disorder.

[c]AUD: alcohol use disorder.

[d]DUD: drug use disorder.

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.

Multimedia Appendix 1
COVID Coach Explorer screenshots.
[PDF File (Adobe PDF File), 2966 KB - mental_v9i4e36217_app1.pdf]

References

1. El-Mohandes A, White TM, Wyka K, Rauh L, Rabin K, Kimball SH, et al. COVID-19 vaccine acceptance among adults in four major US metropolitan areas and nationwide. Sci Rep 2021 Nov 04;11(1):21844 [FREE Full text] [doi: 10.1038/s41598-021-00794-6] [Medline: 34737319]
2. Czeisler ME, Lane RI, Petrosky E, Wiley JF, Christensen A, Njai R, et al. Mental health, substance use, and suicidal ideation during the COVID-19 pandemic - United States, June 24-30, 2020. MMWR Morb Mortal Wkly Rep 2020 Aug 14;69(32):1049-1057 [FREE Full text] [doi: 10.15585/mmwr.mm6932a1] [Medline: 32790653]
3. Ettman CK, Abdalla SM, Cohen GH, Sampson L, Vivier PM, Galea S. Prevalence of depression symptoms in US adults before and during the COVID-19 pandemic. JAMA Netw Open 2020 Sep 01;3(9):e2019686 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.19686] [Medline: 32876685]
4. Vindegaard N, Benros ME. COVID-19 pandemic and mental health consequences: Systematic review of the current evidence. Brain Behav Immun 2020 Oct;89:531-542 [FREE Full text] [doi: 10.1016/j.bbi.2020.05.048] [Medline: 32485289]
5. Xiong J, Lipsitz O, Nasri F, Lui LM, Gill H, Phan L, et al. Impact of COVID-19 pandemic on mental health in the general population: a systematic review. J Affect Disord 2020 Dec 01;277:55-64 [FREE Full text] [doi: 10.1016/j.jad.2020.08.001] [Medline: 32799105]
6. Satre DD, Hirschtritt ME, Silverberg MJ, Sterling SA. Addressing problems with alcohol and other substances among older adults during the COVID-19 pandemic. Am J Geriatr Psychiatry 2020 Jul;28(7):780-783 [FREE Full text] [doi: 10.1016/j.jagp.2020.04.012] [Medline: 32359882]
7. Vital statistics rapid release: Provisional drug overdose death counts. Centers for Disease Control and Prevention National Center for Healthcare Statistics. 2021. URL: https://www.cdc.gov/nchs/nvss/vsrr/drug-overdose-data.htm [accessed 2022-03-23]
8. Vahia IV, Jeste DV, Reynolds CF. Older adults and the mental health effects of COVID-19. JAMA 2020 Dec 08;324(22):2253-2254 [FREE Full text] [doi: 10.1001/jama.2020.21753] [Medline: 33216114]
9. Rutherford BR, Choi CJ, Chrisanthopolous M, Salzman C, Zhu C, Montes-Garcia C, et al. The COVID-19 pandemic as a traumatic stressor: mental health responses of older adults with chronic PTSD. Am J Geriatr Psychiatry 2021 Feb;29(2):105-114 [FREE Full text] [doi: 10.1016/j.jagp.2020.10.010] [Medline: 33153871]
10. Pietrzak RH, Tsai J, Southwick SM. Association of symptoms of posttraumatic stress disorder with posttraumatic psychological growth among US veterans during the COVID-19 pandemic. JAMA Netw Open 2021 Apr 01;4(4):e214972 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.4972] [Medline: 33830230]
11. Lee EE, Depp C, Palmer BW, Glorioso D, Daly R, Liu J, et al. High prevalence and adverse health effects of loneliness in community-dwelling adults across the lifespan: role of wisdom as a protective factor. Int Psychogeriatr 2019 Oct;31(10):1447-1462 [FREE Full text] [doi: 10.1017/S1041610218002120] [Medline: 30560747]
12. Roy J, Jain R, Golamari R, Vunnam R, Sahu N. COVID-19 in the geriatric population. Int J Geriatr Psychiatry 2020 Dec;35(12):1437-1441 [FREE Full text] [doi: 10.1002/gps.5389] [Medline: 32748545]
13. Purcell N, Sells J, McGrath S, Mehlman H, Bertenthal D, Seal KH. "Then COVID happened…": Veterans' health, wellbeing, and engagement in whole health care during the COVID-19 pandemic. Glob Adv Health Med 2021 Nov 03;10:21649561211053828 [FREE Full text] [doi: 10.1177/21649561211053828] [Medline: 35174002]
14. Murphy D, Williamson C, Baumann J, Busuttil W, Fear NT. Exploring the impact of COVID-19 and restrictions to daily living as a result of social distancing within veterans with pre-existing mental health difficulties. BMJ Mil Health 2022 Feb 26;168(1):29-33 [FREE Full text] [doi: 10.1136/bmjmilitary-2020-001622] [Medline: 33243764]
15. Hill ML, Nichter B, Na PJ, Norman SB, Morland LA, Krystal JH, et al. Mental health impact of the COVID-19 pandemic in U.S. military veterans: a population-based, prospective cohort study. Psychol Med 2021 Jun 14:1-12 [FREE Full text] [doi: 10.1017/S0033291721002361] [Medline: 34120667]
16. Nichter B, Hill ML, Na PJ, Kline AC, Norman SB, Krystal JH, et al. Prevalence and trends in suicidal behavior among US military veterans during the COVID-19 pandemic. JAMA Psychiatry 2021 Nov 01;78(11):1218-1227 [FREE Full text] [doi: 10.1001/jamapsychiatry.2021.2332] [Medline: 34431973]
17. Connolly SL, Stolzmann KL, Heyworth L, Sullivan JL, Shimada SL, Weaver KR, et al. Patient and provider predictors of telemental health use prior to and during the COVID-19 pandemic within the Department of Veterans Affairs. Am Psychol 2021 Dec 23:online ahead of print. [doi: 10.1037/amp0000895] [Medline: 34941310]
18. Marini CM, Pless Kaiser A, Smith BN, Fiori KL. Aging veterans' mental health and well-being in the context of COVID-19: the importance of social ties during physical distancing. Psychol Trauma 2020 Aug;12(S1):S217-S219. [doi: 10.1037/tra0000736] [Medline: 32525383]
19. Galea S, Merchant RM, Lurie N. The mental health consequences of COVID-19 and physical distancing: the need for prevention and early intervention. JAMA Intern Med 2020 Jun 01;180(6):817-818 [FREE Full text] [doi: 10.1001/jamainternmed.2020.1562] [Medline: 32275292]

20. Balcombe L, De Leo D. An integrated blueprint for digital mental health services amidst COVID-19. JMIR Ment Health 2020 Jul 22;7(7):e21718 [FREE Full text] [doi: 10.2196/21718] [Medline: 32668402]
21. Mobile Technology and Home Broadband. Pew Research Center. 2021 Jun 03. URL: https://www.pewresearch.org/internet/2021/06/03/mobile-technology-and-home-broadband-2021/ [accessed 2022-03-23]
22. Wang J, Du Y, Coleman D, Peck M, Myneni S, Kang H, et al. Mobile and connected health technology needs for older adults aging in place: cross-sectional survey study. JMIR Aging 2019 May 15;2(1):e13864 [FREE Full text] [doi: 10.2196/13864] [Medline: 31518283]
23. Graham SA, Stein N, Shemaj F, Branch OH, Paruthi J, Kanick SC. Older adults engage with personalized digital coaching programs at rates that exceed those of younger adults. Front Digit Health 2021 Aug 6;3:642818 [FREE Full text] [doi: 10.3389/fdgth.2021.642818] [Medline: 34713112]
24. Portenhauser AA, Terhorst Y, Schultchen D, Sander LB, Denkinger MD, Stach M, et al. Mobile apps for older adults: systematic search and evaluation within online stores. JMIR Aging 2021 Feb 19;4(1):e23313 [FREE Full text] [doi: 10.2196/23313] [Medline: 33605884]
25. Erbes CR, Stinson R, Kuhn E, Polusny M, Urban J, Hoffman J, et al. Access, utilization, and interest in mHealth applications among veterans receiving outpatient care for PTSD. Mil Med 2014 Nov;179(11):1218-1222. [doi: 10.7205/MILMED-D-14-00014] [Medline: 25373044]
26. Connolly SL, Miller CJ, Koenig CJ, Zamora KA, Wright PB, Stanley RL, et al. Veterans' attitudes toward smartphone app use for mental health care: qualitative study of rurality and age differences. JMIR Mhealth Uhealth 2018 Aug 22;6(8):e10748 [FREE Full text] [doi: 10.2196/10748] [Medline: 30135050]
27. Gould CE, Ma F, Scales AN, Juang C, Carlson C, Loup J, et al. Geri-Mobile Health: Development and feasibility of a program to help older veterans use VA mental health mobile apps. Innov Aging 2021;3(Suppl 1):S560-S561 [FREE Full text] [doi: 10.1093/geroni/igz038.2072]
28. Gould CE, Loup JR, Scales AN, Juang C, Carlson C, Ma F, et al. Development and refinement of educational materials to help older veterans use VA mental health mobile apps. Prof Psychol Res Pr 2020 Aug;51(4):414-423 [FREE Full text] [doi: 10.1037/pro0000354] [Medline: 34465939]
29. Rosen CS, Morland LA, Glassman LH, Marx BP, Weaver K, Smith CA, et al. Virtual mental health care in the Veterans Health Administration's immediate response to coronavirus disease-19. Am Psychol 2021 Jan;76(1):26-38 [FREE Full text] [doi: 10.1037/amp0000751] [Medline: 33119331]
30. Boden M, Zimmerman L, Azevedo KJ, Ruzek JI, Gala S, Abdel Magid HS, et al. Addressing the mental health impact of COVID-19 through population health. Clin Psychol Rev 2021 Apr;85:102006 [FREE Full text] [doi: 10.1016/j.cpr.2021.102006] [Medline: 33714167]
31. Jaworski BK, Taylor K, Ramsey KM, Heinz A, Steinmetz S, Pagano I, et al. Exploring usage of COVID Coach, a public mental health app designed for the COVID-19 pandemic: evaluation of analytics data. J Med Internet Res 2021 Mar 01;23(3):e26559 [FREE Full text] [doi: 10.2196/26559] [Medline: 33606656]
32. Baumel A, Muench F, Edan S, Kane JM. Objective user engagement with mental health apps: systematic search and panel-based usage analysis. J Med Internet Res 2019 Sep 25;21(9):e14567 [FREE Full text] [doi: 10.2196/14567] [Medline: 31573916]
33. Carroll JK, Moorhead A, Bond R, LeBlanc WG, Petrella RJ, Fiscella K. Who uses mobile phone health apps and does use matter? A secondary data analytics approach. J Med Internet Res 2017 Apr 19;19(4):e125 [FREE Full text] [doi: 10.2196/jmir.5604] [Medline: 28428170]
34. Stachl C, Au Q, Schoedel R, Gosling SD, Harari GM, Buschek D, et al. Predicting personality from patterns of behavior collected with smartphones. Proc Natl Acad Sci U S A 2020 Jul 28;117(30):17680-17687 [FREE Full text] [doi: 10.1073/pnas.1920484117] [Medline: 32665436]
35. Owen JE, Jaworski BK, Kuhn E, Makin-Byrd KN, Ramsey KM, Hoffman JE. mHealth in the wild: using novel data to examine the reach, use, and impact of PTSD Coach. JMIR Ment Health 2015 Mar 25;2(1):e7 [FREE Full text] [doi: 10.2196/mental.3935] [Medline: 26543913]
36. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug 05;21(3):116-119 [FREE Full text] [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
37. National Center for Veterans Analysis and Statistics. US Department of Veterans Affairs. URL: https://www.va.gov/vetdata/veteran_population.asp [accessed 2022-03-23]
38. Schaeffer K. The changing face of America’s veteran population. Pew Research Center. URL: https://www.pewresearch.org/fact-tank/2021/04/05/the-changing-face-of-americas-veteran-population/ [accessed 2022-03-23]
39. Profile of Veterans: 2017 Highlights. National Center for Veterans Analysis and Statistics. 2017. URL: https://www.va.gov/vetdata/docs/Quickfacts/2017_Veterans_Profile_Fact_Sheet.pdf [accessed 2022-03-22]
40. Reger GM, Harned M, Stevens ES, Porter S, Nguyen J, Norr AM. Mobile applications may be the future of veteran mental health support but do veterans know yet? A survey of app knowledge and use. Psychol Serv 2021 Jun 03:online ahead of print. [doi: 10.1037/ser0000562] [Medline: 34081527]

41. American Community Survey (ACS). United States Census Bureau. 2019. URL: https://www.census.gov/programs-surveys/acs [accessed 2022-03-23]
42. Palo Alto Veterans Institute for Research. COVID Coach Explorer (Version 1). Google Play Store. 2020. URL: https://play.google.com/store/apps/details?id=pavir.explorer.covid&hl=en_US&gl=US [accessed 2022-03-23]
43. Palo Alto Veterans Institute for Research. COVID Coach Explorer (Version 1). Apple Store. URL: https://apps.apple.com/us/app/covid-coach-explorer/id1532843399 [accessed 2022-03-23]
44. VA Technical Reference Model v 22.2. US Department of Veterans Affairs. URL: https://www.oit.va.gov/Services/TRM/TRMHomePage.aspx [accessed 2022-03-23]
45. Finkelhor D, Shattuck A, Turner H, Hamby S. A revised inventory of adverse childhood experiences. Child Abuse Negl 2015 Oct;48:13-21. [doi: 10.1016/j.chiabu.2015.07.011] [Medline: 26259971]
46. VA's National Center for PTSD. The life events checklist for DSM-5 (LEC-5). US Department of Veterans Affairs. 2013. URL: https://www.ptsd.va.gov/professional/assessment/te-measures/life_events_checklist.asp [accessed 2022-03-23]
47. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59(Suppl 20):22-33;quiz 34. [Medline: 9881538]
48. Bovin MJ, Marx BP, Weathers FW, Gallagher MW, Rodriguez P, Schnurr PP, et al. Psychometric properties of the PTSD Checklist for Diagnostic and Statistical Manual of Mental Disorders-Fifth Edition (PCL-5) in veterans. Psychol Assess 2016 Nov;28(11):1379-1391. [doi: 10.1037/pas0000254] [Medline: 26653052]
49. PTSD: National Center for PTSD. The PTSD checklist for DSM-5 (PCL-5). US Department of Veterans Affairs. 2013. URL: https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp [accessed 2022-03-23]
50. Grant B, Dawson D, Hasin D. The Wave 2 National Epidemiologic Survey on Alcohol and Related Conditions Alcohol Use Disorder and Associated Disabilities Interview Schedule DSM-IV version. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism; 2004.
51. Gosling SD, Rentfrow PJ, Swann WB. A very brief measure of the Big-Five personality domains. J Res Person 2003 Dec;37(6):504-528. [doi: 10.1016/s0092-6566(03)00046-1]
52. Kline Institute For Psychiatric Research, National Institute of Mental Health Intramural Research Program Mood Spectrum Collaboration. The CoRonavIruS health impact survey (CRISIS) 2020. National Institutes of Health. URL: https://www.nlm.nih.gov/dr2/CRISIS_Adult_Self-Report_Baseline_Current_Form_V0.3.pdf [accessed 2022-03-23]
53. Kwasny MJ, Schueller SM, Lattie E, Gray EL, Mohr DC. Exploring the use of multiple mental health apps within a platform: secondary analysis of the IntelliCare field trial. JMIR Ment Health 2019 Mar 21;6(3):e11572 [FREE Full text] [doi: 10.2196/11572] [Medline: 30896433]
54. Murphy A, Peltekian S, Gardner DM. Website analytics of a Google Ads campaign for a men's mental health website: comparative analysis. JMIR Ment Health 2018 Dec 13;5(4):e12428 [FREE Full text] [doi: 10.2196/12428] [Medline: 30545812]
55. Patel H. Funnel analysis: find conversion problems. Mixpanel. URL: https://mixpanel.com/blog/introduction-to-analytics-funnel-analysis/ [accessed 2022-03-23]
56. Owen JE, Kuhn E, Jaworski BK, McGee-Vincent P, Juhasz K, Hoffman JE, et al. VA mobile apps for PTSD and related problems: public health resources for veterans and those who care for them. Mhealth 2018 Jul;4:28. [doi: 10.21037/mhealth.2018.05.07] [Medline: 30148141]
57. Meffert BN, Morabito DM, Sawicki DA, Hausman C, Southwick SM, Pietrzak RH, et al. US veterans who do and do not utilize Veterans Affairs health care services: demographic, military, medical, and psychosocial characteristics. Prim Care Companion CNS Disord 2019 Jan 17;21(1):18m02350 [FREE Full text] [doi: 10.4088/PCC.18m02350] [Medline: 30677266]
58. Cuzzone K. The 8 best therapy apps to kick off your mental health journey. Cosmpopolitan. 2021. URL: https://www.cosmopolitan.com/health-fitness/a35896459/best-therapy-apps/ [accessed 2022-03-23]
59. Rameswaram S. A firefighter on battling wildfires [audio podcast episode]. Today, Explained. Vox Media. 2020 Dec. URL: https://www.vox.com/today-explained [accessed 2022-03-23]
60. Kassraie A. 20 mobile apps for veterans and military families. AARP. 2022 Jan 26. URL: https://www.aarp.org/home-family/voices/veterans/info-2022/mobile-apps-military-families.html [accessed 2022-03-23]
61. Women Veterans Network WoVEN. URL: https://www.wovenwomenvets.org/ [accessed 2022-03-23]
62. VA Community Provider Toolkit. URL: https://www.mentalhealth.va.gov/communityproviders/ [accessed 2022-03-23]
63. Tech Into Care Initiative. National Center for PTSD US Department of Veterans Affairs. URL: https://www.ptsd.va.gov/professional/tech-care/index.asp [accessed 2022-03-23]
64. Sutin A, Luchetti M, Stephan Y, Terracciano A. Meaning in life and risk of cognitive impairment: a 9-Year prospective study in 14 countries. Arch Gerontol Geriatr 2020;88:104033 [FREE Full text] [doi: 10.1016/j.archger.2020.104033] [Medline: 32182537]
65. Sutin A, Luchetti M, Terracciano A. Sense of purpose in life and healthier cognitive aging. Trends Cogn Sci 2021 Nov;25(11):917-919. [doi: 10.1016/j.tics.2021.08.009] [Medline: 34538721]
66. Crawford A, Serhal E. Digital health equity and COVID-19: the innovation curve cannot reinforce the social gradient of health. J Med Internet Res 2020 Jun 02;22(6):e19361 [FREE Full text] [doi: 10.2196/19361] [Medline: 32452816]

67. Fulmer R, Reuben DB, Auerbach J, Fick DM, Galambos C, Johnson KS. Actualizing better health and health care for older adults. Health Aff (Millwood) 2021 Feb;40(2):219-225 [FREE Full text] [doi: 10.1377/hlthaff.2020.01470] [Medline: 33476185]
68. Rodriguez JA, Clark CR, Bates DW. Digital health equity as a necessity in the 21st century Cures Act era. JAMA 2020 Jun 16;323(23):2381-2382. [doi: 10.1001/jama.2020.7858] [Medline: 32463421]
69. Sumner J, Chong LS, Bundele A, Wei Lim Y. Co-designing technology for aging in place: a systematic review. Gerontologist 2021 Sep 13;61(7):e395-e409 [FREE Full text] [doi: 10.1093/geront/gnaa064] [Medline: 32506136]
70. Mansson L, Wiklund M, Öhberg F, Danielsson K, Sandlund M. Co-creation with older adults to improve user-experience of a smartphone self-test application to assess balance function. Int J Environ Res Public Health 2020 May 26;17(11):3768 [FREE Full text] [doi: 10.3390/ijerph17113768] [Medline: 32466484]
71. Crowe B, Gaulton JS, Minor N, Asch DA, Eyet J, Rainosek E, et al. To improve quality, leverage design. BMJ Qual Saf 2022 Jan 11;31(1):70-74 [FREE Full text] [doi: 10.1136/bmjqs-2021-013605] [Medline: 34510018]
72. Figueroa C, Luo T, Aguilera A, Lyles C. The need for feminist intersectionality in digital health. Lancet Digit Health 2021 Aug;3(8):e526-e533 [FREE Full text] [doi: 10.1016/S2589-7500(21)00118-7] [Medline: 34325855]
73. Latulippe K, Hamel C, Giroux D. Co-design to support the development of inclusive eHealth tools for caregivers of functionally dependent older persons: social justice design. J Med Internet Res 2020 Nov 09;22(11):e18399 [FREE Full text] [doi: 10.2196/18399] [Medline: 33164905]

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

Edited by J Torous; submitted 06.01.22; peer-reviewed by L Guo, T Hogan; comments to author 01.02.22; revised version received 15.02.22; accepted 01.03.22; published 05.04.22.

Please cite as:
Jaworski BK, Taylor K, Ramsey KM, Heinz AJ, Steinmetz S, Owen JE, Tsai J, Pietrzak RH
Predicting Uptake of the COVID Coach App Among US Military Veterans: Funnel Analysis Using a Probability-Based Panel
JMIR Ment Health 2022;9(4):e36217
URL: https://mental.jmir.org/2022/4/e36217
doi:10.2196/36217
PMID:35245204

Original Paper

Behavioral Health Professionals' Perceptions on Patient-Controlled Granular Information Sharing (Part 2): Focus Group Study

Julia Ivanova[1], MA; Tianyu Tang[2], MS, MD; Nassim Idouraine[3], BS; Anita Murcko[3], MD; Mary Jo Whitfield[4], MSW; Christy Dye[5], MPH; Darwyn Chern[5], MD; Adela Grando[3], PhD

[1]School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, United States

[2]College of Medicine, University of Arizona, Tucson, AZ, United States

[3]College of Health Solutions, Biomedical Informatics, Arizona State University, Scottsdale, AZ, United States

[4]Jewish Family and Children's Services, Phoenix, AZ, United States

[5]Partners in Recovery, Phoenix, AZ, United States

Corresponding Author:
Julia Ivanova, MA
School of Human Evolution and Social Change
Arizona State University
900 Cady Mall
Tempe, AZ, 85281
United States
Phone: 1 480 965 6213
Email: jivanova@asu.edu

Related Article:

Companion article: https://mental.jmir.org/2022/4/e21208

Abstract

Background: Patient-directed selection and sharing of health information "granules" is known as granular information sharing. In a previous study, patients with behavioral health conditions categorized their own health information into sensitive categories (eg, mental health) and chose the health professionals (eg, pharmacists) who should have access to those records. Little is known about behavioral health professionals' perspectives of patient-controlled granular information sharing (PC-GIS).

Objective: This study aimed to assess behavioral health professionals' (1) understanding of and opinions about PC-GIS; (2) accuracy in assessing redacted medical information; (3) reactions to patient rationale for health data categorization, assignment of sensitivity, and sharing choices; and (4) recommendations to improve PC-GIS.

Methods: Four 2-hour focus groups and pre- and postsurveys were conducted at 2 facilities. During the focus groups, outcomes from a previous study on patients' choices for medical record sharing were discussed. Thematic analysis was applied to focus group transcripts to address study objectives.

Results: A total of 28 health professionals were recruited. Over half (14/25, 56%) were unaware or provided incorrect definitions of granular information sharing. After PC-GIS was explained, all professionals demonstrated understanding of the terminology and process. Most (26/32 codes, 81%) recognized that key medical data had been redacted from the study case. A majority (41/62 codes, 66%) found the patient rationale for categorization and data sharing choices to be unclear. Finally, education and other approaches to inform and engage patients in granular information sharing were recommended.

Conclusions: This study provides detailed insights from behavioral health professionals on granular information sharing. Outcomes will inform the development, deployment, and evaluation of an electronic consent tool for granular health data sharing.

(JMIR Ment Health 2022;9(4):e18792) doi:10.2196/18792

KEYWORDS

behavioral health; patient information; granular information; electronic health record; integrated health care; electronic consent tool

Introduction

Patient-directed selection and sharing of health information "granules," that is, the lowest level of health information considered significant (eg, diagnoses, laboratory results, medications), is known as patient-controlled granular information sharing (PC-GIS) [1-6]. With the growth of integrated physical and behavioral health care and its reliance on health data sharing, the Office of the National Coordinator for Health Information Technology has promulgated recommendations for PC-GIS by suggesting the implementation of electronic consent tools [5,6]. This model permits the selection and sharing of health information "granules" with patient-specified institutions or personnel for distinct purposes [7] and creates a foundation of trust and transparency among patients, providers, and data stewards [6,8]. While rudimentary ethical guidelines for PC-GIS exist, more comprehensive research is needed [9-11] to harmonize health professionals' needs with patient choice in an electronically mediated data segmentation environment and, ultimately, a PC-GIS tool.

Literature regarding PC-GIS and granular consent has predominantly focused on the patient perspective [4,7,12]. Outcomes from these studies show that patients view PC-GIS and granular consent positively, especially with respect to sensitive electronic health record (EHR) information [7]. Of the PC-GIS studies, few use actual patient EHR data [7,12-15]. In a 2015 study, Schwartz et al [16] observed how patients enact granular control over their EHR data in a primary care setting. Of the 105 participants given granular sharing control, 45 (43%) chose to limit access to at least one professional, with varying preferences on the type of information restricted. While patients viewed record-sharing control positively (94.3%), they believed that such control may affect their relationship with health professionals (48.6%) [16,17]. Neves et al [15] and Papoutsi et al [18] studied EHR information sharing for research purposes and noted that while patients appear to be concerned about the security of EHR data when shared, health professionals—perhaps incorrectly—worry about patients' perceived unwillingness to share EHR information.

A 2020 systematic literature review by Soni et al [19] found only 8 peer-reviewed articles on PC-GIS and only 1 that considered the health professional perspective [20]. To understand primary care professionals' responses and perceptions, Tierney et al [20] supplemented the study by Schwartz et al [16] by reporting professionals' perceptions and the frequency with which these professionals "broke the glass," or overrode patients' sharing preferences, to access additional patient information. The health professionals in this study were based in a general internal medicine clinic—8 physicians, 4 clinical nurse assistants, 3 physician assistants, 2 nurse practitioners, 5 nurses, and 9 medical assistants. The 31 participating health professionals "broke the glass" 102 times, and 90% of these instances were for patients not enrolled in the study [20]. Professionals "broke the glass" for 14% of the total study patients but never "broke the glass" for patients who did not redact information [20]. Of the 24 professionals who responded to their poststudy survey, 63% responded "strongly agree" to the statement "restricting access to all or part of a patient's EHR will likely reduce the quality of care I deliver" while agreeing that patients should have such control [20]. Although Schwartz et al [16] and Tierney et al [20] provided patient and health professionals' survey responses, they did not provide patient rationale and included minimal provider rationale for the PC-GIS choices.

Prior literature shows that patients may restrict access to potentially sensitive health data because of stigma or fear of discrimination [21-23]. The 2020 research by Soni et al [19,24] reported a mixed-method approach using patients' own EHR information to assess preferences for PC-GIS in behavioral health care settings. The study outlined a card-sorting, semi-structured interview methodology of asking 25 English- and Spanish-speaking patients who were diagnosed as having general behavioral health (GBH) disorder and serious mental illness (SMI) from 2 integrated care clinics to categorize 30 items from their own EHRs as sensitive or nonsensitive. Nonsensitive data included all general physical health items, while the sensitive data groups were based on the Substance Abuse and Mental Health Services Administration sensitivity groups: alcohol use and alcoholism, communicable diseases, drug abuse, genetic information, mental health, other addictions, and sexual and reproductive health [19,24,25]. Participants were asked to classify the 30 items by sensitivity, that is, not sensitive, somewhat sensitive, sensitive, and then to exercise PC-GIS. Participants considered mental health (76%), sexual and reproductive health (75%), and alcohol use (50%) as sensitive categories. They were willing to share items related to other addictions (100%), genetic data (95.8%), and nonsensitive information (90.5%) [19]. Sharing preferences and sensitivity classifications did not significantly correlate [19]. Further, the study showed that participants' understanding and views on sensitivity, categorization, and sharing of information were diverse and that such diversity could impact use of an electronic consent tool [24]. Participants' personal circumstances impacted their sensitivity classifications and sharing preferences, but classification and sharing were approached independently [19,24].

Participants from the study conducted by Soni et al [19,24] and Schwartz et al [16] chose to share their data with all (48%) or some health professionals (52%). Health professionals included primary care providers, specialty care physicians, pharmacists, nurses, case managers, counselors, and medical assistants. The researchers suggested that a study focusing on professionals' perceptions of PC-GIS would enrich their findings and further inform understanding of the elements needed to support PC-GIS.

It is critical to understand professionals' perspectives to develop granular consent systems that balance patient desires with the information needs of health professionals. This study focuses on health professionals employed by the integrated care clinics used by the study conducted by Soni et al [19,24] and uses the study's results in the focus group design [20].

Our study used qualitative data analysis to gain insight into health professionals' perspectives of PC-GIS, specifically the (1) understanding of and opinions about PC-GIS; (2) accuracy in assessing patient-directed redaction of medical information; (3) reactions to patient rationale for health data categorization,

assignment of sensitivity, and sharing choices; and (4) recommendations to improve PC-GIS.

Results and recommendations from our study and Soni et al's research [19,24] will inform the development of a PC-GIS tool, My Data Choices, inspired by similar technology developed by the Substance Abuse and Mental Health Services Administration [25]. My Data Choices will be pilot tested at the same integrated care settings at which this study was conducted.

Methods

Clinical Settings

Study data were collected at 2 integrated care clinics [19,24]. One facility focuses on caring for patients diagnosed as having GBH disorder, and the other facility specializes in caring for those diagnosed as having SMIs. These facilities will be referred to as GBH Facility and SMI Facility throughout the manuscript. The GBH Facility staff had 85% nonprescribers and 15% prescribers, while the SMI Facility staff had 90% nonprescribers and 10% prescribers. Prescriber designation is determined by the Secretary of Labor definition [26,27]. Participants were selected to achieve representative samples at each facility.

Participants

The study was approved by the Arizona State University Institutional Review Board (#00010309). Participants who spoke English, were 21 years or older, worked closely with patients (primary care providers, psychiatrists, nurses, case managers, etc), and are currently or were recently involved in the previous year in consent processes at the 2 facilities were included. The SMI Facility participants were self-selected through a flyer distributed by a facility representative. A representative sample of professionals was solicited at each facility for each focus group. All participants received a $75 gift card as compensation on completing at least an hour and a half of the 2-hour focus group.

Focus Groups

Four 2-hour focus groups were conducted and audio recorded at each facility. Each focus group comprised 7 health professionals. Pre- and postsurveys, adapted from Tierney et al [20], were administered. The 6-section format of the focus groups and corresponding questions are illustrated in Figure 1. Content included PC-GIS didactics and an actual patient case from the Soni et al study [19,24] (Figure 2). In the example case, the patient chose to share information related to depression and diabetes with hospital physicians and the primary care doctor but shared only diabetes information with their dentist. The first author led the focus groups, and 2 observing researchers documented visual and verbal information to ensure all information from focus groups was captured in the final analysis.

Figure 1. Focus group flow by section (1-6) with corresponding target concepts for each section.

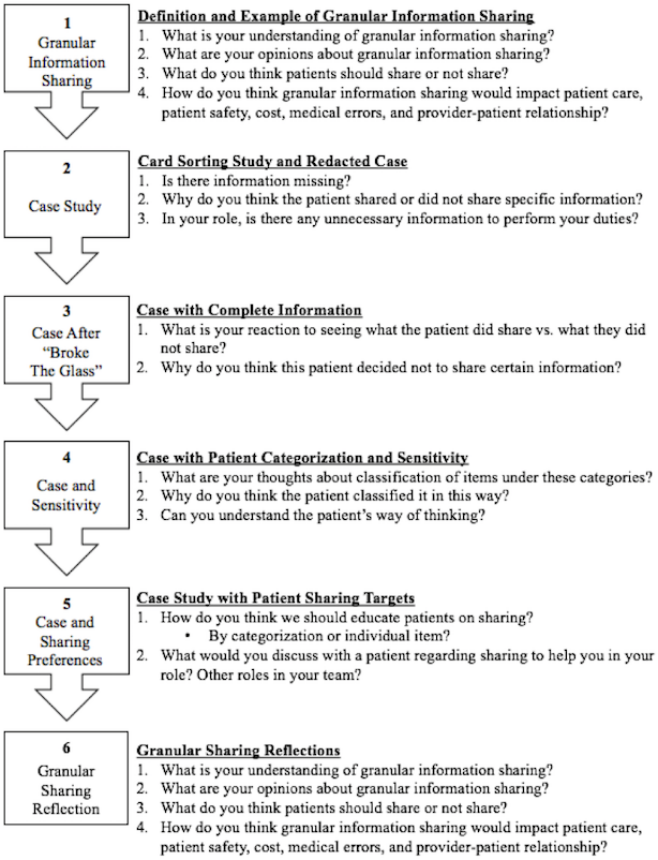
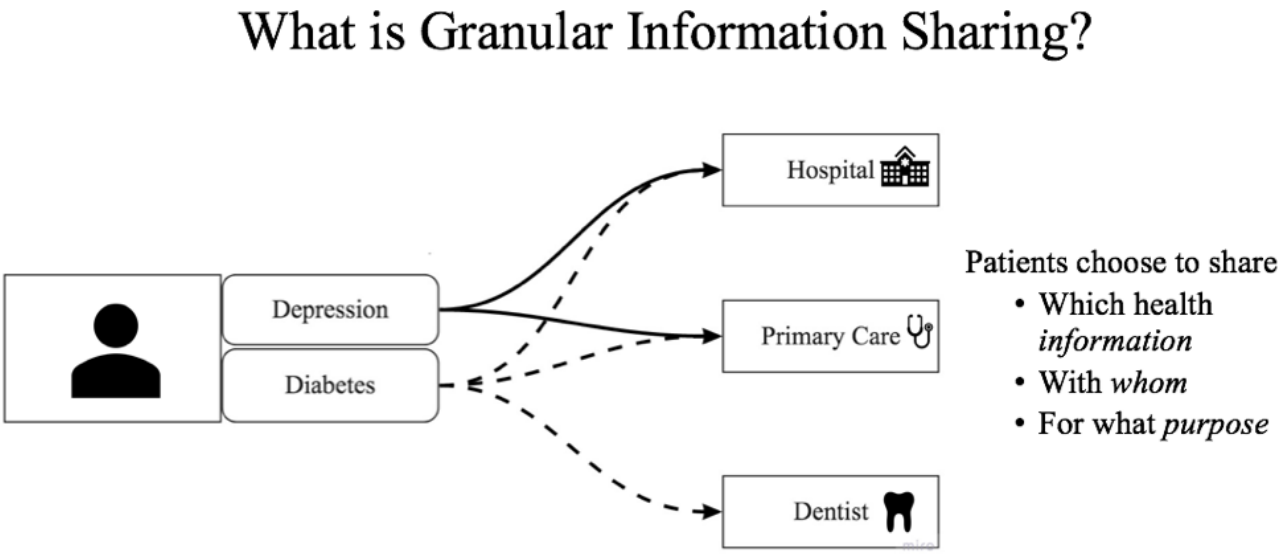


Figure 2. Example used to explain patient-controlled granular information sharing.



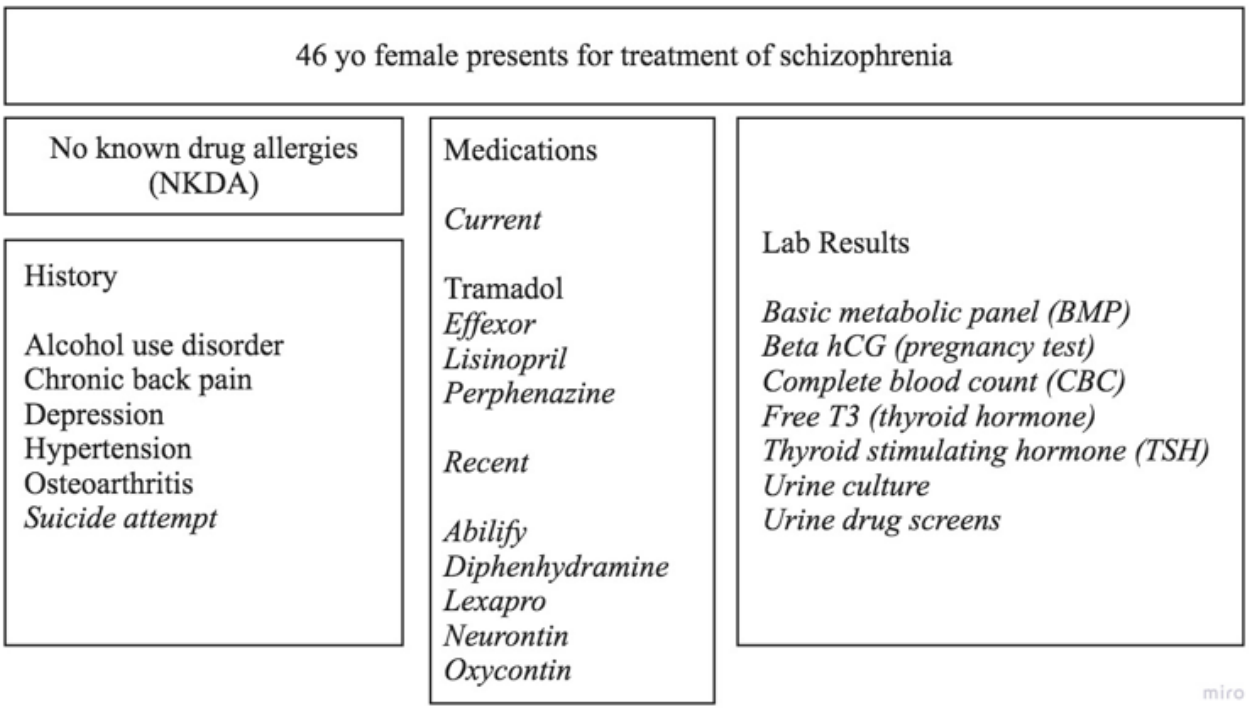
Section 1

Baseline comprehension of granular information sharing was assessed at the start of the focus group (question 1). “Granules” was later defined as “the lowest level of health information considered significant, such as diagnoses, laboratory results, or medications” [1-6] and granular information sharing occurs “when a patient identifies specific health information (granules) to share with or withhold from, specific professionals, entities or organizations, and directs how that information will be used” [2-6,8,27]. Figure 2 was discussed, followed by questions 2-4.

Section 2

The methods and results from the Soni et al study [19,24] were used to demonstrate how actual patients exercised granular information sharing, and a patient-redacted case was presented without disclosing that redaction had occurred (Figure 3). A singular case study was chosen to allow participants enough time to evaluate all aspects of PC-GIS. The specific case study was chosen as an example of how patients who did choose to restrict some information utilized PC-GIS in the Soni et al study [19,24].

Figure 3. Case study from Soni et al [19,24] presented with PC-GIS redaction (Section 2) followed by “breaking the glass” (Section 3), which is simulated by revealing the previously hidden items denoted by italics. PC-GIS: patient-controlled granular information sharing; hCG: human chorionic gonadotropin; T3: triiodothyronine; yo: year-old.



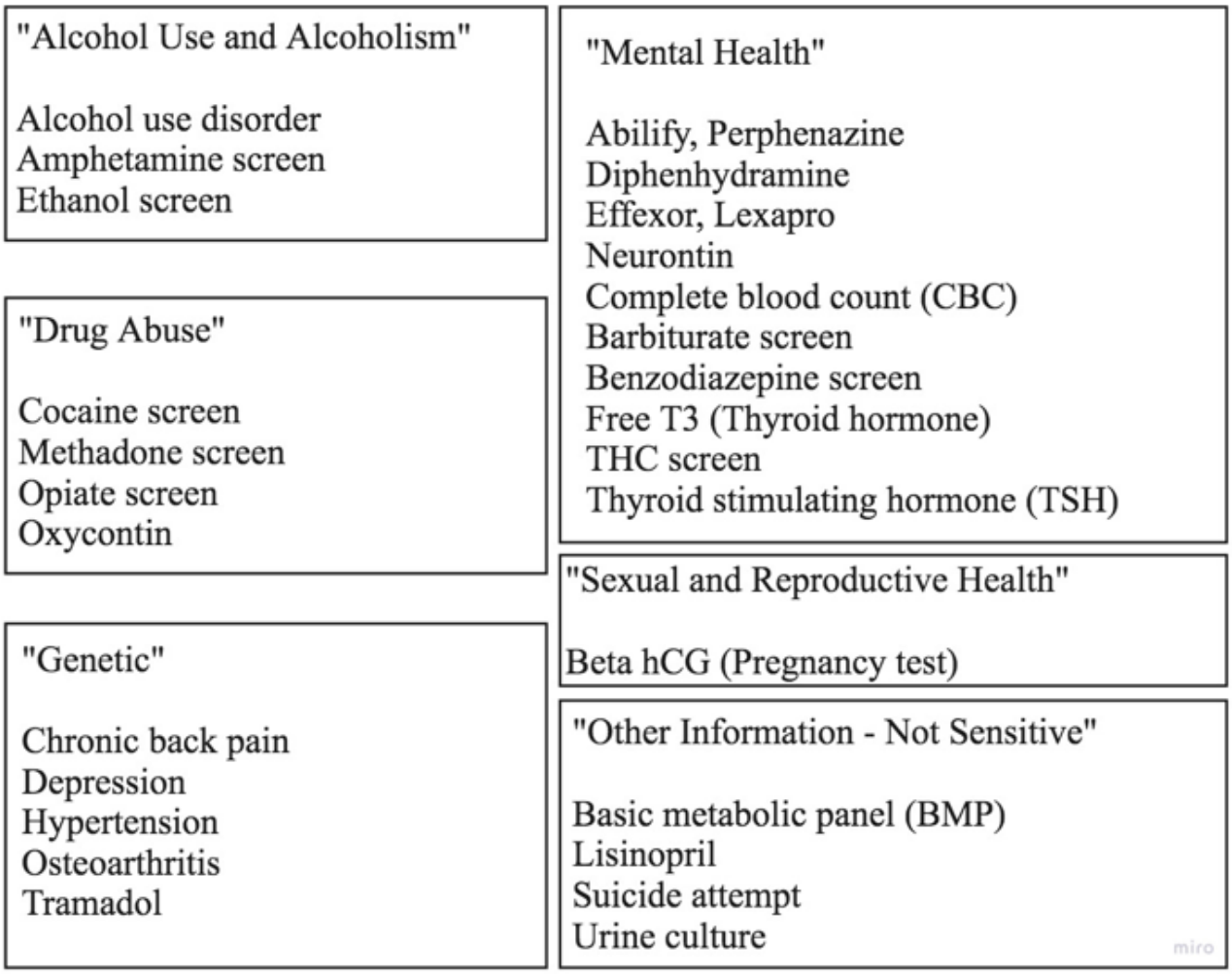
Section 3

The patient case was presented again, this time with the previously redacted health data items [24], and possible patient motivation for withholding this data was discussed (Figure 3).

Section 4

Data categorization and sensitivity classifications by patients were shared [19,24]. The discussion centered on patient rationale for the presented choices (Figure 4).

Figure 4. Classification of data as sensitive and categorization of health information by patients from the Soni et al case study [19,24] presented after the redaction is revealed (Sections 4 and 5). hCG: human chorionic gonadotropin; THC: tetrahydrocannabinol; T3: triiodothyronine.



Section 5

The actual published patient rationale for sharing or withholding specific health data items was presented and discussed. Additional details from the paper on data categorizations, sensitivity classifications, and sharing preferences were provided if requested [19,24].

Section 6

After they reflected on the prior sections and the hypothetical tool presented, participants were asked to share opinions and to make recommendations to those seeking to implement PC-GIS.

Analysis

Focus group audio recordings were transcribed using Transcribe (Wreally LLC) and then sequentially screened by 3 researchers for accuracy. Notes from onsite researchers, that is, notes on nodding and facial expressions, were added to the transcripts. Transcriptions were analyzed using Braun and Clarke's thematic analysis guidelines and anthropological methodology [28,29]. Six iterations of digitally assisted coding (MAXQDA, VERBI GmbH) were performed. The unit of analysis was an individual participant's statement in paragraph form, and themes were identified through repetition and frequency. One researcher coded for and defined emerging themes into a codebook. The output was iteratively revised by 3 researchers and organized according to the focus group segments, as represented in Figure 1. Although the focus group leader prevented dominance by a single participant, transcript codings were calculated per participant to determine whether egalitarian engagement was maintained. Each focus group participant was coded as a separate entity, thus any codes attributed to them could be measured. At conclusion of coding, participants' actual attributed codes were juxtaposed to the expected number of codes per person for each focus group. Coding assessments provided qualitative and quantitative insight into participant rationale for PC-GIS opinions with outcomes. Pre- and postsurvey analysis is described in part 1 of this study [30].

Results

Demographics

The participant group included 23/28 (82%) nonprescribers and 5/28 (18%) prescribers (physicians and nurse practitioners) based on prescriber criteria (Table 1) [26,27]. All participants (28/28, 100%) completed the presurvey and 27/28 (96%) completed the postsurvey (1 participant was on call and departed prior to focus group conclusion).

While the focus groups each had 1 participant who engaged more than anyone else (outside the group average), these participants were individuals who had administration or managerial duties and helped facilitate the discussion of the rest of the group.

Table 1. Health professionals' roles and representation (N=28).

Role type	Professionals, n	GBH[a] facility, n	SMI[b] facility, n
Nonprescribers			
Counselors	4	3	2
Nurses	3	1	2
Rehabilitation specialist	3	1	2
Case managers	3	1	2
Clinical coordinators	3	3	0
Administrators	3	0	3
Peer mentors	1	0	1
Medical assistants	1	0	1
Discharge planners	1	0	1
Social workers	1	1	0
Prescribers			
Physicians	3	2	1
Nurse practitioners	2	2	0

[a]GBH: general behavioral health.

[b]SMI: serious mental illness.

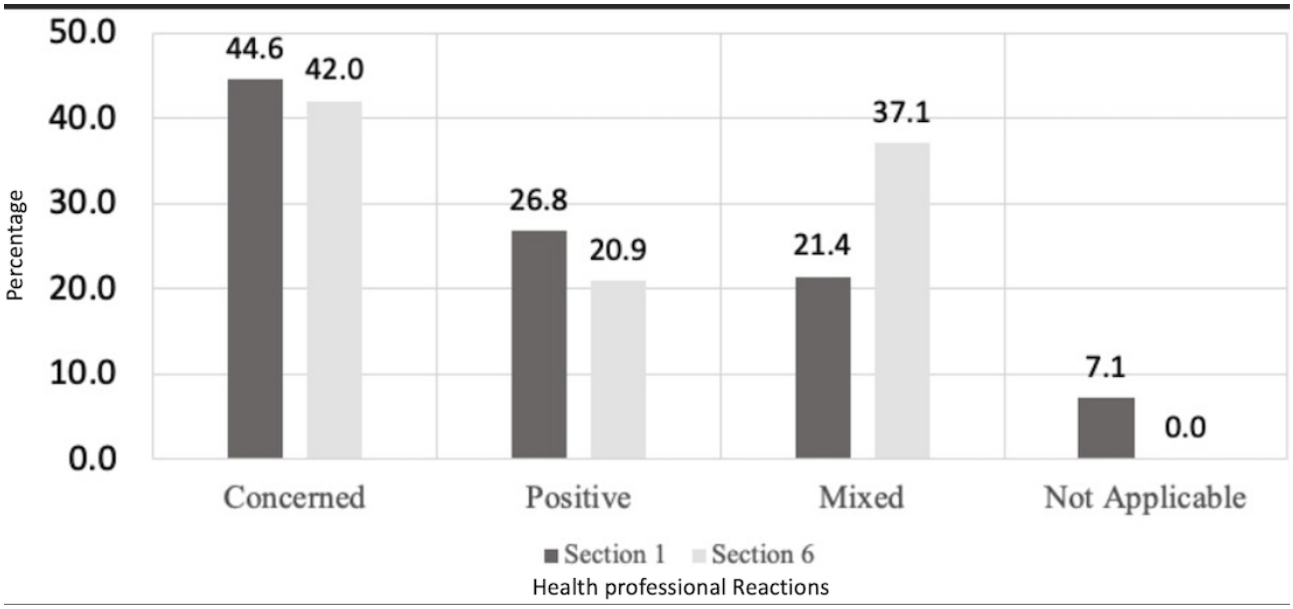
Health Professionals' Understanding and Opinions of Granular Information and Sharing

Thematic analysis coding of individual participant contributions demonstrated overall absence of single participant dominance regarding codes. Thematic analysis results of health professionals' knowledge and perceptions of PC-GIS before (section 1) and after (section 6) the focus group were compared. Prompts from these sections were assessed for comprehension of the granular information concept (unaware of, correct, or incorrect definition) and for opinions (positive, negative, mixed, or not applicable) on how PC-GIS could impact care delivery.

Baseline comprehension of PC-GIS yielded 25 relevant codings in section 1 (Multimedia Appendix 1). Just over half (14/25, 56%) of the participants were unaware of or provided incorrect definitions of granular information sharing: "I don't have any understanding of it." After a brief explanation (Figure 2), participants demonstrated 100% (15 codings) comprehension of PC-GIS. By conclusion of the focus groups (section 6), all participants demonstrated excellent comprehension of PC-GIS (100%, 8 codings) with nuanced discussion (35 codings; Multimedia Appendix 1) of concerned, positive, and mixed opinions.

There was a visible change from the initial to the concluding reactions to PC-GIS within focus groups with respect to mixed opinion (Figure 5). For the "not applicable" codings from section 1, participants expressed neither a definition nor an opinion. They did discuss foreign language and education of patients as issues in granular information sharing. Positive and mixed outlooks focused on patients' rights and choices, alleviation of stigma, rationalization of patients' choices, and streamlining of communication. When asked to imagine PC-GIS from the patient perspective, most participants (20/22, 91% of instances attributed to positive codings) expressed a more supportive view of granular sharing.

Figure 5. Initial and concluding reactions from the focus groups to patient-controlled granular information sharing (Sections 1 and 6, compared). Numbers signify percentage per category. Rounded data do not always add up to 100.



Patient safety and the ability to provide successful treatment were frequently cited (55/99, 56% of codings from sections 1 and 6) in the discussion about patient-directed redaction:

> *It can be life threatening. There are some medications that can be prescribed that if you combine like an MAO inhibitor with a pain medication, it can be fatal…you [patient] may not think it's important, but you're not really trained, you don't have the background to know what's important per se or what isn't. It is difficult to help your patient understand that, a lot of education. (Agreement throughout).*

Concerns about patients' rights and patients' perspectives were raised in both sections 1 and 6, as were complexities of implementing a PC-GIS tool that adequately balanced patients' rights with health professionals' needs. The transition in perception from sections 1 to 6 was striking, with an increase in the proportion of participants with mixed opinions.

> *I actually have a different opinion [than] when I started. I mean, it depends on what angle you're coming at. It really does because if you're a patient, I get it. I get it why you don't want to share everything. But from a provider nurse standpoint, safety becomes a factor. And that's when more information should be shared as much as possible.*

Another participant noted that a granular sharing tool may help reduce the cost of care, if appropriate sharing was enacted. While participants with concerned opinions prioritized health professionals' perspectives, participants with mixed opinions weighed patient and health professionals' perspectives equally.

Health Professionals' Assessments of Patient Health Information Redaction

Section 2 focused on health professionals' recognition of record redaction (32 codings; Multimedia Appendix 1). Most participants (26/32, 81%) accurately identified that some information was missing from the presented health record. Still, 19% (6/32) of codings showed that participants were unaware (3/32, 9%) of the redaction or were uncertain (3/32, 9%) if the information was complete; they explained that they are accustomed to working with incomplete or absent patient information.

All participants responded with varying concern about missing information (Table 2). They identified the data categories that were redacted or incomplete and, in actual practice, would discuss suspected redaction or withholding of information with the patient (51 total codings). Specifically, the patients' discussion (27 codings) with the participants (6/27, 22% of patient discussions) would include rationale for the information: "Tell them [patients] the reason why we're asking, the importance of it, and to help them understand why we need the information." Others (3/27, 13% of patient discussions) included the suggestion that use of standard facility procedures (eg, intake questionnaires) could aid in finding the needed information: "…As a case manager, we do a comprehensive assessment...hey, you don't have to be honest but at least I asked."

Table 2. Health professionals' reactions after redacted material was revealed (N=51; section 2); rounded data do not always add up to 100.

Code	Coding, n (%)	Exemplar quotes
Conducted patient discussion	27 (52)	"I would just be honest and say, so I see that you have a diagnosis here that you're presenting for treatment of schizophrenia but I don't see that you're currently prescribed an anti-psychotic. Are you currently taking one? Have you taken one in the past? Did we possibly forget to list any current medications that you may have forgotten? I would just if it were me, address it pretty upfront." [nonprescriber]
Expressed concern	18 (35)	"Because if she's using, yeah, if she's using both then that could be potentially deadly." [prescriber]
Expressed need for information	6 (11)	"But yeah, there's a lot of information missing. I would want a more complete social, family history, hospitalization history, past medications. It's again, we don't know what's worked, what hasn't worked and we're just kind of now starting from scratch again…if that's all that's there, it's not enough to move forward with treatment without more information." [prescriber]

Health Professionals' Perceptions of Patient Rationale for Sharing Decisions

In sections 3 and 4, we assessed participants' reactions to patient sharing rationale. Participants (48 codings; Multimedia Appendix 1) reacted to the redacted content revelation by expressing the need to know this information (33/48, 68%), surprise (9/48, 18%), and no surprise (6/48, 12%). While many participants noted that they would question patients or use alternate sources to gain need-to-know information, they were skeptical that all the information necessary for optimal and safe treatment of this patient had been identified: "As the therapists doing an assessment, we're not necessarily going to even get to a question that hits on all of the panels that are missing." Similarly, participants (3/9, 33% of those surprised) also noted that the data categorizations chosen by the patient "did not make sense." They were also surprised by how the patient applied these categories to make sharing decisions.

When participants were asked to postulate the patient sharing rationale (Table 3), the results (30 codings) coalesced around stigma and fears, purposeful omission, consideration of data to be irrelevant, lack of clarity on the information that needs to be shared, and symptoms. We then asked the participants to react to the data sensitivity classification assigned by patients (section 4). Participants registered 3 main types of reactions (62 codings): did not understand (41/62, 66%), considered patient incorrect (15/62, 24%), and considered patient correct (6/62, 9%). Of note, all participants who reacted positively to some patient classifications did not agree with the patient's sharing decisions and found the documented patient rationale helpful in understanding the patient's decisions.

Patient data self-categorization using the 6 data categories (77 total codings; Multimedia Appendix 1) resulted in genetic (17/77, 22%), mental health (14/77, 18%), drug use (9/77, 11%), alcohol use (9/77, 11%), sexual and reproductive health (3/77, 3%), and other information (25/77, 32%). Participants focused on the category "other information" that included the topic of attempted suicide (22/25, 88% of all "other information" discussions). Participants found the actual patient explanation for classifying the suicide attempt item into "other information" rather than "mental health" to be particularly relevant: "took a whole bunch of pills" [19]. Participants considered that the patient thinks "they're fixed" (nonprescriber) or "they don't have a good understanding of what mental health is" (nonprescriber). Participants were provided detailed patient rationale as requested, including explanations from other patients in the Soni et al study [24].

Table 3. Health professionals' rationale of the patient's decision to redact (N=30; section 3); rounded data do not always add up to 100.

Code	Coding, n (%)	Exemplar quotes
Stigma and fear	13 (43)	"I don't know the culture of this client, but culturally they might be thinking like, "This person thinks I'm crazy or people will think I'm crazy because I take medication so I'm just not going to say anything." Particularly if it's a court-ordered client, they may be sharing less because they just want to get their mandates over with and get out of services. And the more they share could keep them wrapped up in services for longer than they want." [nonprescriber; nonprescribers nod in agreement]
Purposeful omission	7 (23)	"Well, I'm just saying in general, if I go to the PCP, I'm going for one thing, I don't need 50 other things added on to what I came here for. So, maybe they're just shutting it down. And like, look, this is what I'm here for and this is what I'm giving you." [nonprescriber]
Patient considered data irrelevant	5 (16)	"Or is it with the one-time [suicide] attempt, it really didn't mean nothing. I didn't really want to do it, so I'm okay now. So, it's not important to me. It's not relevant to them." [nonprescriber]
Patient lacks clarity on the information that needs to be shared	3 (10)	"Like six months into treatment, they suddenly randomly talk about a shopping addiction or something like that that they just never mentioned. And so, I'm sure there's some things that they don't realize are important to share [with us]." [nonprescriber]
Symptoms	2 (6)	"There's the possibility that they're not 100% compliant with their medication because again, there's a lot of side effects from medications. And I'm not seeing side effects of medication being prescribed and then there's the drug screen, so we don't know how much the person's self-medicating and taking their meds. So, they may be more symptomatic hence could be more paranoid about sharing the information. So, I'd want to rule that out as well. How symptomatic are they at that particular moment, you know?" [nonprescriber]

Health Professionals' Recommendations to Improve PC-GIS

Participants reacted to the description of a hypothetical PC-GIS tool based on the patient case in section 5 (Table 4). They discussed how a patient should be advised to share certain information (39 total codings). Most participants indicated that ensuring transparency, promoting trust, and fostering understanding are key factors for successful PC-GIS. This might be enhanced by periodical review of sharing decisions in a meeting with adequate time for questions and discussion with the provider.

Participants also addressed the need to simplify education material (4/39, 10%), create role-specific information (3/39, 7%), and use examples (2/39, 5%; Multimedia Appendix 1). Behavioral health professionals stressed that additional resources or a different approach may be needed to elicit informed consent for record sharing from a competent patient with active psychiatric symptoms. Participants also acknowledged that their differing roles and professional preparation necessitate the use of targeted materials to support specific sharing discussions. Finally, they underscored the importance of motivating patients to engage in the data sharing process and to understand its impact on safety.

Table 4. Health professionals' recommendations to improve granular information sharing (N=39; section 5); rounded data do not always add up to 100.

Code	Coding, n (%)	Exemplar quotes
Promote trust and understanding	23 (60)	"Absolutely interesting because again, the client isn't sharing information about their mental health with the people who are designated to help them with their mental health. So again, if that's the theme then trying to (A) understand what is the motivation for that and (B) is there something that can be done to assist with building some trust? If that's in some way, you know, if they don't trust the system or whatever it may be or they're symptomatic, how can we kind of overcome that barrier in order to get that client's unique needs met?" [nonprescriber]
Other	7 (17)	"I would use a similar grid like that grading, because at a glance, you could introduce something every three months, any updates. Are you still sharing with your pharmacist? Are you still sharing with your own specialty care providers, etc.? Have you mentioned that you have an upcoming appointment with PCP? And a bit something of an alert, definitely, you need to work with the team and send an email." [nonprescriber]
Simplify education material	4 (10)	"Even having it written down, sometimes it might be too much for somebody who's having schizophrenia. If I'm hearing voices, I don't have the patience to sit down either listen or read something. I just want to get it done as soon as possible." [nonprescriber]
Provide role-specific information	3 (7)	"We also take time to educate because if we have to educate them on everything, there's thousands of topics to discuss, and we can't educate or try to educate on things that we're not competent in. So, I can't talk to them about medications. I won't [non-prescriber] because I can't. It's not ethical, and it's not a smart decision. So, you know, if they want the education, then they have to go see their doctor or their nurse practitioner, you know? And then it's just more steps. But if they're willing to do it, that's great. But they have to be motivated to do that." [nonprescriber]
Provide examples	2 (5)	"Give an example. Because someone with schizophrenia is not going to have the patience to sit there and listen to what each definition is and where it goes." [non-prescriber; agreement between nonprescribers and prescribers]

Discussion

Main Findings

Our analysis of behavioral health professionals' perceptions of PC-GIS between the start and end of the focus groups demonstrates a shift to mixed opinions from a position of less support (12/56, 21% to 13/35, 37%). Although the terminology and processes of PC-GIS were new to many professionals (14/25, 56%), all participants understood the concept, benefits, and risks associated with PC-GIS after a brief explanation. Additionally, the professionals correctly identified (26/32, 81%) that information was missing (patient-redacted) from the case presentation, with a majority (33/48, 68%) noting that the missing information was necessary for the successful care of the patient. Professionals were perplexed about many patient categorization and sharing decisions (41/62, 66%) and often expressed surprise when the patient rationale for withholding information was shared. Participants' concerns led to a general commitment to improve the consent process with specific recommendations.

The literature suggests that adoption of EHR and health information exchange has accelerated the importance of consent technology. Emanuel and Emanuel [31] noted that patients and professionals, the key stakeholders, must be part of any process change involving the fine balance between care delivery and individual rights. Trust and transparency are key factors in this delicate relationship [14,27,32-36]. Our results support and expand these findings; namely, the health professionals have defined challenges in PC-GIS implementation (28/35, 79% of final concerned and mixed opinions), attempted to understand patient motivation for redaction, aimed for balance between patient and professionals' needs, and underscored the need to access necessary health information for successful care delivery (23/39, 60% of all recommendations). The pre- and postsurvey analysis (described separately) [28] shows significant changes (P<.05) in health professionals' opinions toward concern after the focus group [30]. Our qualitative analysis mirrors and provides insight into the increase of mixed opinions with comprehension of PC-GIS as well as recommendations to mitigate mutual concerns related to PC-GIS and avoid friction in the patient-professional relationship.

While there is research on health professionals' understanding of granular information control, there are no studies measuring baseline PC-GIS knowledge or the effectiveness of an intervention to enhance that knowledge [20,27,35,37]. In our study, behavioral health professionals (14/25, 56%) were either unaware of or provided incorrect definitions of PC-GIS. A brief explanation and example (Figure 2) resulted in 100% (15/15) comprehension of the terms and process. Implementation of PC-GIS requires education of health professionals, and we demonstrated that this can be accomplished with a brief explanation.

Prior literature has focused on whether and how professionals should be notified about patient redactions [9]. The results of the study by Tierney et al [16,20] were inconclusive on provider awareness of redacted information but showed that professionals did not "break the glass" for any of the patients who chose to share all information. In our study, most participants (26/32, 81%) correctly recognized the absence of some data during the case exercise. Therefore, our results suggest that most health professionals may independently surmise that the available data are incomplete; they mentioned that they routinely evaluate

patients with incomplete or fragmented records. When redaction or withholding of information is suspected, the health professionals agreed that such information gaps should prompt timely, directed patient discussion. They further noted that patients place themselves at risk for suboptimal treatment and even injury when health data are incomplete at the point of care.

We asked participants to consider information sharing from the patient perspective. Our study compared the health professionals' postulated rationales with those provided by patients in the Soni et al study [24]. Of the 32 sensitive-information codings from Soni et al [24], half of the patients (50%) said that stigma and fear of discrimination were the reasons for their classifications [19,24]. Our participants agreed that stigma and fear (13/30, 43%) were driving forces for patient data redaction. Soni et al [24] reported that some patients conflate sharing categories with sensitivity by classifying and sharing based not only on comprehension but also on perceived applicability to their own health history. In our study, participants' insights mirrored patient justifications for purposeful omissions (7/30, 23%), omissions for irrelevancy (5/30, 17%), and omissions because of a lack of clarity on the information that needs to be shared (3/30, 10%). Further, participants did not understand (41/62, 66%) the rationale used by the patient to classify and make sharing choices. When provided with patient rationale from the Soni et al study, the health professionals suggested specific techniques for interacting with patients who had unclear or missing health information. Overall, the health professionals were concerned about patients categorizing and classifying information in a manner incongruent to bioscience and health professionals' reasoning.

Recommendations for designing PC-GIS tool features comprised the final segment of the focus groups. Participants emphasized that targeted PC-GIS education for behavioral health and integrated care settings must be appropriate for individuals with lower health literacy (4/39, 10%), permit adaption by profession (3/39, 8%), and contain pertinent, patient-friendly examples (2/39, 5%).

Also important for PC-GIS tool design is consideration of patient and provider time constraints, including the resources needed to engage in face-to-face communication. Most PC-GIS studies employed a "break the glass" option for health professionals to access needed information [16,20]. Our participants expressed the need for such a feature. They also suggested periodic meetings with a trusted care professional and creation of an electronic algorithm within the PC-GIS tool to help patients classify and select information for sharing. Feedback from frontline professionals is integral to the development of a digital tool for informed PC-GIS.

Limitations

The study included a limited number of health professionals, particularly prescribers. Therefore, we cannot comment on differences between prescribers and nonprescribers. However, participant composition in this study reflected the facilities' overall prescriber-nonprescriber ratios and was representative of role demographics for the respective professional population. The resulting demographics of the focus groups for the facilities showed a diverse representation of the entire care team that a patient relies on within integrated care. Focus groups comprised individuals who were actively engaged in the care delivery team. We recommend that future studies consider the prescriber and nonprescriber aspects that may impact the success of a PC-GIS tool.

While this study used a single representative patient case, sharing of the overall results of the Soni et al study [19,24] provided additional context and insight into how other patients with GBH disorder and SMI categorized, classified, and rationalized their decisions.

Focus groups may include participants that tend to dominate a discussion, especially in the context of existing workplace hierarchy. Efforts were made to avoid this situation. Moreover, focus group codes were evaluated per participant and demonstrated no such effects to significantly skew discussions.

Health professional roles (eg, prescriber, nonprescriber) and patient population (eg, GBH, SMI) may impact the interpretation of and opinions about granular information sharing, its potential impact on care, and how best to provide informed consent [27]. Published literature supports the rapidly evolving trend toward integrated care coupled with the need to improve digital sharing processes [38]. This study provides context and recommendations to help achieve this goal.

Future Research

Our results are being incorporated in the design and deployment of a PC-GIS tool, My Data Choices. The participants' recommendations are also being used to develop patient education for the My Data Choices pilot to be launched in several integrated care clinics. The My Data Choices study team is also investigating the impact of trust on PC-GIS.

Conclusions

This study provides detailed insights from behavioral health professionals about granular information sharing, explores scenarios where patients exercise granular consent choices, and includes suggestions to improve patient education and the consent sharing process. The case-based learning intervention during the focus group improved provider comprehension of PC-GIS terminology and process. Health professionals accurately identified the presence of patient-redacted information gaps and provided concrete recommendations to help patients appreciate the risks and benefits associated with PC-GIS. Outcomes of this study are guiding the development, deployment, and evaluation of an electronic granular consent tool.

Acknowledgments

This work was supported by the National Institute of Mental Health through My Data Choices, evaluation of effective consent strategies for patients with behavioral health conditions (R01 MH108992) grant.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Codes from specific sections of the focus groups by code, number of codes, and percentage of codes per section, and example quotes from participants.
[DOCX File , 22 KB - mental_v9i4e18792_app1.docx]

References

1. Roth C, Lim Y, Pevnick J, Asch S, McGlynn E. The challenge of measuring quality of care from the electronic health record. Am J Med Qual 2009;24(5):385-394 [FREE Full text] [doi: 10.1177/1062860609336627] [Medline: 19482968]
2. Saks M, Grando A, Murcko A, Millea C. Granular patient control of personal health information: federal and state law considerations. Jurimetrics 2018;58(4):411-435 [FREE Full text] [Medline: 31798215]
3. The Office of the National Coordinator for Health Information Technology. About ONC. HealthIT.gov. URL: https://www.healthit.gov/topic/about-onc [accessed 2019-08-13]
4. Kim H, Bell E, Kim J, Sitapati A, Ramsdell J, Farcas C, et al. iCONCUR: informed consent for clinical data and bio-sample use for research. J Am Med Inform Assoc 2017 Mar 01;24(2):380-387 [FREE Full text] [doi: 10.1093/jamia/ocw115] [Medline: 27589942]
5. Commonwealth Fund. In Focus: Integrating Behavioral Health and Primary Care. The Commonwealth Fund. URL: https://www.commonwealthfund.org/publications/newsletter-article/2014/aug/focus-integrating-behavioral-health-and-primary-care [accessed 2019-03-19]
6. The Office of the National Coordinator for Health Information Technology. Recommendations to the National Coordinator for Health IT. HealthIT.gov. URL: https://www.healthit.gov/topic/federal-advisory-committees/recommendations-national-coordinator-health-it [accessed 2019-06-11]
7. Caine K, Hanania R. Patients want granular privacy control over health information in electronic medical records. J Am Med Inform Assoc 2013 Jan 01;20(1):7-15 [FREE Full text] [doi: 10.1136/amiajnl-2012-001023] [Medline: 23184192]
8. Grando A, Sottara D, Singh R, Murcko A, Soni H, Tang T, et al. Pilot evaluation of sensitive data segmentation technology for privacy. Int J Med Inform 2020 Jun;138:104121 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104121] [Medline: 32278288]
9. Meslin E, Alpert S, Carroll A, Odell J, Tierney W, Schwartz P. Giving patients granular control of personal health information: using an ethics 'Points to Consider' to inform informatics system designers. Int J Med Inform 2013 Dec;82(12):1136-1143 [FREE Full text] [doi: 10.1016/j.ijmedinf.2013.08.010] [Medline: 24139626]
10. Grando A, Schwab R. Building and evaluating an ontology-based tool for reasoning about consent permission. In: AMIA Annu Symp Proc. 2013 Nov 16 Presented at: AMIA Annual Symposium Proceedings Archive; November 16-20, 2013; Washington, DC p. 514-523 URL: http://europepmc.org/abstract/MED/24551354
11. Grando M, Boxwala A, Schwab R, Alipanah N. Ontological approach for the management of informed consent permissions. 2012 Presented at: IEEE Second International Conference on Healthcare Informatics, Imaging and Systems Biology; September 27 to 28, 2012; San Diego p. 51 URL: https://doi.org/10.1109/HISB.2012.19 [doi: 10.1109/hisb.2012.19]
12. Caine K, Kohn S, Lawrence C, Hanania R, Meslin E, Tierney W. Designing a patient-centered user interface for access decisions about EHR data: implications from patient interviews. J Gen Intern Med 2015 Jan;30 Suppl 1:S7-16 [FREE Full text] [doi: 10.1007/s11606-014-3049-9] [Medline: 25480719]
13. King T, Brankovic L, Gillard P. Perspectives of Australian adults about protecting the privacy of their health information in statistical databases. Int J Med Inform 2012 Apr;81(4):279-289 [FREE Full text] [doi: 10.1016/j.ijmedinf.2012.01.005] [Medline: 22306206]
14. Teixeira P, Gordon P, Camhi E, Bakken S. HIV patients' willingness to share personal health information electronically. Patient Educ Couns 2011 Aug;84(2):e9-12 [FREE Full text] [doi: 10.1016/j.pec.2010.07.013] [Medline: 20724095]
15. Neves A, Poovendran D, Freise L, Ghafur S, Flott K, Darzi A, et al. Health care professionals' perspectives on the secondary use of health records to improve quality and safety of care in England: qualitative study. J Med Internet Res 2019 Sep 26;21(9):e14135 [FREE Full text] [doi: 10.2196/14135] [Medline: 31573898]
16. Schwartz P, Caine K, Alpert S, Meslin E, Carroll A, Tierney W. Patient preferences in controlling access to their electronic health records: a prospective cohort study in primary care. J Gen Intern Med 2015 Jan;30 Suppl 1:S25-S30 [FREE Full text] [doi: 10.1007/s11606-014-3054-z] [Medline: 25480721]
17. Whiddett R, Hunter I, Engelbrecht J, Handy J. Patients' attitudes towards sharing their health information. Int J Med Inform 2006 Jul;75(7):530-541 [FREE Full text] [doi: 10.1016/j.ijmedinf.2005.08.009] [Medline: 16198142]
18. Papoutsi C, Reed J, Marston C, Lewis R, Majeed A, Bell D. Patient and public views about the security and privacy of Electronic Health Records (EHRs) in the UK: results from a mixed methods study. BMC Med Inform Decis Mak 2015 Oct 14;15:86 [FREE Full text] [doi: 10.1186/s12911-015-0202-2] [Medline: 26466787]

19. Soni H, Grando A, Murcko A, Diaz S, Mukundan M, Idouraine N, et al. State of the art and a mixed-method personalized approach to assess patient perceptions on medical record sharing and sensitivity. J Biomed Inform 2020 Jan;101:103338 [FREE Full text] [doi: 10.1016/j.jbi.2019.103338] [Medline: 31726102]
20. Tierney W, Alpert S, Byrket A, Caine K, Leventhal J, Meslin E, et al. Provider responses to patients controlling access to their electronic health records: a prospective cohort study in primary care. J Gen Intern Med 2015 Jan;30 Suppl 1:S31-S37 [FREE Full text] [doi: 10.1007/s11606-014-3053-0] [Medline: 25480720]
21. Lamberg L. Confidentiality and privacy of electronic medical records: psychiatrists explore risks of the "information age". JAMA 2001 Jun 27;285(24):3075-3076 [FREE Full text] [doi: 10.1001/jama.285.24.3075] [Medline: 11427118]
22. Salomon R, Blackford J, Rosenbloom S, Seidel S, Clayton E, Dilts D, et al. Openness of patients' reporting with use of electronic records: psychiatric clinicians' views. J Am Med Inform Assoc 2010;17(1):54-60 [FREE Full text] [doi: 10.1197/jamia.M3341] [Medline: 20064802]
23. Institute of Medicine (US) Committee on Quality Assurance and Accreditation Guidelines for Managed Behavioral Health Care M, Edmunds M, Frank R, Hogan M. Challenges in Delivery of Behavioral Health Care. In: Managing Managed Care Quality: Improvement in Behavioral Health. Washington (DC): National Academies Press (US); 1997:3.
24. Soni H, Ivanova J, Grando A, Murcko A, Chern D, Dye C, et al. A pilot comparison of medical records sensitivity perspectives of patients with behavioral health conditions and healthcare providers. Health Informatics J 2021;27(2):14604582211009925 [FREE Full text] [doi: 10.1177/14604582211009925] [Medline: 33878989]
25. Jia Z. Consent2Share: Patient-driven information consent management. FEI Systems. URL: https://www.feisystems.com/solutions/behavioral-health/consent2share/ [accessed 2019-09-30]
26. Cornell Law. Definition of health care provider. LII / Legal Information Institute. URL: https://www.law.cornell.edu/cfr/text/29/825.125 [accessed 2020-02-13]
27. Grando A, Ivanova J, Hiestand M, Soni H, Murcko A, Saks M, et al. Mental health professional perspectives on health data sharing: Mixed methods study. Health Informatics J 2020 Sep;26(3):2067-2082 [FREE Full text] [doi: 10.1177/1460458219893848] [Medline: 31928295]
28. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101 [FREE Full text] [doi: 10.1191/1478088706qp063oa]
29. Bernard HR. Research Methods in Anthropology: Qualitative and Quantitative Approaches. Maryland, United States: Rowman Altamira, 2011; Dec 1996:91-92.
30. Ivanova J, Tang T, Idouraine N, Murcko A, Whitfield M, Dye C, et al. Behavioral Health Professionals' Perceptions on Patient-Controlled Granular Information Sharing (Part 1): Focus Group Study. JMIR Ment Health 2022;9(4):e21208 [FREE Full text] [doi: 10.2196/21208]
31. Emanuel E, Emanuel LL. Four models of the physician-patient relationship. JAMA 1992;267(16):2221-2226. [Medline: 1556799]
32. Esmaeilzadeh P. The impacts of the perceived transparency of privacy policies and trust in providers for building trust in health information exchange: empirical study. JMIR Med Inform 2019 Nov 26;7(4):e14050 [FREE Full text] [doi: 10.2196/14050] [Medline: 31769757]
33. Dang B, Westbrook R, Njue S, Giordano T. Building trust and rapport early in the new doctor-patient relationship: a longitudinal qualitative study. BMC Med Educ 2017 Feb 02;17(1):32 [FREE Full text] [doi: 10.1186/s12909-017-0868-5] [Medline: 28148254]
34. Guadagnolo B, Cina K, Koop D, Brunette D, Petereit D. A pre-post survey analysis of satisfaction with health care and medical mistrust after patient navigation for American Indian cancer patients. J Health Care Poor Underserved 2011 Nov;22(4):1331-1343 [FREE Full text] [doi: 10.1353/hpu.2011.0115] [Medline: 22080713]
35. Ivanova J, Grando A, Murcko A, Saks M, Whitfield M, Dye C, et al. Mental health professionals' perceptions on patients control of data sharing. Health Informatics J 2020 Sep;26(3):2011-2029 [FREE Full text] [doi: 10.1177/1460458219893845] [Medline: 31912744]
36. Abdelhamid M, Gaia J, Sanders G. Putting the focus back on the patient: how privacy concerns affect personal health information sharing intentions. J Med Internet Res 2017 Sep 13;19(9):e169 [FREE Full text] [doi: 10.2196/jmir.6877] [Medline: 28903895]
37. Grando M, Murcko A, Mahankali S, Saks M, Zent M, Chern D, et al. A study to elicit behavioral health patients' and providers' opinions on health records consent. J Law Med Ethics 2017 Jun;45(2):238-259 [FREE Full text] [doi: 10.1177/1073110517720653] [Medline: 30976154]
38. Baltaxe E, Czypionka T, Kraus M, Reiss M, Askildsen J, Grenkovic R, et al. Digital health transformation of integrated care in Europe: overarching analysis of 17 integrated care programs. J Med Internet Res 2019 Sep 26;21(9):e14956 [FREE Full text] [doi: 10.2196/14956] [Medline: 31573914]

Abbreviations

EHR: electronic health record
GBH: general behavioral health

PC-GIS: patient-controlled granular information sharing
SMI: serious mental illness

Edited by R Kukafka, G Eysenbach; submitted 28.03.20; peer-reviewed by N Levitz, Y Chu; comments to author 05.09.20; revised version received 30.11.20; accepted 28.09.21; published 20.04.22.

Please cite as:
Ivanova J, Tang T, Idouraine N, Murcko A, Whitfield MJ, Dye C, Chern D, Grando A
Behavioral Health Professionals' Perceptions on Patient-Controlled Granular Information Sharing (Part 2): Focus Group Study
JMIR Ment Health 2022;9(4):e18792
URL: https://mental.jmir.org/2022/4/e18792
doi:10.2196/18792
PMID:35442213

Original Paper

Behavioral Health Professionals' Perceptions on Patient-Controlled Granular Information Sharing (Part 1): Focus Group Study

Julia Ivanova[1], MA; Tianyu Tang[2], MS, MD; Nassim Idouraine[3], BS; Anita Murcko[3], MD; Mary Jo Whitfield[4], MSW; Christy Dye[5], MPH; Darwyn Chern[5], MD; Adela Grando[3], PhD

[1]School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, United States

[2]College of Medicine, University of Arizona, Tucson, AZ, United States

[3]College of Health Solutions, Biomedical Informatics, Arizona State University, Scottsdale, AZ, United States

[4]Jewish Family and Children's Services, Phoenix, AZ, United States

[5]Partners in Recovery, Phoenix, AZ, United States

Corresponding Author:
Julia Ivanova, MA
School of Human Evolution and Social Change
Arizona State University
900 Cady Mall
Tempe, AZ, 85281
United States
Phone: 1 480 965 6213
Email: jivanova@asu.edu

Related Article:

Companion article: https://mental.jmir.org/2022/4/e18792

Abstract

Background: Patient-controlled granular information sharing (PC-GIS) allows a patient to select specific health information "granules," such as diagnoses and medications; choose with whom the information is shared; and decide how the information can be used. Previous studies suggest that health professionals have mixed or concerned opinions about the process and impact of PC-GIS for care and research. Further understanding of behavioral health professionals' views on PC-GIS are needed for successful implementation and use of this technology.

Objective: The aim of this study was to evaluate changes in health professionals' opinions on PC-GIS before and after a demonstrative case study.

Methods: Four focus groups were conducted at two integrated health care facilities: one serious mental illness facility and one general behavioral health facility. A total of 28 participants were given access to outcomes of a previous study where patients had control over medical record sharing. Participants were surveyed before and after focus groups on their views about PC-GIS. Thematic analysis of focus group output was paired with descriptive statistics and exploratory factor analysis of surveys.

Results: Behavioral health professionals showed a significant opinion shift toward concern after the focus group intervention, specifically on the topics of patient understanding (P=.001), authorized electronic health record access (P=.03), patient-professional relationship (P=.006), patient control acceptance (P<.001), and patient rights (P=.02). Qualitative methodology supported these results. The themes of professional considerations (2234/4025, 55.5% of codes) and necessity of health information (260/766, 33.9%) identified key aspects of PC-GIS concerns.

Conclusions: Behavioral health professionals agreed that a trusting patient-professional relationship is integral to the optimal implementation of PC-GIS, but were concerned about the potential negative impacts of PC-GIS on patient safety and quality of care.

(JMIR Ment Health 2022;9(4):e21208) doi:10.2196/21208

KEYWORDS

behavioral health professional; granular information; granular information sharing; electronic health record; integrated health care; electronic consent tool

Introduction

Though the terms behavioral health and mental health are often used synonymously, the term behavioral health is broader. Mental health focuses solely on a person's psychological state, whereas behavioral health is a broader umbrella that incorporates physical and mental struggles: eating habits, exercise routines, and alcohol or drug consumption [1,2]. Behavioral care encompasses a variety of health services, including mental health care, psychiatric care, marriage and family counseling, substance use prevention, intervention, treatment and recovery, and others. Behavioral health professionals include, but are not limited to, psychiatrists, psychologists, counselors, clinicians, therapists, social workers, nurse practitioners, and others [3].

Patient-controlled granular information sharing (PC-GIS) allows patients to select "granules" or specific elements of their electronic health records (EHRs) and decide with whom to share this information [4,5]. This paper focuses on the clinical implications of PC-GIS of behavioral health information, long considered to be highly sensitive information by individuals and by US law [2,6-8]. With the advance of integrated physical and behavioral health care delivery, recommendations for expanding patient control of health data have asserted enhanced patient privacy [9-12]. Such suggestions underscore the importance of the PC-GIS process in health care, with patients considering which data to share and which to withhold (eg, "I do not want to share records related to past suicide attempts"). This concept includes the designation of specific data for specific care team members (eg, "I do not want to share past suicide attempt information with my endocrinologist"). Those with a serious mental illness (SMI), defined as an impairment severely interfering with daily activity, are at a higher risk for fragmented care and may, therefore, require different or additional PC-GIS options [13-15]. The literature suggests an evolving tension between patients desiring more access to, and control of, their EHR data and health care professionals who are concerned that such accessibility may negatively impact patient safety, care quality, and cost of care (eg, duplicate labs and diagnostic tests) if critical information is redacted based on patient choice.

Previous studies have established that PC-GIS is attainable using current electronic informed consent systems for care, research, or both, including perspectives on sensitivity and control of information [16-21]. A 394-patient study using granular information sharing for research by Kim et al [17] demonstrated that patients responded positively to granular control, resulting in wide variability of sharing decisions. Caine and Hanania [5] showed that when given the option to exercise granular sharing with various care team participants, all 30 patients chose granular record sharing control over an "all or none" approach. These patients were also most likely (76.7%) to share all information with their primary physicians. A similar 30-patient study found that although PC-GIS on a need-to-know basis (83%) was preferred, patients (20%) admitted they did not understand what items a provider may "need to know" [22]. Soni et al [4,23] evaluated how 25 behavioral health patients would apply PC-GIS using data from their own EHR, comparing patient and health professional sensitivity designation of the same items. Results showed that patients fully (19.3%) and partially disagreed (14.5%) with professionals' characterization of items. Patient rationale for their sharing choices was complex, including fear of discrimination, perceived relevancy to particular provider disciplines, and trust [4,23]. While these studies emphasize patient perceptions of PC-GIS, they also highlight the need for research focused on health professionals' perceptions and recommendations on this topic [4,5,16,18,19,22,24,25].

A few studies have explored how health professionals view PC-GIS and how such control may affect clinical care. In a 6-month prospective study by Tierney et al [26], 105 patients in a primary care clinic with 31 professionals were given PC-GIS capability. Of the 24 professionals who completed the poststudy survey, 63% responded "strongly agree" to the statement that patient restriction of information would reduce quality of care, while 54% of those providers agreed that patients having PC-GIS is "OK," further emphasizing the complexity of PC-GIS [26].

In another study, 20 behavioral health professionals were interviewed about their opinions on PC-GIS and consent [2,15]. Discussion topics were categorized into share, should share, or not share, constituting 100% of professional perceptions. Health professionals noted that patients should share information in cases of medical emergencies (57%), patient history data (52%), and medications and treatments (46%). Health professionals identified certain topics patients seemed reluctant to share: items related to substance use (48%), medical diagnoses (47%), and SMIs (39%) [2]. Overall, the study found that while health professionals agreed patients should have more control over who accesses their EHR (70%), professionals also point out that there is certain information they believe should never be restricted (65%) [15]. Study findings also highlight health professionals' views that trust and patient comprehension may increase patients' sharing of information, especially involving sensitive behavioral health information [2,15].

Previous studies suggest that health professionals have mixed or concerned opinions about the process and impact of PC-GIS for care and research [4,23,26,27]. The research in this paper uses the focus group data of 28 integrated health care professionals collected by Ivanova et al in the part 2 of this study [27] to identify changes in opinions and understandings of behavioral health professionals provided with real patient examples and a full case study of PC-GIS. This study hypothesizes that knowing patients' granular data sharing choices leads to a decrease in behavioral health professionals' support for PC-GIS. To determine whether such effects occur, this study investigates health professionals' views on PC-GIS

before and after the focus groups and explores potential trends based on cohort differences.

Methods

Study Sites and Participants

This study was approved by the Arizona State University Institutional Review Board (No. 00010309). The study sites were two outpatient integrated care facilities using the same EHR platform. One facility treats patients with SMI conditions (SMI facility), while the other predominantly treats patients with general behavioral health (GBH) conditions (GBH facility). This study used definitions from Grando et al [15] for prescribers and nonprescribers. GBH facility professionals were comprised of 15% prescribers and 85% nonprescribers, while SMI facility professionals were comprised of 10% prescribers and 90% nonprescribers.

Four, 2-hour focus groups were conducted at the study sites: two focus groups at each site, with seven behavioral health professionals each. Such a design allows sufficient time for individuals to share their thoughts while providing a small-group environment for conversation [28-30]. Focus group participants were facility employees, 21 years of age or older, who worked closely with patients with behavioral health conditions. Participants from the GBH facility were selected by executive staff, while participants from the SMI facility were recruited using flyers and were self-selected. A representative sample of prescribers and nonprescribers was sought for both facilities.

Survey

Participants were asked to individually complete a survey prior to the focus group. The survey was adapted from Tierney et al [26] and was comprised of nine statements that were rated on a Likert scale with the following responses: “strongly disagree,” “somewhat disagree,” “neutral,” “somewhat agree,” “strongly agree,” and “don’t know/can’t say” (Table 1) [27]. The survey prompts were divided into six specific aims based on measuring concepts and primary impact, patient or professional (Table 1). After each focus group, participants completed the same survey to evaluate changes in opinions of PC-GIS after seeing how actual patients exercise choices. It was hypothesized that discussion of a demonstrative case study would lead to a decrease in behavioral health professionals’ support for PC-GIS. The survey style (Table 1) lends to measuring opinion changes.

The survey analysis was used to determine the presence or absence of directional opinion change following the focus groups. SPSS Statistics for Macintosh (version 27; IBM Corp) was used for descriptive statistics, Wilcoxon signed-rank testing, Cronbach α tests, scree plots, and exploratory factor analysis on pre- and postsurvey results. Exploratory analysis was done with variables, or groups, of interest with appropriate sample sizes of at least 12 for exploratory factor analysis [31-33]. Descriptive statistics were computed using the bootstrapping option (10,000 replicates), and skewness and kurtosis results were used to confirm normal distribution and, thus, verification of data from the survey [34]. Cronbach α tests were run to ensure multidimensionality of the survey; an α value below .80 was considered evidence of a multidimensional scale [35]. Prompt selections of “don’t know/can’t say” were recorded as blank to avoid skewing results. The Likert scale results ranged from 1.00 (“strong disagreement”) to 5.00 (“strong agreement”). For the pre– and post–Likert scale survey results [35,36], the Wilcoxon signed-rank test (power=0.80, α=.05) was applied to each survey prompt to determine significance for changes after the focus group for all participants (N=28) and between participants serving different patient populations (n=14 each).

Intended response concepts of survey prompts were identified based on results from Tierney et al [26] and were labeled as “prompt aim” (Table 1). Exploratory factor analysis showed which aims, or components, were actively measured by the survey and relationships between prompts based on participant response [36,37]. Varimax rotation was used because survey prompts were not designed to correlate [26]. Measured components of the survey were identified by prompt magnitude in the output component matrix. This analysis revealed the component emphasis from pre- to postsurvey that was used to gauge professionals’ perceptions [35]. Scree plots were used to further validate the component results from the factor analysis where viable components had an exponential slope [35].

Table 1. Categorized survey prompts.

Category and prompt aims	Specific survey prompts[a]		
	No.	Phrasing	Directionality
Patient focused			
Patient control acceptance			
	2	I am comfortable with patients restricting my seeing some parts of their EHR[b].	Positive
	4	I think it is OK for patients to have control over who sees what information in their EHR.	Positive
	6	It is a good thing for patients to have control over who sees their EHR.	Positive
Patient understanding			
	1	I believe that patients understand what an EHR is.	Positive
Patient rights			
	8	The patient owns the information in his or her EHR.	Neutral
	9	As a patient, I would like to control the information in my EHR that providers can see.	Positive
Health professional focused			
Authorized EHR access			
	3	My patients' EHRs are viewed only by people who should have access to them.	Positive
Patient-professional relationship			
	5	Patients preventing me from seeing part or all of their EHR could affect my relationships with them.	Neutral
Quality of care			
	7	Restricting access to all or part of a patient's EHR will likely reduce the quality of care I deliver.	Negative

[a]Survey prompts were grouped by overarching themes and classified as positive, neutral, or negative based on framing. Prompt numbers (eg, prompt 2) refer to chronology of the survey, following placement by Tierney et al [26].

[b]EHR: electronic health record.

Focus Group

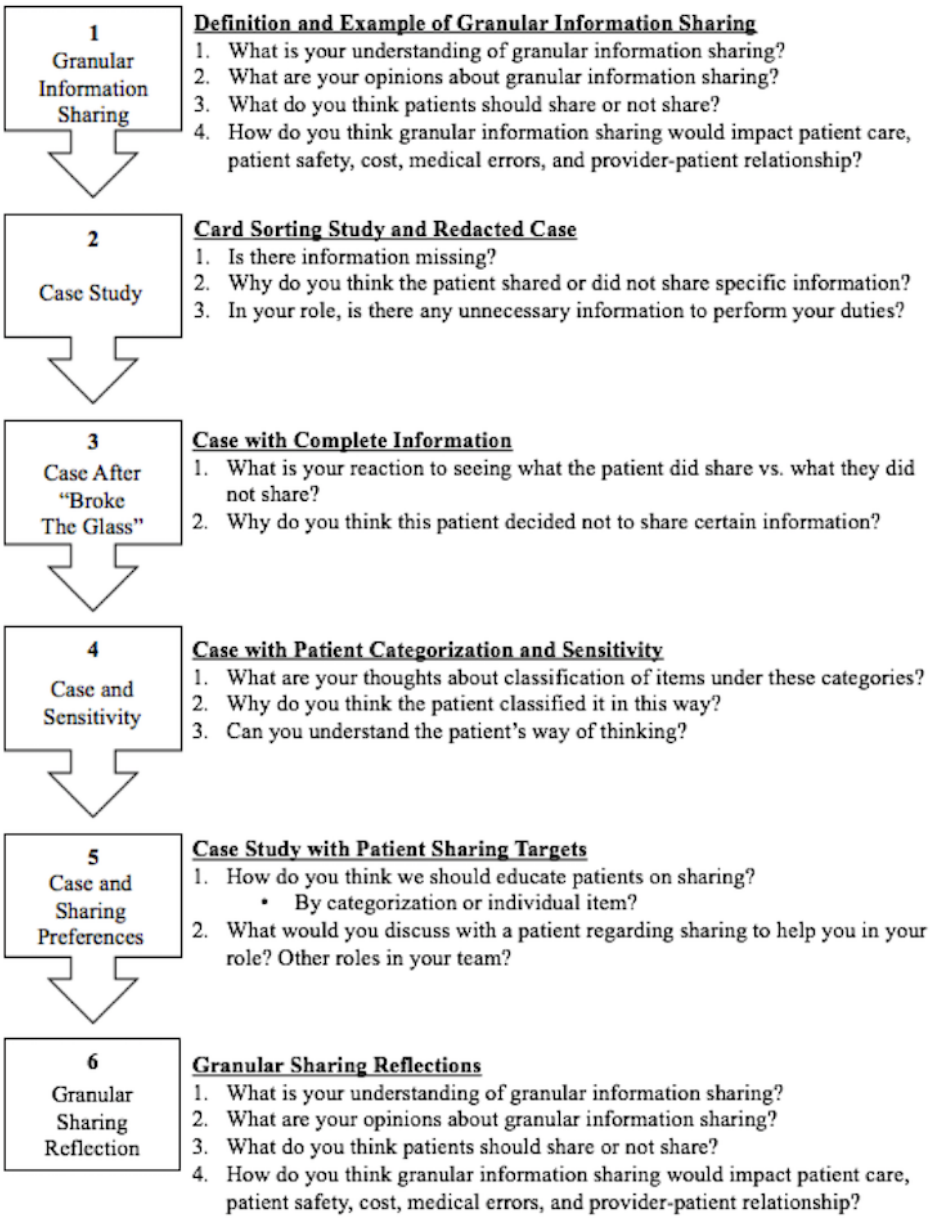
The focus group was presented in six sections (Figure 1) [27]. In Section 1, baseline perceptions of PC-GIS were elicited using examples and explanations of granular information and sharing [9,17,19,27,38,39]. In Section 2, the Soni et al [4,23], or "card sorting," study was explained to participants, and a specific case from the study of granular sharing by an actual patient with patient-executed redactions was presented [27]. Participants were then shown the complete patient data set without redaction in Section 3 as if a provider decided to "break the glass," a term that refers to health care professionals' retrieval of a patient's redacted information in an emergency.

During Section 4, the same patient's data were presented by category (ie, alcohol use and alcoholism, communicable diseases, drug abuse, genetic information, mental health, other addictions, other information, and sexuality and reproductive health) and sensitivity (ie, very sensitive, somewhat sensitive, or not sensitive). Section 5 explored that patient's choices to share those categories with different health care professionals and institutions.

Finally, in Section 6, participants were asked to reflect on their understanding, opinions, and recommendations for PC-GIS.

Two qualitative analysis techniques were applied to focus group outputs to provide insight into the survey results. Thematic analysis was used to find and define emergent topics of importance to participants [40]. Audio recordings of the focus groups were transcribed through a third party [41] and screened for accuracy by three researchers working sequentially by visual annotation. The validated transcriptions were analyzed using Braun and Clarke's [42] thematic analysis guidelines and anthropological methodology through six iterations, resulting in quantifiable codes and themes [35]. The units of analysis were meaningful phrases per participant and themes identified through repetition and frequency in transcripts. MAXQDA software (VERBI GmbH) was used to identify and define emerging themes from transcripts. Themes and codes were defined and refined iteratively by three researchers.

Figure 1. Focus group flow. Section themes are shown on the left with representative questions for each section on the right (numbered). This figure was adapted from Ivanova et al [27].



Survey and Focus Group Integration

The exploratory factor analysis results from the survey were used to organize emergent themes and subthemes from the focus group thematic analysis in the last iteration. This step permitted complementary analysis of qualitative and quantitative results.

After providing descriptive statistics and exploratory factor analysis results, themes and codes were used to interpret these results and provide insight into opinion shifts regarding PC-GIS. To compare themes between patient populations, quote matrices and complex coding queries (intersection-set) were performed.

Results

Demographics

A total of 28 participants were recruited (Table 2). Out of these participants, 5 (18%) prescribers (ie, physicians and nurse practitioners) and 23 (82%) nonprescribers were identified [15]. This was a demographically representative sample of professionals at each site. All 28 participants took the presurvey and 27 (96%) took the postsurvey.

Table 2. Participant roles and population representation.

Role type[a]	Participants, n (%)[b]		
	General behavioral health facility (n=14)	Serious mental illness facility (n=14)	Total (N=28)
Counselors	3 (21)	1 (7)	4 (14)
Nurses	1 (7)	2 (14)	3 (11)
Rehabilitation specialists	1 (7)	2 (14)	3 (11)
Case managers	1 (7)	2 (14)	3 (11)
Clinical coordinators	3 (21)	0 (0)	3 (11)
Administrators	0 (0)	3 (21)	3 (11)
Physicians (prescribers)	2 (14)	1 (7)	3 (11)
Nurse practitioners (prescribers)	2 (14)	0 (0)	2 (7)
Peer mentors	0 (0)	1 (7)	1 (4)
Medical assistants	0 (0)	1 (7)	1 (4)
Discharge planners	0 (0)	1 (7)	1 (4)
Social workers	1 (7)	0 (0)	1 (4)

[a]The table, taken from Ivanova et al [27], groups participants by role types (prescribers are indicated) and patient population.

[b]Percentages may not add up to 100 due to rounding.

Changes in Behavioral Health Professionals' Perceptions: Survey

Comparison of the pre– and post–focus group survey responses demonstrated two significant changes: (1) change from strong agreement to strong disagreement (mean <2.5, including SE) with patient-focused survey prompts and (2) change from strong agreement to strong disagreement (mean >3.5, including SE) with professional-focused survey prompts (Table 3). Descriptive analysis results provided skewness and kurtosis statistics exhibiting normal distribution of data, a validation of the survey results and usability of participant responses.

Drilling into the specific prompts, as defined in Table 1, patient understanding (prompt 1, *P*=.001), patient-professional relationship (prompt 5, *P*=.006), and patient control acceptance (prompt 6, *P*=.005) demonstrated significant change, with increased concern about patient control. Authorized EHR access (prompt 3, *P*=.03) and patient rights (prompt 9, *P*=.02) also showed significant change toward concern, validating the study hypothesis. Of note, patient-professional relationship (prompt 5) is considered a negatively phrased expression [26,27], providing insight into the postsurvey shift in the quality-of-care response.

Table 3. Results of the descriptive statistics for the pre- and postsurveys.

Prompt no.	Prompt aim	Prompt directionality	Presurvey score[a], mean (SE)	Postsurvey score[a], mean (SE)	*P* value[b]
1	Patient understanding	Positive	3.5 (0.2)	2.5[c] (0.2)	.001
2	Patient control acceptance	Positive	3.5 (0.3)	2.3 (0.3)	<.001
3	Authorized EHR[d] access	Positive	4.5[c] (0.2)	3.7[c] (0.3)	.03
4	Patient control acceptance	Positive	3.9[c] (0.3)	2.2[c] (0.2)	<.001
5	Patient-professional relationship	Neutral	3.2 (0.3)	4.0[c] (0.2)	.006
6	Patient control acceptance	Positive	3.8[c] (0.2)	2.7[c] (0.3)	.005
7	Quality of care	Negative	3.6 (0.3)	3.9[c] (0.3)	.16
8	Patient rights	Neutral	3.6 (0.3)	3.3 (0.3)	.40
9	Patient rights	Positive	3.8[c] (0.3)	3.1[c] (0.2)	.02

[a]The survey scores ranged from 1.00 ("strong disagreement") to 5.00 ("strong agreement").

[b]*P* values were based on the pre- to postsurvey change using the Wilcoxon signed-rank test.

[c]These statistics are strongly within overall agreement (mean >3.5, including SE) or disagreement (mean <2.5, including SE).

[d]EHR: electronic health record.

Presurvey results loaded into four main components: patient control acceptance, professional considerations, patient rights, and patient understanding. In postsurvey loadings, exploratory factor analysis revealed only three components present; patient understanding was now absent (Table 4). After pairing with the descriptive statistics, results suggest that after the focus group, participants became more concerned with patient rights and patient control acceptance and the impact of these aspects on professional matters, such as quality of care and patient-professional relationship.

Table 4. Pre- and postsurvey exploratory factor analysis loadings.

Prompt no.	Presurvey component, factor analysis loading[a]				Postsurvey component, factor analysis loading[a]		
	Patient control acceptance	Professional considerations	Patient rights	Patient understanding	Patient control acceptance	Patient rights	Professional considerations
1	0.0	0.0	0.1	0.9[b]	0.4	0.8[b]	0.1
2	0.8[b]	0.1	0.0	–0.1	0.4	0.6[b]	0.3
3	–0.1	–0.8[b]	0.4	–0.2	–0.5	0.7[b]	0.4
4	1.0[b]	0.0	0.1	0.1	0.7[b]	0.4	0.3
5	–0.2	0.7[b]	0.2	–0.4	–0.2	0.1	–0.9[b]
6	0.8[b]	–0.2	0.1	0.3	0.8[b]	0.2	0.3
7	–0.1	0.8[b]	0.4	–0.1	–0.4	–0.3	–0.6[b]
8	0.1	0.1	0.9[b]	0.1	0.2	0.8[b]	–0.1
9	0.9[b]	–0.1	–0.1	–0.2	0.8[b]	0.2	0.2

[a]Negative loadings are due to directionality of prompts and are not significant. Scree plot results ensured overall viability of components.

[b]This value is this prompt's highest absolute loading for this component.

Behavioral Health Professionals' Concerns Around PC-GIS: Focus Group

In the next step of validating the hypothesis, thematic analysis (4025 codes) of focus groups yielded three main themes (Figure 2), complementing the survey components (Table 5). The themes were professional considerations (2234/4025, 55.5%), patient aspects (1046/4025, 26.0%), and PC-GIS technology aspects (745/4025, 18.5%).

The professional considerations theme covers themes that directly impact a provider, including information needed to provide health care services. The patient aspects theme encompasses all topics relating specifically to patient experience and rationale. The PC-GIS technology aspects theme reflects the specific discussion of PC-GIS process and operations (Table 5 and Figure 2).

The survey results show the shift to concern yielding components of patient control acceptance, patient rights, and professional considerations, while thematic analysis shows how professional discussion revolved predominantly around professional considerations, such as necessity of health information (Figure 2). This overall shift toward professional concern around PC-GIS was observed in the focus group discussion and was coded under multiple themes and subthemes; an example quotation is as follows:

> *...what if there were an issue of depression affecting [the patient's] hygiene or dental care, and the dentist doesn't know how to explain that? Similarly, if you had a dentist who was seeing dental care being compromised because you had somebody with an eating disorder, who do they have to collaborate with or even that comfort of making that referral.* [Nonprescriber]

Similarly, professionals quickly pointed to the complexity of PC-GIS in their domain and the potential for negative impact on patient care:

> *...I think if a patient has seen numerous doctors, they all should be on the same page with medications because of any contraindications.* [Nonprescriber]

The survey results reflect understanding and opinions in a quantitative fashion, while the interviews demonstrate how the concepts are linked.

Thematic analysis also conveys the complexity of participant perceptions, with subthemes interweaving patient and professional considerations. Regarding 879 codes, the reactions subtheme included general feedback about PC-GIS (n=255, 29.0%) as well as specific concerns (n=334, 38.0%), predominantly relating to patient safety and health. Patient safety and health encompassed issues ranging from missing medications and the potential for drug-drug interactions to the need for improved physical and mental health integration. A minority (n=149, 17.0%) of participants felt that data sharing in health care as an environment should never be granular: "[This is] not a place to be granular." Other issues surfacing in the health professional concern area included the mismatch of patients' interpretation of information versus health professionals' interpretation.

While the patient aspects theme included two subthemes corresponding to exploratory factor analysis components, the major subtheme of patient perspective considered drew greatly on health professionals' thoughts on patients' reasoning to share or not share health information. Indeed, 65.0% of 722 codes (n=469) within the patient perspective considered subtheme were specifically related to patient reasoning to not share, such as patient fears or fear of discrimination (n=113, 24.0%). Many instances of patient reasoning (n=201, 42.9%) dealt with patients' understanding and comprehension.

To provide further context for the quantitative results, participant opinions from within the reactions subtheme (Section 2 codes: 81/116, 69.8%; Section 4 codes: 35/116, 30.2%) were divided into three groups: concerned, supportive, and mixed opinions. For Section 2 questions, regarding the redacted case study, 46% (37/81) expressed mixed opinions, 36% (29/81) were concerned, and 19% (15/81) were supportive. Within Section 4 questions, regarding patient categorization and sensitivity, 43% (15/35) expressed concerned opinions, 37% (13/35) expressed mixed opinions, and 20% (7/35) were supportive. Of note, PC-GIS unease revolved around patient safety:

> *After you see it in action [Soni et al case study], seeing what they shared versus what was in the chart, I think this safety risk is extremely high.* [Nonprescriber]

To that end, participants suggested prioritizing a risk-benefit analysis tool for patients as an adjunct to professional-patient PC-GIS interactions:

> *...if there can be something figured out...just like the duties to warn, just like this that there is some way that you can obtain that information under certain circumstances...I can see where it could actually improve providers' relations where [the patients are] not going in paranoid that they're judging because, you know, [they] have a mental illness and [professionals] prejudged [them].* [Nonprescriber]

A common theme among all focus group participants was that role-specific, essential health information access should be considered when granting PC-GIS authority:

> *I would be worried that the patient doesn't share the right information with the right provider.* [Prescriber]

The problem is that "essential" information depends on the circumstances of the patient:

> *Okay, I think [PC-GIS is] a good thing because I know if I had depression and there was no good reason for my dentist to know that...now if I have an eating disorder and I'm throwing up all the time, that's going to ruin your teeth. But I would hopefully choose to share that information with my dentist. But it's my choice I guess and that's what's nice.* [Nonprescriber]

Figure 2. Emergent themes and subthemes from focus group thematic analysis on patient-controlled granular information sharing (PC-GIS). There were a total of 4025 codes.

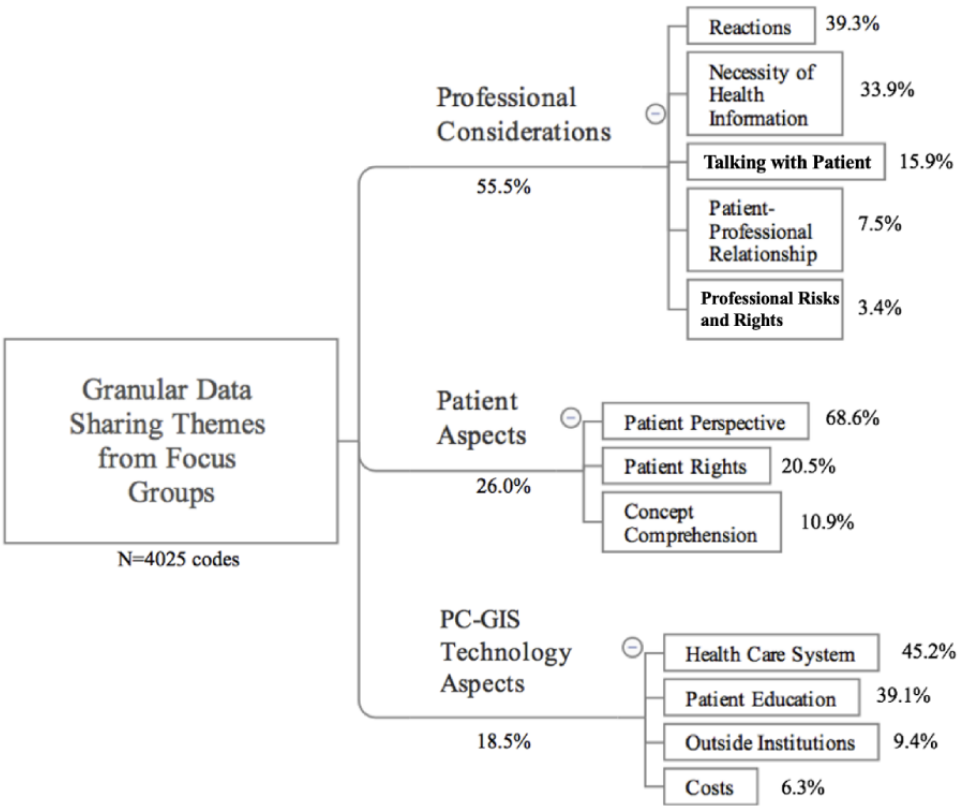


Table 5. Definitions of subthemes for coding and quotations.

Themes and subthemes	Definition	Example quotation
Professional considerations		
Reactions	How professionals react to definitions, case examples, questions, etc	"Yeah, I would be worried that the patient doesn't share the right information with the right provider." [Prescriber]
Necessity of health information	When a professional references the need for pertinent health information at the point of care	"Patients if they did have a mental illness they're taking psychiatric medications they're not going to disclose to the PCP[a] the meds they're taking, you can't check for interactions and then just can't provide good care." [Prescriber]
Talking with patient	How professionals talk with and seek health information from patients	"Just ask what they're being treated for those conditions, and so what are they taking." [Nonprescriber]
Patient-professional relationship	When linking the patient-professional relationship to granular data sharing	"Kind of going back to [the] gender dysphoria thing...That is also something that I would want to know because I would want you to be comfortable, and so, I'd want to make sure that I'm referring to you how you want to be as and using the name you want to be known by. And I'd want my office to do the same thing. So, that kind of stuff is also important to know too." [Prescriber]
Professional risks and rights	When a professional considers their own personal risks related to granular data sharing	"Yeah, be it the actual patient or the provider, life being put in jeopardy by not having certain information. I'm thinking more than HIV AIDS..." [Nonprescriber]
Patient aspects		
Patient perspective	When a professional responds from a patient perspective	"I think more how the patient perceives the information is more sensitive. This is more than perception."[b] [Nonprescriber]
Patient rights	When a professional references federal or state statutes regarding patient rights	"I really think it is hard because I've talked to a lot of people who say that their medical doctors don't understand the behavioral health side. So if they, if it wasn't affecting their behavioral health or their medical health, then I think they should have the right to not talk about it if they don't feel comfortable." [Nonprescriber]
Concept comprehension	When a professional refers to a patient's uninformed or potentially risky data sharing choices	"But I've also seen people's lives be put in jeopardy because maybe, whether it be a paranoia or just not understanding or something, I don't want anything shared or, like provider six said, we've had it—maybe there is a substance abuse issue." [Nonprescriber]
PC-GIS[c] technology aspects		
Health care system	When a professional refers to sharing information for care coordination with others throughout the health care system	"Not necessarily with that topic that I think if a patient has seen numerous doctors, they all should be on the same page with medications because of any contraindications." [Nonprescriber]
Patient education	When a professional describes or references patient education about granular information sharing	"Tell them the reason why we're asking, the importance of it, and to help them understand why we need the information." [Nonprescriber]
Outside institutions	When a professional refers to external institutions and organizations with legal control over health data sharing (eg, Department of Homeland Security, courts, law enforcement, and Department of Public Safety)	"I think that's one of the things that a lot of our patients that they have a legal background or on court-ordered treatment, meaning, they are not necessarily wanting treatment, but the court says that they have to. It is a valid reason for them to be a little nervous and stuff, because 'what are you going to tell, are you just trying to get more information so I can go back to jail...'" [Nonprescriber]
Costs	When a professional highlights the fiscal aspects of granular data sharing (eg, costs to patient, institution, and system)	"I was thinking from a cost perspective. Granular information sharing could increase cost because if you don't give all the information, I could see a provider redoing things that have already began so they can get the information they need to make a good decision. Whereas, if they have that information and knew what the history was, they would know where to start instead of having to start all the time from the beginning." [Nonprescriber]

[a]PCP: primary care physician.

[b]All participants in the focus group agreed with this comment.

[c]PC-GIS: patient-controlled granular information sharing.

Impact of Patient Populations: Survey and Focus Group Integration

Exploration of potential differences between health professionals using descriptive statistics and drilling down on qualitative data led to identification of two distinct patient populations: GBH and SMI. Descriptive statistics and Wilcoxon signed-rank tests were performed separately for each group (Table 6). Differences were observed in the presurvey, where participants treating an SMI population showed agreement with all positively phrased prompts. Meanwhile, those treating the GBH population showed agreement only with prompts 3 and 7: authorized EHR access (positive phrasing) and quality of care (negative phrasing). Differences were compounded in the postsurvey, where participants treating an SMI population showed agreement only with prompt 5: patient-professional relationship (neutral phrasing perceived as negative) [26,27]. Those treating a GBH population agreed with the negatively perceived prompts (ie, patient-professional relationship and quality of care) and disagreed with a positively worded prompt (ie, patient control acceptance; Table 6). Participants treating an SMI population initially perceived PC-GIS positively, then made a significant shift to neutral or mixed opinion with concern over the patient-professional relationship prompt (P=.007). The participants treating a GBH population showed concern over PC-GIS with a shift to concern over the following prompts: patient control acceptance (prompt 2, P=.007; prompt 4, P=.009), quality of care (P=.01), and patient understanding (P=.03).

Table 6. Descriptive statistics by predominant patient population.

Prompt no.	Prompt aim	Prompt directionality	Presurvey score[a], mean (SE)		Postsurvey score[a], mean (SE)			
			GBH[b]	SMI[c]	GBH	*P* value[d]	SMI	*P* value[d]
1	Patient understanding	Positive	3.3 (0.3)	3.7 (0.2)[e]	2.3 (0.3)	.03	2.7 (0.3)	.02
2	Patient control acceptance	Positive	3.1 (0.4)	3.9 (0.3)[e]	1.7 (0.3)[e]	.007	3.0 (0.4)	.02
3	Authorized EHR[f] access	Positive	4.1 (0.4)[e]	4.8 (0.1)[e]	3.8 (0.4)	.47	3.7 (0.4)	.01
4	Patient control acceptance	Positive	3.4 (0.4)	4.5 (0.1)[e]	1.7 (0.3)[e]	.009	2.8 (0.3)	.004
5	Patient-professional relationship	Neutral	3.9 (0.3)	2.5 (0.4)	4.1 (0.3)[d]	.41	3.8 (0.3)[e]	.007
6	Patient control acceptance	Positive	3.4 (0.4)	4.1 (0.2)[e]	2.4 (0.4)	.28	3.2 (0.4)	.16
7	Quality of care	Negative	4.0 (0.3)[e]	3.2 (0.5)	4.4 (0.3)[e]	.01	3.4 (0.4)	.43
8	Patient rights	Neutral	3.6 (0.4)	3.6 (0.4)	2.9 (0.4)	.10	3.8 (0.4)	.72
9	Patient rights	Positive	3.1 (0.4)	4.5 (0.2)[e]	2.4 (0.3)	.19	3.8 (0.3)	.03

[a]The survey scores ranged from 1.00 (“strong disagreement”) to 5.00 (“strong agreement”).

[b]GBH: general behavioral health.

[c]SMI: serious mental illness.

[d]*P* values were based on the pre- to postsurvey change using the Wilcoxon signed-rank test.

[e]These values are in overall agreement (including SE) or disagreement (including SE).

[f]EHR: electronic health record.

Quote matrices were applied to specify differences between GBH and SMI professionals on survey aims, with some subtopics within themes considered when applicable. GBH professionals showed greater frequency of discussing negative reactions (32/44, 73% of codes), positive reactions (14/22, 64%), professional risks and rights (44/67, 66%), and outside institutions (52/69, 75%). SMI professionals more frequently discussed the following subtopics: do not need to know (21/23, 91%), patient-professional relationship (124/180, 68.9%), trust (22/32, 69%), patient aspects (669/1046, 64.0%), and costs (36/46, 78%). These results reflect large differences in frequencies of thematic analysis coded topics. However, while a topic may be suggested, the discourse may not contain an opinion. Therefore, a second layer of qualitative analysis to identify subthemes was performed.

To identify differences in participant perceptions on themes, subthemes, and topics within subthemes, complex coding queries were used to categorize negative, positive, or mixed perception codes. GBH professionals perceived costs and trust negatively, overall (Table 7). SMI professionals referred to professional risks and rights topics with a negative slant, while the other topics were presented in a mixed or positive fashion. The complex coding queries highlight the complexity of the topic and suggest an impact of patient population on subthemes and topics.

Table 7. Complex coding query results of topic perceptions.

Topics, themes, and participant perception	Instances of perceptions for each topic by facility, n (%)	
	General behavioral health facility	Serious mental illness facility
Do not need to know (within the necessity of health information subtheme)		
Negative (n=0)	N/A[a]	N/A
Positive (n=1)	1 (100)	0 (0)
Mixed (n=4)	2 (50)	2 (50)
Patient-professional relationship		
Negative (n=4)	4 (100)	0 (0)
Positive (n=4)	3 (75)	1 (25)
Mixed (n=3)	1 (33)	2 (67)
Trust (within the patient-professional relationship subtheme)		
Negative (n=1)	1 (100)	0 (0)
Positive (n=2)	1 (50)	1 (50)
Mixed (n=0)	N/A	N/A
Professional risks and rights		
Negative (n=15)	11 (73)	4 (27)
Positive (n=3)	3 (100)	0 (0)
Mixed (n=5)	5 (100)	0 (0)
Outside institutions		
Negative (n=0)	N/A	N/A
Positive (n=10)	10 (100)	0 (0)
Mixed (n=3)	3 (100)	0 (0)
Costs		
Negative (n=3)	2 (67)	1 (33)
Positive (n=2)	0 (0)	2 (100)
Mixed (n=3)	0 (0)	3 (100)

[a]N/A: not applicable; there were no instances of this perception regarding this topic.

The mixed methodology analysis focused on the differences between SMI and GBH professionals, where SMI professionals displayed lower levels of concern regarding the process of PC-GIS, more frequently citing the following topics: do not need to know (21/23, 91% of codes), patient-professional relationship (124/180, 68.9%), and trust (22/32, 69%).

Discussion

Principal Findings

Results show that behavioral health professionals had fewer positive views on PC-GIS after the focus group, with a significant opinion shift toward concern on the following topics: patient understanding (*P*=.001), authorized EHR access (*P*=.03), patient-professional relationship (*P*=.006), patient control acceptance (*P*=.005), and patient rights (*P*=.02). Qualitative methodology supported these results, as themes and subthemes, such as professional considerations (2234/4025, 55.5% of codes) and necessity of health information (260/766, 33.9%), identified aspects of PC-GIS concerns; indeed, participant opinions after viewing the case study without redactions (Section 4) showed increased levels of concern (7% overall change).

Mixed methodology results after the focus group showed concerns that PC-GIS could negatively impact behavioral health professionals' ability to deliver optimal care. This perception shift was evident in qualitative results from discussions dominated by patient health and safety topics (combined, 60.3% [70/116] of codes of concerned reaction). Our results show that behavioral health professionals remained highly concerned about patient granular control for a variety of reasons [18]. Health professionals in our study highlighted potential negative effects of granular sharing, including impact on the professional-patient relationship and lack of access to necessary health information, reflected in the professional considerations theme. A minority of professionals (149/879, 17.0%) considered health care as simply "not a place to be granular," while most acknowledged acceptance of the trend toward increasing PC-GIS and offered concrete recommendations for proactive processes that could help ensure patient safety while preserving record sharing choice.

Much of the literature focuses on patient perspectives of PC-GIS, demonstrating that patients respond positively to having granular control over data sharing [5,17,18,24], while casting doubt on a patient's understanding of information relevancy or professionals' "need to know" [5,15]. Our study reflects similar concerns by behavioral health professionals, where subthemes included necessity of health information (260/766, 33.9% of codes) and patient concept comprehension (13/115, 11.3%). Our thematic analysis demonstrated that when professionals are shown results of a patient's granular sharing choices, they view the choices from the patient perspective, while expressing apprehension that necessary role-specific information may not be appropriately shared [4]. Employing shared decision-making using specialty-tailored methods may help alleviate such concerns [5,15,27].

A trusting bond is important in health care delivery and is a continued underlying basis in quality of care and patient outcomes in health literature [2,43-46]. Our qualitative results refer to trust when subthemes such as talking with patient, patient-professional relationship, and patient education directly deal with strengthening the relationship and understanding between health professionals and patients. A recent study by Esmaeilzadeh [47] showed that patients' trust in providers influences their trust in information sharing technology. Our qualitative results exemplify some processes that health professionals may use to strengthen trust: "Tell them [patients] the reason why we're asking, the importance of it, and to help them understand why we need the information." Indeed, this type of approach to strengthening the professional-patient relationship was a common recommendation in our study, as well as in existing literature, to alleviate professionals' worry over patients not sharing appropriate information [15,27,44,46,47].

Proceeding further into the topic of education, professionals involved in PC-GIS must have the knowledge, background, and tools to assist patients in making safe sharing choices. This is exemplified in the case where a professional supports a patient's choice to withhold behavioral health diagnoses and medications from a patient's dentist. In reality, oral health and behavioral health have many important intersections, including substance use disorder and eating disorders [48]. Therefore, organizations and institutions must ensure that their PG-GIS process and professionals are prepared to provide sound advice to ensure patient safety. While health care institutions need to consider PC-GIS use in integrated and coordinated care, attention should be paid to critical policies, such as Title 42 CFR (Code of Federal Regulations) Part 2 and the CARES (Coronavirus Aid, Relief, and Economic Security) Act [8,38,49], and safe implementation of health care technologies relevant to health information exchange and patient EHRs [11,22]. Professionals must be actively engaged in the creation, implementation, and monitoring of data sharing policies that integrate relevant statutes as well as advances in technology and biomedicine. Our results show that when provided with an in-depth explanation of tools affecting health care delivery, health care professionals can begin the necessary dialogue regarding concerns and recommendations for improvement [27]. Identification of concerns for all stakeholders is necessary for successful implementation of health care technologies, such as PC-GIS.

While some research exists that considers health professionals' opinions on granular sharing and consent, few studies have explored the behavioral health realm, where care generates and uses highly sensitive patient information [26,50,51]. Our study has added to knowledge on behavioral health professionals' perceptions [27], while exposing divergence between behavioral health professionals who treat differing patient populations. Results comparing behavioral health professionals' predominant patient populations indicate that those working with an SMI population displayed lower levels of concern and focused on patient aspects (669/1046, 64.0% of codes), as compared to participants working with the GBH population (patient control acceptance, P=.004; patient understanding, P=.02).

Our results provide a perspective on the relevance of studies from physical health settings application to behavioral and integrated health groups [13]. Based on a pilot study where some patients were given PC-GIS capabilities, Tierney et al [26] showed that 63% of the providers strongly agreed that granular information restriction will likely reduce the quality of care delivered. Our study found that after the focus group intervention, behavioral health professionals agreed overall with Tierney et al's findings, moving from strong agreement to strong disagreement regarding patient-focused survey prompts. Such outcomes show that behavioral health professionals have similar levels of concern about PC-GIS as compared to other health professionals. However, their concerns may differ in scope and application, as behavioral health patients may be more vulnerable to addictions, discrimination, and influence from outside institutional pressures [52-54].

The divergence observed became more visible when looking at the two facilities. Survey analysis showed that participants from the SMI facility not only viewed PC-GIS more positively than those from the GBH facility, but they also displayed differences in what survey aims (ie, patient-professional relationship and patient rights) we found to have significant change from pre- to postsurvey. Professionals from the SMI facility displayed a significant (P=.007) shift from neutral to mixed opinions regarding the patient-professional relationship, an aim falling within the patient aspects category. Results from the quote matrices reinforced these outcomes, as the conversation by SMI facility professionals focused more heavily on patient aspects (669/1046, 64.0% of codes). On the contrary, GBH facility participants showed consistent agreement (Table 6) on professional-focused prompts and disagreement with patient-focused prompts. Health professionals working with an SMI population may view PC-GIS with less concern because they typically interact with their patients more frequently over longer periods of time [55,56]. One approach to alleviating PC-GIS concerns may be in bolstering patient-professional relationships and communication.

Integrated care that emphasizes transparency and patient-centeredness are health system goals [9,10]. This study highlights key aspects of PC-GIS that must be considered for its broader and deeper integration in the care environment. Improving process transparency benefits patients and

professionals by exposing gaps in care, differences in patient safety and outcomes, and drivers of costs [2,9,15,18,57]. Understanding diverse professional perspectives is critical in developing granular consent systems that balance patient-professional information needs. Results of this study will be used in the development of a PC-GIS tool, My Data Choices, to allow patients with behavioral health conditions to choose which medical records (eg, mental health information) to share with whom (eg, behavioral health providers) and for what purpose (eg, health care).

Limitations and Future Research

While our study had a small number of participants, it was a demographically representative sample within each integrated care study site, capturing the essence of the integrated health care team and their role-inspired concerns and needs. The sample size may affect the study replicability; further research with a larger sample size and in a variety of behavioral health and integrated care settings is needed. Another limitation in our study is the pragmatic difference in recruitment methods to accommodate facility leadership preference. Future work may consider semistructured interviews to more clearly identify differences in health professionals' needs. Finally, our study presented a single exemplar case derived from Soni et al [4,23]. Future studies could leverage additional patient data sharing scenarios.

Conclusions

This study enhanced what is known about PC-GIS by systematically exploring the rationale behind behavioral health professionals' perceptions, using results from a study of PC-GIS by real patients using their own data. Outcomes show that as health care professionals learn about PC-GIS implementation, they develop greater levels of concern. However, professionals balanced their concern with material recommendations for PC-GIS process improvement that ranged from patient and staff education to strengthening patient trust. Participants agreed that an informed and transparent system for health information sharing is needed to foster the mutual trust required to implement robust PC-GIS.

Acknowledgments

This work was supported by a grant from the National Institute of Mental Health for the following project: My Data Choices, Evaluation of Effective Consent Strategies for Patients With Behavioral Health Conditions (grant R01 MH108992). We thank our collaborating site members, Dr Veronica Sabater and Byron Hoston, for their support in data collection. We are also grateful for the thoughtful comments received from the journal reviewers.

Conflicts of Interest

None declared.

References

1. Neese B. Behavioral health vs mental health. Alvernia University. 2016. URL: https://online.alvernia.edu/program-resources/behavioral-health-vs-mental-health/ [accessed 2019-12-12]
2. Ivanova J, Grando A, Murcko A, Saks M, Whitfield M, Dye C, et al. Mental health professionals' perceptions on patients control of data sharing. Health Informatics J 2020 Sep;26(3):2011-2029 [FREE Full text] [doi: 10.1177/1460458219893845] [Medline: 31912744]
3. Types of mental health professionals. NAMI: National Alliance on Mental Illness. URL: https://www.nami.org/About-Mental-Illness/Treatments/Types-of-Mental-Health-Professionals [accessed 2021-02-12]
4. Soni H, Grando A, Murcko A, Diaz S, Mukundan M, Idouraine N, et al. State of the art and a mixed-method personalized approach to assess patient perceptions on medical record sharing and sensitivity. J Biomed Inform 2020 Jan;101:103338 [FREE Full text] [doi: 10.1016/j.jbi.2019.103338] [Medline: 31726102]
5. Caine K, Hanania R. Patients want granular privacy control over health information in electronic medical records. J Am Med Inform Assoc 2013 Jan 01;20(1):7-15 [FREE Full text] [doi: 10.1136/amiajnl-2012-001023] [Medline: 23184192]
6. Bringing Behavioral Health into the Care Continuum: Opportunities to Improve Quality, Costs and Outcomes. Washington, DC: American Hospital Association; 2012 Jan. URL: https://www.aha.org/guidesreports/2012-01-26-bringing-behavioral-health-care-continuum-opportunities-improve-quality [accessed 2019-03-19]
7. Campbell A, McCarty D, Rieckmann T, McNeely J, Rotrosen J, Wu L, et al. Interpretation and integration of the federal substance use privacy protection rule in integrated health systems: A qualitative analysis. J Subst Abuse Treat 2019 Feb;97:41-46 [FREE Full text] [doi: 10.1016/j.jsat.2018.11.005] [Medline: 30577898]
8. McCarty D, Rieckmann T, Baker R, McConnell K. The perceived impact of 42 CFR Part 2 on coordination and integration of care: A qualitative analysis. Psychiatr Serv 2017 Mar 01;68(3):245-249 [FREE Full text] [doi: 10.1176/appi.ps.201600138] [Medline: 27799017]
9. Klein S, Hostetter M. In focus: Integrating behavioral health and primary care. The Commonwealth Fund. 2014 Aug 28. URL: https://www.commonwealthfund.org/publications/newsletter-article/2014/aug/focus-integrating-behavioral-health-and-primary-care [accessed 2019-03-20]
10. Recommendations to the National Coordinator for Health IT. HealthIT.gov. URL: https://www.healthit.gov/topic/federal-advisory-committees/recommendations-national-coordinator-health-it [accessed 2019-06-11]

11. Wiljer D, Urowitz S, Apatu E, DeLenardo C, Eysenbach G, Harth T, Canadian Committee for Patient Accessible Health Records. Patient accessible electronic health records: Exploring recommendations for successful implementation strategies. J Med Internet Res 2008 Oct 31;10(4):e34 [FREE Full text] [doi: 10.2196/jmir.1061] [Medline: 18974036]
12. Leonard KJ, Wiljer D. Patients are destined to manage their care. Healthc Q 2007;10(3):76-78 [FREE Full text] [doi: 10.12927/hcq.2007.18928] [Medline: 17626550]
13. Institute of Medicine (US) Committee on Quality Assurance and Accreditation Guidelines for Managed Behavioral Health Care. Challenges in delivery of behavioral health care. In: Edmunds M, Frank R, Hogan M, McCarty D, Robinson-Beale R, Weisner C, editors. Managing Managed Care: Quality Improvement in Behavioral Health. Washington, DC: The National Academies Press; 1997.
14. Mental health by the numbers. NAMI: National Alliance on Mental Illness. URL: https://www.nami.org/Learn-More/Mental-Health-By-the-Numbers [accessed 2019-03-19]
15. Grando A, Ivanova J, Hiestand M, Soni H, Murcko A, Saks M, et al. Mental health professional perspectives on health data sharing: Mixed methods study. Health Informatics J 2020 Sep;26(3):2067-2082 [FREE Full text] [doi: 10.1177/1460458219893848] [Medline: 31928295]
16. Kim J, Kim H, Bell E, Bath T, Paul P, Pham A, et al. Patient perspectives about decisions to share medical data and biospecimens for research. JAMA Netw Open 2019 Aug 02;2(8):e199550 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.9550] [Medline: 31433479]
17. Kim H, Bell E, Kim J, Sitapati A, Ramsdell J, Farcas C, et al. iCONCUR: Informed consent for clinical data and bio-sample use for research. J Am Med Inform Assoc 2017 Mar 01;24(2):380-387 [FREE Full text] [doi: 10.1093/jamia/ocw115] [Medline: 27589942]
18. Bell SK, Mejilla R, Anselmo M, Darer J, Elmore J, Leveille S, et al. When doctors share visit notes with patients: A study of patient and doctor perceptions of documentation errors, safety opportunities and the patient-doctor relationship. BMJ Qual Saf 2017 Apr;26(4):262-270 [FREE Full text] [doi: 10.1136/bmjqs-2015-004697] [Medline: 27193032]
19. Grando MA, Murcko A, Mahankali S, Saks M, Zent M, Chern D, et al. A study to elicit behavioral health patients' and providers' opinions on health records consent. J Law Med Ethics 2017 Jun;45(2):238-259 [FREE Full text] [doi: 10.1177/1073110517720653] [Medline: 30976154]
20. Grando A, Schwab R. Building and evaluating an ontology-based tool for reasoning about consent permission. In: Proceedings of the AMIA Annual Symposium. 2013 Presented at: AMIA Annual Symposium; November 16-20, 2013; Washington, DC p. 514-523 URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3900195/pdf/amia_2013_symposium_514.pdf
21. Soni H, Grando A, Murcko A, Bayuk M, Chandrashekar P, Mukundan M, et al. Current state of electronic consent processes in behavioral health: Outcomes from an observational study. In: Proceedings of the AMIA Annual Symposium. 2018 Presented at: AMIA Annual Symposium; November 3-7, 2018; San Francisco, CA p. 1607-1616 URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5977724/pdf/2731718.pdf
22. Caine K, Kohn S, Lawrence C, Hanania R, Meslin E, Tierney W. Designing a patient-centered user interface for access decisions about EHR data: Implications from patient interviews. J Gen Intern Med 2015 Jan;30 Suppl 1:S7-S16 [FREE Full text] [doi: 10.1007/s11606-014-3049-9] [Medline: 25480719]
23. Soni H, Ivanova J, Grando A, Murcko A, Chern D, Dye C, et al. A pilot comparison of medical records sensitivity perspectives of patients with behavioral health conditions and healthcare providers. Health Informatics J 2021;27(2):1-19 [FREE Full text] [doi: 10.1177/14604582211009925] [Medline: 33878989]
24. Schwartz P, Caine K, Alpert S, Meslin E, Carroll A, Tierney W. Patient preferences in controlling access to their electronic health records: A prospective cohort study in primary care. J Gen Intern Med 2015 Jan;30 Suppl 1:S25-S30 [FREE Full text] [doi: 10.1007/s11606-014-3054-z] [Medline: 25480721]
25. Campos-Castillo C, Anthony DL. The double-edged sword of electronic health records: Implications for patient disclosure. J Am Med Inform Assoc 2015 Apr;22(e1):e130-e140. [doi: 10.1136/amiajnl-2014-002804] [Medline: 25059953]
26. Tierney W, Alpert S, Byrket A, Caine K, Leventhal J, Meslin E, et al. Provider responses to patients controlling access to their electronic health records: A prospective cohort study in primary care. J Gen Intern Med 2015 Jan;30 Suppl 1:S31-S37 [FREE Full text] [doi: 10.1007/s11606-014-3053-0] [Medline: 25480720]
27. Ivanova J, Tang T, Idouraine N, Murcko A, Whitfield MJ, Dye C, et al. Behavioral Health Professionals' Perceptions of Patient-Controlled Granular Information Sharing (Part 2): Focus Group Study. JMIR Ment Health 2022;9(4):e18792 [FREE Full text] [doi: 10.2196/18792]
28. Morrison E, George V, Mosqueda L. Primary care for adults with physical disabilities: Perceptions from consumer and provider focus groups. Fam Med 2008 Oct;40(9):645-651 [FREE Full text] [Medline: 18830840]
29. Witry M, Doucette W, Daly J, Levy B, Chrischilles E. Family physician perceptions of personal health records. Perspect Health Inf Manag 2010 Jan 01;7:1d [FREE Full text] [Medline: 20697465]
30. King T, Brankovic L, Gillard P. Perspectives of Australian adults about protecting the privacy of their health information in statistical databases. Int J Med Inform 2012 Apr;81(4):279-289. [doi: 10.1016/j.ijmedinf.2012.01.005] [Medline: 22306206]
31. Mundfrom D, Shaw D, Ke T. Minimum sample size recommendations for conducting factor analyses. Int J Test 2005 Jun;5(2):159-168. [doi: 10.1207/s15327574ijt0502_4]

32. Williams B, Onsman A, Brown T. Exploratory factor analysis: A five-step guide for novices. Australas J Paramed 2010 Aug 02;8(3):990399 [FREE Full text] [doi: 10.33151/ajp.8.3.93]
33. Henson R, Roberts J. Use of exploratory factor analysis in published research. Educ Psychol Meas 2016 Jul 02;66(3):393-416. [doi: 10.1177/0013164405282485]
34. George D, Mallery P. SPSS for Windows Step by Step: A Simple Guide and Reference, 10.0 Update. 3rd edition. Boston, MA: Allyn & Bacon, Inc; 2000.
35. Bernard HR. Research Methods in Anthropology: Qualitative and Quantitative Approaches. 5th edition. Lanham, MD: Rowman Altamira; 2011.
36. Simonian SJ, Tarnowski KJ, Park A, Bekeny P. Child, parent, and physician perceived satisfaction with pediatric outpatient visits. J Dev Behav Pediatr 1993;14(1):8-12. [doi: 10.1097/00004703-199302000-00002]
37. Fitzgerald J, Stansfield R, Tang T, Oh M, Frohna A, Armbruster B, et al. Patient and provider perceptions of diabetes: Measuring and evaluating differences. Patient Educ Couns 2008 Jan;70(1):118-125 [FREE Full text] [doi: 10.1016/j.pec.2007.09.005] [Medline: 17997265]
38. Saks M, Grando A, Murcko A, Millea C. Granular patient control of personal health information: Federal and state law considerations. Jurimetrics 2018;58(4):411-435 [FREE Full text] [Medline: 31798215]
39. Roth C, Lim Y, Pevnick J, Asch S, McGlynn E. The challenge of measuring quality of care from the electronic health record. Am J Med Qual 2009;24(5):385-394. [doi: 10.1177/1062860609336627] [Medline: 19482968]
40. Montero-Marín J, Prado-Abril J, Botella C, Mayoral-Cleries F, Baños R, Herrera-Mercadal P, et al. Expectations among patients and health professionals regarding web-based interventions for depression in primary care: A qualitative study. J Med Internet Res 2015 Mar 10;17(3):e67 [FREE Full text] [doi: 10.2196/jmir.3985] [Medline: 25757358]
41. Transcribe. URL: https://transcribe.wreally.com/ [accessed 2019-10-25]
42. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
43. Schoenthaler A, Montague E, Baier Manwell L, Brown R, Schwartz M, Linzer M. Patient-physician racial/ethnic concordance and blood pressure control: The role of trust and medication adherence. Ethn Health 2014;19(5):565-578 [FREE Full text] [doi: 10.1080/13557858.2013.857764] [Medline: 24266617]
44. Schneider I, Konijn E, Righetti F, Rusbult C. A healthy dose of trust: The relationship between interpersonal trust and health. Pers Relatsh 2011;18:668. [doi: 10.1111/j.1475-6811.2010.01338.x]
45. Pugh PM, Perrin P, Rybarczyk B, Tan J. Racism, mental health, healthcare provider trust, and medication adherence among Black patients in safety-net primary care. J Clin Psychol Med Settings 2021 Mar;28(1):181-190. [doi: 10.1007/s10880-020-09702-y] [Medline: 32008136]
46. Kisekka V, Giboney J. The effectiveness of health care information technologies: Evaluation of trust, security beliefs, and privacy as determinants of health care outcomes. J Med Internet Res 2018 Apr 11;20(4):e107 [FREE Full text] [doi: 10.2196/jmir.9014] [Medline: 29643052]
47. Esmaeilzadeh P. The impacts of the perceived transparency of privacy policies and trust in providers for building trust in health information exchange: Empirical study. JMIR Med Inform 2019 Nov 26;7(4):e14050 [FREE Full text] [doi: 10.2196/14050] [Medline: 31769757]
48. Pitułaj A, Kiejna A, Dominiak M. Negative synergy of mental disorders and oral diseases versus general health. Dent Med Probl 2019;56(2):197-201 [FREE Full text] [doi: 10.17219/dmp/105253] [Medline: 31099501]
49. Knopf A. Confidentiality of substance use disorder treatment records: Update. Brown Univ Child Adolesc Psychopharmacol Update 2020 Jun;22(6):1-3. [doi: 10.1002/cpu.30496]
50. Wright A, Soran C, Jenter C, Volk L, Bates D, Simon S. Physician attitudes toward health information exchange: Results of a statewide survey. J Am Med Inform Assoc 2010 Jan 01;17(1):66-70 [FREE Full text] [doi: 10.1197/jamia.m3241]
51. Patel V, Abramson E, Edwards A, Malhotra S, Kaushal R. Physicians' potential use and preferences related to health information exchange. Int J Med Inform 2011 Mar;80(3):171-180. [doi: 10.1016/j.ijmedinf.2010.11.008] [Medline: 21156351]
52. Mays V, Jones A, Delany-Brumsey A, Coles C, Cochran S. Perceived discrimination in healthcare and mental health/substance abuse treatment among Blacks, Latinos, and Whites. Med Care 2017;55:173. [doi: 10.1097/mlr.0000000000000638]
53. Ross C, Goldner E. Stigma, negative attitudes and discrimination towards mental illness within the nursing profession: A review of the literature. J Psychiatr Ment Health Nurs 2009 Aug;16(6):558-567. [doi: 10.1111/j.1365-2850.2009.01399.x] [Medline: 19594679]
54. Knaak S, Mantler E, Szeto A. Mental illness-related stigma in healthcare: Barriers to access and care and evidence-based solutions. Healthc Manage Forum 2017 Mar;30(2):111-116 [FREE Full text] [doi: 10.1177/0840470416679413] [Medline: 28929889]
55. Khatib A, Farraj F. Differences in the level of trust in mental health providers among individuals with severe mental illnesses. Int J Health Sci 2014;2(4):33-46. [doi: 10.15640/ijhs.v2n4a3]
56. Maura J, Weisman de Mamani A. Mental health disparities, treatment engagement, and attrition among racial/ethnic minorities with severe mental illness: A review. J Clin Psychol Med Settings 2017 Dec;24(3-4):187-210. [doi: 10.1007/s10880-017-9510-2] [Medline: 28900779]

57. Meslin E, Alpert S, Carroll A, Odell J, Tierney W, Schwartz P. Giving patients granular control of personal health information: Using an ethics 'Points to Consider' to inform informatics system designers. Int J Med Inform 2013 Dec;82(12):1136-1143 [FREE Full text] [doi: 10.1016/j.ijmedinf.2013.08.010] [Medline: 24139626]

Abbreviations

CARES: Coronavirus Aid, Relief, and Economic Security
CFR: Code of Federal Regulations
EHR: electronic health record
GBH: general behavioral health
PC-GIS: patient-controlled granular information sharing
SMI: serious mental illness

Edited by G Eysenbach, R Kukafka; submitted 15.06.20; peer-reviewed by E Ford, C Fincham; comments to author 12.08.20; revised version received 17.10.20; accepted 28.09.21; published 20.04.22.

Please cite as:
Ivanova J, Tang T, Idouraine N, Murcko A, Whitfield MJ, Dye C, Chern D, Grando A
Behavioral Health Professionals' Perceptions on Patient-Controlled Granular Information Sharing (Part 1): Focus Group Study
JMIR Ment Health 2022;9(4):e21208
URL: https://mental.jmir.org/2022/4/e21208
doi:10.2196/21208
PMID:35442199

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org