frontiers | Research Topics

Digital mental health research: Understanding participant engagement and need for user-centered assessment and interventional digital tools

Edited by
Abhishek Pratap, Patricia A. Arean, Benjamin Nelson, Brenna Renn and Abigail Ortiz

Published in
Frontiers in Psychiatry
Frontiers in Digital Health

ISSN 1664-8714
ISBN 978-2-8325-3195-2
DOI 10.3389/978-2-8325-3195-2

About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

Digital mental health research: Understanding participant engagement and need for user-centered assessment and interventional digital tools

Topic editors

Abhishek Pratap — University of Toronto, Canada
Patricia A. Arean — University of Washington, United States
Benjamin Nelson — Meru Health Inc., United States
Brenna Renn — University of Nevada, Las Vegas, United States
Abigail Ortiz — University of Toronto, Canada

Citation

Pratap, A., Arean, P. A., Nelson, B., Renn, B., Ortiz, A., eds. (2023). *Digital mental health research: Understanding participant engagement and need for user-centered assessment and interventional digital tools*.
Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-3195-2

Table of contents

05 HEARTSMAP-U: Adapting a Psychosocial Self-Screening and Resource Navigation Support Tool for Use by Post-secondary Students
Punit Virk, Ravia Arora, Heather Burt, Anne Gadermann, Skye Barbic, Marna Nelson, Jana Davidson, Peter Cornish and Quynh Doan

23 Delivering Remote Measurement-Based Care in Community Addiction Treatment: Engagement and Usability Over a 6-Month Clinical Pilot
Kevin A. Hallgren, Eliza B. Cohn, Richard K. Ries and David C. Atkins

36 mHealth Solutions for Mental Health Screening and Diagnosis: A Review of App User Perspectives Using Sentiment and Thematic Analysis
Erin Lucy Funnell, Benedetta Spadaro, Nayra Martin-Key, Tim Metcalfe and Sabine Bahn

53 A Mixed-Methods Analysis of Mobile ACT Responses From Two Cohorts
Sydney Hoel, Amanda Victory, Tijana Sagorac Gruichich, Zachary N. Stowe, Melvin G. McInnis, Amy Cochran and Emily B. K. Thomas

66 Digital Strategies to Accelerate Help-Seeking in Youth With Psychiatric Concerns in New York State
Michael L. Birnbaum, Chantel Garrett, Amit Baumel, Nicole T. Germano, Cynthia Lee, Danny Sosa, Hong Ngo, Kira H. Fox, Lisa Dixon and John M. Kane

80 Building Emotional Awareness and Mental Health (BEAM): A Pilot Randomized Controlled Trial of an App-Based Program for Mothers of Toddlers
Anna L. MacKinnon, Kaeley M. Simpson, Marlee R. Salisbury, Janelle Bobula, Lara Penner-Goeke, Lindsay Berard, Charlie Rioux, Gerald F. Giesbrecht, Ryan Giuliano, Catherine Lebel, Jennifer L. P. Protudjer, Kristin Reynolds, Shannon Sauer-Zavala, Melanie Soderstrom, Lianne M. Tomfohr-Madsen and Leslie E. Roos

96 Feasibility, engagement, and preliminary clinical outcomes of a digital biodata-driven intervention for anxiety and depression
Charalampos Tsirmpas, Dimitrios Andrikopoulos, Panagiotis Fatouros, Georgios Eleftheriou, Joaquin A. Anguera, Konstantinos Kontoangelos and Charalabos Papageorgiou

117 Digital therapeutics for mental health: Is attrition the Achilles heel?
Adaora Nwosu, Samantha Boardman, Mustafa M. Husain and P. Murali Doraiswamy

122 Predicting symptom response and engagement in a digital intervention among individuals with schizophrenia and related psychoses
George D. Price, Michael V. Heinz, Matthew D. Nemesure, Jason McFadden and Nicholas C. Jacobson

132 Meaningful engagement: A crossfunctional framework for digital therapeutics
Gabriel Strauss, Jessica E. Flannery, Elise Vierra, Xin Koepsell, Emily Berglund, Ian Miller and Jessica I. Lake

140 Acceptability of virtual therapy for postpartum women during COVID-19: A national mixed methods study
Carmen Gonzalez, Magaly Ramirez, Felicia Mata-Greve, Autumn Diaz, Miriana C. Duran, Morgan Johnson, Nancy Grote and Patricia A. Areán

148 The design and development of an experience measure for a peer community moderated forum in a digital mental health service
Charlotte Mindel, Lily Mainstone-Cotton, Santiago de Ossorno Garcia, Aaron Sefi, Georgia Sugarman, Louisa Salhi, Holly Brick, Katherine Jackson and Terry Hanley

170 Therapeutic components of digital counseling for chronic heart failure
Gabriel C. Fezza, Stephanie Sansone and Robert P. Nolan

frontiers in Psychiatry

ORIGINAL RESEARCH
published: 22 February 2022
doi: 10.3389/fpsyt.2022.812965

HEARTSMAP-U: Adapting a Psychosocial Self-Screening and Resource Navigation Support Tool for Use by Post-secondary Students

Punit Virk[1,2], Ravia Arora[2], Heather Burt[1,2], Anne Gadermann[1,2,3], Skye Barbic[3,4], Marna Nelson[5], Jana Davidson[2,6], Peter Cornish[7] and Quynh Doan[1,2,8]*

[1] School of Population and Public Health, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada, [2] BC Children's Hospital Research Institute, Vancouver, BC, Canada, [3] Centre for Health Evaluation and Outcome Sciences, Providence Health, Vancouver, BC, Canada, [4] Department of Occupational Science and Occupational Therapy, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada, [5] Student Health Service, University of British Columbia, Vancouver, BC, Canada, [6] Division of Child and Adolescent Psychiatry, Department of Psychiatry, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada, [7] Student Counselling Services, University of California, Berkeley, Berkeley, CA, United States, [8] Division of Emergency Medicine, Department of Pediatrics, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

OPEN ACCESS

Edited by:
Abhishek Pratap, Centre for Addiction and Mental Health, Canada

Reviewed by:
Lauren McGillivray, University of New South Wales, Australia
Kim Mathiasen, University of Southern Denmark, Denmark

***Correspondence:**
Punit Virk
pvirk@bcchr.ca

Specialty section:
This article was submitted to Digital Mental Health, a section of the journal Frontiers in Psychiatry

Received: *10 November 2021*
Accepted: *24 January 2022*
Published: *22 February 2022*

Citation:
Virk P, Arora R, Burt H, Gadermann A, Barbic S, Nelson M, Davidson J, Cornish P and Doan Q (2022) HEARTSMAP-U: Adapting a Psychosocial Self-Screening and Resource Navigation Support Tool for Use by Post-secondary Students. Front. Psychiatry 13:812965. doi: 10.3389/fpsyt.2022.812965

Background: Mental health challenges are highly prevalent in the post-secondary educational setting. Screening instruments have been shown to improve early detection and intervention. However, these tools often focus on specific diagnosable conditions, are not always designed with students in mind, and lack resource navigational support.

Objective: The aim of this study was to describe the adaptation of existing psychosocial assessment (HEARTSMAP) tools into a version that is fit-for-purpose for post-secondary students, called HEARTSMAP-U.

Methods: We underwent a three-phase, multi-method tool adaptation process. First, a diverse study team proposed a preliminary version of HEARTSMAP-U and its conceptual framework. Second, we conducted a cross-sectional expert review study with Canadian mental health professionals ($N = 28$), to evaluate the clinical validity of tool content. Third, we conducted an iterative series of six focus groups with diverse post-secondary students ($N = 54$), to refine tool content and language, and ensure comprehensibility and relevance to end-users.

Results: The adaptation process resulted in the HEARTSMAP-U self-assessment and resource navigational support tool, which evaluates psychosocial challenges across 10 sections. In Phase two, clinician experts expressed that HEARTSMAP-U's content aligned with their own professional experiences working with students. In Phase three, students identified multiple opportunities to improve the tool's end-user relevance by calling for more "common language," such as including examples, definitions, and avoiding technical jargon.

Conclusions: The HEARTSMAP-U tool is well-positioned for further studies of its quantitative psychometric properties and clinical utility in the post-secondary educational setting.

Keywords: mental health, screening, validity, post-secondary students, focus groups

INTRODUCTION

In recent years, post-secondary students have reported increasing levels of mental health challenges including psychological distress and diagnosed conditions (e.g., anxiety, depression) (1). While the post-secondary years are often a period of self-exploration and interpersonal growth (2), they have also been associated with high stress, peer pressure, and greater responsibilities with reduced social supports (3, 4). For young adults, this period coincides with significant physiological, psychological, and social development (5, 6). In 2019, Canadian data from the National College Health Assessment ($N = 55{,}284$) showed that, within the last 12-months, most post-secondary students reported experiencing overwhelming anxiety (68.9%) and at least half reported functionally impairing depression (51.6%) (7). Among the sample, 16.4% of students endorsed active suicidal ideation in the last 12-months, compared to 2.5% of the general Canadian adult population and 6% of young adults (ages 15–24 years) that same year (8, 9). During the COVID-19 pandemic, the rate of mental health concerns escalated in the student population, one study ($N = 1{,}388$) reported a 30-day anxiety and/or depressive symptom prevalence of 75% among Canadian students during the pandemic's first-wave (up till May 2020) (10). Similarly, the Healthy Minds survey ($N = 18{,}764$) saw increased prevalence of depression and lower levels of resiliency among American students compared to pre-pandemic estimates (11). The pandemic has compounded psychological and social challenges (psychosocial stressors) (12, 13), magnifying an already severe campus mental health crisis (14).

Students experience individual- and system-level barriers that may impede timely access to age-appropriate care. Low mental health literacy, poor system navigation support, and service saturation (e.g., wait-times) all impede help-seeking (15–18). National Canadian standards for student mental health and well-being call for institutions to have early identification and preventative infrastructures (19), which can improve long-term mental health outcomes and timely connectedness into services (20). Universal mental health screening and navigational support tools can address challenges institutions experience with identifying mental health concerns and supporting connectivity to care. Such measures have been successfully integrated within post-secondary health systems (21–23). Digital screening tools may alleviate the need for in-person intake assessment/triaging and more seamlessly bridge in-person and digital resources (3, 24, 25). Digital self-reporting of psychosocial challenges also shows higher disclosure rates and may be preferred over clinician-administered or paper-based assessment (26–28), offering users privacy, time, and space to articulate needs.

Notwithstanding the potential of screening, existing scales often focus exclusively on common psychological issues, such as the PHQ-9 (depression), GAD-7 (anxiety), AUDIT (substance use), and SBQ-R (suicidality) (29–32). These tools are diagnoses-specific, have not been developed with student engagement, and generally lack comprehensive validity evidence in student populations (33–36). However, several instruments have been developed or adapted with students' unique contextual (e.g., academic stress, social autonomy) and clinical needs (e.g., emerging adulthood) in mind. Downs et al. (37) previously developed the 34-item Symptoms and Assets Screening Scale specifically for college students to self-screen on common mental health challenges (e.g., eating disorder, substance abuse, anxiety, depressive symptoms) and generalized distress (37). Similarly, Alschuler et al. (38) developed the 11-item College Health Questionnaire, which facilitates behavioral screening of psychological (e.g., anxiety, depression) and social concerns (e.g., academic problems, relationships, finances) (38). Other post-secondary-specific screening and assessment measures include the Counseling Center Assessment of Psychological Symptoms and the Mental Health Continuum model. However, these assessment tools lack an actionable, resource navigational component, which may support students' help-seeking and contribute to the utility of screening (39–41).

Our team has previously developed, validated, and implemented psychosocial instruments for the pediatric population. The clinical HEARTSMAP assessment and management guiding tool supports pediatric acute care providers with psychosocial interviewing and disposition planning (42). MyHEARTSMAP is a self-administered version allowing self-/proxy-screening, to facilitate universal screening by youth and parents (43). Both instruments have demonstrated evidence for strong psychometric properties (42–45), high clinical utility (46, 47), and user acceptability (48). These instruments expand on the seminal HEADSS psychosocial interview and history-taking tool (49, 50) and assess ten broad psychosocial sections: Home, Education and activities, Alcohol and drugs, Relationships and bullying, Thoughts and anxiety, Safety, Sexual Health, Mood, Abuse, and Professional resources. These psychosocial issues are clinically significant and theoretically supported within human development and socio-ecological models. According to Maslow's Hierarchy of Needs, individuals work up from physiological (e.g., Home, Safety) and psychological needs (e.g., "Relationships") toward self-fulfillment-oriented needs (e.g., "Education and activities) (51, 52). Within socio-ecological models, these psychosocial areas demonstrate how youths' mental well-being is shaped through the interplay of individual (e.g., Mood, Thoughts and anxiety, Safety risk), interpersonal (e.g., Relationships, Abuse, Sexual Health), institutional (e.g., Education and activities), and community factors (e.g., Professionals and resources) (19, 53, 54). We provide further details on the HEARTSMAP tools' measurement model, assessment structure, and resource recommendation decision-making algorithm in **Web-Appendix A**.

Adapted specifically for post-secondary students, HEARTSMAP-U is a brief, digital self-administered psychosocial screening tool. Similar to previous HEARTSMAP versions, HEARTSMAP-U assesses ten psychosocial areas ranging from Housing to Abuse. For each section, students first score their concerns on a 4-point Likert-type scale ranging from 0 (no concern) to 3 (severe concern), using anchor descriptions for each scoring option. Second, student's score whether they have previously accessed services pertaining to this section (yes/no). After students have answered these questions for all 10 sections,

their responses feed into a built-in algorithm, triggering urgency-specific resource recommendations for identified mental health needs (13, 16, 19).

The current paper describes the three-phase process by which previously developed HEARTSMAP tools were adapted into HEARTSMAP-U, a version that is fit-for-purpose for the post-secondary student population, to help students self-identify psychosocial support needs. Our study will serve as a foundational paper on the HEARTSMAP-U tool and its preliminary adaptation. We will collect multi-faceted evidence of instrument validity and reliability in an ongoing manner and report it in later studies.

METHODS

Our tool adaptation process includes three phases and has been informed by established guidelines for developing patient-reported outcome measures in the literature (55–58), and expertise from diverse stakeholders including clinical experts and student end-users. We used an iterative, multi-method approach, outlined in **Figure 1**. For each phase we describe the design, study procedures, and analytic approach. We obtained approval from our institutional research ethics board for Phase two and three, in which research participants were recruited.

Phase One: Collaborative Working Meetings

Design

We conducted virtual working meetings between November 2018 and April 2019 with a diversely assembled study team of students and co-investigators. The purpose of our one-on-one student consultations was to generate ideas on how HEARTSMAP-U needs to be adapted for fitness for purpose in the university context, through a collaborative and consensus-based process. Our co-investigators included a family physician, clinical psychologist, a youth psychiatrist, addiction psychiatrist, patient-reported outcome measurement expert, and a graduate student researcher. The purpose of our co-investigator meetings was to formalize HEARTSMAP-U's intended use and conceptual framework. This included ensuring the tool assessed relevant psychosocial stressors (e.g., student-specific, age-related), and that its resource recommendations were accessible and match desired clinical flow (e.g., how/when specific supports should be accessed).

Study Procedure

Prior to co-investigator meetings, we had a group of gender and racially diverse research students (medical/undergrad/ graduate) review the pediatric MyHEARTSMAP tool and change language and content to be suitable for the post-secondary student population. We did not put restrictions or parameters on student researchers proposed modifications. This exercise resulted in the first HEARTSMAP-U version.

Co-investigators used the first HEARTSMAP-U version and existing HEARTSMAP conceptual framework as a starting point for tool modification. Discussions were free-flowing and open-ended, and investigators' feedback/suggestions were not constrained to the measurement model and conceptual framework of existing HEARTSMAP tools. We used a consensus-based decision-making process. Proposed tool changes required 100% investigator consensus. When we could not reach consensus, we held discussions until all investigators came to agreement. The lead investigator (PV) took comprehensive notes documenting all team decision-making and made approved tool modifications between each meeting. We held meetings until the team collectively felt a clinically and contextually relevant tool version had been reached.

Analysis

Throughout all meetings, we summarized and reported general impressions and key discussion points. We made necessary tool modifications between co-investigator team meetings.

Phase Two: Clinical Expert Review

Design

We conducted a cross-sectional survey study with Canadian mental health clinicians who support post-secondary students, guided by an expert review methodology (60).

Study Recruitment

We recruited a convenience and snowball sample of participants through our professional networks, until data saturation was reached. Participation was self-paced and took place remotely, over our secure study website from July 2019 to September 2019.

Study Procedure

Participants watched a mandatory 3-min instructional video, explaining study procedures, the digital platform, and HEARTSMAP-U (purpose, structure). Second, we asked participants to reflect on their professional experience and formulate a fictional clinical vignette describing a student presenting to their practice in psychosocial distress (mild to severe). Clinicians were expected to provide a brief description of their vignette and used this information as they progressed throughout the tool.

Next, for each tool section, clinicians reviewed all HEARTSMAP-U guiding questions and scoring criteria, scored their fictional students' concerns (if any), and completed a survey item asking *"Do HEARTSMAP-U's [guiding questions/scoring descriptors] sufficiently capture the full range of [section]-related stressors that youth in your practice might experience? (yes/no)"* As a follow-up item, irrespective of their prior response, all participants were asked to provide a qualitative response to *"what could be added or changed so the [guiding questions/scoring descriptors] better capture the range of concerns students may experience in relation to [section]?"* Clinicians also provided high-level feedback (e.g., tool impressions, content suggestions). All qualitative responses were collected through open-ended survey questions (textbox response).

After scoring all sections, clinicians reviewed tool-generated support recommendations and assessed whether they over- or underestimated fictional students' needs. Clinicians had the choice of completing a second evaluation with a new vignette. Upon study completion, the core research team analyzed all

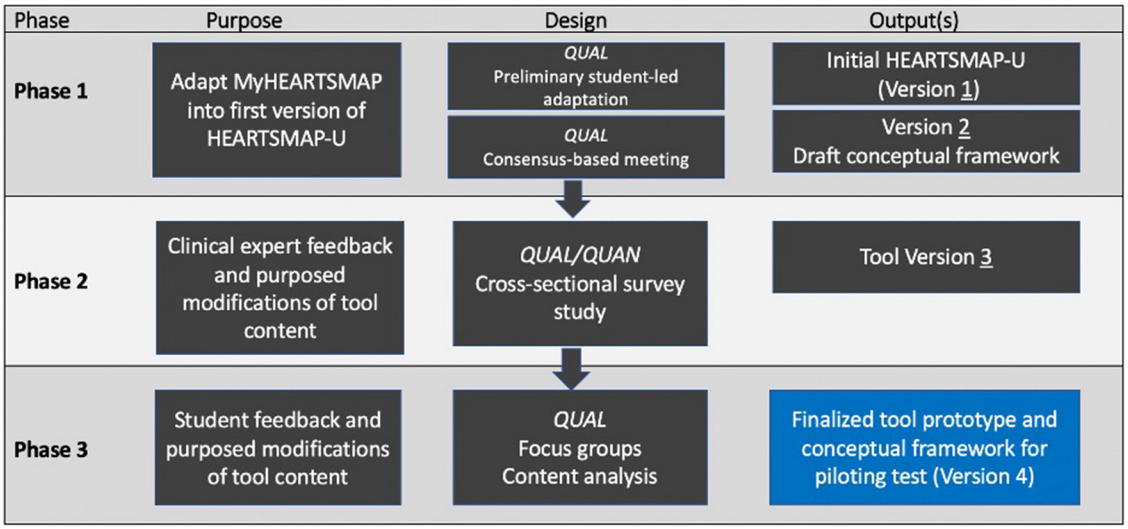


FIGURE 1 | Schematic outlining our reported multiphasic tool adaptation process. Figure adapted from Riff et al. (59). QUAN, quantitative; QUAL-qualitative.

feedback and found opportunities to further adapt each tool section (e.g., content, language), to ensure it covers a full range of concern severity, both in terms of distress and functional impairment. We used the HEARTSMAP-U version resulting from Phase two as a starting point for Phase three student focus group discussions.

Analysis

We summarized clinician demographics and responses to dichotomous survey items (yes/no) as counts and proportions. A blended/abductive approach to qualitative content analysis was taken to synthesize and analyze all qualitative responses (61). Based on an initial, holistic exploration of the raw data (inductive process) and existing healthcare measurement literature (deductive process) (62), we developed a tentative coding framework that would encompass participants' qualitative responses (e.g., content coverage, context of use, etc.). We coded qualitative data in three cycles, each introducing an added layer of interpretation and data abstraction. Our research team used reflective memos documented throughout the data collection stage to support the coding process and interpretation. First, we conducted attribute coding, whereby all qualitative survey feedback was structurally coded and organized by tool section, to support feedback interpretation. Second, we conducted descriptive coding and, for each tool section, mapped all clinician responses to our pre-defined coding framework/categories. We separately analyzed and coded guiding question and scoring descriptor feedback. Third, we performed pattern coding to explore variations and sub-categories within existing codes. For clinicians who responded "no" to whether guiding questions and/or scoring descriptors aligned with their professional experience, we coded their qualitative responses into the most appropriate feedback category. For each tool section and feedback category, we report count data on the total number of clinicians/responses that map to them. Two investigators conducted qualitative coding, HB (first cycle) and PV (second, third cycle). We conducted analyses using Microsoft Office Excel and NVivo 12.0.

Phase Three: Student Focus Groups

Design

We conducted a qualitative study with UBC-Vancouver students, guided by cognitive testing and iterative design methodologies (63–65). Similar to Phase two, we incorporated a variation of verbal probing, asking participants targeted questions on tool content and functionality. Through a series of sequential focus groups, we iteratively modified HEARTSMAP-U based on participants' feedback on guiding questions, scoring criteria, tool language (e.g., unclear, insensitive), and other suggestions (e.g., new tool section, format/structure). Focus groups took place between November 2019 and May 2020. Initially, we held in-person sessions, but later made them virtual, to allow remote participation and compliance with COVID-19 restrictions.

Study Recruitment

We recruited students through an existing partnership with university administration, health centers, and student organizations. Prospective participants completed an online expression of interest and demographic form. Using this information, we recruited a purposive sample of UBC-Vancouver students ages 17 years and older and setup heterogenous focus groups. We strived for proportional representation of the overall UBC student population across demographics: age, gender and sexual identity, program-type, year of study, race/ethnicity, international/domestic status, and lived mental health experiences (**Web-Appendix B**). We excluded students uncomfortable with being audiotaped.

Study Procedures

During each focus group, we first supplied participants with a high-level introduction to HEARTSMAP-U (e.g., purpose, components). Next, we reviewed tool components (guiding questions, scoring criteria), for each tool section. During this time, we asked participants to share their first impressions and engage in a dialogue around the tool's (1) comprehensiveness (*issues important to you and your peers*). (2) Relevance (*realistic content reflecting your experiences).* (3) Understandability (*easily understood language).* We encouraged participants to suggest tool modifications for the study team's consideration, either through group discussion or written feedback. We audiotaped focus groups, had them professionally transcribed (verbatim), and compared them against the original audio to confirm accuracy.

Analysis

We conducted two sets of analyses using focus group data. Consistent with analytic guidelines, we treat the focus group as our unit of analysis (66). First, between each focus group, the core research team reviewed RA notes documenting tool modifications proposed by students. For each comment or suggested modification, we took into consideration the general response from other focus group members (e.g., endorsed, objected) and whether it was consistent with clinical guidelines and earlier focus groups. Focus groups were held until a point of data saturation was achieved, whereby no new feedback was received that investigators had not already considered or considerations were mostly minor (e.g., word choice, grammar) (66).

After reaching sufficient data saturation, we performed an in-depth, abductive qualitative content analysis, with inductive and deductive components, using verbatim transcripts and research memos. First, an investigator (RA) deductively conducted attribute-based coding, to organize and sort all student comments by session and tool section. A second investigator (PV) interpretatively performed descriptive coding using Stewart et al.'s framework to categorize sectional feedback as either content or format/interface-related (67). Tool content-related feedback and modification suggestions were further analyzed through pattern coding using two additional frameworks. The COSMIN content validity framework and it's operational definitions for content relevance, representativeness, and understandability were used to analyze and characterize students' proposed modifications with respect to these categories (55). Coons et al.'s framework was used to assess modifications as either (1) minor, those not expected to change content or meaning (e.g., switching format from paper to online). (2) Moderate, subtle content/meaning changes (e.g., item wording, ordering). (3) Substantial, extensive content/ meaning changes (e.g., changing response options, new guiding questions) (68). Inductive, descriptive coding was also performed to characterize and report comments and feedback that did not fit within our a priori analytic frameworks.

For each tool section, we report representative quotes for each modification-type and inductively derived category, and reference quotes by focus group number (*FG X)*. We summarize participant sociodemographics using descriptive statistics and conduct the Chi-square test of independence (alpha = 0.05) to compare the demographic profile of participating students with those who expressed interest but did not take part in the study (e.g., not invited, declined).

RESULTS

Phase One: Collaborative Working Meetings

A total of five students took part in preliminary tool adaptation activities, two undergraduate students and three medical students. Subsequently, we had five co-investigators who took part in three rounds of discussion and iterative tool modification, at which time all co-investigators agreed on HEARTSMAP-U prototype content. One clinical investigator took part and contributed feedback outside of organized group discussions.

Conceptual Framework

We largely retained MyHEARTSMAP's conceptual framework, recognizing universality of the measured constructs, however several sections were redefined. MyHEARTSMAP's "Home" section only measures the safety and supportiveness of the home environment, which may not encompass the transient nature of student housing. For HEARTSMAP-U, we modified this section into "Housing arrangements and finances" to include an assessment of housing stability and ease of managing housing-related responsibilities (e.g., paying bills, cleaning, cooking, etc.), in addition to housing safety/supportiveness. Finalized construct definitions are reported in **Web-Appendix C** and our conceptual framework is illustrated in **Figure 2**.

Tool Content and Resource Recommendations

Investigators decided MyHEARTSMAP's severity scoring spectrum (none to severe) required modification to accurately reflect the student population. "Alcohol and drugs" needed to reflect the social acceptability of leisurely drinking and marijuana usage among young adults. For several sections, investigators agreed that two different concepts were being measured together (e.g., thought disturbances and anxiety) which needed to be consistently assessed and delineated across all severity levels using "OR" Boolean operators. The team modified HEARTSMAP-U's resource recommendations so that they reflected the appropriate tier of resources/services as outlined by the post-secondary institution (69). Investigators identified opportunities to incorporate strength-building recommendations, triggered when students report no more than mild concerns. Feedback across all three working group sessions is summarized by tool section in **Web-Appendix C**.

Phase Two: Clinical Expert Study

Participating mental health clinicians ($N = 28$) mostly identified as women (89%) and worked at large-size Canadian post-secondary institutions (96%). Most clinicians were either registered counselors (32.1%) or psychologists (32.1%) and affiliated with their institutions counseling (60.9%) and/or health services (30.4%). Complete demographic details are summarized in **Table 1**.

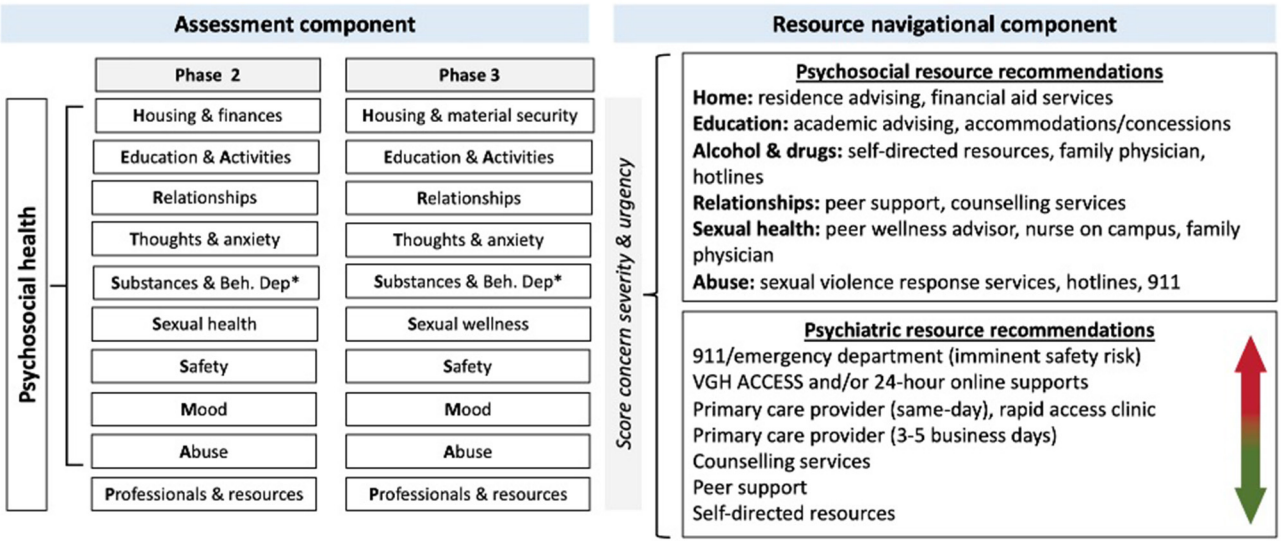


FIGURE 2 | Conceptual framework of the finalized HEARTSMAP-U tool version, following adaptation among Canadian mental health professionals (Phase 2) and post-secondary students (Phase 3).

TABLE 1 | Demographic characteristics of phase two participating clinicians.

Characteristics	N_{total} = 28 (%)
Gender, Female	25 (89.3%)
Provider type	
Psychologist	9 (32.1%)
Registered counselor	9 (32.1%)
Social workers	6 (21.4%)
Mental health nurse	2 (7.1%)
Family physician	2 (7.1%)
Campus provider, yes	28 (100)
Affiliated services[a]	
Counseling service	28 (60.9)
Student health services	14 (30.4)
More than one	4 (8.7)
Institution size[b]	
Medium	2 (7.1%)
Large	26 (92.9%)

[a] *Total proportion exceeds 100% as several clinicians held multiple affiliations.*
[b] *Large-size institutions were defined as those with a student population larger than 30,000, mid-size institutions were defined as having a student population between 10,000 and 30,000 students.*

Fictional Vignettes

Clinician-prepared vignettes scored across severity levels (0–3) for all sections, except "Alcohol and drugs" and "Abuse" which were only assessed on no (0) to moderate concerns (2). Of the 46 completed fictional cases, most described mild (46%) or moderate (44%) psychosocial concerns. Of cases reporting psychological challenges, participants assessed 20% as being severe, compared to only 2–4% of cases reporting on other psychosocial issue. A total of 18 (64%) clinicians decided to complete a second vignette evaluation and 17 (61%) expressed interest in referring a colleague to join the study.

Section-Level Review

Participating clinicians felt that HEARTSMAP-U's guiding questions (46–86%) and scoring criteria (54–82%) aligned with their own clinical characterization of each tool section. A majority felt the tool was *"very thorough,"* guiding questions were *"simple yet broad"* and scoring criteria were *"easy"* to understand and there was *"nothing to add."* Conversely, 14% (Housing; Professionals and resources) to 54% (Education and activities) and 18% (Housing; Abuse) to 46% (Sexual Health) of clinicians felt that HEARTSMAP-U's guiding questions and scoring descriptors, respectively, required more characterization to match their observations of each psychosocial construct. From clinicians' qualitative responses, we derived four categories that feedback was related to: (1) coverage of concern severity, consistent with the tool's intended use; (2) tool suitability in the clinician's own context-of-use; (3) minor language/wording issues with minimal impact on sectional content/meaning; and (4) content that clinicians perceived as missing but was elsewhere in the tool. We elaborate on each of these themes below. Counts and proportions summarizing participants' feedback by coding category are summarized in **Tables 2**, **3**, for guiding questions and scoring descriptors, respectively.

Sectional Content Coverage

Respectively, 23 and 39% of all guiding question and scoring criteria comments focused on how well sections captured behaviors and experiences necessary for students to be able to self-evaluate the presence of concerns, across the entire spectrum of severity. Two major sub-categories emerged from these comments: improving scale gradation and broadening content. Clinicians felt scoring descriptors needed to accommodate students who may fit *"in-between"* existing criteria. For example, in the "Relationships" section, one participating clinician suggested we:

TABLE 2 | A breakdown of phase two clinician's feedback on HEARTSMAP-U's guiding questions.

Total number of clinicians (N_{Total} = 28)	Guiding question alignment[a]		Clinician count per feedback category[b]				
Tool section	**Yes**	**No**	**Concept coverage**	**Context of use**	**Language**	**Covered elsewhere**	**Unclear**
Housing	24	4	1	3	0	0	0
Education and activities	13	15	3	6	0	6	0
Alcohol and drugs	18	10	2	7	0	1	0
Relationships	17	11	4	5	0	2	0
Thoughts and anxiety	17	11	2	4	0	5	0
Safety	16	12	4	0	1	6	1
Sexual health	14	14	0	7	1	6	0
Mood	14	14	5	3	1	5	0
Abuse	16	12	3	7	0	2	0
Professional and resource	24	4	1	2	1	0	0
Total[c]		**107**	**25**	**44**	**4**	**33**	**1**
			23%	**41%**	**4%**	**31%**	**1%**

Each count represents a unique clinician. Under each "feedback category", counts represent the number of unique clinician's whose qualitative response mapped to that respective category. Each clinician's qualitative response mapped to a single feedback category, based on the focus of their concerns.

[a] Clinician response to whether guiding questions captures the full range of section-related concerns seen in their own practice (yes/no).

[b] Number of clinicians who felt guiding questions did not align with their professional experience (responded "no"), stratified by the feedback category most closely relating to their comments/suggestions.

[c] Total counts and percentages of qualitative responses (i.e., clinicians) per feedback category.

TABLE 3 | A breakdown of phase two clinician's feedback on HEARTSMAP-U's scoring descriptors.

Total number of clinicians (N_{Total} =28)	Scoring descriptor alignment[a]		Clinician count per feedback category[b]				
Tool section	**Yes**	**No**	**Concept coverage**	**Context of use**	**Language**	**Covered elsewhere**	**Unclear**
Housing	23	5	2	0	0	3	0
Education and activities	16	12	3	4	0	5	0
Alcohol and drugs	19	9	5	4	1	0	0
Relationships	22	6	3	2	0	1	0
Thoughts and anxiety	22	6	1	2	1	2	0
Safety	20	8	2	3	2	3	0
Sexual health	15	13	2	5	0	6	0
Mood	18	10	6	2	0	0	2
Abuse	23	5	2	1	0	2	0
Professional and resource	19	9	6	0	3	0	0
Total[c]		**83**	**32**	**23**	**7**	**22**	**2**
			39%	**28%**	**8%**	**27%**	**2%**

Each count represents a unique clinician. Under each 'feedback category', counts represent the number of unique clinician's whose qualitative response mapped to that respective category. Each clinician's qualitative response mapped to a single feedback category, based on the focus of their concerns.

[a] Clinician response to whether scoring descriptors captures the full range of section-related concerns seen in their own practice (yes/no).

[b] Number of clinicians who felt scoring descriptors did not align with their professional experience (responded "no"), stratified by the feedback category most closely relating to their comments/suggestions.

[c] Total counts and percentages of qualitative responses (i.e., clinicians) per feedback category.

> "*Address [the] situation where someone is not losing connections but is working onbuilding confidence to have romantic connections.*"

We changed the score 1 descriptor to include instances where students may have emotionally supportive connections but may struggle to build or maintain them. For "Education and activities," clinicians indicated two instances where partial criteria could be met, and students may struggle to score themselves:

> *"Need to capture that mental health concerns are impacting academic performance, but student is still actively engaging in studies"*
> *"Need options that capture languishing in one area only. Academics and activities are separate constructs. You can be functioning in one and not in the other."*

We used an "OR" Boolean operator to create two scoring pathways across scores 1–3, distinguishing academics from other activities, and allowing students to select the most severe score applicable to their situation. Under score 1, we have also taken into consideration instances where students may be engaged in class, but their academic performance may be declining.

Feedback often focused on broadening certain criteria and guiding questions to encompass a larger cross-section of the general student population, examples include:

> *"What about behavioural addictions (e.g., gambling, gaming)"* (Alcohol and drugs)
> *"Include family relationships"* (Relationships)
> *"Financial abuse is not listed - some family members have taken a client's student loanand used it for themselves"* (Abuse)
> *"Needs to encompass more range of emotions - anger and shame in particular aremissing."* (Mood)

The "Alcohol and drugs" section was expanded to include additional substances (marijuana, prescribed medication, illicit substances) and behavioral addictions (e.g., excessive exercise or sex, gambling).

Context-of-Use

A sizable proportion of guiding questions (41%) and scoring criteria feedback (28%) focused on introducing a diagnostic level of detail and specificity to each section's content. Clinicians requested the tool assess sub-categories of its existing broad psychosocial areas. For example, in the "Relationships" section, clinicians felt that HEARTSMAP-U did not explore specific relationship types or problems and they proposed guiding questions that consider:

> *"Parental expectations to perform or excel impacting relationships, being able tocommunicate with one's parents."*
> *"Break-ups specifically"*
> *"Friends nearby, versus those only met through social media (and not physicallyavailable)"*

HEARTSMAP-U's general assessment of relationship challenges at mild, moderate, and severe levels is suitable for its intended use, as a multi-domain psychosocial screen. Feedback calling for added detail and subcategorization were deemed by the study team as most relevant to the clinician's own assessment context, rather than initial screening purposes. A lengthier tool may also reduce usability and increase respondent burden.

Content Covered Elsewhere and Language

A large portion of concerns raised with guiding question (31%) and scoring descriptors (27%) had already been addressed in different tool sections, that participants may not have yet reviewed. In the "Thoughts and anxiety" section, participants expressed that:

> *"I'm not sure if this is coming later in the questionnaire but adding more depressive symptom questions. Perhaps that will be in the mood section I haven't come to yet."*

Language-related concerns made up a small proportion of guiding questions (4%) and scoring descriptor feedback (8%). These comments flagged language that students may misinterpret or find confusing such as *"psychosocial," "intoxication,"* and *"intrusive thoughts."* In another instance, participants felt the tool's singular use of *"partner,"* may stigmatize students in polyamorous relationships.

Resource Recommendations

Participants rated the appropriateness of 265 triggered recommendations and perceived that most recommendations (70%) were consistent with the fictional students' support needs. A smaller portion of participants felt tool-generated recommendations underestimated (18%) or overestimated (12%) support needs. Participants also expressed concerns with recommending emergency services (e.g., 911, emergency department) in the absence of imminent safety risk. Rather, participants considered same-day primary care, rapid access clinics, and 24/7 e-counseling as appropriate supports.

Phase Three: Student Focus Groups

Demographic distributions did not significantly differ ($P > 0.05$) between participating students and those not invited to a focus group session (non-participating students). A total of 54 students took part in 6 focus group sessions, each 2 hours in length. We had nearly equal proportions of student's aged 18–21 (50%) and 22–25 years (48.1%). Approximately two-thirds of participants identified as female and undergraduate students. Most participants were in their first or second year (61.1%), living off-campus (57.4%), identified as straight (72.2%) and as part of a visible ethnic minority (53.7%). Over 80% reported experiencing mental health challenges in the past (72.2%) and/or present (55.6%). Demographic details are summarized in **Table 4**.

Earlier focus groups emphasized substantial content-related modifications (*FG 1-3*) relating to relevance and representativeness of HEARTSMAP-U's content. In later sessions, students raised mild/moderate content suggestions. The proportion of focus group participants engaged in group discussion remained consistent across sessions (**Web-Appendix D**).

Interface Modifications

Participants suggested multiple interface-related modifications summarized below. First, a privacy disclaimer at the beginning of the tool, so the user is aware of the scope, intended purpose, and confidentiality implications associated with completing HEARTSMAP-U. Second, a progress bar with a coordinated color scheme (e.g., green complete, orange=in-progress), to motivate users in completing the tool. Third, users should have the ability to download screening results to potentially share with their care provider, and that tool recommendations link to service information the user can directly act on. Finally, participants felt that pairing the tool with calendar apps would ease repeat screening and booking appointments.

TABLE 4 | Demographic characteristics of phase three student participants and non-participating students who expressed interest.

Characteristics	Study participants (*N* = 54)	Non-Participants (*N* = 152)
	N (%)	*N* (%)
Age (years)		
18–21	27 (50.0)	89 (58.6)
22–25	26 (48.1)	52 (34.2)
26 and older	1 (1.85)	11 (7.2)
Program of study		
Undergraduate	36 (66.7)	116 (76.3)
Graduate	9 (16.7)	19 (12.5)
Professional program	9 (16.7)	17 (11.2)
Year of study		
1 and 2	33 (61.1)	77 (50.7)
3 and 4	16 (29.6)	59 (38.8)
5 +	5 (9.26)	14 (9.2)
Living arrangements, on-campus	23 (42.6)	65 (42.8)
Ethnicity		
Visible ethnic minority	29 (53.7)	73 (48.0)
Aboriginal Person	1 (1.9)	4 (2.6)
Caucasian	24 (44.4)	75 (49.3)
Gender identity		
Female	35 (64.8)	117 (77.0)
Male	19 (35.2)	31 (20.4)
Non-binary	0 (0.0)	4 (2.6)
Sexual Orientation		
Straight	39 (72.2)	118 (77.6)
Queer/questioning	15 (27.8)	34 (22.4)
Type of student		
International student	18 (33.3)	45 (29.6)
Domestic student	36 (66.7)	107 (70.4)
Mental health concerns[a]		
Past	39 (72.2)	105 (69.1)
Present	30 (55.6)	88 (57.9)

[a] *Total proportion exceeds 100% as participants could check-off multiple options.*

Content Modifications

Representativeness-Related

When probed, students did not name any novel psychosocial concepts that were completely missing from the tool. However, in session one and two, participants felt that a student's financial situation is a crucial stressor that contributes to their mental well-being, however its assessment in the tool was limited to housing-related finances (e.g., bills, rent). One participant summarized the issue as:

> *"Regarding finances, this is quite broad, perhaps distinguish between living affordability (house/shelter, food, health) and school (tuition); perhaps a better term would be security or financial stability/security." (FG1)*

After the first focus group session, study investigators revised the overall concept to measure 'Housing and Material Security,' shifting the focus away from strictly housing and financial difficulties and assessing whether necessities in general were met or not. **Figure 2** displays our conceptual framework prior to and following focus groups.

Relevance-Related

Across all focus group sessions, most students felt that HEARTSMAP-U's psychosocial areas applied to their lived experiences and captured the challenges they experience within and outside the post-secondary educational context. One student described the tool's multi-dimensional nature as:

> *"Going into the different facets could be really helpful… people sometimes underestimate how much other stuff can really influence their mental health. Like if you're reallystruggling with school or rent money, that really has an impact on mental health. Butsometimes we don't realize it. We just think oh, it's because I'm just having a hardtime." (FG 4)*

Participants found the graded scoring spectrum to be an important attribute as it recognizes a middle ground, which could allow more students to see themselves in the options. One participant expressed:

> *"I like the use of "but" in [the] sections, a lot of questionnaires have all or nothing questions when sometimes you do struggle with the problems but have implemented coping skills." (FG 3)*

However, many students expressed concerns that the scoring gradation was not always clearly delineated in psychiatric sections such as "Mood" and "Thoughts and Anxiety," as participants felt that descriptors for scores 0 and 1 were *"blurred"* and they *"had a little bit of trouble distinguishing them."* Students also felt that descriptors should emphasize functional impairment and *"refer more to actions"* associated with various levels of concern severity, as opposed to just focusing on how students are feeling. Participants also found score 0 to be strength oriented whereas the remaining options reflected a gradient of deficits. They felt the score 0 language should be more neutral, and unassuming that the student is flourishing. One participant suggested:

> *"Resolve the language for 0 since it seems— it sounds a little idealistic for students. Instead of the word 'satisfied', … say like, "I'm keeping up and maintaining my academics and activities"… I think that would be a better capture of the baseline." (FG 2)*

Where applicable, the research team changed the scoring criteria to have a more consistent pattern across sections. A score 0 would indicate no perceived challenges (neutral), a 1 would

indicate challenges with no to minimal functional impairment or distress (i.e., can still go about self-care/daily activities), a 2 would indicate challenges with moderate functional impact (i.e., difficulties going about self-care/daily activities), and a 3 representing challenges with severe impairment/distress, preventing self-care and daily activities. Participants in later sessions affirmed and supported these changes.

For each section, students highlighted opportunities to refine content and improve its relevance in assessing the concept it maps to. For example, in session one and two, participants expressed that engagement level and satisfaction should be included in "Education and Activities" to help evaluate how academics and extracurricular activities interact with students' mental well-being. One student expressed that being "*engaged and [also] unsatisfied should be included [in the tool] because in that way [the University] can measure how meaningful or successful the activities [on] campus are for students." (FG 6)*. In another instance, participants felt the "Mood" section overly focused on sad or anger-related emotions and needed to incorporate situations where students may perceive *"no emotions or numb."* For "Sexual wellness," students felt that score 0 (healthy sexual relationships) needed to clearly reflect protection-less, consensual sex between long-term, responsible intimate partners, and score 3 (high-risk behavior) needed to integrate discussion of capacity to consent. **Table 5** reports representative quotes and corresponding modifications relating to student's perception of HEARTSMAP-U's relevance.

Understandability-Related

Students agreed that overall, HEARTSMAP-U's scoring descriptors, guiding questions, and purpose were clear and easily understood. Guiding questions were perceived helpful and provided additional *"clarification"* on the section to be scored. Students suggested multiple modifications to improve content understandability. In session one, participants felt many terms and phrases (e.g., control over thoughts, basic needs, emotional support) were unfamiliar or ambiguous. Participants expressed the need for a "*common language*" between the tool user and researcher, so students comprehend questions and scoring criteria as intended, *"that way connotations aren't playing as much of a role."* In response to this, we introduced a 'hover-over' feature for any term or phrase students expressed uncertainty or confusion. In sessions 2-6, students consistently expressed approval of this feature and built a library of concise definitions with student-friendly language. Students stressed the need for a clear instructions page at the beginning of the tool, to ensure students knew how to approach each section. Participants felt if the user is uncertain between two scoring options (e.g., score 1 or 2), they should select the most conservative/higher score that applies to their situation. For example, under "Relationship," *"some relationships might be fine, but others aren't. Then you're basing [your score] on the struggling ones."* **Table 6** summarizes participants comprehension-related feedback for each tool section, followed by the study teams agreed upon modifications. Overall, we found students in later sessions (5-6) affirmed and supported content modifications made in response to concerns raised in earlier sessions (1-4).

DISCUSSION

We document the multiphasic adaptation of previously developed pediatric psychosocial assessment tools into HEARTSMAP-U, a version fit-for-purpose for the post-secondary student population. In Phase one, the study team arrived at a prototype considered clinically and contextually suitable for post-secondary students. In Phase two, participants saw alignment between HEARTSMAP-U's content and their clinical experiences. Of those who offered constructive feedback, most called for a diagnostic level of content detail and specificity (28–41%), which may not be relevant for screening purposes. Between 23 and 39% clinicians provided modifications/feedback related to sectional content and severity coverage, as per the tool's intended use as a self-administered screener. In Phase three, students provided feedback for improving the content relevance and understandability. Modifications focused on creating a common language between tool users and researchers, as well as ensuring scoring options were realistic and distinguishable. Students did not propose novel psychosocial domains that HEARTSMAP-U does not already directly or indirectly measure.

Our tool adaptation process and methods built on existing screening literature and prior student-specific, rapid screening tools described in the literature. The Symptoms and Assets Screening Scale is a lengthier (34-item) instrument, and its content focuses mostly on psychological concerns. In the absence of more generalized psychosocial screening, students' resource needs may be underestimated or only partially understood. The College Health Questionnaire addresses these concerns and allows for multiple-domain screening. Both previous instruments display promising reliability and construct validity evidence. However, reporting of their development process is limited and describes a traditional "top–down" approach, with little mention of student and/or clinical expert (non-investigators) involvement. Engaging the target population has important implications in refining tool content, language, and instructions, which would contribute evidence toward the instrument's content validity, helping to ensure the measure reflects students' lived experiences and vernacular (63). While not intended for screening or assessing mental health issues, the Post-Secondary Stressor Index (PSSI) is an institution-facing tool that evaluates students' exposure to stress and supports targeted mental health intervention/programming (70, 71). In developing and evaluating the PSSI through an extensive process of student engagement, Linden et al. noted that their tool saw markedly stronger psychometric properties compared to similar tool's previously developed without involving students. The ISPOR Patient-Reported Outcomes Good Research Practices Task Force Report highlights the critical contribution that end-user engagement makes to the content validity argument of an instrument and its quantitative psychometric properties (72, 73). In line with this literature, HEARTSMAP-U's adaptation was closely informed by students, content experts of their own lived experiences and the collective experience of being a post-secondary student. While previously described measures have focused exclusively on assessment, scoring on HEARTSMAP-U feeds into a

TABLE 5 | Content-relevance related phase three student feedback with representative quotes and tool modifications.

Tool section	Key feedback	Representative quotes	Tool modifications
Housing and material security	Take into consideration that while needs are met, the student may not be satisfied with how well or easily they are being met.	*"If I can't pay for my housing, my parents are always… going to have my back…But my finances are a completely different situation…like, not being able to go out with friends because I'm just thinking about the future and, oh man, I'm going to have to put money towards this and that.…[this describes] a lot of I think first year [students]…especially people living on campus." (FG 3)*	Scoring criteria: assess how easily and satisfactorily students perceive their needs being met.
	Expand on what falls under finances and material needs	*"Housing, food, rent, tuition, those are the only four you're interested in? Are there additional ones that we're supposed to know are material needs? I just don't know if those are just examples and there's many, or if those are just, like, specifically those four." (FG 4)*	Term definitions: add "housing, food, rent, tuition, insurance, medication" as examples of needs
Education and activities	High GPA does not mean a student is engaged with or enjoying what they are studying	*"I feel like this has a lot to do with motivation, like whether you like or not. I mean, I did do well. I wouldn't say I was struggling, it's just I didn't feel motivated to do it, but I was doing it either ways. Because I need my GPA but I wasn't happy." (FG 1)*	Guiding questions: add-in "Do I feel motivated to engage in my academics and activities?"
	Motivation may not be the best word choice for this section.	*"In my experience, my friend's experience, a lot of it is just life gets overwhelming and it's not like… you've lost motivation in your schoolwork. It's that there's just so many other things going on that your academics start slipping...It's not because you don't want it to...Maybe you have, like, a breakup going on or maybe you're moving… there's just a lot of other factors…grouping it under lost motivation and having it underlined and bolded, doesn't really do justice." (FG 3)*	Scoring criteria: captures engagement and satisfaction with engagement, rather than strictly motivation.
	Resolve language to reflect that the student is unable to engage in their academics, not by choice.	*"Instead of saying, I have completely stopped engaging with academics or activities, etc., I would more lean towards the side of, like, I have been unable to. Yeah, because it's not really like the student choosing to completely stop, right. It's like them not being able to anymore." (FG 5)*	Scoring criteria: change "I have completely stopped" to "I have been unable to" engage with…
Relationships	Clarify how relationships is being defined and assessed.	*" Relationships are often thought of as more intimate, and I think that changing [relationships] to 'social connections' would work well." (FG 6)*	Word choice: change relationships" to "social connections"
		"For zero it says, I am emotionally supported and satisfied with my social connections. But for one, two and three it focuses on the word "relationships." …Social connections seem more broad, whereas relationships seems like they are referring to something more personal… If the goal is to be more broad, maybe social connections would reach a wider audience." (FG 6)	
Thoughts and anxiety	Frequency and time frame can make it easier for students to place themselves on a score.	*"Frequency is a good measure of the intensity of someone who has anxiety disorder. I don't have anxiety disorder, but I think it'd be really helpful if someone the person...getting their assessments, and if they feel comfortable filling that out [could] differentiate between their intensity." (FG 2)*	Scoring criteria: incorporate frequency descriptors (e.g., sometimes, often)
	Language can be resolved to sound less accusatory.	*"The language of losing control feels a little accusatory. If it could be more like you feel out of control. That way that things like clearly an emotion you are in control of and not something you're doing wrong." (FG 5)*	Scoring criteria: "I am losing control" changed to "I feel like I am losing control

(Continued)

TABLE 5 | Continued

Tool section	Key feedback	Representative quotes	Tool modifications
Substances and behavioral dependencies	Recreational substance use and addictive substance use should be differentiated.	*"Especially in a university culture, there's a difference between – abusing drugs, and recreationally using it, and addiction. If you just want to focus on the dependence and addictive behaviors, I think it'd be good to clarify or specify that these drugs – or some drugs are getting in the way of my life (FG 2)*	Scoring criteria: incorporates concepts of dependence and levels of functional impairment (e.g., disruption to daily activities/self-care)
	University "norms" may not be healthy and can be excessive.	*"I would kind of hesitate against using norms because I think in university it can be a norm to kind of drink quite excessively. And that is still problematic in and of itself." (FG 4)*	Term definitions: "norms" was removed.
	Non-suicidal self-harm may be more severe than how it is currently recognized.	*"People who would put themselves in number one, [could also] end up in number three, like they have a suicide plan, but maybe they're self-harming as well...So then that might just create a discrepancy." (FG 4)*	Clarification: users score the most conservative/severe scoring options that applies to them.
Safety	Non-suicidal self-harm can be therapeutic, however past history is strong predictor of suicidal behavior.	*"Non-suicidal self-harm is actually a really therapeutic coping mechanism to the patient. Because it's kind of like their way of dealing with it, and if you take it away, then they kind of may progress to doing worse things. So, I see why it is a one… Add like past history of attempted suicide [to score three]. Because I know that's a huge risk for future suicide." (FG 5)*	Scoring criteria: maintain non-suicidal self-harm as score 1 and add "previous suicide attempts" to score 3.
Sexual wellness	Include scenario where one partner may not use protection.	*"Maybe under one, you could say, 'I always use protection but I'm unsure or I know that my partner doesn't use one."' (FG 2)*	Scoring criteria: incorporate uncertainty around partner's sexual wellness or risk-taking.
Mood	Clarify that changes in daily activities/self-care are in relation to mood changes.	*"[In score 3] I'd be good to specify because of elevated or low moods…I feel like you can have sleep, energy, diet changes that prevents you from going about your day for at least a week if you're stressed." (FG 2)*	Scoring criteria: clarify connection between functional impact and mood.
	Include perceived numbness/lack of emotion.	*"The flat affect that some people can get when they're depressed is not really being captured all the time in any of these kinds of four questions…Because they might not feel down, but just the things that make them happy, no longer make them happy." (FG 4)*	Scoring criteria and guiding questions: incorporate numbness and flat affect.
Abuse	Section will prevent students from slipping through the cracks.	*"I think what's important about this tool is that someone who would be saying that they believe that this is happening to them and scoring a one, two or three would hopefully then get the resources where they'd be able to talk about it further or more… Not have people slip through the cracks of the questionnaire. So I think that's great." (FG 6)*	No substantial modifications were made.
Professionals and resources	Consider the situation where some but not all needs are met.	*"I am supported in some of my mental health needs but need further support." (FG 1)*	Scoring criteria: recognize partial resource connection into score 1.
	Commenting on helpfulness of existing care	*"I like the inclusion of 'I didn't find it helpful'." (FG 5)*	No modifications made.

TABLE 6 | Comprehension-related phase three student feedback with representative quotes and tool modification.

Tool section	Key feedback	Representative quotes	Tool modifications
Housing and material security	Unclear whether relational stressors are relevant here.	*"With the pandemic, I've read many articles that say that a lot of students that are a part of the LGBTQ community…have to go back home. That is not considered, like, a safe space for them because their parents reject them, or they suffer domestic violence… I thought [this section] meant a little bit more of that." (FG 6)*	Scoring criteria: assess how easily and satisfactorily students perceive their needs being met.
		"When you say safe and secure, does that also include, like a stressed home kind of thing? Because I know sometimes at home, yeah, you can feel safe, but it can be very stressed from time to time." (FG 4)	Term definitions: added "housing, food, rent, tuition, insurance, medication" as examples of stressors associated with material security
	Appropriate to assessing needs.	*"Language is pretty good, and the categories are clear. Since this is meant to be a broad scanning, I do not think more information should be asked." (FG 6)*	No modifications
Education and activities	Idealistic language should be avoided.	*"Language for [score zero]... sounds a little idealistic for students,... instead of the word 'satisfied', we could maybe say like, "I'm keeping up and maintaining with my academics and activities'... a better capture of the baseline." (FG 1)*	Word choice: changed to reflect a neutral perception toward one's academic situation.
	Normal versus overwhelming academic stress	*"I'm thinking that maybe there needs to be some clarification between just a normal level of overwhelmed with university work and feeling so overwhelmed you're paralyzed or whatever. Some differentiation between when it becomes mental illness and what's just normal levels of a lot of stress." (FG 5)*	Point of clarification: feeling "paralyzed" is meant to be captured through functional impairment in relation to academics and extracurriculars.
Relationships	Wording in this section needs clarity.	*"I had an initial confusion in reading the first one. The, "I have emotionally supportive connections, but they're hard to build, maintain, or sometimes cause conflict." I think my initial confusion was that it didn't make sense to me at the beginning...How can you have emotional supporting connections if they're hard to maintain, or if they're hard to build in the first place, or if they cause conflict? I feel like the language can be changed there." (FG 2)*	Word choice: changed throughout the entire section to reflect those students who may feel supported but still struggle with relationships. Ex. Score 1 changed to "I feel emotionally supported but feel challenged building/maintaining social connections."
	Good use of "overwhelm"	*"I like overwhelmed. I actually really like the fact that you kind of have two dimensions where it's, like, whether or not you're feeling supported and kind of capturing whether or not you're actually able to kind of maintain those relationships. Because sometimes you can have a very good support network but just feel so overwhelmed with things, that you don't feel like you can access it or it's really hard to kind of maintain that. And I really, really like… and I have never seen anybody kind of ask it in that way." (FG 4)*	No modifications
	"Overwhelmed" it can be positive or negative	*"Overwhelmed could be, like, positive, but it could also be negative. So in this case it would be negative I assume. Maybe specify, like, what exactly is overwhelming and again, defining building and maintaining relationships might be helpful." (FG 4)*	Added hover over: for overwhelmed. Wording of this section was changed and overwhelmed was removed from scoring and added to a hover over.
Thoughts and anxiety	Clarify score 1 and self-care activities.	*"I think that number one needs some clarification. I feel like I sometimes lose control of my thoughts, but I can go about self-care daily activities. Do you mean that even though you sometimes lose control of your thoughts, with self-care activities, you can handle them? Or what do you mean?" (FG 6)*	Word choice: changed "I can go about" to "I can keep up with" self-care/daily activities. Removed "always" from "always in control of my thoughts" (absolute language/unrealistic).

(Continued)

TABLE 6 | Continued

Tool section	Key feedback	Representative quotes	Tool modifications
Substances and behavioral dependencies	Provide dependency examples	*"I think the more examples and the more definitions, it makes it more... accessible to people, by... explaining it in many ways as possible for people to identify and try to find where they want to scale themselves." (FG 2)*	Additional examples: added "excessive sex/gambling/gaming/ exercise/eating/spending" as dependency examples to the hover-over.
Safety	Safety alert could be more supportive.	*"Saying 'you need to connect to the crisis line' may sound a bit scary? Would people feel reluctant to do so?" (FG 6)*	Word choice: changed to "Immediate help is available. Click here to connect with a crisis responder now" (hyperlinked text).
Sexual wellness	Clarify methods of protection	*"Maybe specify what is meant by protection. Is it just physical, the condom, or birth control pills?" (FG 2)*	Additional examples: added "condoms, dental dams, contraception" to the hover-over.
	Consent should be discussed more	*"I noticed that consent was in the description for healthy sexual decisions. But I feel like it could also be more visible, because it's also… consent is a big part of sexual wellness." (FG 6)*	Scoring criteria: consent incorporated into score 3: "at least one of us does not have the capacity to consent."
Mood	Clarify mood changes	*"For Score 3, could it be changed to say "have been swinging..."?" (FG 6)*	Word choice: instead of saying "mood swings", change to " swinging between the two extreme" low/numb and elevated/elated.
Abuse	Recognize effective coping with past abuse.	*"[change to] working through [past abuse] effectively instead of able to work." (FG 3)*	Word choice: added "effectively" to score 1 and added a hover over with clarification.
Professionals and resources	Avoid absolute language	*"Say 'satisfied" instead of supported in all' mental health resources (FG 1)*	Word choice: changed "supported with all" to "satisfied" with mental health needs.

complex decision-making algorithm to generate severity and urgency-specific recommendations for both psychiatric and social/functional resources. The tool's action-oriented approach to assessment may help avoid "run-around" and potentially unnecessary referrals to already scarce psychiatric services.

During Phase two expert review, a fraction of clinician's (23–39%) identified opportunities to improve severity coverage across HEARTSMAP-U's sections, particularly of concerns that fit in-between "none" and "mild" scoring options. Capturing subthreshold and milder cases is a critical challenge with existing self-report measures (74). If transient and non-severe issues are not explicitly reflected in the scoring criteria, these cases may go underreported due to stigma and remain unmanaged until crisis situations. Recently, transdiagnostic clinical staging models of mental illness have received great attention as an improved means of characterizing the progression of mental disorders into adulthood. HEARTSMAP-U's symptomatic and functional characterization of low to high severity concerns may support its screening utility for mental disorders at their earliest stages, from non-specific to subthreshold symptoms (75). A sizable proportion of clinician's (28–41%) provided feedback more suited for their own practice and context of use (e.g., diagnostic-level probing), which would not be consistent with HEARTSMAP-U's intended use as a brief screener. These comments may reflect outstanding assessment needs and challenges in the post-secondary counseling settings, where validated, standardized intake procedures/measures are infrequently used, difficult to interpret, and can be time-consuming (76). A number of clinician's (18%) scored the tool's resource recommendations as underestimating the support needs of their fictional case. This may have been an artificial finding reflecting our online survey setup, where we asked participants to assess the appropriateness of each individual recommendation. By design, HEARTSMAP-U pairs intensive and lower tier resources, recognizing that multiple treatment and self-management modalities can help students cope with the long-wait times associated with scarcely available psychiatric resources (77). We believe that if clinicians had been asked to holistically assess the appropriateness of their case's service recommendations all-together (low tier and intensive options), support needs would have been perceived as sufficiently met. Future studies with a modified data collection instrument would help verify this was a methodological flaw.

In Phase three focus group sessions, student's felt the severity gradation (impairment, frequency, intensity) needed to be more distinguishable across scoring options. These comments are unsurprising given that internal, emotional states can be difficult to concretely self-score and numeric scales often have arbitrary scaling, with unclear distance between answer options (78). Student feedback also allowed us to revise tool language and build-in mechanisms (e.g., hover-overs, examples) to avoid assumptions and gender- and culturally-specific references, and use person first language where possible (79). Future validation studies will confirm whether students interpret and respond to tool content as intended.

Post-secondary student mental well-being is a growing national and international priority, with recent standards calling for the integration of student-centeredness within campus mental health strategies, to ensure responsiveness to students' perceived needs and experiences (19, 25). In striving toward these principles, our work demonstrates the development of early detection capacities built for, by, and with post-secondary students. Growing research demonstrates the potential for campus-based mental health screening interventions in helping students identify unmet support needs and initiate resource-seeking (33, 35, 80). Unfortunately, measuring what matters most to end-users/patients has not been traditionally prioritized in the psychological instrument development literature (81). Diverse student engagement was a key strength of the current study. Purposive sampling allowed us to ensure focus groups reflected student voices across a range of socially co-created realities, who may have differing experiences with respect to stigma, mental health literacy, barriers to care, and systemic challenges (e.g., oppression, discrimination). Another methodological strength is our use of vignettes during Phase two expert review, allowing us to interactively engage clinicians and elicit their feedback on tool content, given they could not self-administer the tool.

We note several study limitations. Phase one discussions and outputs may have been biased by the study team's proximity to the project. However, subsequent feedback and insights from clinicians and students offered additional perspectives and opportunities to further refine HEARTSMAP-U's content. In Phase two, we did not outline clear parameters for vignette development. As a result, no vignettes evaluated the tool's scoring criteria on severe "Alcohol and drug" and "Abuse" concerns. However, clinical investigators reviewed the tool's service provisions mapping to these severe scores, and found they matched current clinical safety protocols. Additionally, we restricted focus groups to students of a single, large-size post-secondary institution in Western Canada. Students from smaller institutions (e.g., community colleges, vocational schools), rural regions, and francophone communities may see the need for further tool content modification for alignment with their experiences and learning environment (82, 83). Still, our findings may be transferable to other similarly large, research-intensive institutions.

HEARTSMAP-U has undergone a rigorous, systematic, and multi-stage tool adaptation process with clinical experts and student end-users. Later validity investigations will report evidence of HEARTSMAP-U's measurement properties, which will be crucial in gauging the tool's suitability for universal screening utility and the early detection of students' mental health needs.

DATA AVAILABILITY STATEMENT

The datasets presented in this article are not readily available because the study participants did not agree for their data to be shared publicly, due to the nature of the research. Requests to access the datasets should be directed to Punit Virk, pvirk@bcchr.ca.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by University of British Columbia Behavioural Research Ethics Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

PV and QD contributed to conception and design of the study. PV and RA conducted data collection. PV, RA, and HB organized the database. PV performed the analysis and wrote the first draft of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

FUNDING

This work was supported by The University of British Columbia Faculty of Medicine Strategic Investment Fund and The University of British Columbia Office of the Vice-President, Students. Funding sources had no involvement in study design, data collection, data analysis, writing report, or in the decision to submit for publication.

ACKNOWLEDGMENTS

We would like to express our deepest gratitude to Dr. Heather Robertson for her insights and guidance throughout Phase one of this work. We are grateful for all the time, support, and expertise of our student researchers Paula Gosse, Lea Separovic, Vanessa Samuel, Allison Lui, Max LeBlanc, Andy Lui, and Bohan Hans Yang, during the early stage of this work. We extend our deepest gratitude to Ms. Karly Stillwell for her coordination and administrative support. We thank Caitlin Pugh for her critical review of our draft manuscript. We would also like to acknowledge Gurm Dhugga, Naveen Karduri, and Vincent Lum for their IT support. We are also thankful of the University of British Columbia Mental Health Network for their support of our work and their commitment to advocating for student mental health.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyt.2022.812965/full#supplementary-material

REFERENCES

1. Linden B, Boyes R, Stuart H. Cross-sectional trend analysis of the NCHA II survey data on Canadian post-secondary student mental health and wellbeing from 2013 to 2019. *BMC Public Health.* (2021) 21:590. doi: 10.1186/s12889-021-10622-1
2. Ibrahim R, Sarirete A. Student development in higher education: a constructivist approach. In: Lytras MD, Ordonez De Pablos P, Avison D, et al., editors. *Technology Enhanced Learning. Quality of Teaching and Educational Reform*. Berlin, Heidelberg: Springer Berlin Heidelberg (2010). p. 467–73. doi: 10.1007/978-3-642-13166-0_66
3. Pedrelli P, Nyer M, Yeung A, Zulauf C, Wilens T. College students: mental health problems and treatment considerations. *Acad Psychiatry.* (2015) 39:503–11. doi: 10.1007/s40596-014-0205-9
4. Giamos D, Young Soo A, Suleiman A, Heather S, Shu-Ping C. Understanding campus culture and student coping strategies for mental health issues in five Canadian colleges and universities. *Can J High Educ.* (2017) 47:136–51. doi: 10.47678/cjhe.v47i3.187957
5. Kessler RC, Amminger P, Aguilar-Gaxiola S, Alonso J, Lee S, Ustün TB. Age of onset of mental disorders: a review of recent literature. *Curr Opin Psychiatry.* (2007) 20:359–64. doi: 10.1097/YCO.0b013e32816ebc8c
6. Singh SP, Tuomainen H. Transition from child to adult mental health services: needs, barriers, experiences and new models of care. *World Psychiatry.* (2015) 14:358–61. doi: 10.1002/wps.20266
7. American College Health Association. *National College Health Assessment II: Canadian Consortium Executive Summary Spring 2019*. Silver Spring, MD (2019).
8. Public Health Agency of Canada. Suicide in Canada: Key statistics (infographic). Available at: https://www.canada.ca/en/public-health/services/publications/healthy-living/suicide-canada-key-statistics-infographic.html. Published 2019 (accessed October 6, 2020).
9. Findlay L. *Depression and Suicidal Ideation among Canadians Aged 15 to 24*. Vol 28 (2017). Available at: http://www.ncbi.nlm.nih.gov/pubmed/28098916
10. Vigo D, Jones L, Munthali R, et al. Investigating the effect of COVID-19 dissemination on symptoms of anxiety and depression among university students. *BJPsych Open.* (2021) 7:e69. doi: 10.1192/bjo.2021.24
11. The Healthy Minds Network, American College Health Association. *The Impact of COVID-19 on College Student Well-Being* (2020).
12. Vizzotto ADB, de Oliveira AM, Elkis H, Cordeiro Q, Buchain PC. Psychosocial characteristics. In: Gellman MD, Turner JR, editors. *Encyclopedia of Behavioral Medicine*. New York, NY: Springer New York (2013). p. 1578–80. doi: 10.1007/978-1-4419-1005-9_918
13. Jenei K, Cassidy-Matthews C, Virk P, Lulie B, Closson K. Challenges and opportunities for graduate students in public health during the COVID-19 pandemic. *Can J Public Heal.* (2020) 111:408–9. doi: 10.17269/s41997-020-00349-8
14. Lederer AM, Hoban MT, Lipson SK, Zhou S. Eisenberg D. More Than Inconvenienced: The Unique Needs of US College Students During the COVID-19 Pandemic. *Heal Educ Behav.* (2020) 48:14–9. doi: 10.1177/1090198120969372
15. Bray SR, Born HA. Transition to University and Vigorous Physical Activity: Implications for Health and Psychological Well-Being. *J Am Coll Heal.* (2004) 52:181–8. doi: 10.3200/JACH.52.4.181-188
16. Mahmoud JSR, Staten RT, Hall LA, Lennie TA. The relationship among young adult college students' depression, anxiety, stress, demographics, life satisfaction, and coping styles. *Issues Ment Health Nurs.* (2012) 33:149–56. doi: 10.3109/01612840.2011.632708
17. Eisenberg D, Golberstein E, Gollust SE. Help-seeking and access to mental health care in a university student population. *Med Care.* (2007) 45:594–601. doi: 10.1097/MLR.0b013e31803bb4c1
18. Czyz EK, Horwitz AG, Eisenberg D, Kramer A, King CA. Self-reported barriers to professional help seeking among college students at elevated risk for suicide. *J Am Coll Heal.* (2013) 61:398–406. doi: 10.1080/07448481.2013.820731
19. Canadian Standards Associations & Mental Health Commission of Canada. *National Standard of Canada (Z2003:20): Mental Health and Well-Being for Post-Secondary Students*. Toronto (2020). Available at: https://www.mentalhealthcommission.ca/English/studentstandard
20. Hetrick SE, Parker AG, Hickie IB, Purcell R, Yung AR, McGorry PD. Early identification and intervention in depressive disorders: towards a clinical staging model. *Psychother Psychosom.* (2008) 77:263–70. doi: 10.1159/000140085

21. Frick MG, Butler SA, deBoer DS. Universal suicide screening in college primary care. *J Am Coll Heal.* (2019) 69:1–6. doi: 10.1080/07448481.2019.1645677
22. Klein MC, Ciotoli C, Chung H. Primary care screening of depression and treatment engagement in a university health center: a retrospective analysis. *J Am Coll Health.* (2011) 59:289–95. doi: 10.1080/07448481.2010.503724
23. Forbes F-JM, Whisenhunt BL, Citterio C, Jordan AK, Robinson D, Deal WP. Making mental health a priority on college campuses: implementing large scale screening and follow-up in a high enrollment gateway course. *J Am Coll Heal.* (2019) 69:1–8. doi: 10.1080/07448481.2019.1665051
24. Canady VA. University mental health screening kiosk promotes awareness, prevention. *Ment Heal Wkly.* (2015) 25:1–5. doi: 10.1002/mhw.30334
25. Linden B, Grey S, Stuart H. *National Standard for Psychological Health and Safety of Post-Secondary Students—Phase I: Scoping Literature Review. Kingston, ON* (2018).
26. Bradford S, Rickwood D. Young People's Views on Electronic Mental Health Assessment: Prefer to Type than Talk? *J Child Fam Stud.* (2015) 24:1213–21. doi: 10.1007/s10826-014-9929-0
27. Bradford S, Rickwood D. Acceptability and utility of an electronic psychosocial assessment (myAssessment) to increase self-disclosure in youth mental healthcare: a quasi-experimental study. *BMC Psychiatry.* (2015) 15:305. doi: 10.1186/s12888-015-0694-4
28. Murrelle L, Ainsworth BE, Bulger JD, Holliman SC, Bulger DW. Computerized Mental Health Risk Appraisal for College Students: User Acceptability and Correlation with Standard Pencil-and-Paper Questionnaires. *Am J Heal Promot.* (1992) 7:90–2. doi: 10.4278/0890-1171-7.2.90
29. Garlow SJ, Rosenberg J, Moore JD, et al. Depression, desperation, and suicidal ideation in college students: Results from the American Foundation for Suicide Prevention College Screening Project at Emory University. *Depress Anxiety.* (2008) 25:482–8. doi: 10.1002/da.20321
30. Osman A, Bagge CL, Gutierrez PM, Konick LC, Kopper BA, Barrios FX. The suicidal behaviors questionnaire-revised (SBQ-R): Validation with clinical and nonclinical samples. *Assessment.* (2001) 8:443–54. doi: 10.1177/107319110100800409
31. Campbell CE, Maisto SA. Validity of the AUDIT-C screen for at-risk drinking among students utilizing university primary care. *J Am Coll Heal.* (2018) 66:774–82. doi: 10.1080/07448481.2018.1453514
32. Bártolo A, Monteiro S, Pereira A. Factor structure and construct validity of the Generalized Anxiety Disorder 7-item (GAD-7) among Portuguese college students. *Cad Saude Publica.* (2017) 33:e00212716. doi: 10.1590/0102-311x00212716
33. Williams A, Larocca R, Chang T, Trinh NH, Fava M, Kvedar J, et al. Web-based depression screening and psychiatric consultation for college students: a feasibility and acceptability study. *Int J Telemed Appl.* (2014) 2014:580786. doi: 10.1155/2014/580786
34. Chung H, Klein MC, Silverman D, Corson-Rikert J, Davidson E, Ellis P, et al. *A Pilot for Improving Depression Care on College Campuses: Results of the College Breakthrough Series-Depression (CBS-D) Project. Vol 59).* (2011). doi: 10.1080/07448481.2010.528097
35. Mackenzie S, Wiegel JR, Mundt M, Brown D, Saewyc E, Heiligenstein E, et al. Depression and suicide ideation among students accessing campus health care. *Am J Orthopsychiatry.* (2011) 81:101–7. doi: 10.1111/j.1939-0025.2010.01077.x
36. Khubchandani J, Brey R, Kotecki J, Kleinfelder J, Anderson J. The psychometric properties of PHQ-4 depression and anxiety screening scale among college students. *Arch Psychiatr Nurs.* (2016) 30:457–62. doi: 10.1016/j.apnu.2016.01.014
37. Downs A, Boucher LA, Campbell DG, Dasse M. Development and initial validation of the Symptoms and Assets Screening Scale. *J Am Coll Health.* (2013) 61:164–74. doi: 10.1080/07448481.2013.773902
38. Alschuler KN, Hoodin F, Byrd MR. Rapid Assessment for psychopathology in a college health clinic: utility of college student specific questions. *J Am Coll Heal.* (2009) 58:177–9. doi: 10.1080/07448480903221210
39. Bewick BM, West R, Gill J, O'May F, Mulhern B, Barkham M, et al. Providing Web-based feedback and social norms information to reduce student alcohol intake: A multisite investigation. *J Med Internet Res.* (2010) 12:e1461. doi: 10.2196/jmir.1461
40. Ebert DD, Franke M, Kählke F, Küchler AM, Bruffaerts R, Mortier P, et al. Increasing intentions to use mental health services among university students. Results of a pilot randomized controlled trial within the World Health Organization's World Mental Health International College Student Initiative. *Int J Methods Psychiatr Res.* (2019) 28:e1754. doi: 10.1002/mpr.1754
41. King CA, Eisenberg D, Zheng K, Czyz E, Kramer A, Horwitz A, et al. Online suicide risk screening and intervention with college students: a pilot randomized controlled trial. *J Consult Clin Psychol.* (2015) 83:630–6. doi: 10.1037/a0038805
42. Lee A, Deevska M, Stillwell K, Black T, Meckler G, Park D, et al. A psychosocial assessment and management tool for children and youth in crisis. *Can J Emerg Med.* (2019) 21:87–96. doi: 10.1017/cem.2018.1
43. Virk P, Laskin S, Gokiert R, Richardson C, Newton M, Stenstrom R, et al. MyHEARTSMAP: development and evaluation of a psychosocial self-assessment tool, for and by youth. *BMJ Paediatr Open.* (2019) 3:1–9. doi: 10.1136/bmjpo-2019-000493
44. Doan Q, Wright B, Atwal A, et al. Utility of MyHEARTSMAP for Psychosocial screening in the Emergency Department. *J Pediatr.* (2020) 219:56–61. doi: 10.1016/j.jpeds.2019.12.046
45. Li BCM, Wright B, Black T, Newton AS, Doan Q. Utility of MyHEARTSMAP in Youth Presenting to the Emergency Room with Mental Health Concerns. *J Pediatr.* (2021) 235:124–9. doi: 10.1016/j.jpeds.2021.03.062
46. Koopmans E, Black T, Newton A, Dhugga G, Karduri N, Doan Q. Provincial dissemination of HEARTSMAP, an emergency department psychosocial assessment and disposition decision tool for children and youth. *Paediatr Child Heal.* (2019) 24:359–65. doi: 10.1093/pch/pxz038
47. Gill C, Arnold B, Nugent S, et al. Reliability of HEARTSMAP as a tool for evaluating psychosocial assessment documentation practices in emergency departments for pediatric mental health complaints. *Acad Emerg Med.* (2018) 25:1375–84. doi: 10.1111/acem.13506
48. Virk P, Atwal A, Wright B, Doan Q. Exploring parental acceptability of psychosocial screening in pediatric emergency departments. *J Adolesc Heal.* (2021) Submitted.
49. van Amstel LL, Lafleur DL, Blake K. Raising Our HEADSS: adolescent psychosocial documentation in the emergency department. *Acad Emerg Med.* (2004) 11:648–55. doi: 10.1197/j.aem.2003.12.022
50. Cohen E, Mackenzie RG, Yates GL, HEADSS. a psychosocial risk assessment instrument: Implications for designing effective intervention programs for runaway youth. *J Adolesc Heal.* (1991) 12:539–44. doi: 10.1016/0197-0070(91)90084-Y
51. Gawel JE. Herzberg's theory of motivation and Maslow's hierarchy of needs. *Pract Assessment, Res Eval.* (1997) 5:1996–7.
52. Freitas FA, Leonard LJ. Maslow's hierarchy of needs and student academic success. *Teach Learn Nurs.* (2011) 6:9–13. doi: 10.1016/j.teln.2010.07.004
53. Versaevel LN. *Canadian Post-Secondary Students, Stress, and Academic Performance—A Socio-Ecological Approach.* (2014). Available at: https://ir.lib.uwo.ca/cgi/viewcontent.cgi?article=4054&context=etd
54. Woodgate RL, Gonzalez M, Stewart D, Tennent P, Wener P. A Socio-Ecological Perspective on Anxiety Among Canadian University Students. *Can J Couns Psychother.* (2020) 54:486–519. Available at: https://cjc-rcc.ucalgary.ca/article/view/67904
55. Terwee CB, Prinsen CAC, Chiarotto A, et al. COSMIN methodology for evaluating the content validity of patient-reported outcome measures: a Delphi study. *Qual Life Res.* (2018) 27:1159–70. doi: 10.1007/s11136-018-1829-0
56. American Educational Research American Psychological Association National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing. *Standards for Educational and Psychological Testing* (2014).
57. *Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims.* (2009). Available at: https://www.fda.gov/media/77832/download.
58. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL. Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Front Public Heal.* (2018) 6:149. doi: 10.3389/fpubh.2018.00149
59. Wong Riff KW, Tsangaris E, Goodacre T, Forrest CR, Pusic AL, Cano SJ, et al. International multiphase mixed methods study protocol to develop a

cross-cultural patient-reported outcome instrument for children and young adults with cleft lip and/or palate (CLEFT-Q). *BMJ Open.* (2017) 7:1–9. doi: 10.1136/bmjopen-2016-015467
60. Davis LL. Instrument review: Getting the most from a panel of experts. *Appl Nurs Res.* (1992) 5:194–7. doi: 10.1016/S0897-1897(05)80008-4
61. Assarroudi A, Heshmati Nabavi F, Armat MR, Ebadi A, Vaismoradi M. Directed qualitative content analysis: the description and elaboration of its underpinning methods and data analysis process. *J Res Nurs.* (2018) 23:42–55. doi: 10.1177/1744987117741667
62. Barbic S, Celone D, Dewsnap K, Duque SR, Osiecki S, Valevicius D, et al. *Mobilizing Measurement into Care: A Guidebook for Service Providers Working with Young Adults Aged 12–24 Years* (2021).
63. Rothman M, Burke L, Erickson P, Leidy NK, Patrick DL, Petrie CD. Use of existing patient-reported outcome (PRO) instruments and their modification: The ISPOR good research practices for evaluating and documenting content validity for the use of existing instruments and their modification PRO task force report. *Value Heal.* (2009) 12:1075–83. doi: 10.1111/j.1524-4733.2009.00603.x
64. Geisen E, Romano Bergstrom J. Think aloud and verbal-probing techniques. In: Geisen E, Romano Bergstrom JBT-UT for SR, editors. *Usability Testing for Survey Research*. Boston: Morgan Kaufmann (2017). p. 131–61. doi: 10.1016/B978-0-12-803656-3.00006-3
65. Alwashmi MF, Hawboldt J, Davis E, Fetters MD. The iterative convergent design for mobile health usability testing: mixed methods approach. *JMIR mHealth uHealth.* (2019) 7:e11656. doi: 10.2196/11656
66. Carlsen B, Glenton C. What about N? A methodological study of sample-size reporting in focus group studies. *BMC Med Res Methodol.* (2011) 11:26. doi: 10.1186/1471-2288-11-26
67. Stewart AL, Thrasher AD, Goldberg J, Shea JA. A framework for understanding modifications to measures for diverse populations. *J Aging Health.* (2012) 24:992–1017. doi: 10.1177/0898264312440321
68. Coons SJ, Gwaltney CJ, Hays RD, Lundy JJ, Sloan JA, Revicki DA, et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO good research practices task force report. *Value Heal.* (2009) 12:419–29. doi: 10.1111/j.1524-4733.2008.00470.x
69. UBC Student Health and Wellbeing. *Student Health & Wellbeing (Green Folder): Supporting Students in Distress in a Virtual Learning Environment*. (2021). Available at: https://wellbeing.ubc.ca/sites/wellbeing.ubc.ca/files/u9/2021.01_Green Folder_with_date.pdf
70. Linden B, Boyes R, Stuart H. The post-secondary student stressors index: proof of concept and implications for use. *J Am Coll Health*. (2020). p. 1–9. doi: 10.1080/07448481.2020.1754222
71. Linden B, Stuart H. Psychometric assessment of the Post- Secondary Student Stressors Index (PSSI). *BMC Public Health.* (2019) 19:1139. doi: 10.1186/s12889-019-7472-z
72. Brod M, Pohlman B, Tesler Waldman L. Qualitative research and content validity. In: Michalos AC, editor. *Encyclopedia of Quality of Life and Well-Being Research*. Dordrecht: Springer Netherlands (2014). p. 5257–65. doi: 10.1007/978-94-007-0753-5_3848
73. Patrick DL, Burke LB, Gwaltney CJ, Leidy NK, Martin ML, Molsen E, et al. Content validity—Establishing and reporting the evidence in newly developed patient-reported outcomes (PRO) instruments for medical product evaluation: ISPOR PRO good research practices task force report: Part 2—Assessing respondent understanding. *Value Heal.* (2011) 14:978–88. doi: 10.1016/j.jval.2011.06.013
74. Tannenbaum C, Lexchin J, Tamblyn R, Romans S. Indicators for measuring mental health: towards better surveillance. *Healthc Policy*. (2009) 5:177–86. doi: 10.12927/hcpol.2013.21180
75. Shah JL, Scott J, McGorry PD, Cross SPM, Keshavan MS, Nelson B, et al. Transdiagnostic clinical staging in youth mental health: a first international consensus statement. *World Psychiatry.* (2020) 19:233–42. doi: 10.1002/wps.20745
76. Broglia E, Millings A, Barkham M. Challenges to addressing student mental health in embedded counselling services: a survey of UK higher and further education institutions. *Br J Guid Counc.* (2018) 46:441–55. doi: 10.1080/03069885.2017.1370695
77. Lipson SK, Lattie EG, Eisenberg D. Increased rates of mental health service utilization by U.S. college students: 10-year population-level trends (2007–2017). *Psychiatr Serv*. (2019) 70:60–3. doi: 10.1176/appi.ps.201800332
78. Blanton H, Jaccard J. Arbitrary metrics in psychology. *Am Psychol.* (2006) 61:27–41. doi: 10.1037/0003-066X.61.1.27
79. Thomas J, McDonagh D. Shared language:towards more effective communication. *Australas Med J.* (2013) 6:46–54. doi: 10.4066/AMJ.2013.1596
80. Shepardson RL, Funderburk JS. Implementation of universal behavioral health screening in a university health setting. *J Clin Psychol Med Settings.* (2014) 21:253–66. doi: 10.1007/s10880-014-9401-8
81. Krause KR, Bear HA, Edbrooke-Childs J, Wolpert M. Review: what outcomes count? a review of outcomes measured for adolescent depression between 2007 and 2017. *J Am Acad Child Adolesc Psychiatry*. (2019) 58:61–71. doi: 10.1016/j.jaac.2018.07.893
82. Hussain R, Guppy M, Robertson S, Temple E. Physical and mental health perspectives of first year undergraduate rural university students. *BMC Public Health.* (2013) 13:848. doi: 10.1186/1471-2458-13-848
83. Cooke M, Juliet Huntly. *White Paper on Postsecondary Student Mental Health*. (2015). Available at: https://occccco.files.wordpress.com/2015/05/ccvps-white-paper-on-postsecondary-student-mental-health-april-2015.pdf

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Psychiatry

ORIGINAL RESEARCH
published: 07 April 2022
doi: 10.3389/fpsyt.2022.840409

Delivering Remote Measurement-Based Care in Community Addiction Treatment: Engagement and Usability Over a 6-Month Clinical Pilot

Kevin A. Hallgren, Eliza B. Cohn, Richard K. Ries and David C. Atkins*

Department of Psychiatry and Behavioral Sciences, University of Washington, Seattle, WA, United States

OPEN ACCESS

Edited by:
Abigail Ortiz,
University of Toronto, Canada

Reviewed by:
Anette Søgaard Nielsen,
University of Southern Denmark, Denmark
Victor Mark Tang,
University of Toronto, Canada
Antoni Gual,
Hospital Clínic de Barcelona, Spain
Nele A. J. De Witte,
Thomas More University of Applied Sciences, Belgium
James McKay,
University of Pennsylvania, United States

***Correspondence:**
Kevin A. Hallgren
khallgre@uw.edu

Specialty section:
This article was submitted to Digital Mental Health, a section of the journal Frontiers in Psychiatry

Received: *21 December 2021*
Accepted: *28 February 2022*
Published: *07 April 2022*

Citation:
Hallgren KA, Cohn EB, Ries RK and Atkins DC (2022) Delivering Remote Measurement-Based Care in Community Addiction Treatment: Engagement and Usability Over a 6-Month Clinical Pilot. Front. Psychiatry 13:840409. doi: 10.3389/fpsyt.2022.840409

Objective: Measurement-based care (MBC) is an evidence-based practice in which patients routinely complete standardized measures throughout treatment to help monitor clinical progress and inform clinical decision-making. Despite its potential benefits, MBC is rarely used in community-based substance use disorder (SUD) treatment. In this pilot study, we evaluated the feasibility of incorporating a digital and remotely delivered MBC system into SUD treatment within a community setting by characterizing patients' and clinicians' engagement with and usability ratings toward the MBC system that was piloted.

Methods: A pilot study was conducted with 30 patients receiving SUD treatment and eight clinicians providing SUD treatment in a large, publicly funded addiction and mental health treatment clinic. Services as usual within the clinic included individual psychotherapy, case management, group therapy, peer support, and medication management for mental health and SUD, including buprenorphine. Patients who enrolled in the pilot continued to receive services as usual and were automatically sent links to complete a 22-item questionnaire, called *the weekly check-in*, *via* text message or email weekly for 24 weeks. Results of the weekly check-in were summarized on a clinician-facing web-based dashboard. Engagement was characterized by calculating the mean number of weekly check-ins completed by patients and the mean number times clinicians logged into the MBC system. Ratings of the MBC system's usability and clinical utility were provided by patients and clinicians.

Results: Patient participants (53.3% male, 56.7% white, 90% Medicaid enrolled) completed a mean of 20.60 weekly check-ins (i.e., 85.8% of the 24 expected per patient). All but one participating clinician with a patient enrolled in the study logged into the clinician-facing dashboard at least once, with an average of 12.20 logins per clinician. Patient and clinician ratings of usability and clinical utility were favorable: most patients agreed with statements that the weekly check-in was easy to navigate and aided self-reflection. All clinicians who completed usability questionnaires agreed with statements indicating that the dashboard was easy to navigate and that it provided meaningful information for SUD treatment.

Conclusions: A digital and remotely delivered MBC system can yield high rates of patient and clinician engagement and high ratings of usability and clinical utility when added into SUD treatment as usual. The success of this clinical pilot may be attributable, in part, to the user-centered design processes that were used to develop and refine the MBC system that was piloted. Future efforts may focus on strategies to test whether MBC can be sustainably implemented and offers clinical benefits to patients in community SUD treatment settings.

Keywords: addiction, measurement-based care (MBC), recovery, routine outcome monitoring (ROM), user-centered design (UCD)

INTRODUCTION

Over 2.5 million US adults engage in treatment for substance use disorder (SUD) annually (1), each of whom experiences a unique clinical course and outcome. *Measurement-based care* (MBC) is a clinical method in which clinicians routinely administer standardized measures to systematically monitor their patients' responsiveness to treatments over time and inform clinical decision-making (2–4). MBC has been tested many times in non-SUD mental health treatment settings, where it is associated with several benefits, including larger treatment effect sizes (3) and a better ability for clinicians to detect non-improvement and adjust treatment approaches accordingly (5). As a result of these findings, a growing body of research has aimed to improve the implementation of MBC in mental health treatment settings (3). Research testing the use of MBC in specialty SUD treatment settings has been limited (6, 7), even though it is possible that the benefits of MBC observed in mental health treatment settings could be extendable to SUD treatment settings. SUD treatment settings often have unique workflows, treatment approaches, clinician training requirements, and patient populations compared to non-SUD mental health treatment settings, warranting research on the development and testing of MBC systems specific to the context of SUD treatment settings.

In a previous effort to inform the design of a MBC system for outpatient adult community SUD treatment settings, members of our team conducted formative research in partnership with three community SUD treatment clinics. Through this collaboration, we aimed to understand clinicians' ideas, concerns, and preferences related to the MBC system designs, workflows, and content (8). Results of that work indicated that clinicians saw several potential benefits of MBC, including opportunities for improved treatment delivery, patient self-reflection, and communication between patients and their providers about clinical progress. Clinicians noted that MBC systems would be particularly helpful in their settings if they (a) include options for personalization to individual patients (e.g., include questions about patients' goals when asking about their progress, include questions that allow open-ended/free-text responses), (b) minimize burden to clinicians and patients (e.g., use technology to automatically administer and score questionnaire results, utilize patients' smartphones rather than adding devices to clinic waiting areas, allow clinicians to access MBC results using existing their organization's existing login credentials), and (c) measure clinical domains that reflect positive outcomes that clinicians often directly target in SUD treatments (e.g., self-efficacy, use of positive coping skills, and engagement in valued activities) as opposed to exclusively measuring negative outcomes that patients often feel stigmatized when reporting (e.g., substance use and relapse).

Informed by these perspectives, we developed a prototype of a MBC system intended for use in outpatient adult community SUD treatment settings. Following a user-centered design framework (9), the prototype was iteratively refined based on five rounds of usability testing with feedback from patients and clinicians in a large community-based SUD treatment clinic (10). This work resulted in a fully functional MBC system with two primary components: a patient-facing MBC questionnaire, called *the weekly check-in*, and a web-based clinician-facing dashboard for reviewing MBC results, called *the clinician dashboard*. The current pilot study evaluated the feasibility of using this MBC system when it is added onto SUD treatment as usual for up to 6 months. In this paper, we report outcomes related to clinicians' and patients' engagement with the MBC system and their assessments of its usability and clinical utility when used in conjunction with SUD treatment as usual.

MATERIALS AND METHODS

Setting and Participants

All study procedures were approved by the University of Washington Institutional Review Board. Clinician and patient participants were recruited from two treatment teams within a large, publicly funded addiction and mental health treatment clinic owned by King County in Washington State and managed by the University of Washington. Services available in the clinic included individual psychotherapy, case management, group therapy, peer support, and medication management for mental health and SUD (including buprenorphine). Clinician participants were recruited through verbal announcements at team meetings and invitation letters placed in staff mailboxes. Clinicians who expressed interest in participating were given more information about study procedures and provided written informed consent to participate.

Patients were recruited using flyers posted in clinic waiting areas and paper handouts that participating clinicians could

distribute to their patients. Patient eligibility criteria included: receiving treatment for SUD from a clinician who was also participating in the study, having a smartphone, self-reporting speaking and reading English, ≥18 years old, and reporting past year unhealthy alcohol use [measured by an Alcohol Use Disorders Identification Test-Consumption version (AUDIT-C) score ≥3 or 4 for women or men, respectively (11, 12)] and/or past-year use of illicit or non-prescribed drugs (13). Patients were ineligible if they anticipated leaving the region or becoming incarcerated within the next 6 months. Patients who were interested in participating called the study phone number listed on the flyer or handout and completed a brief eligibility screen during the phone call. Eligible participants then completed a baseline appointment, described below. The recruitment period was October 2019–June 2021, with a pause in recruitment between March and June 2020 to accommodate necessary protocol changes due to the emerging COVID-19 pandemic.

Procedures

Eligible patients attended a baseline appointment with a research coordinator in-person or by phone to provide informed consent for all study procedures and to complete research assessments. Patient participants met with the research coordinator again at 6-, 12-, and 24-week follow-ups to complete research assessments and structured interviews (described below). Patients received $50 for each research appointment they completed but were not compensated for completing weekly check-ins. A visual timeline for patient participants is shown in **Figure 1**. Clinicians received no compensation for participating.

At baseline, patient participants were non-randomly assigned to one of two conditions based on the time they enrolled in the study. Patients who enrolled between October 2019 and February 2020 were assigned to the "weekly check-in only" condition, where they completed weekly check-in questionnaires (described below), but their clinicians did not have access to the results of the weekly check-in. The purpose of this condition was to ensure that all research protocols and technologies were fully functional and acceptable before providers accessed the results to MBC questionnaires that would potentially impact patient care. Specifically, we utilized this condition to ensure that most patients were able to complete the weekly check-in and provide opportunity for them to report on its usability. During this phase of the study, we also ensured that the clinician dashboard was correctly displaying patients' responses to the weekly check-in questionnaire. Patients who enrolled between July 2020 and June 2021 were assigned to a "weekly check-in + clinician dashboard" condition, in which patients completed weekly check-ins and the results of those weekly check-ins were accessible to clinicians through the clinician dashboard (described below). Participants in both conditions were informed as to whether their weekly check-in results would be viewable to their clinicians.

Patient Weekly Check-In

All patient participants were sent weekly invitations to complete a brief questionnaire called *the weekly check-in* (see **Figure 2**). Invitations were sent automatically each week *via* text message or email (based on patient preference) using REDCap software (14). The first weekly check-in was completed during the baseline research appointment with a research coordinator present in-person or by phone who encouraged patients to ask for assistance or clarification when needed. For patients in the weekly check-in+ clinician dashboard condition, the research coordinator encouraged patients to answer weekly check-in questionnaires with the understanding that their clinician would review their responses. When patients asked how to interpret potentially ambiguous items on the weekly check-in (e.g., whether using a specific substance counted as "drug use"), they were encouraged to answer in a way that would be most meaningful to them and most useful for communicating about their treatment progress with their clinician. At the baseline appointments, the research coordinator encouraged patients to complete weekly check-ins as early and as often as possible. The research coordinator monitored weekly check-in completion throughout the 24-week study period, and during the first 12 weeks of the study the research coordinator contacted patients to offer support completing weekly check-ins if they were not completed.

The weekly check-in assessed 8 clinical domains using 22 questions derived from existing assessment instruments (15–19). Two clinical domains asked about past-week drinking and other drug use (**Figure 2A**). Six clinical domains assessed areas that reflect hypothesized mechanisms of change in SUD treatment and were previously identified by SUD treatment clinicians as particularly helpful to measure as part of MBC, including past-week experiences with craving, coping skills, abstinence self-efficacy, depression symptoms, positive outlook on life, and therapeutic alliance [**Figure 2B**; see also (8)]. Six questions asked about goals for the upcoming week with respect to reducing substance use, reducing cravings, learning more effective coping skills, increasing abstinence self-efficacy, working on mental health, and having a more positive outlook on life (**Figure 2C**). Two optional questions invited patients to provide open-ended/free-text narratives describing additional goals for the upcoming week and additional information that they may wish to relay to their clinician (**Figure 2D**).

Clinician Dashboard

Clinician participants were given access to a secure web-based dashboard on which they could review summarized results from the weekly check-ins completed by patients in the weekly check-in + clinician dashboard condition. The dashboard displayed line graphs to illustrate change over time for each domain measured by the weekly check-in (**Figure 3A**), text-based summaries of changes in domains over time (**Figure 3B**) bar graphs showing the most recent responses to each question (**Figure 3C**), and a table displaying answers to all questions from previous weekly check-ins (**Figure 3D**). Clinicians received email reminders to review the dashboard every 2 weeks while they had patients enrolled in the study.

Measures

Demographics

Patients completed a questionnaire to self-report their age, gender, race, ethnicity, highest level of education, marital status, employment, annual income, housing status, insurance,

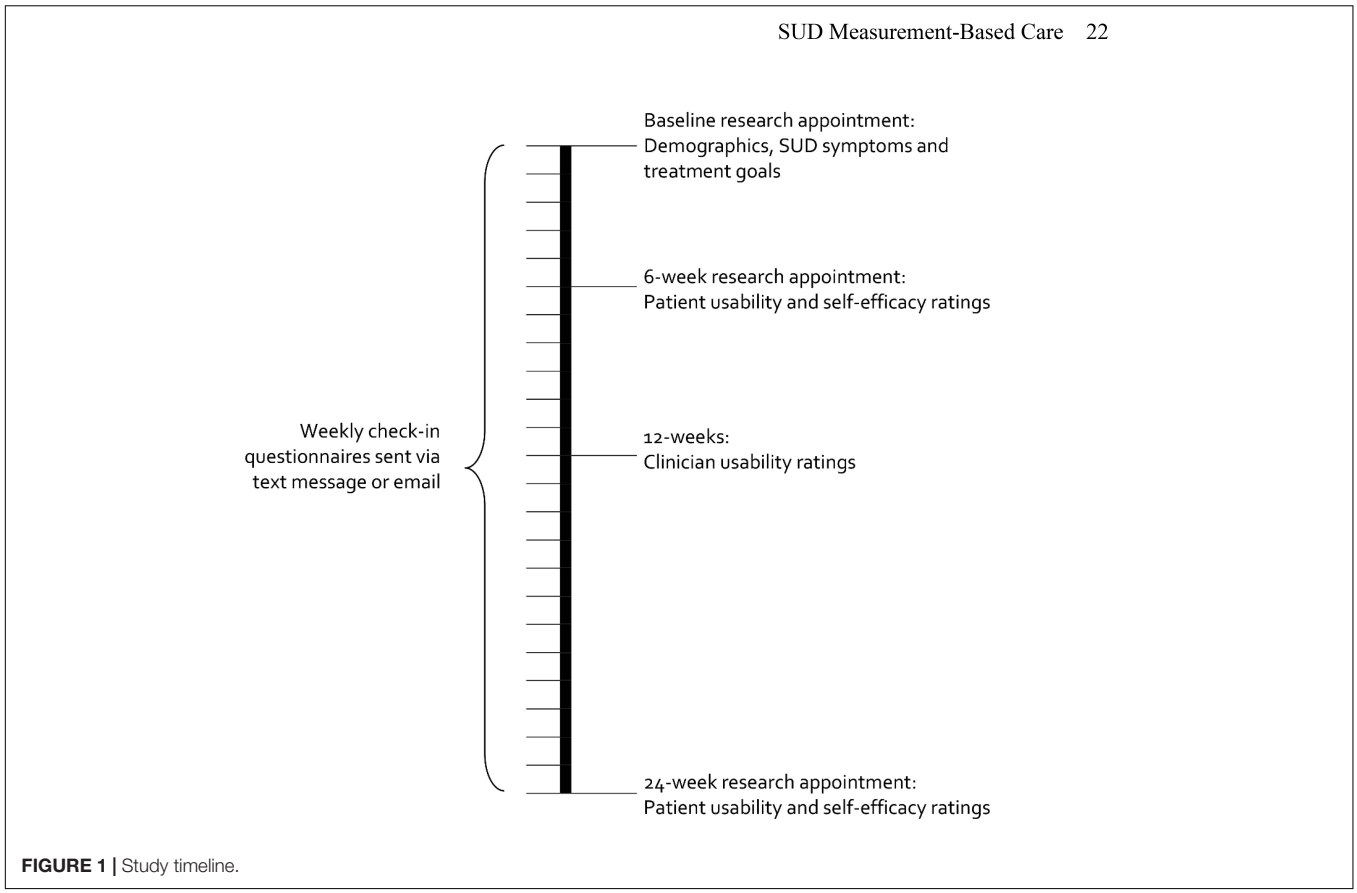


FIGURE 1 | Study timeline.

and current legal involvement. Clinicians completed a brief questionnaire to report their age, gender, race, ethnicity, highest education, number of years working in the current clinical setting, and typically used treatment approaches.

Substance Use Disorder Symptoms and Treatment Goals

Patients self-reported which substances they were addressing in treatment and completed symptom checklists (20, 21) on which they self-reported the presence or absence of each of the 11 SUD criteria for those substances, as defined by the Diagnostic and Statistical Manual, 5th Edition (DSM-5) (22). SUD severity was then categorized based on the number of symptoms reported at baseline, categorized as severe SUD (6–11 symptoms), moderate SUD (4–5 symptoms), mild SUD (2–3 symptoms), or no SUD (0–1 symptoms). Patients also self-reported whether they had a goal of abstinence, reduced use, or no specific goal to change their use of alcohol and other drugs.

Engagement

Patient engagement was characterized using data automatically recorded when weekly check-ins were completed. The primary engagement metrics included the mean number of weekly check-ins completed per patient over the full 24-week study period as well as the mean number of weekly check-ins completed per patient during weeks 1–12 (when the research coordinator proactively contacted patients when they did not complete the weekly check-in) and weeks 13–24 (when the research coordinator would not contact patients). We also calculated the number of weekly check-ins in which patients provided a written response to either of the two optional, open-ended questions that asked about additional goals or additional information the patient would like to relay to their clinician. We estimated the mean length of time it took to complete each weekly check-in by computing differences in timestamps for when the weekly check-in was first opened and when it was submitted, excluding durations that appeared unrealistically long (>30 min; 6.3% of weekly check-ins) as these likely reflected times when patients completed the weekly check-in over two or more sittings (i.e., patients could partially complete the weekly check-in and return to it later).

Clinician engagement with the dashboard was characterized using login and page-visit data that was automatically recorded upon logging into the clinician dashboard. We identified the number of clinicians who logged into the dashboard at least once, the mean number of dashboard logins per clinician, and the mean duration that the dashboard remained open per login session.

Usability and Clinical Utility

Usability and clinical utility of the weekly check-in was self-reported by patients at research appointments. On the usability questionnaire, patients were asked to rate their level of agreement

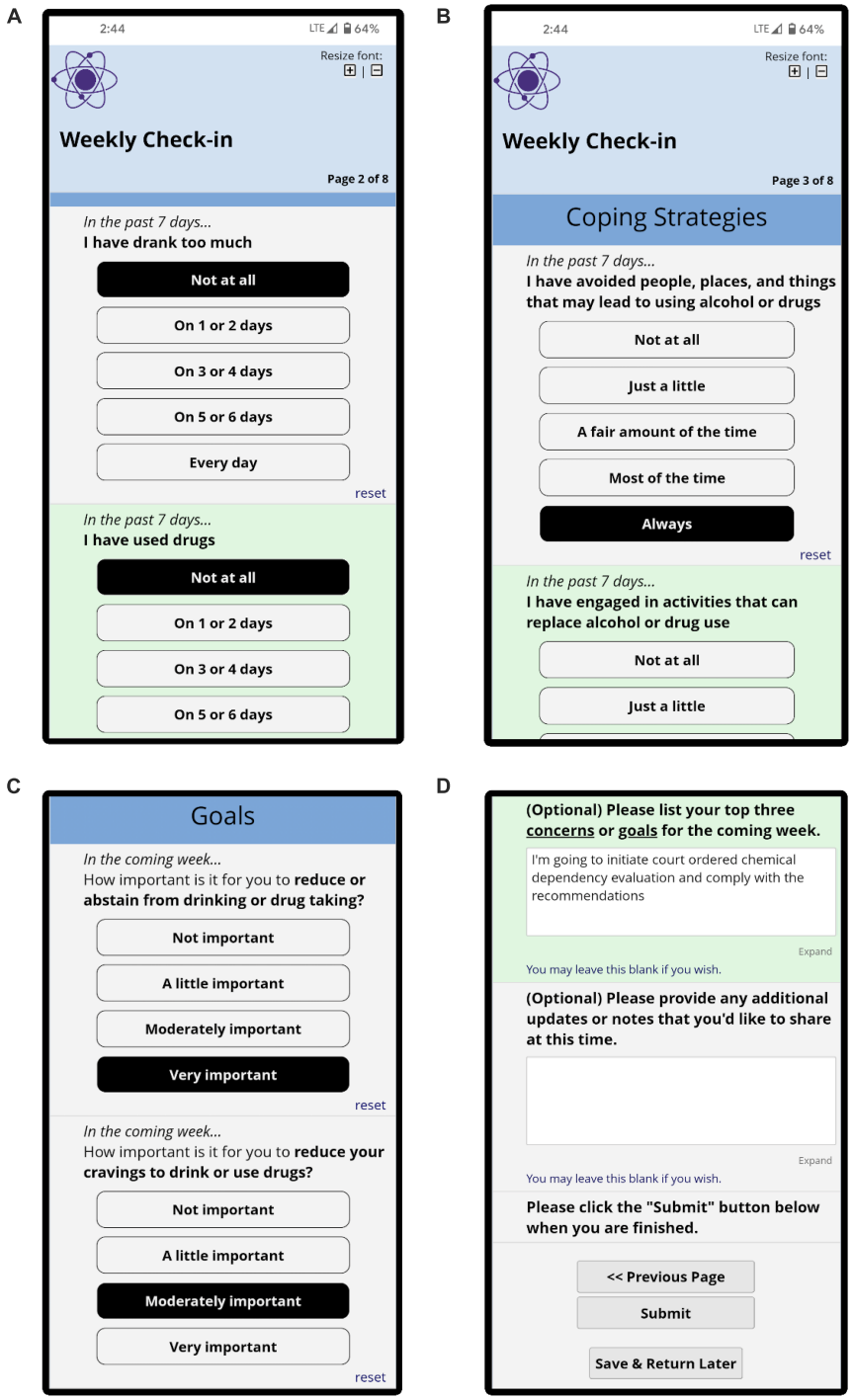


FIGURE 2 | Screenshots showing selected sections of the weekly check-in completed by patients, including questions about substance use **(A)**, mechanisms of change **(B)**, next-week goals **(C)**, and optional open-ended/free-text questions **(D)**.

with several statements about the usability of the weekly check-in (example item: "I can easily find my way on the weekly check-in") and the clinical utility of the weekly check-in (example item: "The weekly check-in can help me reflect on what I want"). Response options for these questions were on a 5-point Likert scale ranging from strongly disagree to strongly agree. An additional question asked patients to report whether the length of the weekly check-in survey was too long, too short, or "just right."

Patients also completed a 5-item questionnaire asking about their confidence in their ability to complete weekly check-ins

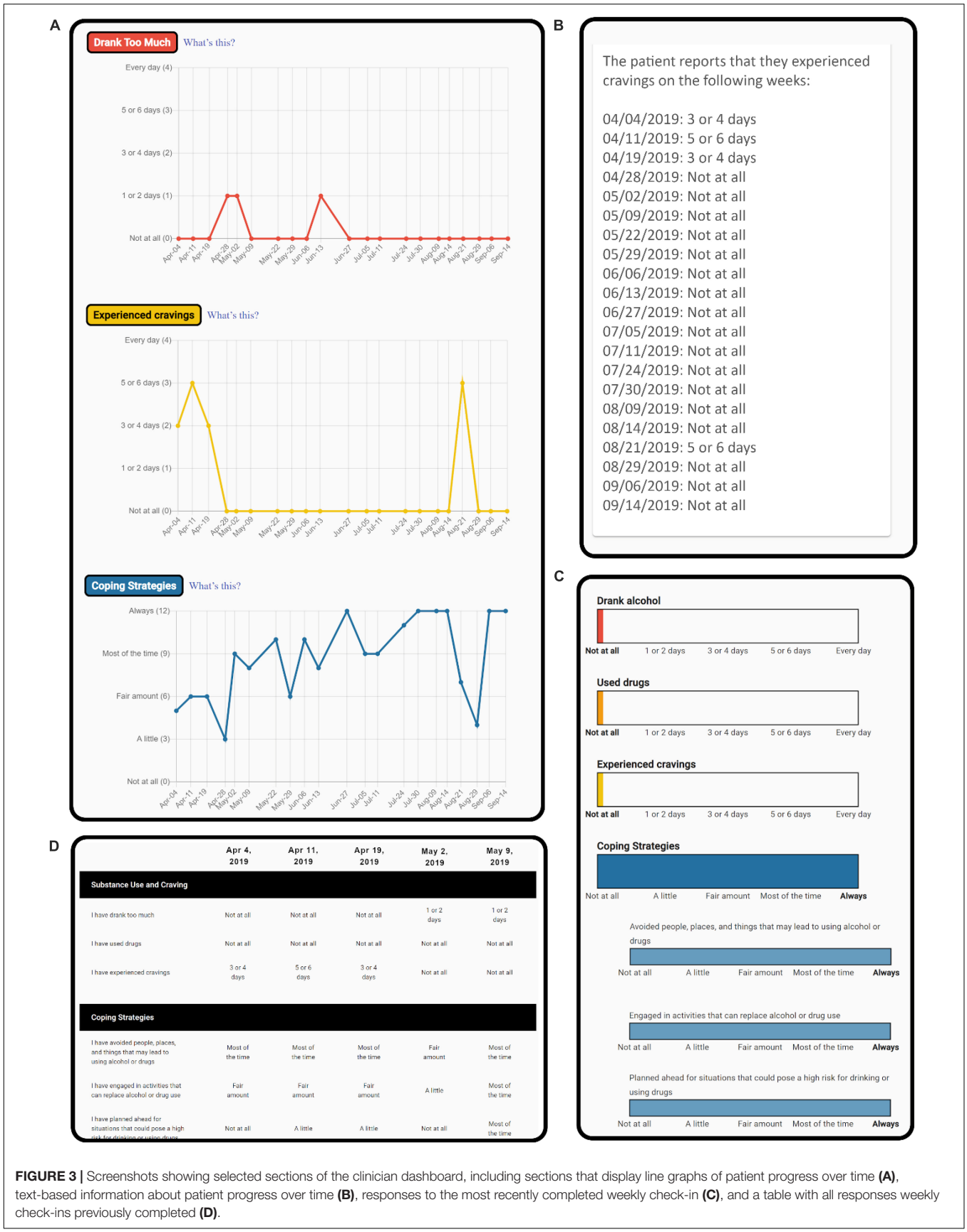


FIGURE 3 | Screenshots showing selected sections of the clinician dashboard, including sections that display line graphs of patient progress over time **(A)**, text-based information about patient progress over time **(B)**, responses to the most recently completed weekly check-in **(C)**, and a table with all responses weekly check-ins previously completed **(D)**.

independently or during a session with their clinician. Response options were on a 5-point Likert scale ranging from strongly disagree to strongly agree.

Clinicians were invited to complete a questionnaire rating their experiences using the dashboard with each of their patients in the weekly check-in + clinician dashboard condition. The questionnaire included items about the dashboard's usability (example item: "I could easily find my way on the dashboard") and clinical utility (example item: "The dashboard provided me with useful information"). Response options were on a 5-point Likert scale ranging from strongly disagree to strongly agree. The questionnaire was emailed to clinicians approximately 12 weeks after their patient enrolled in the study.

Analysis Plan

Descriptive statistics were used to characterize patients' SUD severity, substance use goals, and the measures of patient and clinician demographics, engagement, and usability that were described above. Rates of weekly check-in completion across the 24-week study period for all patients were estimated overall and by study period (weeks 1–12, weeks 13–24) and for each study week. Questionnaire results for the usability and clinical utility of the weekly check-in were analyzed descriptively for the 6- and 24-week time points to understand perceptions of usability and clinical utility earlier and later in the span of using the weekly check-in.

RESULTS

Description of Samples

Sixty-three individuals called the study phone number to inquire about participating in the study. Sixty-two completed the eligibility screening, of which 33 were eligible to participate, 28 were ineligible, and 1 was eligible but declined to participate. Of the 28 ineligible individuals, 21 were ineligible because they were not receiving care in the participating clinic and/or receiving care from a participating clinician, and 7 were ineligible because they did not report any past-year unhealthy alcohol use or any past-year drug use. A total of 30 patients completed a baseline enrolled appointment, including 16 in the weekly check-in only condition and 14 in the weekly check-in + clinician dashboard condition. Patient participants are described in **Table 1**. The distributions of age, gender, race, and ethnicity were similar to that of the full clinic population, according to electronic health care record data from the clinic. Most patients were aged 35–54 (n = 19), male (n = 16), and white (n = 17) and most had an associate's degree, trade degree, or other higher education degree (n = 23). Most patients were not currently employed (n = 27), just over half (n = 16) were homeless, in transitional, temporary, or other housing, or living in a house someone else owned or leased. Most reported symptoms consistent with severe SUD (n = 23). Patients reported that their treatment was addressing use of stimulants (n = 18), opioids (n = 16), alcohol (n = 15), cannabis (n = 5), sedatives (n = 4), and hallucinogens (n = 1). With regard to alcohol, patients reported goals of abstinence (n = 12), reduced drinking (n = 5), or had no specific goal for changing alcohol use (n = 13). With regard to other drugs, patients reported goals of abstinence (n = 16), reduced use (n = 5), or had no specific goal for changing drug use (n = 9). All 30 patients completed the baseline and 6-week research appointments; 29 patients completed the 12- and 24-week research appointments.

TABLE 1 | Characteristics of patient participants (N = 30).

	n	(%)
Age		
25–34	8	(26.7%)
35–44	9	(30.0%)
45–54	10	(33.3%)
55–65	3	(10.0%)
Gender		
Female	11	(36.7%)
Male	16	(53.3%)
Non-binary	2	(6.7%)
Prefer not to say	1	(3.3%)
Race		
American Indian or Alaska Native	2	(6.7%)
Asian	1	(3.3%)
Black or African American	4	(13.3%)
Native Hawaiian or Pacific Islander	0	(0.0%)
White or Caucasian	17	(56.7%)
Another race not listed	6	(20.0%)
Hispanic or Latino (any race)	2	(6.7%)
Highest education		
Less than high school	1	(3.3%)
High school diploma or equivalent	6	(20.0%)
Some college, associate's degree, or trade degree	17	(56.7%)
Bachelor's degree or higher	6	(20.0%)
Employed currently (part time or full time)	3	(10.0%)
Income below federal poverty level for single person household	17	(56.7%)
Housing		
In a home owned or leased by participant	14	(46.7%)
In a home someone else owns or leases	8	(26.7%)
Transitional, temporary, other housing, or homeless	8	(26.7%)
Married or in a committed relationship	6	(20.0%)
Medicaid enrolled	27	(90.0%)
Current legal system involvement*	5	(16.7%)
SUD symptoms, past year		
0–1	3	(10.0%)
2–3 (mild SUD)	4	(13.3%)
4–5 (moderate SUD)	0	(0.0%)
6+ (severe SUD)	23	(76.7%)

**Current legal system involvement including drug court, probation, parole, current legal charges, house arrest, court-mandated treatment, or awaiting sentencing.*

Eight clinicians enrolled in the study. Their age, gender, race/ethnicity, education, experience working in the clinical setting, and treatment approaches are described in **Table 2**. Seven clinicians had at least 1 patient enroll in the study (median = 4 patients per clinician, range: 1–8). Six clinicians had at least 1 patient enroll in the weekly check-in + clinician dashboard condition (median = 2 patients per clinician, range: 1–4), and thus these six clinicians were able to review their patients'

TABLE 2 | Characteristics of clinician participants (n = 8).

Clinicians (N = 8)	N	(%)
Age		
25–44	3	(37.5%)
45–64	5	(62.5%)
Gender		
Female	3	(37.5%)
Male	5	(62.5%)
Race and ethnicity		
Black or African American, non-Hispanic	1	(12.5%)
White or Caucasian, non-Hispanic	7	(87.5%)
Highest education		
Bachelor's degree	2	(25.0%)
Master's degree	6	(75.0%)
Number of years worked in the current clinical setting, median (range)	5	(2 to 18)
Clinical approaches used		
Case management	7	(87.5%)
Client-centered/humanistic counseling	5	(62.5%)
Cognitive-behavioral therapy	4	(50.0%)
Family or couples therapy	1	(12.5%)
Motivational interviewing	5	(62.5%)
Twelve-step based treatment	2	(25.0%)
Psychodynamic/psychoanalytic	2	(25.0%)
Relapse prevention	5	(62.5%)
Medication management	2	(25.0%)
Other approaches	2	(25.0%)

progress on the clinician dashboard and were invited to the complete dashboard usability questionnaire.

Patient and Clinician Engagement

Twenty-nine patients elected to receive weekly check-in prompts *via* text message and one elected to receive them *via* email. Patient engagement metrics are described in the upper half of **Table 3**. Rates of weekly check-in completion for all patients over the 24-week pilot are shown in **Figure 4**. On average, patient participants completed 20.60 weekly check-ins (85.8% of the 24 available to each patient). All patients completed the first 2 weekly check-ins, and the proportion of patients completing the weekly check-in decreased slightly over time until week 12, at which time 80% of patients completed the weekly check-in (**Figure 4**). Between weeks 13–24, rates of weekly check-in completion remained stable with approximately 80% of patients completing it each week (**Figure 4**). Patients provided a write-in response to either or both of the optional, open-ended questions on a mean of 9.17 (SD = 7.90) of the weekly check-ins that were completed (44.5% of completed weekly check-ins). Patients in the weekly check-in only condition and the weekly check-in + clinician dashboard condition did not differ in the number of weekly check-ins completed (p = 0.33) or the number of weekly check-ins with a write-in response (p = 0.94). The mean estimated time to complete each weekly check-in was 4.99 min (SD = 4.46).

Five out of six clinicians who had a patient in the weekly check-in + clinician dashboard condition logged into the clinician dashboard at least once. Among them, there was a mean of 12.20 logins per clinician (SD = 9.33, range = 3–25). On average, each login session lasted 2.30 min.

Usability and Clinical Utility

Usability ratings were favorable at the 6- and 24-week time points (**Table 4**), with most patients (86.2–100%) agreeing or strongly agreeing with statements that the weekly check-in was helpful for reflecting on their substance and recovery, that they would be willing to use the weekly check-in in the future, and that they would recommend the weekly check-in to others. Most patients (86.2%) described the length of the weekly check-in as "just right" at both time points. Most patients also reported feeling confident in their ability to complete weekly check-ins independently and/or during treatment sessions (82.8–100%) and few reported that they would feel stress completing weekly check-ins independently or during treatment sessions (6.9–14.3%).

Five out of six clinicians with patients in the weekly check-in + clinician dashboard condition completed a dashboard usability questionnaire (mean = 2 questionnaires per clinician). Usability and clinical utility ratings were favorable (**Table 5**), with all clinicians reporting that the dashboard was easy to navigate, that the information was meaningful and could be helpful to clinicians who offer alcohol or drug treatment, and that they would be willing to use the dashboard in the future. Most clinicians also said that they would be able to use the dashboard during sessions with patients and that the information included on it was helpful to their patients.

DISCUSSION

Results from this clinical pilot provide preliminary support for the feasibility of incorporating a digital, remotely delivered MBC system into SUD treatment as usual in a community SUD treatment setting. Among patients and clinicians who consented to participate in this 6-month pilot, rates of engagement with the MBC system were high for patients (e.g., patients completed 85.8% of weekly check-ins, with optional free-text responses included in 44.5% of the weekly check-ins that were completed) and for clinicians (e.g., clinicians logged into the dashboard a mean of 12.20 times). Further, the system seemed to impose minimal time burden to patients and clinicians, who on average took less than 5 min to complete weekly check-ins and less than 3 min to review MBC results on the clinician dashboard, respectively. Usability ratings were favorable, with most patients reporting that the weekly check-in was interesting, helpful for self-reflection, and something they would be willing to continue using, and most clinicians reporting that the information on the dashboard was helpful and that they discussed the information with their patient.

While SUD treatment providers have been previously shown to report positive attitudes toward MBC (8, 23), studies have also identified numerous barriers that can impede the implementation of MBC in SUD treatment settings (5, 23–25). Notably, implementation barriers may vary between

TABLE 3 | Patient and clinician engagement metrics.

Patient engagement (*N* = 30)	**M**	**(SD)**
Number of weekly check-ins completed per patient (full 24-week period)	20.60	(5.54)
Number of weekly check-ins completed per patient (weeks 1–12)	10.80	(2.23)
Number of weekly check-ins completed per patient (weeks 13–24)	9.80	(3.46)
Number of weekly check-ins with an open-text response, per patient	9.17	(7.90)
Time to complete each weekly check-in (min.)[a]	4.99	(4.46)
Clinician engagement (*N* = 5 clinicians with ≥ 1 patient in weekly check-in + clinician dashboard condition who logged into the clinician dashboard)		
Number of dashboard login sessions per clinician	12.20	(9.33)
Time spent using dashboard per login session (min.)	2.30	(4.61)

[a]*Estimated based on the difference in time between when the weekly check-in was first opened and when it was submitted.*

different SUD treatment settings and may be associated with patient-, provider-, and system-level factors (23). For example, patients may have difficulty completing measures due to illness, disability, or distress; they may perceive completing MBC questionnaires as not being personally meaningful or useful; and they may experience difficulty using technology or experience usability-related barriers (24). Clinicians may perceive information provided by MBC as impersonalized or unreliable; they may experience additional workload associated with administering, scoring, and reviewing measures; and they may feel uncomfortable or uncertain about how to integrate MBC into their clinical practice (8, 23). Healthcare systems may also lack adequate structures to support MBC due to a lack of training and technical support for MBC; payment models that do not reimburse for time spent using MBC; and limited integration of MBC systems with other technologies that are used by patients and clinicians (25, 26).

Despite numerous potential barriers, the current pilot study found that digital, remotely delivered MBC was feasible to incorporate into SUD treatment as usual with high rates of engagement and high ratings of usability and clinical utility reported by patients and clinicians. It is possible that the positive findings obtained here are partly attributable to the user-centered design methods that informed the specific designs, workflows, and contents included in the MBC system that was piloted in the current study (8, 10). For example, informed by stakeholder input, we designed the MBC system to allow patients to complete weekly check-ins on their personal smartphones (in contrast to our initial idea of using tablet computers or paper questionnaires in clinic waiting rooms), which may address some implementation barriers cited above by integrating the weekly check-in with existing technologies used by patients and by eliminating the need for clinicians to administer and score MBC measures. The user-centered design approach also led us to measure domains that clinicians identified as most clinically helpful, to include questions about patients' treatment goals, and to include open-ended questions that invited optional free-text responses, potentially reducing barriers related potential lack of personalization in MBC. Many of the clinical domains in the weekly check-in reflected positive outcomes (e.g., self-efficacy, use of coping skills, and positive outlook on life) rather than focusing more exclusively on outcomes that are perceived as negative or stigmatized (e.g., substance use, relapse, and SUD symptoms), potentially helping patients and clinicians reflect on positive experiences and growth, rather than focusing more exclusively on negative experiences (e.g., substance use and SUD symptoms). Usability testing also helped us iteratively improve the format of the weekly check-in (e.g., optimizing the layout for mobile devices, improving the consistency of wording used in questions and response options) and the clinician dashboard (e.g., displaying results graphically and in text formats), potentially reducing usability-related barriers.

Seeking and incorporating input from clinicians and patients in SUD treatment settings may be critical to high rates of engagement, and adequate usability of MBC systems, which in turn may be key facilitators to implementing MBC into routine care (27). Findings from another study are consistent with this emphasis on stakeholder engagement; for example, in recent pilot study, Russell and colleagues (28) successfully pilot tested the use of a MBC system in a 15-bed residential adolescent SUD treatment setting after conducting multiple rounds of stakeholder engagement and collaborative development of MBC workflows and questionnaire items. Developing clinical technologies through user-centered design approaches may be necessary for producing clinical technologies that are more usable, engaging, and sustainable in clinical settings (29). In addition, this user-centered design approach honors the lived experiences and expertise patients and clinicians in SUD treatment settings and helps their voices be heard in clinical research and technology development.

Delivering MBC using digital technologies that are accessible from any location (including outside of the clinic) may provide several advantages in SUD treatment settings that may have also contributed to the high engagement and usability observed in this study (30). For example, patients can continue to complete weekly check-ins from any location, including when they might have irregular or infrequent contact with the clinic. This might occur when scheduled treatment sessions are scheduled infrequently, conducted virtually, missed, or inaccessible due to barriers to attendance (e.g., difficulty traveling to clinic and COVID-19-related restrictions). Digital platforms also could potentially help minimize burden to clinicians by allowing weekly check-in reminders to be automatically sent to patients and for the data in the weekly check-ins to be automatically

TABLE 4 | Patient ratings of usability, clinical utility, and self-efficacy completing weekly check-ins.

	Week 6 (*n* = 29)					Week 24 (*n* = 29)				
Usability and clinical utility of the weekly check-in	**Strongly disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly agree**	**Strongly disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly agree**
I can easily find my way on the weekly check-in.	1		1	2	24	1			4	23
I am satisfied with the language used on the weekly check-in.			2	8	19			1	8	20
The weekly check-in survey is interesting.	1	1	8	13	5		2	9	8	9
The weekly check-in survey does not contain distracting elements.		1	1	8	17			3	7	18
I find the weekly check-in helpful.		2	2	13	12		1	3	9	16
The weekly check-in can help me reflect on what I want.		1	2	14	12				12	17
The weekly check-in helps me reflect on my substance use and recovery.			1	10	18				10	19
I can imagine myself discussing the information on the weekly check-in with my clinician.	1	4	6	7	11			10	7	12
I can imagine the weekly check-in being helpful to others.			2	9	18			1	12	16
I would be willing to use the weekly check-in in the future.		1		11	17		1	3	5	20
I would recommend the weekly check-in to others.		1	2	11	15			2	8	19
Length of the weekly check-in	**Much too short**	**Too short**	**About right**	**Too long**	**Much too long**	**Much too short**	**Too short**	**About right**	**Too long**	**Much too long**
The length of the weekly check-in is:		1	25	3			3	25	1	
	Week 6 (*n* = 29)					**Week 24 (*n* = 28)**				
Self-efficacy for completing the weekly check-in	**Strongly disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly agree**	**Strongly disagree**	**Disagree**	**Neutral**	**Agree**	**Strongly agree**
I have been able to understand the questions that were asked in the weekly check-in.	1			7	21				5	23
I would feel confident in my ability to answer similar questions if I were completing the weekly check-in at home by myself.	1	1		6	21				6	22
I would feel confident in my ability to answer similar questions if I were completing the weekly check-in during my treatment.		1	3	5	19				8	20
I would feel stressed if I were asked to complete the weekly check-in while I was at home by myself.	15	10	1	1	2	18	5	2	1	2
I would feel stressed if I were asked to complete the weekly check-in while I was in a treatment session with my clinician.	12	8	6	2		14	7	3	3	1

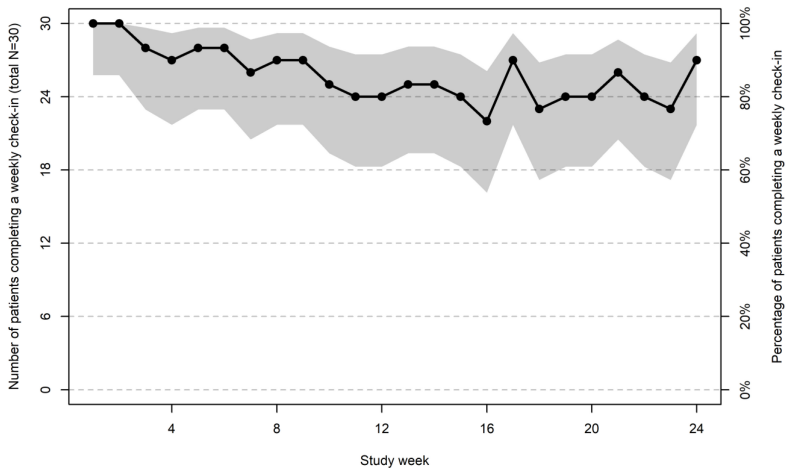


FIGURE 4 | Number **(left axis)** and percentage **(right axis)** of patients completing a weekly check-in during each week of the clinical pilot. The shaded region reflects the 95% CI of the estimated percentage for each week.

TABLE 5 | Clinician ratings of usability and clinical utility (n = 5)*.

Usability and clinical utility of the clinician dashboard	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I could easily find my way on the dashboard.				1	4
I was satisfied with the language used on the dashboard.				2	3
The dashboard provided me with meaningful information.				3	2
The information on the dashboard was helpful to my patient.			1	2	1
The information on the dashboard can be helpful to clinicians who offer alcohol or drug treatment.				2	3
I was able to discuss the information on the dashboard with my patient.				3	
I would be willing to use the dashboard in the future.				1	4
I would be able to use the dashboard during sessions with patients.				3	2

**Clinicians could complete a usability questionnaire for each patient they had enrolled in the weekly check-in + dashboard condition. When a participant completed multiple questionnaires, the average ratings across questionnaires for that participant were used.*

scored, stored, and presented on the clinician dashboard in numerous formats.

Measuring potential mechanisms of change in SUD treatment—including craving, coping skills, abstinence self-efficacy, engagement in valued activities, depression symptoms, and therapeutic alliance—may be a valuable contrast to the common practice for outcome measures to focus on substance use as the primary treatment outcome. For example, many clinical trials of SUD treatments measure abstinence and/or reductions in substance use as the primary clinical endpoint. Likewise, in real-world clinical practice it is common for patients to discuss how long they have been abstinent from alcohol and/or drugs or how often they use substances to gauge their treatment progress. However, gauging substance use treatment progress by focusing primarily on substance use and abstinence may reinforce existing stigma and black-and-white thinking related to substance use (e.g., a person is either succeeding or failing based on whether they are drinking or using drugs), while also failing to capture a more complete or holistic understanding of patient progress across a range of clinical domains during treatment. In contrast, assessing multiple clinical domains, including measures that do not directly reflect substance use or abstinence, may help patients and clinicians better understand treatment progress more holistically. It also may potentially deemphasize abstinence or reduced substance use as the sole purpose of SUD treatment and instead help emphasize that SUD treatment can potentially impact multiple dimensions within a person's life.

There are important limitations to this study. The MBC system and procedures were tested within a single, large, publicly funded addiction and mental health treatment program affiliated with an academic medical center; therefore, results may not generalize to other types of settings. By design, we recruited a small sample for this pilot study, which precluded us from conducting subgroup analyses that could evaluate whether engagement, usability, and clinical utility ratings varied between specific subgroups. Our sample of clinicians also was small and predominantly white and non-Hispanic. The sample only included patients with smartphones, and while most patients in SUD treatment have smartphones (31, 32), the approaches used here would not be accessible to all patients in SUD treatment. All patients in the sample elected to participate in a research study focusing on MBC, received payments for attending research interviews (but not for completing MBC questionnaires), and were supported by a research coordinator

for the first 3 months of the 6-month trial, and thus the results may not fully capture feasibility and engagement of the MBC system if it were implemented for all patients outside of a research study context. The clinical domains that were assessed in the weekly check-in were informed by the preferences of clinicians from the same setting; however, the clinical utility of these domains could vary across treatment settings and other measures that have been proposed for MBC in SUD treatment were not tested here, such as the Brief Addiction Monitor (33–35), the Outcome Questionnaire-45 (36, 37), or measures based on SUD symptoms (20, 21). Finally, research testing the impact of MBC on patients' clinical outcomes in SUD treatment has been limited to date (6, 7), although one study has suggested that MBC may help some patients in SUD treatment make faster reductions in their alcohol use (36).

There are also important strengths of this study. The MBC system was tested within a community treatment setting added onto treatment as usual, bolstering the external validity of the findings. Patients in the sample were diverse with respect to age, gender, race, education, and housing and reflected the demographic distribution of patients within the clinic. Clinicians in the study reported using multiple types of treatment approaches, suggesting that engagement and usability results are not contingent on providers using a specific treatment modality. Data on engagement, usability, and clinical utility were obtained from multiple modalities (automatically generated engagement measures and self-report usability measures) and from both clinician and patient participants, providing multiple perspectives about the reactions to the MBC system that was tested.

CONCLUSION

Results from this clinical pilot suggest that the MBC system tested here can potentially be feasibly incorporated into existing SUD treatment settings with high rates of patient and clinician engagement, high usability and clinical utility, and minimal clinical disruptiveness. These findings lend support for additional efforts to test methods for implementing MBC into routine care in SUD treatment settings and to evaluate the impact of MBC on SUD treatment processes (e.g., therapeutic alliance, shared decision making, patient empowerment, and stigma reduction). Future studies should further evaluate the impact of MBC on patient outcomes, including outcomes related and unrelated to substance use (e.g., treatment engagement, goal attainment, and patient experience). Additionally, future research should evaluate strategies for implementing MBC as part of standard of care for patients in SUD treatment, including across clinical settings that offer different treatment modalities and that serve diverse patient populations who may have different requirements for successful MBC implementation.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the University of Washington Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

KH contributed to the study design, oversight, data analysis, and writing of the results. RR and DA contributed to the study design and writing of the results. EC contributed to the data collection and writing of the results. All authors contributed to the article and approved the submitted version.

FUNDING

Research reported in this publication was supported by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) of the National Institutes of Health (NIH) under award number K01AA024796. The use of REDCap software was supported by NIH grants UL1TR002319, KL2TR002317, and TL1TR002318.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyt.2022.840409/full#supplementary-material

REFERENCES

1. SAMHSA. *Results From the 2018 National Survey on Drug Use and Health*. (2019). Available online at: https://www.samhsa.gov/data/ (accessed January 14, 2021).
2. Fortney JC, Unützer J, Wrenn G, Pyne JM, Smith GR, Schoenbaum M, et al. A tipping point for measurement-based care. *Psychiatr Serv.* (2017) 68:179–88. doi: 10.1176/appi.ps.201500439
3. Lewis CC, Boyd M, Puspitasari A, Navarro E, Howard J, Kassab H, et al. Implementing measurement-based care in behavioral health: a review. *JAMA Psychiatry.* (2019) 76:324–35. doi: 10.1001/jamapsychiatry.2018.3329
4. Scott K, Lewis CC. Using measurement-based care to enhance any treatment. *Cogn Behav Pract.* (2015) 22:49–59. doi: 10.1016/j.cbpra.2014.01.010
5. Boswell JF, Kraus DR, Miller SD, Lambert MJ. Implementing routine outcome monitoring in clinical practice: benefits, challenges, and solutions. *Psychother Res.* (2015) 25:6–19. doi: 10.1080/10503307.2013.817696
6. Goodman JD, McKay JR, DePhilippis D. Progress monitoring in mental health and addiction treatment: a means of improving care. *Prof Psychol Res Pract.* (2013) 44:231–46. doi: 10.1037/a0032605
7. Carlier IVE, van Eeden WA. Routine outcome monitoring in mental health care and particularly in addiction treatment: evidence-based clinical and research recommendations. *J Addict Res Ther.* (2017) 8:1–7.

8. Tauscher JS, Cohn EB, Johnson TR, Diteman KD, Ries RK, Atkins DC, et al. What do clinicians want? Understanding frontline addiction treatment clinicians' preferences and priorities to improve the design of measurement–based care technology. *Addict Sci Clin Pract.* (2021) 16:1–15. doi: 10.1186/s13722-021-00247-5
9. IDEO. *The Field Guide to Human–Centered Design*. Palo Alto, CA: IDEO (2015).
10. Hallgren KA, Diteman KD, Atkins DC. Designing measurement-based care technology for frontline addiction treatment settings: usability testing and iterative design improvement. *Res Soc Alcohol.* (2020). Available online at: https://doi.org/10.1111/acer.14358 (accessed December 3, 2021).
11. Bradley KA, DeBenedetti AF, Volk RJ, Williams EC, Frank D, Kivlahan DR. AUDIT-C as a brief screen for alcohol misuse in primary care. *Alcohol Clin Exp Res.* (2007) 31:1208–17. doi: 10.1111/j.1530-0277.2007.00403.x
12. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. ambulatory care quality improvement project (ACQUIP). alcohol use disorders identification test. *Arch Intern Med.* (1998) 158:1789–95. doi: 10.1001/archinte.158.16.1789
13. Hearon BA, Pierce CL, Björgvinsson T, Fitzmaurice GM, Greenfield SF, Weiss RD, et al. Improving the efficiency of drug use disorder screening in psychiatric settings: validation of a single-item screen. *Am J Drug Alcohol Abuse.* (2015) 41:173–6. doi: 10.3109/00952990.2015.1005309
14. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* (2009) 42:377–81. doi: 10.1016/j.jbi.2008.08.010
15. Neale J, Vitoratou S, Finch E, Lennon P, Mitcheson L, Panebianco D, et al. Development and validation of "SURE": a patient reported outcome measure (PROM) for recovery from drug and alcohol dependence. *Drug Alcohol Depend.* (2016) 165:159–67. doi: 10.1016/j.drugalcdep.2016.06.006
16. Kroenke K, Spitzer RL, Williams JBW. The patient health questionnaire-2: validity of a two-item depression screener. *Med Care.* (2003) 41:1284–92. doi: 10.1097/01.MLR.0000093487.78664.3C
17. Litt MD, Kadden RM, Cooney NL, Kabela E. Coping skills and treatment outcomes in cognitive-behavioral and interactional group therapy for alcoholism. *J Consult Clin Psychol.* (2003) 71:118. doi: 10.1037//0022-006x.71.1.118
18. Breslin FC, Sobell LC, Sobell MB, Agrawal S. A comparison of a brief and long version of the situational confidence questionnaire. *Behav Res Ther.* (2000) 38:1211–20. doi: 10.1016/s0005-7967(99)00152-7
19. Munder T, Wilmers F, Leonhart R, Linster HW, Barth J. Working alliance inventory-short revised (WAI-SR): psychometric properties in outpatients and inpatients. *Clin Psychol Psychother.* (2010) 17:231–9. doi: 10.1002/cpp.658
20. Sayre M, Lapham GT, Lee AK, Oliver M, Bobb JF, Caldeiro RM, et al. Routine assessment of symptoms of substance use disorders in primary care: prevalence and severity of reported symptoms. *J Gen Intern Med.* (2020) 35:1111–9. doi: 10.1007/s11606-020-05650-3
21. Hallgren KA, Matson TE, Oliver M, Witkiewitz K, Bobb JF, Lee AK, et al. Practical assessment of alcohol use disorder in routine primary care: performance of an alcohol symptom checklist. *J Gen Intern Med.* (2021). Available online at: https://doi.org/10.1007/s11606-021-07038-3 doi: 10.1007/s11606-021-07038-3 (accessed December 3, 2021).
22. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders (DSM-5§)*. Washington, DC: American Psychiatric Pub (2013).
23. Revill AS, Anderson LE, Kidd C, Gullo MJ. Drug and alcohol practitioners' attitudes toward the use of standardized assessment. *Addict Behav.* (2022) 128:107231. doi: 10.1016/j.addbeh.2021.107231
24. Nielsen AS, Kidholm K, Kayser L. Patients' reasons for non-use of digital patient-reported outcome concepts: a scoping review. *Health Inform J.* (2020) 26:2811–33. doi: 10.1177/1460458220942649
25. McCarty D. Measurement—based care: the implementation challenge. *Addiction.* (2019) 114:1354–5. doi: 10.1111/add.14590
26. Veerbeek MA, Voshaar RCO, Pot AM. Clinicians' perspectives on a web-based system for routine outcome monitoring in old-age psychiatry in the Netherlands. *J Med Internet Res.* (2012) 14:e1937. doi: 10.2196/jmir.1937
27. Martin-Cook K, Palmer L, Thornton L, Rush AJ, Tamminga CA, Ibrahim HM. Setting measurement-based care in motion: practical lessons in the implementation and integration of measurement-based care in psychiatry clinical practice. *Neuropsychiatr Dis Treat.* (2021) 17:1621–31. doi: 10.2147/NDT.S308615
28. Russell KC, Gillis HL, Law L, Couillard JA. Pilot study examining outcomes associated with the implementation of progress monitoring at a substance use disorder treatment program for adolescents. *Child Youth Care Forum.* (2018) 47:403–19. doi: 10.1007/s10566-018-9437-2
29. Lyon AR, Brewer SK, Areán PA. Leveraging human-centered design to implement modern psychological science: return on an early investment. *Am Psychol.* (2020) 75:1067. doi: 10.1037/amp0000652
30. Hallgren KA, Bauer AM, Atkins DC. Digital technology and clinical decision making in depression treatment: current findings and future opportunities. *Depress Anxiety.* (2017) 34:494–501. doi: 10.1002/da.22640
31. Ashford RD, Lynch K, Curtis B. Technology and social media use among patients enrolled in outpatient addiction treatment programs: cross-sectional survey study. *J Med Internet Res.* (2018) 20:e84. doi: 10.2196/jmir.9172
32. Milward J, Day E, Wadsworth E, Strang J, Lynskey M. Mobile phone ownership, usage and readiness to use by patients in drug treatment. *Drug Alcohol Depend.* (2015) 146:111–5. doi: 10.1016/j.drugalcdep.2014.11.001
33. Cacciola JS, Alterman AI, DePhilippis D, Drapkin ML, Valadez C, Fala NC, et al. Development and initial evaluation of the brief addiction monitor (BAM). *J Subst Abuse Treat.* (2013) 44:256–63. doi: 10.1016/j.jsat.2012.07.013
34. Nelson KG, Young K, Chapman H. Examining the performance of the brief addiction monitor. *J Subst Abuse Treat.* (2014) 46:472–81. doi: 10.1016/j.jsat.2013.07.002
35. Gifford E, Fuller A, Stephens R, Matthews K, DePhilippis D, Hawkins E, et al. Implementation outcomes in context: leadership and measurement based care implementation in VA substance use disorder programs. *Implement Sci.* (2015) 10:A74. doi: 10.1186/s13012-016-0428-0
36. Crits-Christoph P, Ring-Kurtz S, Hamilton JL, Lambert MJ, Gallop R, McClure B, et al. A preliminary study of the effects of individual patient-level feedback in outpatient substance abuse treatment programs. *J Subst Abuse Treat.* (2012) 42:301–9. doi: 10.1016/j.jsat.2011.09.003
37. Lambert MJ, Gregersen AT, Burlingame GM. The Outcome questionnaire-45. 3rd ed. *The Use of Psychological Testing for Treatment Planning and Outcomes Assessment: Instruments for Adults*. (Vol. 3), Mahwah, NJ: Lawrence Erlbaum Associates Publishers (2004). p. 191–234.

Author Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of NIAAA or the NIH.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Psychiatry

ORIGINAL RESEARCH
published: 27 April 2022
doi: 10.3389/fpsyt.2022.857304

mHealth Solutions for Mental Health Screening and Diagnosis: A Review of App User Perspectives Using Sentiment and Thematic Analysis

Erin Lucy Funnell[1], *Benedetta Spadaro*[1], *Nayra Martin-Key*[1], *Tim Metcalfe*[2] *and Sabine Bahn*[1]*

[1] *Cambridge Centre for Neuropsychiatric Research, Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge, United Kingdom,* [2] *Independent Researcher, Cambridge, United Kingdom*

Mental health screening and diagnostic apps can provide an opportunity to reduce strain on mental health services, improve patient well-being, and increase access for underrepresented groups. Despite promise of their acceptability, many mental health apps on the market suffer from high dropout due to a multitude of issues. Understanding user opinions of currently available mental health apps beyond star ratings can provide knowledge which can inform the development of future mental health apps. This study aimed to conduct a review of current apps which offer screening and/or aid diagnosis of mental health conditions on the Apple app store (iOS), Google Play app store (Android), and using the m-health Index and Navigation Database (MIND). In addition, the study aimed to evaluate user experiences of the apps, identify common app features and determine which features are associated with app use discontinuation. The Apple app store, Google Play app store, and MIND were searched. User reviews and associated metadata were then extracted to perform a sentiment and thematic analysis. The final sample included 92 apps. 45.65% (n = 42) of these apps only screened for or diagnosed a single mental health condition and the most commonly assessed mental health condition was depression (38.04%, n = 35). 73.91% (n = 68) of the apps offered additional in-app features to the mental health assessment (e.g., mood tracking). The average user rating for the included apps was 3.70 (SD = 1.63) and just under two-thirds had a rating of four stars or above (65.09%, n = 442). Sentiment analysis revealed that 65.24%, n = 441 of the reviews had a positive sentiment. Ten themes were identified in the thematic analysis, with the most frequently occurring being performance (41.32%, n = 231) and functionality (39.18%, n = 219). In reviews which commented on app use discontinuation, functionality and accessibility in combination were the most frequent barriers to sustained app use (25.33%, n = 19). Despite the majority of user reviews demonstrating a positive sentiment, there are several areas of improvement to be addressed. User reviews can reveal ways to increase performance and functionality. App user reviews are a valuable resource for the development and future improvements of apps designed for mental health diagnosis and screening.

Keywords: app users, app reviews, digital health, mental health, mHealth, sentiment analysis, thematic analysis

OPEN ACCESS

Edited by:
Abhishek Pratap,
University of Toronto, Canada

Reviewed by:
Stephen Schueller,
University of California, Irvine, United States
Kelsey Lynett Ford,
Lirio, United States
Neelamadhav Gantayat,
IBM Research, India

***Correspondence:**
Sabine Bahn
sb209@cam.ac.uk

Specialty section:
This article was submitted to Digital Mental Health, a section of the journal Frontiers in Psychiatry

Received: *18 January 2022*
Accepted: *21 March 2022*
Published: *27 April 2022*

Citation:
Funnell EL, Spadaro B, Martin-Key N, Metcalfe T and Bahn S (2022) mHealth Solutions for Mental Health Screening and Diagnosis: A Review of App User Perspectives Using Sentiment and Thematic Analysis. Front. Psychiatry 13:857304. doi: 10.3389/fpsyt.2022.857304

INTRODUCTION

The prevalence of mental health concerns and disorders has increased following the COVID-19 pandemic (1–6). Despite this, during the initial lockdown in the United Kingdom contact with mental health services fell (6). Therefore, an increase in the demand for mental healthcare is predicted (6). The increased demand on mental healthcare services alongside workforce shortages (7) pose major obstacles to timely and effective mental healthcare provision (8, 9). This is concerning as long wait times in mental health are associated with poorer outcomes including increased suicidal risk, poorer social adjustment, decreased treatment responses and a higher risk of comorbidities (10–17). In contrast, early intervention and at-home treatment for mental health can minimize hospital admissions, shorten hospital stays, and result in cost savings for healthcare providers (18–21). Therefore, finding faster ways to assess, triage and treat mental health patients is vital. Mental health screening can provide a fast way to identify patients who could benefit from additional, more comprehensive mental health assessments (22). Thus, screening could also identify patients whose mental health could be managed with self-help strategies and do not require formal treatment (23). Screening could additionally reduce strain on primary care services (24) which is vital considering that, as of 2018, General Practitioners (GPs) report that 40% of appointments are related to mental health concerns (22, 25). This could be accomplished *via* signposting to other services following mental health screening, as a case study conducted in a group of GP surgeries demonstrated that active signposting frees up 80 additional appointments per surgery each week (26). Additionally, screening could minimize the risk of overlooking the presence of a mental health condition, which could delay access to treatment and worsen their prognosis (27).

In this regard, mobile health (mHealth) tools, such as applications (apps), could reduce strain on and increase access to mental health support or services. Additionally, they can facilitate early identification of mental health disorders and support self-management (28). mHealth tools are convenient, instant, and scalable (24, 29), as well as empowering individuals in managing their mental health (28–30), without the restrictions imposed by traditional mental health services (i.e., lack of access and long waiting times) (28, 29). Apps can also aid in engaging typically hard-to-reach patient populations by reducing stigma and increasing help-seeking behaviors (30). This is increasingly important following the recent COVID-19 pandemic, during which already present health disparities have only widened (31). mHealth for screening and management of mental health issues have gained popularity in recent years (32). The current estimates for the number of mental health apps available for public use stands at between 10,000 and 20,000 (33). In addition, the recent COVID-19 pandemic has further highlighted the benefits of mHealth apps for mental health by offering patients the opportunity to access a variety of mental health support during the COVID-19 pandemic, when traditional face-to-face care was not possible (34).

Furthermore, several studies have demonstrated the acceptability of digital technologies for improving the accessibility of mental health care and screening (35, 36). In a recent cross-sectional study involving over 8,000 users of a depression-screening app, it was found that a high percentage (73.90%) of app users completed the screening questionnaire (37). A second, multinational study of a depression screening app showed that, of those who downloaded the app, over two thirds completed a depression screening (38). These findings indicate that questionnaires completed *via* apps are a potentially feasible method of delivery for mental health assessments.

However, there is evidence for low user engagement (39), and a high drop-out rate for mental health apps (40, 41), with the drop-out rate appearing to be higher in real-world settings compared to clinical trials (42). One possible explanation for high drop-out could be issues related to usability, which are widely recognized in mental health apps (43, 44). In a systematic review of engagement in digital mental health interventions, issues with usability are a substantial barrier (45). ORCHA, a digital health compliance reviewer, found that 29.6% of the reviewed mental health apps do not meet their quality thresholds (46). These quality thresholds include usability issues. In addition to lack of compliance with clinical standards and data protection guidelines (46). Additionally, other factors are put forward as being related to low app engagement including concerns about the security of user data (44, 45), the app not adequately meeting the users' needs, and the app being considered untrustworthy by users (44).

Publicly available app reviews and ratings can provide a wealth of information regarding user perspectives and usability issues (47–55). Indeed, ratings are a key decision-making tool for whether a user downloads and uses an app (49) and reviews can highlight key issues which may not be reflected in ratings alone (50). In studies analyzing user reviews of mental health apps including cognitive behavioral therapy (CBT) and mood monitoring apps, main findings emphasize usability and visual appeal (43, 44, 51–55). In addition, complaints reported by users seem to center on poor design, bugs, and issues related to content with a lack of clear information on how to use the app. In extreme cases, some users report that usability issues caused them to immediately stop using the app (43).

Insights from app user reviews of general mental health apps identified a mental health assessment as a feature users perceived as positive (55). Despite this however, to our knowledge, no studies to date have focused on user perceptions of screening and/or diagnostic assessments included within apps designed for mental health. Considering how essential early screening is, both for patient outcomes (10–21) and potentially for reducing strain on care services, understanding user perception of mental health apps which offer a screening and/or diagnostic assessment is essential.

Therefore, we set out to conduct a review of publicly available apps which offer a self-administered mental health screening and/or diagnostic assessment. Additionally, we aimed to investigate the content of user reviews, with a focus on themes related to the mental health screening and/or diagnostic assessment offered within the app.

Sentiment analysis was employed to determine user perception of their experiences using the app. Furthermore, thematic analysis was used to identify both app feature themes

related to the app in general and those specific to the included mental health screening and/or diagnostic assessment. Thematic analysis utilized additionally to identify which app feature themes were associated with discontinuation in app use.

Whilst there is an overall lack of consensus on how to analyze app reviews, a combination of sentiment and thematic analysis was the favored analysis method in previous studies (51, 54, 55). Sentiment analysis can be used to identify the feelings and attitudes expressed by an individual in relation to a specific area. It is a popular method of analysis for user reviews (56) as it can determine the overall opinion, either positive, negative, or neutral, within short, informal text passages. On the other hand, thematic analysis allows a rich investigation of themes and their frequency within the data (57), thereby aiding our understanding of key app features as reported by users. The methods and findings from this study can inform future research efforts in mining large app review datasets as well as informing a user-centric design of future mental health apps which offer a screening or diagnostic assessment.

METHODS

Due to the focus of the current study being on user perspectives of publicly available apps which offer a mental health assessment, searches of app stores were conducted. However, app stores are not designed for rigorous, reproducible searches (58). In an effort to address this, the current study was inspired by a scoping review approach to improve the transparency and reproducibility. The preferred reporting items for systematic reviews and meta-analyses extension for scoping reviews (PRISMA-ScR) checklist was used to improve reporting of the methods implemented and searches which were conducted (see **Supplementary Material 1**) (59). The current study was also registered as a scoping review with the Open Science Framework [OSF; (60)].

The scope of the review included mental health apps whose intended user populations were adults (18+) who were searching for an app which included a self-administered screening for or diagnosis of common mental health disorders. Apps of interest offered a self-administered question- and answer-based digital screening or diagnostic tools.

Search Strategy

The current study used two different app search strategies, performed in July 2021. A manual search of the Apple and Google Play app stores was conducted by the first author (EF). The app store searches were performed using the search terms: (1) "Mental health assessment", (2) "Mental health test", (3) "Mental health symptom checker" and (4) "Mental health check-up". The number of results were capped at a maximum of 200 apps per search term to provide a comprehensive view of the current landscape of available apps, while still being feasible for manual analysis.

The decision was made to use broad search terms to capture the experience of app searchers who are seeking a general mental health assessment of common mental health disorders, whilst still identifying apps designed for assessment of specific disorders. This is in line with previous similar literature focused on analysis of user reviews of mental health apps, which also favored general search terms when performing app store searches (51, 54, 55).

Additionally, the M-health Index and Navigation Database (MIND) (61) was searched to identify apps of interest. MIND is a publicly available database of mental health apps which have been reviewed against the American Psychiatric Association's App Evaluation Model. The filter, "Assessments/Screening", was applied to the application library to narrow the search to apps of interest to the current study. Any relevant apps identified during the search of MIND were then accessed *via* the Apple and/or Google Play app store.

App Selection Criteria

The inclusion criteria (**Table 1**) were developed in consultation with a practicing psychiatrist (SB). After the searches were performed (EF), duplicate apps from different search terms were identified and removed (EF). Independent reviewers (EF/BS/NMK) then screened all the identified app's app store descriptions against the inclusion criteria (**Table 1**). Apps were labeled as "exclude," "include," or "maybe". Any disagreements regarding the labeling were discussed among the reviewers until a consensus was reached. After discussion, any apps which were still labeled as "maybe" ($n = 7$) were downloaded and checked for suitability (EF/BS).

If the same app was identified in both the app stores, both were screened using the app store description from each store for suitability against the inclusion criteria. If both apps from the different app stores were deemed relevant, both were included in the dataset. Both were included to provide a complete set of user reviews between both stores and to account for any between-store app differences.

Data Analysis

Descriptive Information and App Features

Descriptive information was manually retrieved for each included app from the description provided in the app store. This included information about the cost and in-app purchases, additional app functions (i.e., if the app offers self-help advice or strategies), the number of mental health conditions screened in the app, which specific mental health condition(s) the app assessed and whether the app had a medical device certification. The apps identified using the MIND database were found in the relevant app stores, and descriptive data was collected from app store descriptions. All the descriptive app data was collated into an Excel spreadsheet (see **Supplementary Material 2.1**).

App Review Extraction

The app user review selection followed a method utilized by previous studies with a similar focus (48, 51). Reviews can be organized by using filters in both of the app stores included in the current study. Given the current study's focus on users' perspectives of mental health screening and/or diagnostics apps, it was decided to only analyze a subset of app reviews filtered by "most helpful" and by date. "Most helpful" reviews are determined within app stores by users up- or down-voting other users' reviews as either "Helpful" or "Not helpful". Therefore, it is likely that the resulting sample includes reviews that users are

TABLE 1 | Inclusion and exclusion criteria for apps in the current study.

	Inclusion criteria	Exclusion criteria
Accessibility of the app	Available for download through the official Apple app store or Google Play app store, without a referral. Either available for free or at cost, which may or may not offer in-app purchases	Not publicly available (i.e., requires a referral from a healthcare provider to access)
Intended population for the app	Intended for use in the general adult population (18+)	Intended for use in a specific population (i.e., pregnant individuals or individuals in the perinatal period, veterans/ active service members, refugees)
Assessment offered within the app	Any app which offers a self-administered, question-and-answer based mental health assessment (i.e., a questionnaire, conversational agent)	An app which does not offer a self-administered question-and answer-based mental health assessment (e.g., Rorschach test or designed to be administered by a healthcare professional)
Mental health condition assessed in the app	Offers screening and/or diagnostic assessment for any of the following conditions/symptoms: Bipolar disorder (BD), Major depressive disorder (MDD), Obsessive compulsive disorder (OCD), Generalized anxiety disorder (GAD), Agoraphobia, Social phobia, Panic disorder, Insomnia, Schizophrenia, Psychosis, Eating disorders (e.g., bulimia nervosa, anorexia nervosa), Personality disorders, Alcohol abuse, Substance abuse, Post-traumatic stress disorder (PTSD), Acute stress disorder, Adjustment disorder, Autism spectrum disorders (ASDs), Attention hyperactivity deficit disorder (ADHD), Self-harm, Suicidal thoughts and/or suicidality risk	Offer screening and/or diagnostic assessment for: neuropsychiatric disorders (e.g., dementias), any disorders that are due to clinically confirmed temporary or permanent dysfunction of the brain, physical health disorder or measure (e.g., a heart rate monitor) OR The app does not exclusively screen for or diagnose a mental health condition (i.e., also screens for physical health conditions)

interacting with and find most relevant. Additionally, a previous similar study also sorted reviews by "helpfulness" in an attempt to ensure there was a mix of both positive and negative reviews (54). However, differing from the previous study, the scope of included reviews was also limited to reviews submitted within the last 6 months. This was done in an effort to retrieve reviews which were likely relevant to the current app version.

In order to extract the "most helpful" user reviews and relevant metadata for these apps, scraping of the app stores was performed using open-source code (62) for the Google Play app store, and was performed in Node.js using the app-store-scraper module (63) for the Apple app store. The review extraction from the app stores was performed in August and September of 2021. If an app from the dataset was available for download on both the Google Play and Apple app stores then any relevant reviews from both app stores were scraped.

Sentiment Analysis

A sentiment analysis was manually conducted within an excel spreadsheet. The sentiment of each review was determined through consensus of at least two independent reviewers. Each review was manually labeled as either "positive", "negative" or "neutral" (EF) depending on their sentiment. The reviews were manually re-analyzed under blinded conditions (BS/NMK). Any disagreements on the sentiment labeling of the reviews were discussed by all authors until a consensus was reached. Any reviews not written in English or reviews where a sentiment could not be determined (e.g., "Never really used this app much".) were labeled as "unclear". Any reviews which were not relevant to the focus of the study (i.e., questions to app developers, information about their mental health symptoms with no reference to the app, a review written on the behalf of someone else, a review of the clinician or service rather than the app itself) were labeled as "not relevant". Any reviews labeled as "unclear" or "not relevant" were removed from the dataset.

Thematic Analysis

Following the sentiment analysis, any reviews lacking enough data to perform a thematic analysis (< 5 words) were removed from the dataset. The thematic analysis was manually conducted in an excel spreadsheet following the Braun and Clarke framework (57).

The reviews were read and re-read until the first author (EF) was familiar with them and any initial ideas were noted. Initial codes were created (EF) and added to a coding framework with brief descriptions for each code. The reviews were then manually allocated codes under blinded conditions (EF/BS/NMK) using this coding framework. Any inconsistencies in the code allocations between the authors (EF/BS/NMK) were discussed until a consensus was reached. During the thematic analysis, every review included in the dataset received its final coding based on the consensus of at least two independent reviewers.

The identified codes were then grouped into broader themes, independently by two reviewers (EF/BS), which were then discussed with the third reviewer (NMK) until consensus was reached. Once the thematic labeling was finalized and code/theme frequencies had been calculated, theme co-occurrence was calculated in Excel (see **Supplementary Materials 2.3–2.6**). This included determining which themes were commonly identified in combination within the user reviews.

To identify app features which were associated with app use discontinuation, reviews which referred to app use discontinuation were labeled during the thematic analysis

(EF/BS/NMK). User reviews were labeled as commenting on app use discontinuation if the user stated so either explicitly (i.e., the user stated they stopped using the app, deleted the app, uninstalled the app or found a different app to use) or if the user implied they would not use the app again (i.e., the user review describing the app as a waste of time or the user not being able to use the app at all).

Additionally, the thematic analysis was compared against the sentiment analysis in Excel. This analysis was performed in order to determine the context in which specific app features were mentioned in the review.

RESULTS

Description of Included Apps and Review Extraction

The final app sample included 92 apps, comprising 69 apps from the Google Play app store and 23 apps from the Apple app store (**Figure 1**). Twelve of the identified apps (13.04%) were available for download on both the Google Play and Apple app stores.

Of the apps identified in the searches of the app stores (N = 1,378), 6.67% (n = 92) were relevant to the focus of the study, 1,286 apps were excluded as they did not meet the inclusion criteria of the study (**Figure 2**).

The majority of the included apps offered mental health screening (91.30%, n = 84) and only one of the included apps offered a diagnostic assessment. In a subset of the apps (7.61%, n = 7) it was unclear whether they offered mental health screening or diagnosis. Just under half of the included apps (46.74%, n = 43) offered a disclaimer in the app description (i.e., to consult a doctor after receiving their results, that the result is not a diagnosis, or only to use the results for educational purposes).

Of the included apps, the same proportion of apps assessed a single mental health condition (e.g., depression) as compared to apps which assessed more than one mental health condition (**Figure 3**). The most commonly assessed mental health conditions within the included apps was depression (38.04%, n = 35), followed by anxiety and/or anxiety spectrum disorders (i.e., generalized anxiety disorder, social anxiety disorder, panic disorder; 30.43%, n = 28) and bipolar disorder (11.96%, n =1 1). Eight apps (8.70%) were unclear as to which conditions they assessed within their app store description. Of the apps which assessed more than one condition, 13 (30.95%) did not list all the conditions they assessed in their app store description.

All of the included apps were free to download, with 22.83% (n = 21) of these apps offering in-app purchases. The majority of the apps (73.91%, n = 68) offered additional features in addition to the mental health assessment (e.g., the ability to track changes in symptoms over time, self-help exercises, the ability to connect virtually with a clinician). The most common additional app feature identified using the app store descriptions was a tracking or journaling feature which allowed the user to save and monitor inputted data (i.e., mood, symptoms, thought patterns) over time (42.39%, n = 39 apps with tracking functionality). This was followed by information and/or psychoeducation (39.71%, n = 27) and self-help strategies (35.29%, n = 24). Twenty-four of the identified apps (35.29%) offered only a mental health assessment with no additional features. See **Supplementary Material 2.1** for the full list of app features and information.

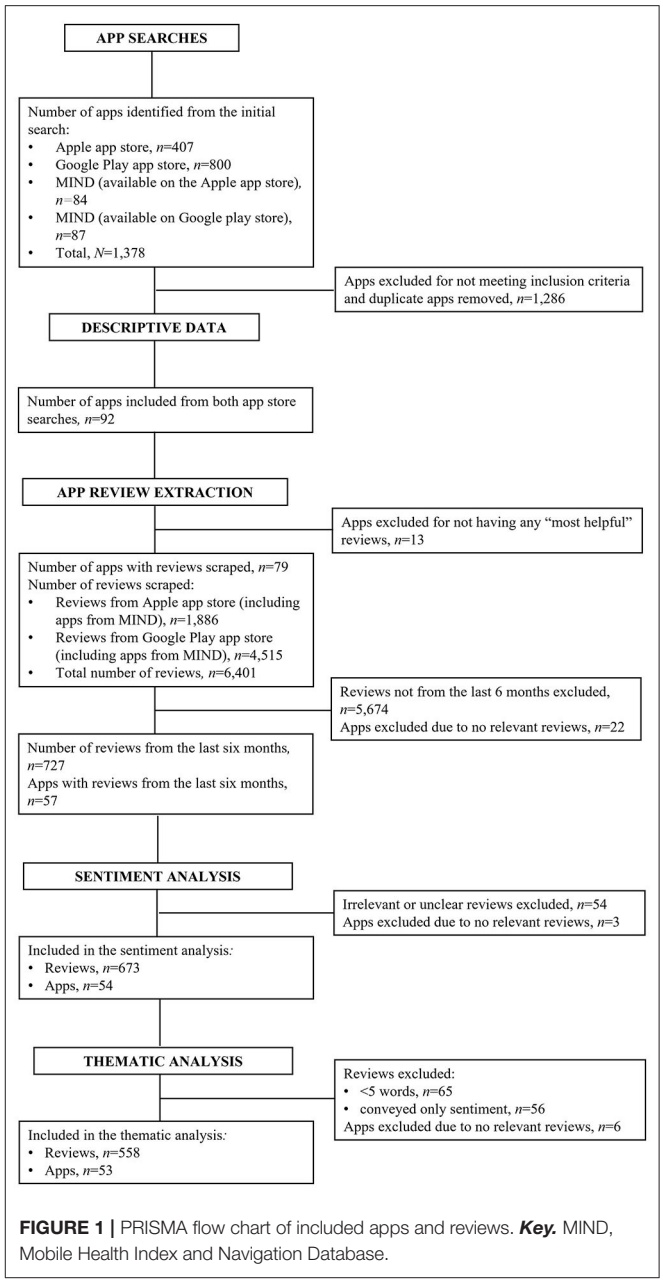


FIGURE 1 | PRISMA flow chart of included apps and reviews. ***Key.*** MIND, Mobile Health Index and Navigation Database.

Once the app reviews were extracted, apps with no relevant reviews (i.e., the code did not scrape them from the app store or reviews which were not relevant once the filters for data selection were applied) were excluded (n = 35). Seven hundred and twenty-seven reviews were identified for inclusion from the remaining 57 apps. Of these reviews, 16 were categorized as "unclear" and 35 were categorized as "not relevant" and were thus removed from the dataset. Once the "unclear" and "not

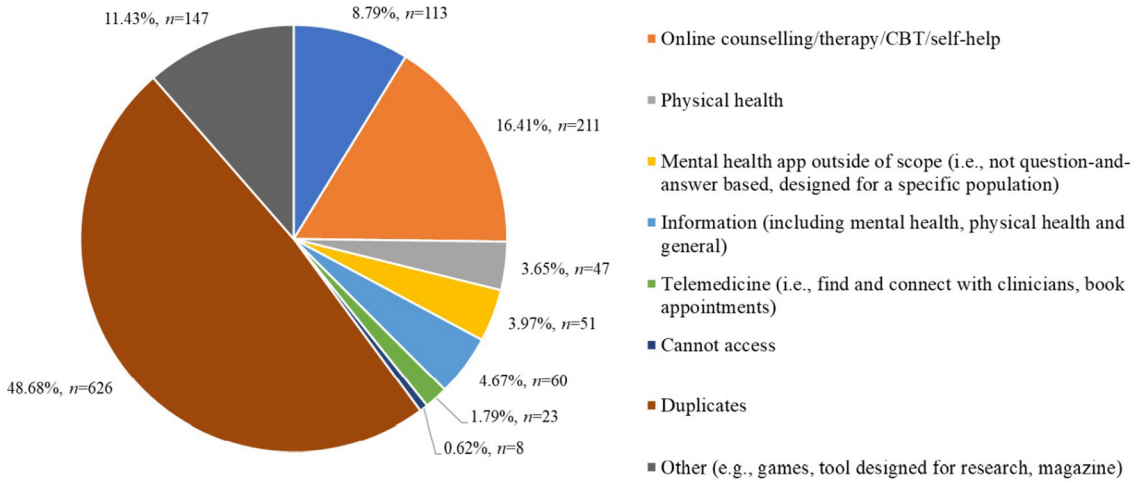


FIGURE 2 | Number of apps excluded within a given exclusion category (n = 1,286). ***Key.*** CBT, cognitive behavioral therapy.

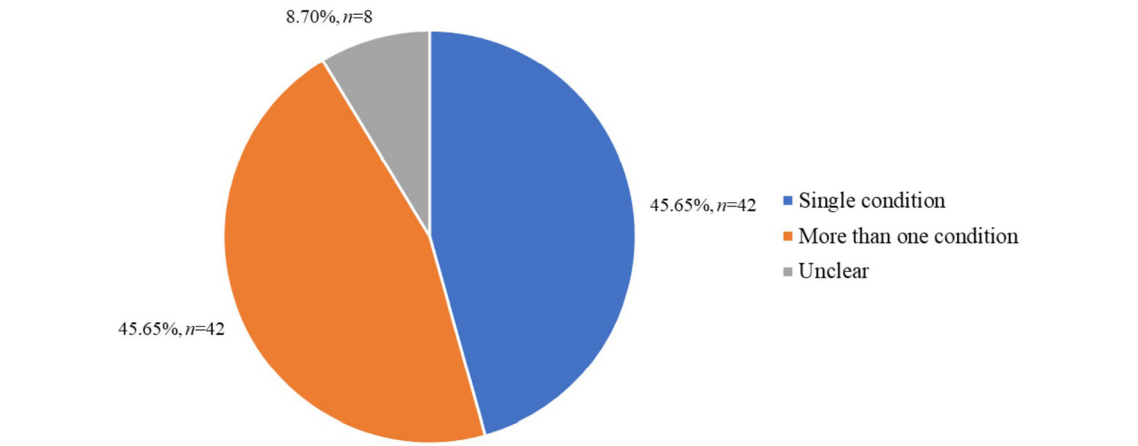


FIGURE 3 | Apps by number of conditions assessed (N = 92).

relevant" reviews were removed, 676 reviews were included in the sentiment and star rating analysis from 54 apps (**Figure 1**). Please see **Supplementary Material 2.8** to see a distribution breakdown of the number of reviews per app.

Sentiment Analysis and Star Ratings

Just under two-thirds of the reviews had a positive sentiment (65.24%, n = 441) and just over a third of the reviews had a negative sentiment (33.28%, n = 225), the minority of the reviews had a neutral sentiment (1.48%, n = 10; **Figure 4**).

The average user star rating for the included apps was 3.70 (SD = 1.63) and just under two-thirds of the included apps had a rating of 4 stars or above from users (65.09%, n = 442).

A mismatch rate between the star rating and the sentiment analysis was calculated by considering a rating of 3 to be neutral, with a rating lower than 3 considered to be negative and a rating of above 3 to be positive. If the review and rating did not convey the same sentiment (i.e., a negative sentiment in the review but a rating of 4) then it was labeled as a mismatch. There was a mismatch between the review sentiment and the star ratings provided by the user in 9.02% (n = 61) of the included reviews. (For a full breakdown of the sentiment analysis, star ratings and mismatches see **Supplementary Material 2.2**).

Thematic Analysis

Any reviews with <5 words from the sentiment analysis dataset were removed for thematic analysis (n = 59), leaving a dataset of 617 reviews. Of these, 58 reviews only conveyed sentiment and did not comment on a specific app feature or on app discontinuation. Therefore, only the

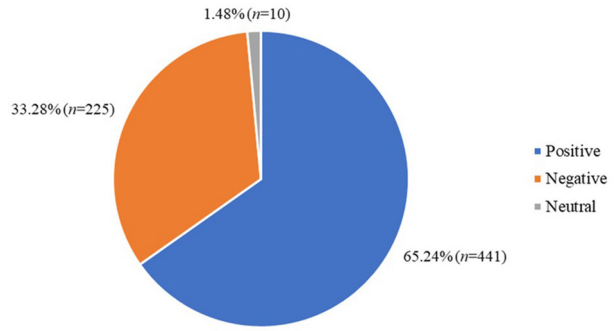


FIGURE 4 | Frequencies of different sentiments of the reviews included in the sentiment analysis dataset ($n = 676$).

remaining 559 user reviews, from 53 apps, had codes assigned to them.

Ten themes were identified within the included user reviews (**Table 2**, **Figure 5**), comprising 64 individual codes (see **Supplementary Material 2.3**). 93.01% ($n = 519$) of the reviews were assigned to more than one theme.

In the quotations included in the thematic analysis, *R* and the following number refers to a number assigned to each review included in the thematic analysis. (See **Supplementary Material 2.3** for the list of numbered reviews with their relevant codes). The review text included in the results was taken verbatim from the dataset so any spelling errors are as intended.

Performance

41.32% of the user reviews commented on app performance ($n = 231$). Half of the reviews which commented on the app performance mentioned that the app promoted mental health understanding and/or responsibility (50.65%, $n = 117$; *"Love that you can find so much out about yourself"* [R64]). The other most frequently observed indicators of performance in this theme were related to the perceived quality (i.e., good, or bad) of the assessment and/or questions included in the app (20.83%, $n = 60$; *"Simple to the point tests. Easy and they give rescources, which is nice"* [R105]) and the perceived accuracy of the app (10.07%, *n*

TABLE 2 | Frequency of the identified themes ($n = 559$; see **Supplementary Material 2.3** for the breakdown of each theme into its codes and code frequencies).

Theme	Description of the theme	Example review	Frequency, *n* (%)
Performance	Refers to the quality of the screening and/or diagnostic assessment and results included within the app as well as the fit to app purpose (i.e., if the app is detrimental to mental health)	*"Awesome apl. You learn a lot about yourself through all of the tests they have available."* [R134]	231 (41.40)
Functionality	Includes app features related to usability (i.e., ease of navigation), visual appeal, and technical issues	*"This update sucks. You have to search for the im feeling. When"m having a hard time that makes it worse. I do"t whos feedback was used but that was not a good idea."* [R269]	219 (39.25)
Therapeutic alliance and treatment	Includes app features which offer the user the ability to connect with a clinician (i.e., find a local therapist, video call with a therapist) or a treatment intervention (i.e., self-administered cognitive behavioral therapy)	*"Amazing app lessons in CBT section worked to change my life."* [R80]	140 (25.09)
Additional features and engagement style	Includes non-treatment-related additional in-app features other than the mental health screening and/or diagnostic assessment (i.e., tracking or logging of mood). Also includes features designed to increase engagement with the app (i.e., reminders to use the app)	*"This app is amazing it helps me to keep track of how"m feeling so her"s a 5 star review!!:-)"* [R165]	114 (20.43)
Accessibility	Related to the ease of accessing the app content including a paywall, cost, and inclusivity	*"Was perfect back when it was usable. Now every is behind a 100 buck a year paywall. Absolutely heartbreaking and has actively damaged my mental health progress."* [R218]	103 (18.46)
Information quality	Either poor (i.e., in-accurate) or high (i.e., detailed) quality of the information provided within the app	*"Wonderful read and great info..So thankful to the author for writing this."* [R87]	51 (9.14)
Customer service	Includes the quality of customer service provided by the app team and requests for additional app features within the app reviews	*"Unable to install application. It gives error. Tried callingriesterr but no response."* [R98]	38 (6.81)
Language	Either poor (i.e., offensive) or high (i.e., professional) quality of language used within the app	*"Horrible app! So many misspellings, I could"t count them all! The"resul" of your tests make no sense, and you HAVE TO upgrade to get any information."* [R222]	25 (4.48)
Advertisements (adverts)	The presence of advertisements (adverts) within the app	*"There are quite a few ads, but overall it was enjoying"* [R74]	21 (3.76)
Concerns related to privacy and security	User concerns related to the privacy/security or terms and conditions of the app (i.e., requests for unnecessary data)	*"Suspicious App Permission Requirementl. Why do you need access to my gallery? Why do you need to know my exact gps location?"* [R184]	6 (1.08)

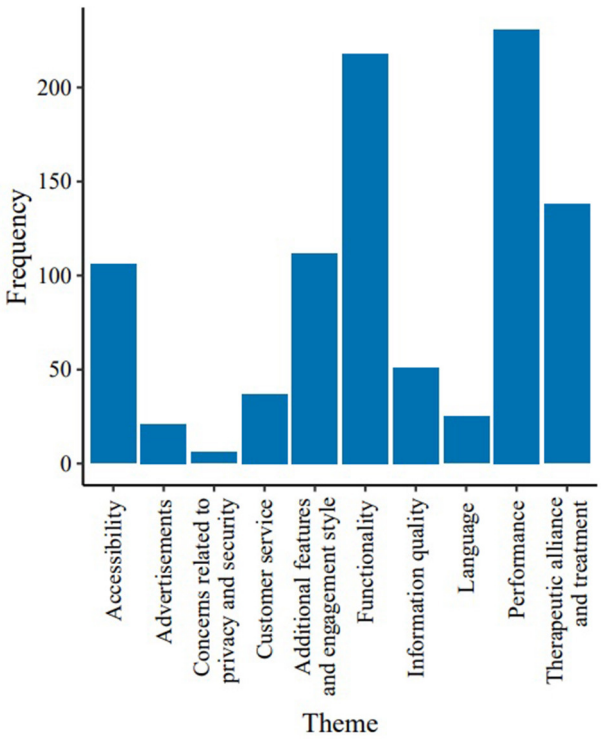


FIGURE 5 | Frequencies of different themes identified in the thematic analysis ($n = 558$).

$= 29$; *"Very few questions in the test, and they are poorly written, hence its very inaccurate."* [R119]).

Functionality

Over a third of the user reviews (39.18%, $n = 219$) commented on the functionality of the app. Just over a third of reviews in the functionality theme (35.62%, $n = 78$) were related to a bad app update, such as an update which made the app harder to use or introduced a paywall (*"I LOVE Youper but please fix the lag and confusing, cluttered environment from the recent updates!"* [R133]). Other frequent aspects of the functionality theme included the quality of the visual design (36.53%, $n = 80$) and whether the ease of app use (30.14%, $n = 66$).

Therapeutic Alliance and Treatment

Just over a quarter of the user reviews commented on therapeutic alliance and treatment (25.04%, $n = 140$). The majority of these reviews (66.43%, $n = 93$), commented on treatment features included in the app (e.g., CBT, therapy, or a prescription service) with self-help being identified as the most frequently offered treatment from the reviews commenting on a treatment feature (51.43%, $n = 72$; *"What a great tool for improving mental health. I love doing my daily personalized brain trainings!"* [R527]). Just over a quarter of user reviews commenting on therapeutic alliance and treatment (27.54%, $n = 38$) referenced functionality which could connect a user with a clinician (e.g., GP, therapist, crisis hotline; *"It's a good way to link in with your Doctor. It's nerve racking to do on a personal aspect, but the app is easy and smooth to use."* [R229]), with an additional 9.42% ($n = 13$) of reviews referencing functionality which gave users the ability to share their app data with their clinician(s) (*"Extremely helpful! Its great to be able to connect with my dietician and have the accountability piece. Its also really helpful to be able to get feedback so that I have more confidence in my ability to portion"* [R294]).

Additional Features and Engagement Style

20.39% ($n = 114$) of the user reviews referred to additional in-app features and engagement style offered in addition to the screening and/or diagnostic assessment. Tracking/logs/journaling was the most commonly identified feature within the theme of features and engagement style, present in 69.64% ($n = 78$; *"Helps me to keep track of my anxiety easily"* [R42]) of user reviews. Additional features designed to increase app engagement (e.g., rewards, motivational quotes or affirmations, reminders and/or app notifications) were present in 33.93% ($n = 38$) of reviews (*"[. . .] Every time I log a meal it gives me a coping skill or a positive statement or a cute cat picture or some other reward [. . .]"* [R429]; *"[. . .] The reminders really help keep me on track."* [R118]).

Accessibility

18.43% ($n = 103$) of the user reviews commented on the accessibility of the app. The most frequently identified app feature related to accessibility was a paywall 73.79% ($n = 76$; *"An app that used to be so useful, is literally garbage now. The purpose of app seems silly now that's only available behind a paywall, considering that most people from target audience cant afford/manage that. Very disappointed. Please request the creators to make the app free again. Even the older, simpler version would work for free users"* [R370]). Additionally, 24.27% ($n = 25$) of reviews related to accessibility commented on the app being non-accessible or non-inclusive (i.e., the app being overly expensive, the app not recognizing all gender identities and only acknowledging biological sex, the app design being unsuitable for individuals who are visually impaired or neurodivergent).

Information Quality

9.12% ($n = 51$) of reviews commented on information quality. Over three-quarters of reviews related to information quality were positive (76.47%, $n = 39$; *"This is an amazing app, very informative, covers a broad spectrum of mental health issues. Great format easy to use and easy to understand. Thank u"* [R210]). The majority of app reviews which commented on the app including poor information ($n = 11$; *"[. . .] Then when you actually start the program is literally a joke and the same info google could teach you."* [R389]) stated that the information was too basic (72.73%, $n = 8$).

Customer Service

6.80% ($n = 38$) of the user reviews commented on the customer service offered by the app team. A request for an app feature (e.g., to simplify the app design, to allow the user to save their app data, allow more flexible tracking/logs/journaling) was the most frequently identified code included within the customer service theme (63.16%, $n = 24$; *"I wish I could create an account to save the data."* [R72]). 26.32% ($n = 10$) of reviews identified

as commenting on customer service mentioned the quality of a response from the app development team (e.g., either helpful: *"[…] after I got charged prior to payment, it got fixed and I'm thankful they were able to quickly get the situation sorted out and give me a refund."* [R303]; or no response: *"Unable to install application. It gives error. Tried calling developer but no response."* [R98]).

Language

4.47% ($n = 25$) of the user reviews commented on the quality of language used within the app. Over three quarters (76%, $n = 19$) of these reviews stated that the language used within the app was poor, most commonly due to poor translation or being written by a seemingly non-proficient English speaker (36.84, $n = 7$;*"very inarticulate, as though the dev doesn't speak English"* [R44]). In addition, 21.05% ($n = 4$) of reviews commenting on poor in-app language stated that the app included offensive or inappropriate language (*"We we're insulted to read within the first paragraph of D.I.D [...]"* [R288]). The use of offensive or inappropriate language was only reported in reviews of apps designed for personality disorders and was more frequent in apps for dissociative identity disorder (DID; 75.00%, $n = 3$).

Advertisements

3.76% ($n = 21$) of the user reviews commented on advertisements included within the app. Of these, 71.43% ($n = 15$) stated that the app contained too many adverts (*"Too many adds per test. There were 4 in my test 1 banner ad. And 3 that took me out of the test. I may use again but over time this would cause me to yeet"* [R284]). In contrast, 28.57% ($n = 6$) of user reviews stated that the app had no or well-placed adverts (i.e., they are well-placed within the flow of the app or assessment and so are not overly distracting; *"Not a bad lil app here. Even with the ads BECAUSE and only because they are spaced almost perfectly in your tests."* [R186]).

Concerns Related to Privacy and Security

1.07% ($n = 6$) of reviews indicated concerns related to privacy and security, including suspicious data requirements or terms of service (*"Can't opt out of an extremely nefarious privacy policy. No thanks."* [R84]). The majority of the reviews which mentioned concerns about privacy and security (3.33%, $n = 2$) commented on "suspicious" requests to access phone data outside of the app which users perceived as unnecessary (e.g., access to the user's location or their photographs).

Theme Co-occurrence

The most common theme co-occurrence was accessibility and functionality (4.84%, $n = 27$) (**Table 3**).

Several themes were more commonly identified in the review set in combination with another theme than on their own. The additional features and engagement style theme was more frequent in combination with the performance theme (4.12%, $n = 23$; *"Very good app. I love the quote of the day and the tests are fun and accurate"*. [R144]) than the additional features or engagement style theme alone (3.41% $n = 19$). The customer service theme was more frequent in combination with the functionality theme (1.97%, $n = 11$; *"I downloaded this app and I answered about your questions, Then I went to create an account and every time I press on create account it does nothing. So I'm not going to be able to save my progress. I sent feedback to the app developers with my technical issue but who knows when it will be resolved."* [R565]) than the customer service theme alone (1.79%, $n = 10$). Finally, the language theme was more frequently identified in combination with the performance theme (1.43%, $n = 8$; *"The test may be good but the English version has many questions that need a more accurate translation. Some of the questions are impossible to understand."* [R209]) than in isolation (0.72%, $n = 4$).

TABLE 3 | Frequency of theme combinations with five or more instances in the dataset ($n = 559$; see **Supplementary Materials 2.4–2.5** for the remaining theme co-occurrences).

Theme combination	Frequency (*n*, %)
Accessibility AND Functionality	27 (4.83)
Functionality AND Performance	25 (4.47)
Additional features and engagement style AND Performance	23 (4.11)
Therapeutic alliance and treatment AND Performance	19 (3.40)
Therapeutic alliance and treatment AND Functionality	13 (2.33)
Therapeutic alliance and treatment AND Additional features and engagement style AND Performance	12 (2.15)
Additional features and engagement style AND Functionality	11 (1.97)
Functionality AND Customer service	11 (1.97)
Information quality AND Performance	9 (1.61)
Language AND Performance	8 (1.43)
Information quality AND Functionality	7 (1.25)
Therapeutic alliance and treatment AND Additional features and engagement styles	7 (1.25)
Therapeutic alliance and treatment AND Functionality AND Performance	7 (1.25)
Accessibility AND Performance	6 (1.07)
Additional features and engagement style AND Functionality AND Therapeutic alliance and treatment	6 (1.07)
Therapeutic alliance and treatment AND Additional features and engagement style AND Performance	5 (0.89)

Association Between Themes and Sentiment

Of the reviews included in the thematic analysis ($n = 617$), just under two-thirds had a positive sentiment (64.02%, $n = 395$). Just over a third of the reviews had a negative sentiment (34.85%, $n = 215$) and the remaining reviews had a neutral sentiment (1.13%, $n = 7$).

When comparing the review sentiment against themes (see **Table 4**), positive sentiment most frequently occurred in combination with performance (16.71%, $n = 66$). On the other hand, negative sentiment was more commonly identified in combination with the functionality theme (13.95%, $n = 30$). In addition, negative sentiment frequently occurred alongside the accessibility theme, in isolation (7.91%, $n = 17$), and in combination with both the functionality theme and app

TABLE 4 | Frequency of themes associated with either a positive or negative review sentiment with five or more instances in the data set (positive reviews, $n = 395$; negative reviews, $n = 215$; see **Supplementary Material 2.7** for the remaining theme and sentiment co-occurrences).

Sentiment	Theme	Frequency, *n* (%)
Positive	Performance of the assessment	66 (16.71)
	Therapeutic alliance and treatment	35 (8.86)
	Functionality	26 (6.58)
	Additional features and engagement style **AND** performance	22 (5.57)
	Therapeutic alliance and treatment **AND** performance	18 (4.56)
	Additional features and engagement style	18 (4.56)
	Functionality **AND** performance	12 (3.04)
	Therapeutic alliance and treatment **AND** additional features and engagement style **AND** performance	12 (3.04)
	Therapeutic alliance and treatment **AND** functionality	10 (2.53)
	Information quality	9 (2.28)
	Customer service	9 (2.28)
	Additional features and engagement style **AND** functionality	8 (2.03)
	Therapeutic alliance and treatment **AND** additional features and engagement style	8 (2.03)
	Information quality **AND** performance	7 (1.77)
	Therapeutic alliance **AND** functionality **AND** performance	7 (1.77)
	Information quality **AND** functionality	5 (1.27)
	Additional features and engagement style **AND** functionality **AND** performance	5 (1.27)
Negative	Functionality	30 (13.95)
	Accessibility **AND** functionality **AND** app discontinuation	19 (8.84)
	Accessibility	17 (7.91)
	Performance of the assessment	11 (5.12)
	Functionality **AND** app discontinuation	11 (5.12)
	Accessibility **AND** app discontinuation	10 (4.65)
	Functionality **AND** performance	9 (4.19)
	Accessibility **AND** functionality	8 (3.72)
	Language **AND** performance	6 (2.79)
	Functionality **AND** customer service	5 (2.33)
	Accessibility **AND** performance	5 (2.33)

discontinuation (8.84%, $n = 19$; *"I prefer the old version, the new update is just bad and as someone who is broke and can't afford the subscription I can't even talk to the AI for free, I'm forced to go look for other applications."* [R309]).

When considering themes in isolation, positive sentiment was more frequent across the identified themes than negative sentiment (**Table 5**). The highest proportion of positive sentiment compared to negative sentiment was identified in the therapeutic alliance and treatment theme (positive reviews: 86.43%, $n = 121$; negative reviews: 13.57%, $n = 19$).

Negative sentiment was more frequent in 4 themes: functionality (negative reviews: 54.79%, $n = 120$; positive reviews: 44.75%, $n = 98$), accessibility (negative reviews: 75.73%, $n = 78$; positive reviews: 23.30%, $n = 24$), language (negative reviews: 72.00%, $n = 18$; positive reviews: 28.00%, $n = 7$), and advertisements (negative reviews: 52.38%, $n = 11$; positive reviews: 42.86%, $n = 9$), with the highest proportion of negative sentiment over positive sentiment identified in the accessibility and language themes (**Table 5**).

Themes Associated With App Discontinuation

Seventy-five (13.42%) of the user reviews included in the thematic dataset ($n = 559$) were labeled as either explicitly (*"Really sad about the direction this app has taken since the last update :(Hope the developers actually listen to the hundreds of displeased customers. I'll be uninstalling"* [R233]) or implicitly commenting on app use discontinuation (i.e., the user uninstalling the app, stating that installing the app is worthless: *"Waste of tax payer money, all negative reviews on here are 100% true. DON'T WASTE YOUR TIME."* [R135]; *"I have adhd and i took the test and it said i dont so it does not work DO NOT DOWNLOAD"* [R173]).

The most common themes associated with app use discontinuation were accessibility and functionality in combination (**Table 6**; 25.00%, $n = 19$; *"It started as an awesome app, now it gets worse with every update. The subscription price hit the ceiling and now they even hide the simple single-answer emotion tracking behind the paywall. Greedy owners trying to turn it into a cashcow. Uninstalling."* [R314]). In reviews which were labeled as mentioning app use discontinuation, functionality, and accessibility, the most commonly identified codes were a paywall (52.63%, $n = 10$; *"[. . .] But now it's all just one big paywall that is impossible for me to use anymore. [. . .]* [R424]) and a poor-quality update (42.11%, $n = 8$; *"[. . .] ever since you guys started with the new update(s), it's gotten so much worse. [. . .]* [R445]).

DISCUSSION

The current study, inspired by a scoping review methodology, conducted searches of popular app stores to identify mental health apps which offered a screening and/or diagnostic assessment. Following these searches, we aimed to understand app user perceptions with a particular focus on the included assessment, *via* a qualitative analysis of the app's written reviews.

Overview of App Landscape and Sentiment

The current study demonstrated that the majority of apps resulting from the store searches were not relevant. This finding illustrates the difficulties of identifying apps using app store searches from both a user and a research perspective, due to the nature of app stores which base results on factors beyond the search terms employed (64) and allow for search results to be influenced by App Store Optimization (65). Shen et al. (66) reported similar findings: when using the search term "depression" in app stores, over a quarter of the results are

TABLE 5 | Sentiment associated with themes (see **Supplementary Material 2.7** for the remaining theme and sentiment co-occurrences).

Theme	Number of reviews with positive sentiment, *n* (%)	Number of reviews with negative sentiment, *n* (%)	Number of reviews with a neutral sentiment, *n* (%)
Accessibility ($n = 103$)	24 (23.30)	78 (75.73)	1 (0.97)
Additional features and engagement style ($n = 114$)	98 (85.96)	14 (12.28)	2 (1.75)
Advertisements (adverts) ($n = 21$)	9 (42.86)	11 (52.38)	1 (4.76)
Concerns related to privacy and security ($n = 6$)	0	6 (100.00)	0
Customer service ($n = 38$)	20 (52.63)	17 (44.74)	1 (2.63)
Functionality ($n = 219$)	98 (44.75)	120 (54.79)	1 (0.46)
Information quality ($n = 51$)	39 (76.47)	12 (23.53)	0
Language ($n = 25$)	7 (28.00)	18 (72.00)	0
Performance ($n = 231$)	173 (74.89)	57 (24.68)	1 (0.43)
Therapeutic alliance and treatment ($n = 140$)	121 (86.43)	19 (13.57)	0

TABLE 6 | Frequency of themes associated with app use discontinuation with three or more instances in the data set ($n = 75$; see **Supplementary Material 2.6** for the remaining theme co-occurrences).

Theme(s) associated with app use discontinuation	Frequency (*n*, %)
Accessibility **AND** Functionality	19 (25.33)
Functionality	11 (14.67)
Accessibility	11 (14.67)
Functionality **AND** Performance	4 (5.33)
Accessibility **AND** Functionality AND Performance	4 (5.33)
Functionality **AND** Customer service	3 (4.00)

apps not related to depression. Additionally, they showed that a quarter of the excluded apps did not mention depression in either the app title or the app store description (66). The proportion of irrelevant apps was even higher in another study, finding that under a third of the apps identified in an app store search for apps related to depression were relevant to the condition (67).

Of the apps that were relevant to the inclusion criteria of this review, many lacked sufficient information for the user regarding the content of the app in the store description. Despite the majority of included apps only offering mental health screening, just under half of the included apps offered a disclaimer in their app description. This is potentially concerning, as without a disclaimer users of the app may consider the results of a screening assessment to be a formal mental health diagnosis and, thus may not consult with a clinician. In addition, whilst most apps reported the mental health conditions they assessed, just under 10% of the included apps were unclear as to which mental health disorders they assessed. Considering the overwhelming proportion of irrelevant apps identified in the initial search, the lack of clear information further contributes to the difficulties faced by users searching the app stores. Therefore, app developers should strive to provide more detailed and accurate information within app store descriptions. This would help users in both finding an app and in ensuring its suitability from the information provided.

With respect to the user perspectives, just under two-thirds of the included app reviews received star ratings of 4 stars or above. This indicates overall satisfaction with the apps currently available for mental health screening and/or diagnostic assessment based on star rating. Additionally, the majority of written reviews for mental health screening and/or diagnostic apps conveyed a positive sentiment. The sentiment analysis also revealed that a very small minority of reviews conveyed a neutral sentiment. This could be due to app users being more likely to leave a review when they have had a particularly good or poor experience (68; *"I don't usually review apps unless I am obsessed with them or absolutely hate them. […]"* [R464]). Therefore, whilst there seems to be an overall positive user experience reported within app reviews, some information on app features which users find acceptable but neither particularly positive or negative may be missed. Hence, app developers should explore further methods, aside from only ratings and reviews, such as in-app user surveys to capture a broader spectrum of user experiences (68). Previous research has demonstrated that asking users about their experiences directly increases the likelihood of collecting feedback (68), suggesting this is a viable option for collecting neutral feedback.

Whilst star rating and written review sentiment were both overall positive, the current study revealed that star ratings should not be considered in isolation. The rate of mismatch between rating and review sentiment reported in the current study, while low, shows that star ratings may not fully capture sentiment. In addition, by analyzing the written user reviews, app developers can uncover a wealth of insights beyond what app star ratings alone can provide. This is demonstrated in the results of the thematic analysis performed in the current study, which identified 10 distinct themes which are important to users. In fact, despite the majority of reviews having a positive sentiment and high star rating by engaging in qualitative analysis of the written reviews, negative feature themes (i.e., functionality and accessibility issues) were identified, which otherwise may not have been captured by these metrics alone. Additionally, almost all of the reviews included in the thematic analysis mentioned multiple themes, which indicates how much information is provided by the user in written reviews. Similar

findings are reported in previous literature reports (53) indicating the complexity of features which users appreciate in a mental health app.

Identified Themes

In line with previous literature of mental health app reviews, we identified similar themes including the quality of information provided and language used, the presence of advertisements and customer service amongst other larger themes. Additionally, we also identified themes not exclusively reported for mental health apps including functionality (69), accessibility (70), and concerns about privacy and security (69, 71–73). However, our analysis reveals theme frequencies and patterns which differ from previous findings.

In the current study performance was the most commonly mentioned theme by users suggesting that, when focusing on the screening and/or diagnostic assessment aspect of mental health apps, performance is considered an important feature by users. A commonly reported facet of the performance theme was a self-reported increase in the user's understanding of or responsibility for their mental health following completion of the mental health assessment. This finding builds on previous reports demonstrating how mhealth tools designed for mental health can be employed in order to empower individuals to self-manage their mental health (28–30, 74) or encourage a user to seek help from a healthcare professional (38). In addition to increasing the user's understanding and/or responsibility for their mental health, another facet of the performance theme was the quality of the diagnostic and/or screening assessment and the accuracy as perceived by the user. These dimensions of the performance theme are intertwined, as a high accuracy of the assessment is essential for ensuring that the insights the user is gaining about their mental health are correct and that any actions that are taken (i.e., self-help, seeking help from a healthcare professional) are appropriate for their specific needs. This is important to note as, while the current study only reported on self-perceived accuracy as determined by the users themselves, a recent systematic review and meta-analysis demonstrated that the accuracy of apps currently publicly available for mental health assessment is mixed, with some demonstrating poor discriminatory performance (75). Therefore, whilst the ability of apps which offer a mental health diagnostic and/or screening assessment to increase understanding and promote responsibility for mental health is considered an important feature by users, these insights may not be as accurate as users perceive them to be. Reassuringly, when investigating the relationship between sentiment and feature themes, positive sentiment is observed in just under three-quarters of the user reviews which comment on performance.

Whilst previous studies analyzing reviews of mental health apps instead found usability (44) to be the most common theme above themes such as accuracy, in the current study, the performance theme was closely followed by the functionality theme. Previous research has determined that usability issues (referred to as functionality issues in the current study) constitute the key weakness of mental health apps (44). Additionally, issues related to usability are the main fix request made by users (44). Within the current study, we found that the most commonly reported functionality theme aspect was related to quality of updates implemented. Many users reported that the update worsened the app by introducing functionality issues (i.e., becoming less visually appealing, harder to use and introducing in-app bugs) and accessibility issues, chiefly a paywall. Previous work shows that users appreciate regular updates to improve and update app content (44). However, similarly to the current study's findings, updates can also be a cause of frustration for users by introducing issues (43, 44). In fact, in a survey of 654 app users, just under a third expressed hesitation before updating an app, with just under half also reporting they had experienced issues with an app after updating it (76). Users reported issues related to app crashing, low app speed, changes to features included within the app, and bugs as the largest issues following an update (76). Additionally, the quality of visual design was found to be a major aspect of the functionality theme in the current study. The majority of these reported that apps were poorly designed, characterized by flaws such as the interface being too cluttered and overwhelming. Visual design has been identified as a key area of usability issues before (43), with users preferring a "clean" design (53). Within the current study, negative sentiment was frequently observed alongside the functionality theme. This finding supports previous literature findings which demonstrates that usability is most commonly mentioned in a negative context within user reviews (43, 44, 53, 55).

Previous studies had identified accessibility as the most frequent theme mentioned in user reviews (53). The current study instead identified accessibility being reported by users less frequently. In this dataset, I largest dimension of accessibility was the app having a paywall, perhaps because the majority of the apps included in this study were free with in-app purchases. A paywall seems to be a common theme across reviews of mental health apps, including CBT apps, in which users often requested increased access to free features within the app (77). Users of mood monitoring apps described, *via* reviews, feelings of frustration when they paid for an app which they then determined was unsuitable to their needs (53). This again, highlights the importance of including accurate information about the services provided and features included in the app, as well as which features the user will be expected to pay for.

Unlike previous work, the current study identified the theme of therapeutic alliance and treatment as a frequent theme within the user reviews. Particularly, many users commented on self-help aspects of the included apps (i.e., self-guided meditation, breathing exercises, coping skill programs). This is perhaps explained by self-help features being the most commonly offered treatment-related feature within apps. This finding demonstrates the ability of mental health apps to encourage self-management of some conditions (28). An additional dimension of the therapeutic alliance theme, is the ability to connect with and share data with clinicians. Users often reported that being able to share the data collected within their app with their clinician increased their clinician's insight into their condition. This attitude was also present in user reviews of CBT apps for depression (51), of apps designed for bipolar disorder (52), and mood monitoring apps (53).

In terms of the additional features and engagement style theme, a tracking/log/journaling function was the most commonly mentioned feature within user reviews. Recent research on people's use and perspectives on mood tracking support the view this app feature is generally perceived as useful (78) and positive (55) by users (78). Furthermore, tracking was the most commonly requested feature within apps designed for bipolar disorder (52). Additionally, tracking has been identified as a facilitator to user engagement in digital mental health intervention apps (45). Tracking can promote self-reflection (79), which was also observed in the current study. The results of the thematic analysis conducted in the current study revealed an overlap between the themes of additional features and performance. Specifically, the dimension of performance which was most frequently identified in combination with tracking/log/journaling, as this was increasing understanding and/or responsibility for mental health. Therefore, due to the positive perception of tracking, we recommend that mental health apps which include a screening and/or diagnostic assessment consider also adapting the assessment into a longitudinal tracking tool. The addition of this tool may help facilitate user engagement *via* long-term symptom monitoring (45), as well as further increasing the ability of an app to offer mental health understanding and/or responsibility. Additionally, a differing finding from previous studies is the frequency with which concerns about privacy or security were raised by users. Privacy is often mentioned as a foremost concern of mental health and non-mental health app users (44, 55, 71–73), but this was not reflected in our dataset. A study looking at mood monitoring apps intended for use in young people showed that the proportion of reviews commenting on privacy and security was 5.66% (53). Furthermore, user reviews of bipolar apps commented that concerns related to privacy and security were considered "dealbreakers" by users (52). By contrast, we report very few instances of users commenting on the privacy or security of the app within their reviews (1%). However, as mental health apps which include a screening and/or diagnostic assessment would capture potentially sensitive symptom and demographic data, any concerns related to privacy and security are worrying.

App Discontinuation

Our findings suggest that reviews can also offer insights into reasons for app use discontinuation. With regard to our observations of app use discontinuation reported in the user reviews, this was higher than in previous reports (48). However, this difference could be due to our broader definition which also included implicit references to app use discontinuation. Considering that negative sentiment was commonly observed alongside functionality and accessibility, it is perhaps unsurprising that these themes were also key factors reported in app use discontinuation; in particular, a bad update and a paywall.

Complaints mentioning app use discontinuation related to poor-quality updates usually commented on the app becoming less visually appealing, harder to use or the update introducing in-app bugs (i.e., lagging or freezing). Previous literature focused on mental health apps also reports that issues with usability may lead to users discontinuing app usage (44). This also extends to non-mental health related apps, with surveys of app users revealing that 53% of respondents reported they would uninstall an app following severe functionality issues (80). However, regular app updates are important for increasing user engagement and may help avoid drop-out if they are of high quality (43). Therefore, app developers should aim to perform a comprehensive app testing period before deploying any app updates. Ideally, this would also include a period of A-B testing of the old vs. the new version of the app in consultation with a subset of active app users or other individuals in the population of interest, to ensure the new update is functional and acceptable.

Reviews mentioning app use discontinuation in relation to a paywall were related to the addition of a subscription fee needed to access some or all of the features offered within the app. In some cases, these paywalls were not disclosed at all within the app store descriptions before downloads, or the extent to which they would impact the users access to the app was not disclosed (i.e., the user could access all features but one). Previous findings revealed that these hidden costs were frustrating to users (53) and even indicate that a paywall may encourage users to search for other apps which offer the same features at a lower cost or no cost (55). Interestingly, when exploring themes that co-occurred with app use discontinuation, accessibility (paywall) and functionality (bad update) in combination were more common than either theme alone. Hence, in our dataset, functionality issues alone were not the most common reason for app use discontinuation despite previous literature indicating this is a key factor. Additionally, this finding suggests that paywalls are considered more of a barrier to sustained app use when introduced as part of an update than if present from the initial download. Users may view a paywall introduced in an app update as unexpected and a hidden app cost: known to be a barrier to health app usage (81). In light of these findings, app developers should notify app users if and when an app update will also include the introduction of a paywall. This, along with an explanation of why the paywall introduction is required, may promote users' acceptance of the update and prevent use discontinuation. Additionally, app developers should consider, where practical, to allow users continued access to a scaled back version of the app for free following an update.

Recommendations

- App developers should aim to expand the analysis of user feedback to incorporate written reviews alongside star ratings in order to obtain a more comprehensive picture of user perspectives to increase engagement and avoid app use discontinuation.
- App developers should consider additional methods to collect user feedback data in order to capture the full spectrum of user experiences, including those of users who may not leave an app store review.
- Users seem to value promoting understanding of or responsibility for mental health when using apps which include a mental health screening and/or diagnostic

assessment. With this in mind, providing an accurate picture of user's mental health is of paramount importance. Therefore, we recommend that app developer implement high quality, validated assessments within mental health apps to ensure a high level of screening and/or diagnostic accuracy.
- Before implementation of a new update, extensive testing should be performed to ensure the quality of the app is still high and its functionality is intact. Additionally, if a paywall is to be introduced, work should be undertaken to ensure users do not feel blindsided by the introduction of in-app purchases; or, an old or basic version of the app should remain available for those users who do not wish to pay to access the app.
- If possible, app developers should consider adapting the core screening and/or diagnostic assessment into a longitudinal tracking tool in an effort to increase user engagement as well as increased mental health understanding and/or responsibility.

Limitations

A limitation of the current study is related to performing searches in app stores. App stores are not designed for robust, rigorous searches unlike electronic journal databases (58). Searching in app stores may introduce potential challenges to reproduce any findings (58). In an effort to address this, the PRISMA-ScR checklist was used to improve reporting of the methods implemented and searches which were conducted (see **Supplementary Material 1**) (58, 59).

Additionally, whilst we employed a thematic analysis in the current study for identifying theme frequencies and allowing for theme comparisons, we recognize that the results of any analysis may have been influenced by possible reviewer bias. We attempted to minimize any bias during the qualitative analysis by performing a dual independent review process at each stage of the analysis.

Furthermore, due to the feasibility constraints imposed by performing a manual analysis with multiple independent reviewers, only a subset of all app store reviews were analyzed. However, by only considering the "most helpful" reviews (i.e., upvoted by users for their helpfulness) as per previous similar literature (54) and filtering reviews from the last 6 months, we hope to have compiled a dataset that depicts an up-to-date and relevant picture of app users' perspectives.

Other limitations are beyond the control of the authors, for instance the results from app stores searches are based on factors beyond search terms, such as whether the app offers in-app purchases and the number of downloads (64). There are also limitations related to the accuracy of the data, as the descriptive app information is directly provided by the app developers (82) without any information of accuracy checks being performed by the app stores. Furthermore, app developers are able to vary the information provided in the app store descriptions (i.e., in-app screenshots, app keyword, app description) based on geographical locations (83, 84).

CONCLUSIONS

To conclude, app reviews remain a valuable yet underutilized resource which offer an abundance of insights and actionable information provided directly by patients and users. Considering information provided within the user reviews can inform app design, and ensure the app is suitable for purpose as determined by end users themselves. Overall, apps which include a mental health screening and/or diagnostic assessment are perceived positively, with very few users reporting app use discontinuation and many users valuing an apps' ability to increase understanding of their mental health. However, there are clear areas of improvement which can be considered by app development teams to avoid negative user experiences and app use discontinuation. These include avoiding the implementation of an unexpected paywall and extensive app testing before an update is released. In addition, consideration of the quality of assessments delivered *via* mental health apps should be undertaken. Principally, ensuring that the included assessment is high-quality, validated, and confers a high degree of accuracy. In doing so, developers will contribute to an increased likelihood that the app will provide an accurate picture of the user's mental health, which was identified as the most commonly cited indicator of app performance rating by users.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

EF and NM-K conceived the study focus. The searches of the app stores and inclusion decisions were made by EF, BS, and NM-K. Reviews from the app stores were scraped by EF and TM. Code was developed for scraping of the Apple app store by TM. The qualitative analysis was performed by EF as first reviewer and BS and NM-K as the second reviewers. EF prepared the manuscript with revisions from BS, NM-K, TM, and SB. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by the Stanley Medical Research Institute (grant 07R-1888).

ACKNOWLEDGMENTS

Many thanks to Jiri Benacek and Dr. Jakub Tomasik for their advice on the data extraction from the app stores.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyt.2022.857304/full#supplementary-material

REFERENCES

1. Office for National Statistics. Coronavirus and depression in adults, Great Britain: July to August 2021 (2021). Available online at: https://www.ons.gov.uk/peoplepopulationandcommunity/wellbeing/articles/coronavirusanddepressioninadultsgreatbritain/julytoaugust2021 (accessed December 1, 2021).
2. Armitage R. Antidepressants, primary care, and adult mental health services in England during COVID-19. *Lancet Psychiatry.* (2021) 8:e3. doi: 10.1016/S2215-0366(20)30530-7
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) 3:e2019686. doi: 10.1001/jamanetworkopen.2020.19686
4. Xiong J, Lipsitz O, Nasri F, Lui LMW, Gill H, Phan L, et al. Impact of COVID-19 pandemic on mental health in the general population: a systematic review. *J Affect Disord.* (2020) 277:55–64. doi: 10.1016/j.jad.2020.08.001
5. Office for National Statistics. Are we facing a mental health pandemic? (2021). Available online at: https://blog.ons.gov.uk/2021/05/05/are-we-facing-a-mental-health-pandemic/ (accessed December 3, 2021).
6. Nuffield Trust. What impact had Covid-19 had on mental health services? (2020). Available online at: https://www.nuffieldtrust.org.uk/news-item/what-impact-has-covid-19-had-on-mental-health-services (accessed December 3, 2021).
7. British Medical Association. Measuring progress: Commitments to support and expand the mental health workforce in England (2019). Available online at: https://www.bma.org.uk/media/2405/bma-measuring-progress-of-commitments-for-mental-health-workforce-jan-2020.pdf (accessed August 18, 2021).
8. British Medical Association. The impact of COVID-19 on mental health in England; Supporting services to go beyond parity of esteem (2020). Available online at: https://www.bma.org.uk/media/2750/bma-the-impact-of-covid-19-on-mental-health-in-england.pdf (accessed August 18, 2021).
9. Abbas MJ, Kronenberg G, McBride M, Chari D, Alam F, Mukaetova-Ladinska E, et al. The early impact of the COVID-19 pandemic on acute care mental health services. *Psychiatr Serv.* (2021) 72:242–6. doi: 10.1176/appi.ps.202000467
10. Reichert A, Jacobs R. The impact of waiting time on patient outcomes: Evidence from early intervention in psychosis services in England. *Health Econ.* (2018) 27:1772–87. doi: 10.1002/hec.3800
11. Goldberg JF, Ernst CL. Features associated with the delayed initiation of mood stabilizers at illness onset in bipolar disorder. *J Clin Psychiatry.* (2002) 63:985–91. doi: 10.4088/JCP.v63n1105
12. Altamura AC, Dell'Osso B, Berlin HA, Buoli M, Bassetti R, Mundo E. Duration of untreated illness and suicide in bipolar disorder: a naturalistic study. *Eur Arch Psychiatry Clin Neurosci.* (2010) 260:385–91. doi: 10.1007/s00406-009-0085-2
13. Post RM, Leverich GS, Kupka RW, Keck PE, McElroy SL, Altshuler LL, et al. Early-onset bipolar disorder and treatment delay are risk factors for poor outcome in adulthood. *J Clin Psychiatry.* (2010) 71:864–72. doi: 10.4088/JCP.08m04994yel
14. Altamura AC, Dell'Osso B, Mundo E, Dell'Osso L. Duration of untreated illness in major depressive disorder: a naturalistic study. *Int J Clin Pract.* (2007) 61:1697–700. doi: 10.1111/j.1742-1241.2007.01450.x
15. Albert U, Barbaro F, Bramante S, Rosso G, De Ronchi D, Maina G. Duration of untreated illness and response to SRI treatment in obsessive-compulsive disorder. *Eur Psychiatry.* (2019) 58:19–26. doi: 10.1016/j.eurpsy.2019.01.017
16. Altamura AC, Dell'Osso B, Vismara S, Mundo E. May duration of untreated illness influence the long-term course of major depressive disorder? *Eur Psychiatry.* (2008) 23:92–6. doi: 10.1016/j.eurpsy.2007.11.004
17. Patel R, Shetty H, Jackson R, Broadbent M, Stewart R, Boydell J, et al. Delays before diagnosis and initiation of treatment in patients presenting to mental health services with bipolar disorder. *PLoS ONE.* (2015) 10:e0126530. doi: 10.1371/journal.pone.0126530
18. McGorry PD, Ratheesh A, O'Donoghue B. Early intervention—an implementation challenge for 21st century mental health care. *JAMA Psychiatry.* (2018) 75:545. doi: 10.1001/jamapsychiatry.2018.0621
19. NHS Oxford Health. Study finds that early intervention services can save NHS millions (2016). Available online at: https://www.oxfordhealth.nhs.uk/news/study-finds-that-early-intervention-services-can-save-nhs-millions/ (accessed August 18, 2021).
20. Correll CU, Galling B, Pawar A, Krivko A, Bonetto C, Ruggeri M, et al. Comparison of early intervention services vs treatment as usual for early-phase psychosis. *JAMA Psychiatry.* (2018) 75:555. doi: 10.1001/jamapsychiatry.2018.0623
21. National Mental Health Development Unit. Factfile 3 The costs of mental ill health (2021). Available online at: https://www.networks.nhs.uk/nhs-networks/regional-mental-health-workshop-mids-east/documents/supporting-materials/nmhdu-factfile-3.pdf (accessed August 18, 2021).
22. Cano A, Sprafkin RP, Scaturo DJ, Lantinga LJ, Fiese BH, Brand F. Mental health screening in primary care. *Prim Care Companion J Clin Psychiatry.* (2001) 03:206–10. doi: 10.4088/PCC.v03n0503
23. Sin J, Galeazzi G, McGregor E, Collom J, Taylor A, Barrett B, et al. Digital interventions for screening and treating common mental disorders or symptoms of common mental illness in adults: systematic review and meta-analysis. *J Med Internet Res.* (2020) 22:e20581. doi: 10.2196/20581
24. Spadaro B, Martin-Key NA, Bahn S. Building the digital mental health ecosystem: opportunities and challenges for mobile health innovators. *J Med Internet Res.* (2021) 23:e27507. doi: 10.2196/27507
25. Mind. 40 per cent of all GP appointments about mental health (2018). Available online at: https://www.mind.org.uk/news-campaigns/news/40-per-cent-of-all-gp-appointments-about-mental-health/ (accessed December 3, 2021).
26. NHS. Active signposting frees up 80 inappropriate GP appointments a week. Available online at: https://www.england.nhs.uk/gp/case-studies/active-signposting-frees-up-80-inappropriate-gp-appointments-a-week/ (accessed January 4, 2022).
27. Schmitz N, Kruse J, Tress W. Improving screening for mental disorders in the primary care setting by combining the GHQ-12 and SCL-90-R subscales. *Compr Psychiatry.* (2001) 42:166–73. doi: 10.1053/comp.2001.19751
28. Bucci S, Schwannauer M, Berry N. The digital revolution and its impact on mental health care. *Psychol Psychother Theory, Res Pract.* (2019) 92:277–97. doi: 10.1111/papt.12222
29. Torous J, Haim A. Dichotomies in the development and implementation of digital mental health tools. *Psychiatr Serv.* (2018) 69:1204–6. doi: 10.1176/appi.ps.201800193
30. Schueller SM, Hunter JF, Figueroa C, Aguilera A. Use of digital mental health for marginalized and underserved populations. *Curr Treat Options Psychiatry.* (2019) 6:243–55. doi: 10.1007/s40501-019-00181-z
31. GOV.UK. Policy paper COVID-19 mental health and wellbeing recovery action plan (2021). Available online at: https://www.gov.uk/government/publications/covid-19-mental-health-and-wellbeing-recovery-action-plan/covid-19-mental-health-and-wellbeing-recovery-action-plan (accessed December 3, 2021).
32. Naslund JA, Gonsalves PP, Gruebner O, Pendse SR, Smith SL, Sharma A, et al. Digital innovations for global mental health: opportunities for data science, task sharing, and early intervention. *Curr Treat Options Psychiatry.* (2019) 6:337–51. doi: 10.1007/s40501-019-00186-8
33. Torous J, Roberts LW. Needed innovation in digital health and smartphone applications for mental health transparency and trust. *JAMA Psychiatry.* (2017) 74:437–8. doi: 10.1001/jamapsychiatry.2017.0262
34. ORCHA. COVID-19: digital health trends & opportunities for 2021 (2021). Available online at: https://orchahealth.com/wp-content/uploads/2021/01/COVID_Report_Jan_2021_final-version.pdf (accessed August 20, 2021).
35. Chan AHY, Honey MLL. User perceptions of mobile digital apps for mental health: acceptability and usability-an integrative review. *J Psychiatr Ment Health Nurs.* (2022) 29:147–68. doi: 10.1111/jpm.12744
36. Naslund JA, Marsch LA, McHugo GJ, Bartels SJ. Emerging mHealth and eHealth interventions for serious mental illness: a review of the literature. *J Ment Heal.* (2015) 24:321–32. doi: 10.3109/09638237.2015.1019054
37. BinDhim NF, Shaman AM, Trevena L, Basyouni MH, Pont LG, Alhawassi TM. Depression screening *via* a smartphone app: cross-country user characteristics and feasibility. *J Am Med Informatics Assoc.* (2015) 22:29–34. doi: 10.1136/amiajnl-2014-002840

38. BinDhim NF, Alanazi EM, Aljadhey H, Basyouni MH, Kowalski SR, Pont LG, et al. Does a mobile phone depression-screening app motivate mobile phone users with high depressive symptoms to seek a health care professional's help? *J Med Internet Res.* (2016) 18:e156. doi: 10.2196/jmir.5726
39. 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 Heal.* (2018) 21:116–9. doi: 10.1136/eb-2018-102891
40. 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) 20:e199. doi: 10.2196/jmir.9275
41. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. *J Affect Disord.* (2020) 263:413–9. doi: 10.1016/j.jad.2019.11.167
42. Meyerowitz-Katz G, Ravi S, Arnolda L, Feng X, Maberly G, Astell-Burt T. Rates of attrition and dropout in app-based interventions for chronic disease: systematic review and meta-analysis. *J Med Internet Res.* (2020) 22:e20283. doi: 10.2196/20283
43. Alqahtani F, Orji R. Usability issues in mental health applications. *Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization*. New York, NY, USA: ACM (2019). p. 343–8. doi: 10.1145/3314183.3323676
44. Alqahtani F, Orji R. Insights from user reviews to improve mental health apps. *Health Informatics J.* (2020) 26:2042–66. doi: 10.1177/1460458219896492
45. Borghouts J, Eikey E, Mark G, De Leon C, Schueller SM, Schneider M, et al. Barriers to and facilitators of user engagement with digital mental health interventions: systematic review. *J Med Internet Res.* (2021) 23:e24387. doi: 10.2196/24387
46. ORCHA. What is the Health of Mental Health Apps? (2020). Available online at: https://orchahealth.com/health-of-mental-health-apps/ (accessed December 6, 2021).
47. Iacob C, Harrison R. Retrieving and analyzing mobile apps feature requests from online reviews. *2013 10th Working Conference on Mining Software Repositories (MSR)*. San Francisco, CA: IEEE (2013). p. 41–4. doi: 10.1109/MSR.2013.6624001
48. Iacob C, Veerappa V, Harrison R. What are you complaining about?: a study of online reviews of mobile applications. In: *Proceedings of the 27th International BCS Human Computer Interaction Conference* (London). (2013). doi: 10.14236/ewic/HCI2013.38
49. Apptentive. Mobile App Ratings and Reviews: Where to Start and How to Win (2021). Available online at: https://go.apptentive.com/rs/170-TZF-108/images/Ratings-Reviews-1.pdf?mkt_tok=MTcwLVRaRi0xMDgAAAF-bWzonyVyyjyamC8dcN4qIcF5cja-_kc-1qaVfvX67Khl1kvIsNiattXD2udRHQYw1SubrrqHWt0U5GpG-Sb5RypGFT3ALcRTThaxlOMWfQ (accessed August 20, 2021).
50. Al-Natour S, Turetken O. A comparative assessment of sentiment analysis and star ratings for consumer reviews. *Int J Inf Manage.* (2020) 54:102132. doi: 10.1016/j.ijinfomgt.2020.102132
51. Stawarz K, Preist C, Tallon D, Wiles N, Coyle D. User experience of cognitive behavioral therapy apps for depression: an analysis of app functionality and user reviews. *J Med Internet Res.* (2018) 20:e10120. doi: 10.2196/10120
52. Nicholas J, Fogarty AS, Boydell K, Christensen H. The reviews are in: a qualitative content analysis of consumer perspectives on apps for bipolar disorder. *J Med Internet Res.* (2017) 19:e105. doi: 10.2196/jmir.7273
53. Widnall E, Grant CE, Wang T, Cross L, Velupillai S, Roberts A, et al. User perspectives of mood-monitoring apps available to young people: qualitative content analysis. *JMIR mHealth uHealth.* (2020) 8:e18140. doi: 10.2196/18140
54. Stawarz K, Cox AL, Blandford A. Don't forget your pill! In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM (2014). p. 2269–78. doi: 10.1145/2556288.2557079
55. Oyebode O, Alqahtani F, Orji R. Using machine learning and thematic analysis methods to evaluate mental health apps based on user reviews. *IEEE Access.* (2020) 8:111141–58. doi: 10.1109/ACCESS.2020.3002176
56. Pang B, Lee L. Opinion mining and sentiment analysis. *Found Trends® Inf Retr.* (2008) 2:1–135. doi: 10.1561/1500000011
57. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol.* (2006) 3:77–101. doi: 10.1191/1478088706qp063oa
58. Grainger R, Devan H, Sangelaji B, Hay-Smith J. Issues in reporting of systematic review methods in health app-focused reviews: a scoping review. *Health Informatics J.* (2020) 26:2930–45. doi: 10.1177/1460458220952917
59. 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) 169:467–73. doi: 10.7326/M18-0850
60. Funnell E. User reviews of mental health screening and diagnostic apps: a sentiment and thematic analysis (OSF registered study plan). (2021). doi: 10.17605/OSF.IO/DZBPJ
61. M-Health Index and Navigation Database. Framework (2020). Available online at: https://mindapps.org/ (accessed August 11, 2021).
62. Colaboratory. Create a Dataset for Sentiment Analysis. Available online at: https://colab.research.google.com/github/curiousily/Getting-Things-Done-with-Pytorch/blob/master/07.create-dataset-for-sentiment-analysis.ipynb (accessed August 18, 2021).
63. npm Inc. app-store-scraper (2021). Available online at: https://www.npmjs.com/package/app-store-scraper (accessed September 4, 2021).
64. Store Maven. Ranking Factors for App Store and Google Play (2020). Available online at: https://www.storemaven.com/academy/app-store-and-google-play-algorithms/ (accessed February 18, 2022).
65. App Radar. What is App Store Optimization: Ultimate Guide to ASO in 2022 (2022). Available online at: https://appradar.com/academy/what-is-app-store-optimization-aso (accessed February 24, 2022).
66. Shen N, Levitan M-J, Johnson A, Bender JL, Hamilton-Page M, Jadad A, et al. Finding a depression app: a review and content analysis of the depression app marketplace. *JMIR mHealth uHealth.* (2015) 3:e16. doi: 10.2196/mhealth.3713
67. Larsen ME, Nicholas J, Christensen H. Quantifying app store dynamics: longitudinal tracking of mental health apps. *JMIR mHealth uHealth.* (2016) 4:e96. doi: 10.2196/mhealth.6020
68. Apptentive. Mobile App Ratings and Reviews Aren't Enough–Here's Why (2022). Available online at: https://www.apptentive.com/blog/2022/01/27/mobile-app-ratings-and-reviews-arent-enough-heres-why/ (accessed February 24, 2022).
69. Minen MT, Gopal A, Sahyoun G, Stieglitz E, Torous J. The functionality, evidence, and privacy issues around smartphone apps for the top neuropsychiatric conditions. *J Neuropsychiatry Clin Neurosci.* (2021) 33:72–9. doi: 10.1176/appi.neuropsych.19120353
70. Alshayban A, Ahmed I, Malek S. Accessibility issues in Android apps. In: *Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering*. New York, NY, USA: ACM (2020). p. 1323–34. doi: 10.1145/3377811.3380392
71. Adhikari R, Richards D, Scott K. Security and privacy issues related to the use of mobile health apps. In: *Proceedings of the 25th Australasian Conference on Information Systems, 8th-10th December*. Auckland, New Zealand: ACIS (2014).
72. Minen MT, Stieglitz EJ, Sciortino R, Torous J. Privacy issues in smartphone applications: an analysis of headache/migraine applications. *Headache J Head Face Pain.* (2018) 58:1014–27. doi: 10.1111/head.13341
73. Scoccia GL, Peruma A, Pujols V, Malavolta I, Krutz DE. Permission issues in open-source android apps: an exploratory study. In: *2019 19th International Working Conference on Source Code Analysis and Manipulation (SCAM)*. Cleveland, OH: IEEE (2019). p. 238–49. doi: 10.1109/SCAM.2019.00034
74. Mirea D-M, Martin-Key NA, Barton-Owen G, Olmert T, Cooper JD, Han SYS, et al. Impact of a web-based psychiatric assessment on the mental health and well-being of individuals presenting with depressive symptoms: longitudinal observational study. *JMIR Ment Heal.* (2021) 8:e23813. doi: 10.2196/23813
75. Martin-Key NA, Spadaro B, Funnell E, Barker EJ, Schei TS, Tomasik J, et al. The current state and validity of digital assessment tools for psychiatry: systematic review. *JMIR Ment Heal.* (2022) 9:e32824. doi: 10.2196/32824
76. Nayebi M, Adams B, Ruhe G. Release Practices for Mobile Apps–What do Users and Developers Think? In: *2016 IEEE 23rd International Conference on Software Analysis, Evolution, and Reengineering (SANER)*. Osaka: IEEE (2016). p. 552–62. doi: 10.1109/SANER.2016.116
77. Thach KS. User's perception on mental health applications: a qualitative analysis of user reviews. In: *2018 5th NAFOSTED Conference on Information and Computer Science (NICS)*. Ho Chi Minh City: IEEE (2018). p. 47–52. doi: 10.1109/NICS.2018.8606901

78. Kelley C, Lee B, Wilcox L. Self-tracking for Mental Wellness. In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM (2017). p. 629–41. doi: 10.1145/3025453.3025750
79. Schueller SM, Neary M, Lai J, Epstein DA. Understanding people's use of and perspectives on mood-tracking apps: interview study. *JMIR Ment Heal.* (2021) 8:e29368. doi: 10.2196/29368
80. RAYGUN. Mobile application abandonment: common reasons (2016). Available online at: https://raygun.com/blog/mobile-application-abandonment/ (accessed February 24, 2022).
81. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. *JMIR mHealth uHealth.* (2015) 3:e101. doi: 10.2196/mhealth.4924
82. Paglialonga A, Lugo A, Santoro E. An overview on the emerging area of identification, characterization, and assessment of health apps. *J Biomed Inform.* (2018) 83:97–102. doi: 10.1016/j.jbi.2018.05.017
83. Grundy QH, Wang Z, Bero LA. Challenges in assessing mobile health app quality: a systematic review of prevalent and innovative methods. *Am J Prev Med.* (2016) 51:1051–9. doi: 10.1016/j.amepre.2016.07.009
84. App Store Connect Help. Required, localizable, and editable properties (2022). Available online at: https://help.apple.com/app-store-connect/#/devfc3066644 (accessed February 18, 2022).

Conflict of Interest: SB is a director of Psynova Neurotech Ltd and Psyomics Ltd. SB has financial interests in Psyomics Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Digital Health

ORIGINAL RESEARCH
published: 12 May 2022
doi: 10.3389/fdgth.2022.869143

A Mixed-Methods Analysis of Mobile ACT Responses From Two Cohorts

Sydney Hoel[1†], *Amanda Victory*[2†], *Tijana Sagorac Gruichich*[1], *Zachary N. Stowe*[1], *Melvin G. McInnis*[2], *Amy Cochran*[3*] and *Emily B. K. Thomas*[4]

[1] Psychiatry, University of Wisconsin-Madison, Madison, WI, United States, [2] Psychiatry, University of Michigan, Ann Arbor, MI, United States, [3] Population Health Sciences and Mathematics, University of Wisconsin-Madison, Madison, WI, United States, [4] Psychological and Brain Sciences, University of Iowa, Iowa City, IA, United States

Background: Mobile transdiagnostic therapies offer a solution to the challenges of limited access to psychological care. However, it is unclear if individuals can actively synthesize and adopt concepts and skills via an app without clinician support.

Aims: The present study measured comprehension of and engagement with a mobile acceptance and commitment therapy (ACT) intervention in two independent cohorts. Authors hypothesized that participants would recognize that behaviors can be flexible in form and function and respond in an ACT process-aligned manner.

Methods: Mixed-methods analyses were performed on open-ended responses collected from initial participants (n = 49) in two parallel micro-randomized trials with: 1) first-generation college students (FGCSs) (n = 25) from a four-year public research university and 2) individuals diagnosed with bipolar disorder (BP) (n = 24). Twice each day over six weeks, participants responded to questions about mood and behavior, after which they had a 50-50 chance of receiving an ACT-based intervention. Participants identified current behavior and categorized behavior as values-based or avoidant. Interventions were selected randomly from 84 possible prompts, each targeting one ACT process: engagement with values, openness to internal experiences, or self-awareness. Participants were randomly assigned to either exploratory (10 FGCS, 9 BP) or confirmatory (15 FGCS, 15 BP) groups for analyses. Responses from the exploratory group were used to inductively derive a qualitative coding system. This system was used to code responses in the confirmatory group. Coded confirmatory data were used for final analyses.

Results: Over 50% of participants in both cohorts submitted a non-blank response 100% of the time. For over 50% of participants, intervention responses aligned with the target ACT process for at least 96% of the time (FGCS) and 91% of the time (BP), and current behavior was labeled as values-based 70% (FGCS) and 85% (BP) of the time. Participants labeled similar behaviors flexibly as either values-based or avoidant in different contexts. Dominant themes were needs-based behaviors, interpersonal and family relationships, education, and time as a cost.

OPEN ACCESS

Edited by:
Abigail Ortiz,
University of Toronto, Canada

Reviewed by:
Anabel De la Rosa-Gómez,
National Autonomous University of Mexico, Mexico
Michael Musker,
University of South Australia, Australia

***Correspondence:**
Amy Cochran
cochran4@wisc.edu

[†] *These authors share first authorship*

Specialty section:
This article was submitted to Digital Mental Health, a section of the journal Frontiers in Digital Health

Received: *03 February 2022*
Accepted: *25 April 2022*
Published: *12 May 2022*

Citation:
Hoel S, Victory A, Sagorac Gruichich T, Stowe ZN, McInnis MG, Cochran A and Thomas EBK (2022) A Mixed-Methods Analysis of Mobile ACT Responses From Two Cohorts. Front. Digit. Health 4:869143. doi: 10.3389/fdgth.2022.869143

Conclusions: Both cohorts were engaged with the app, as demonstrated by responses that aligned with ACT processes. This suggests that participants had some level of understanding that behavior can be flexible in form and function.

Keywords: mobile health (mHealth), acceptance and commitment therapy (ACT), engagement, bipolar disorder, first-generation college students (FGCS), psychological flexibility, mixed methods, research methodology

INTRODUCTION

Given the increasing use of technology in daily life, mobile health apps offer solutions for filling in gaps in mental health care (1, 2). Mobile health apps contribute to the management of several mental health conditions, including bipolar disorder (BP) (3, 4), borderline personality disorder (5), major depressive disorder (6, 7), anxiety disorders (8), and posttraumatic stress disorder (9–11). In addition, they offer techniques to monitor internal thoughts and emotions outside of *in-person* care sessions and in the relevant moments of daily experience; personal awareness of such internal experiences is often a therapeutic goal. In the current mental health care model, such awareness is often retrospective and conveyed in the clinical sessions. Mobile health apps directly address many gaps in health systems, including clinician availability, constraints of transportation, health insurance, and cost, among others. For example, Tondo et al. reported that increasing access to care can improve even severe symptoms of depression such as suicidality (12). To build on these successes, we sought to investigate the quality of engagement with an intervention delivered in a mental health app without the support of a clinician to help navigate the intervention.

Realizing the promise of mental health apps requires addressing several barriers. The majority of mental health apps have yet to be evaluated for their efficacy (13, 14). Further, apps that target specific psychiatric diagnoses or operate around a specific treatment may not be useful to an individual who is unable to access care and obtain a diagnosis or treatment recommendation. There are groups who are at a high risk for experiencing psychological distress yet not characterized by a psychiatric diagnosis and may benefit from a mental health app. For example, numerous studies have identified college students as needing mental health care but not seeking it out (15, 16). Mental health apps may be a viable option for this population. Choosing among the multitude of apps available may be overwhelming and negatively impact engagement. For example, many apps allow users to monitor their symptoms over time. This can be helpful for some psychiatric diagnoses, but potentially detrimental to others without additional clinical support (17, 18). High attrition rates are common among mHealth interventions (19). Zucchelli et al. noted 53% of participants completed four out of six app-delivered 30-min ACT sessions in a study focused on alleviating psychosocial appearance concerns of those with atypical appearances via ACT (20). However, 60% of participants reported finding the interventions helpful, 88.6% said they were easy to understand, and exit interviews revealed daily reminders were important in encouraging app usage (20). Nevertheless, there is limited research on how best to characterize engagement among mHealth ACT interventions, and more work is needed to understand what predicts greater engagement with digital health interventions. Overcoming these issues may require mental health apps that provide additional clinical support beyond symptom monitoring, deliver interventions across diagnoses, and have been evaluated empirically.

A critical metric is the actual engagement with the process targeted by the app (e.g. monitoring or intervention). Clinicians will actively encourage treatment adherence and guide understanding of therapeutic processes. However, those who seek treatment through an app while not concurrently in the care of a clinician may not engage with the app in the manner expected to achieve the desired outcome. Measures of success include a) the success of the user experiencing the desired outcomes (i.e., symptom reduction, an increase in health-promoting behaviors), and b), successfully engaging users with the app itself. One would assume that desired outcomes are caused by high engagement, but this may not be the case, and if it is, the *degree* of engagement required to produce a positive effect may not be clear.

Acceptance and commitment therapy (ACT) is a transdiagnostic mindfulness-based therapy that targets experiential avoidance and encourages openness to internal experiences (e.g., emotions, thoughts), awareness of the function of behavior, and behavioral engagement with one's values (21). The driving principle of ACT is that pursuit of personally chosen values, vitality, and personal fulfillment are attainable, even when living with distressing experiences. ACT aims to increase psychological flexibility, the ability to engage in behavior that is consistent with one's values even when challenging or distressing (21). An important component of ACT is to reduce reliance on experiential avoidance, which is the inability or unwillingness to experience thoughts, emotions, physical sensations, or memories. Avoidance may provide relief in the moment, but in the long-term, it reduces contact with valued life directions and worsens the intensity and duration of the avoided stimulus. In contrast, psychological flexibility is associated with an increase in well-being and a reduction in symptoms (22).

ACT is effective in treating a variety of study populations (23), including cancer patients experiencing psychological symptoms (24) and substance abuse disorders (25, 26). In a study on nicotine addiction, Bricker et al. utilized the fundamental approach of ACT—to *accept* smoking triggers in adults attempting to quit rather than to avoid smoking triggers—when designing their smartphone application, iCanQuit (26). At the 12-month follow-up mark, the participants who used iCanQuit had 1.49 greater odds of quitting smoking when compared to a second group of participants who used an app called QuitGuide created

by the National Cancer Institute, which focuses on avoidance ($p < 0.001$). Mobile ACT interventions have also been used concurrently with in-person ACT, as is the case with the ACT Daily app prototype, which was used with 14 patients with depression as they received treatment from an ACT clinician (27). In another study, a sample of college students showed improvement in depressive symptoms after completing an online, guided ACT intervention (28).

Further support for the efficacy of ACT when delivered virtually comes from positive outcomes when internet ACT, or iACT, is studied among individuals with depression (29, 30). Two commonalities among these previous two studies are especially relevant to this intervention: the use of college students and the use of those on a waitlist to receive care as the control group members. Recent research suggests that first-generation college students (FGCS) experience more anxiety and depression than non-FGCSs (31, 32). These two commonalities are relevant because they highlight a population of students who would potentially benefit from an ACT intervention and by bringing to attention the possibility of an ACT-based intervention to improving access to care.

To summarize, mental health apps clearly expand access to care; however, the actual engagement of the user is not well scrutinized, leaving the question of which components of health care apps contribute to efficacy. As an effective transdiagnostic treatment, ACT addresses the accessibility gap, but the question still remains: can individuals in need of care independently learn from an ACT-based mHealth intervention? This is especially relevant to those who have access to mobile technology, but not a health care provider. Further, could such an intervention be effective for individuals with a range of diagnoses and needs?

The present study sought to investigate engagement with and learning from an ACT-based mental health app in two cohorts:

First-Generation College Students

FGCSs experience unique and significant distress compared to non-first-generation students. FGCSs indicate a lesser sense of belonging on average and poorer mental health on average than non-FGCSs (32); and needing but not using counseling and/or psychological services at a greater rate than non-first-generation students (32). This supports findings from the 2012 National Postsecondary Student Aid Study (NPSAS:12), a prospective study examining a sample of students who began postsecondary education in the 2011–2012 academic year (33). In 2014, follow-up data collection of over 24,000 students found that FGCSs (14%) utilized campus health services < non-FGCSs (29%) (34). First-generation status is associated with known risk factors for mental illness, such as coming from a low socioeconomic status (SES) household or belonging to a historically marginalized racial or ethnic group (35–37).

Bipolar Disorder

BP is a chronic mood disorder characterized by dynamic episodes of depression and mania or hypomania (38). BP affects an estimated 45 million people worldwide (39), with one-third to one-half of those with BP experiencing a suicide attempt at least once in their lifetime (40). Clinical manifestations and patterns of BP are highly variable and often require a combination of medication and psychotherapy for treatment (41). However, those with BP may be more likely to have limited access to healthcare and therefore go longer before initially receiving mental health care (42). Despite the possibility of psychotherapy and medication to treat BP, clinical care often remains fragmented with a lack of clinical integration and reduced access to care (12, 42, 43). Non-adherence to medication and inconsistent access to care, including psychosocial interventions, are common obstacles leading to a worsened disease course (43).

The two-cohort model herein was used to evaluate whether the same intervention content was learned similarly by two diagnostically and demographically distinct samples. In the intervention, participants were tasked with independently learning complex emotional and psychological phenomena and applying the underlying ACT concepts to their own lives in order to make behavioral change. We predicted participants would engage with the intervention prompts to develop increased self-awareness, as observed by the content of responses. Via twice-daily assessments, we anticipated that flexibility would be observed in both behavioral form and function of the identified behaviors.

MATERIALS AND METHODS

Cohorts

First-Generation College Students

First-generation college students (FGCSs) are defined in this study as students whose parent(s) or legal guardian(s) have attained less education than a bachelor's degree. Participants in this sample were recruited from the University of Wisconsin-Madison (UW) during the Fall 2019 and Spring 2020 academic terms. Recruitment methods included sending a mass email to first- and second-year undergraduate students, posting flyers on the UW campus, and brief presentations to students in UW lecture-style classes. Interested individuals completed an online eligibility screening. To be included in the study, individuals had to 1) be aged 18–19, 2) be enrolled as a freshman or sophomore undergraduate student at UW, 3) have access to a smartphone, 4) be a FGCS, and 5) endorse a subjectively high level of distress at the time of screening. Recruitment was ongoing at the time of the present analysis, and data from 25 participants were randomly selected for analysis in this study. This study has been approved by the Health Sciences Institutional Review Board at the University of Wisconsin-Madison [2019-0819] and is registered at clinicaltrials.gov (NCT04081662).

Bipolar Disorder

Participants with a diagnosis of either type I BP (BPI) or type II BP (BPII) were recruited from the Prechter Longitudinal Study of Bipolar Disorder (44) for a 6-week study. Recruitment began in September of 2019 and ended in August of 2020. Recruitment was ongoing at the time of the present analysis, and data from the first 24 participants who submitted app data were analyzed. Eligibility criteria included a diagnosis of BPI or BPII, consent to be contacted for future research, and access to a smartphone. This study was approved by institutional

review boards at the University of Michigan (HUM126732) and the University of Wisconsin (2017-1322) and is registered at clinicaltrials.gov (NCT04098497).

Study Design

A detailed description of the study methodology has been published (45). Here, only that which is relevant to current analyses is discussed.

Participants in each cohort completed a consent discussion with a member of the research team before reviewing and signing the informed consent document. They then completed baseline demographic and psychometric questionnaires. After doing so, they received instructions to download and use a free mobile app. Twice a day, participants were prompted to complete a brief log consisting of questions about current mood and behavior. The behavioral assessment asked participants to write what behavior they were engaged in at that moment (behavioral form) and to categorize it as either "toward" (motivated by values) or "away" (motivated by avoiding negative internal experiences) (behavioral function). Participants also had a 50% chance of receiving an ACT-based intervention prompt each time they completed the assessment log. When an ACT-based prompt was delivered, it was selected randomly from a list of 84 possible prompts, with the possibility of questions being repeated. The prompts were evenly divided across three core principles of ACT: 1) openness to internal experiences, 2) engagement with values, and 3) awareness of internal experiences. Participants responded to both the current behavior item and ACT-based prompts in a free-text format. The logging functionality allowed for participants to skip assessment items by submitting blank fields, and text responses had no minimum or maximum word limit.

Qualitative Analysis

We performed qualitative analysis on intervention and behavioral assessment data from the FGCS ($n = 25$) and BP ($n = 24$) cohorts. Participants were randomly assigned to either the exploratory (FGCS $n = 10$, BP $n = 9$) or confirmatory (FGCS $n = 15$, BP $n = 15$) group. To be included in the analysis, participants must have downloaded the study app and responded at least once to logging prompts (showing that they knew how to respond).

The exploratory dataset was used to inductively establish a preliminary coding system. Two primary coders, co-first authors SH and AV, independently completed in-depth reviews of the exploratory data. Following the initial review, the research team met as a group to discuss themes and concepts of interest in the data. After identifying these data elements, we developed an initial coding system, which SH and AV then used to independently code the exploratory dataset. Results from both coders were compared and discussed by the research team to evaluate the quality of the coding system and further refine it. SH and AV then applied the refined coding system to 30 randomly selected intervention responses and 30 randomly selected behavior responses from the exploratory dataset, and once again compared results. The research team discussed final revisions to the coding system. In the final coding system, the behavioral qualitative data was coded across 5 categories: work-related behaviors, leisure behaviors, self-care behaviors, activity level, and social behaviors. Qualitative data for the interventions coded for response alignment, values, negative internal experiences, and contexts.

The primary coders then independently coded the confirmatory data using the finalized coding system, comparing results upon completion. Codes with discrepancies more than 1/3 of the time were removed from the analysis. Any remaining discrepancies were resolved by author TSG.

Metrics of Engagement

To evaluate the quality of participant engagement with the intervention, the following metrics were recovered to describe each response: 1) submitted response, 2) non-blank response, 3) identification of the function of behaviors, 4) process alignment, 5) word count, and 6) qualitative content. A submitted response refers to a response to a prompt that is submitted by a person in the study app. A non-blank response refers to a submitted response that had any amount of text provided in the text field. We hypothesized that participants will submit non-blank responses to a majority of prompts. These first two metrics were calculated separately for behavioral and intervention responses.

Identification of the function of behaviors was determined for each behavioral response based on each individual's categorization of their current behavior as either moving them toward what matters ("values-based") or away from negative internal experiences ("avoidant"). We predicted that participants would demonstrate flexibility in the function of behaviors in terms of being able to categorize the same or similar behaviors as both values-based and avoidant over the course of their intervention period (e.g., categorizing an academic behavior as values-based at one time point, and categorizing another academic behavior as avoidant at a different time point).

Process alignment was determined for each intervention response during the coding process. Each intervention prompt was designed to align with one of three core ACT processes: openness, awareness, or engagement. Responses were coded to reflect whether or not the responses were "process-aligned," meaning that participants addressed the intended process in their response. Process-alignment was thought to indicate meaningful engagement; we hypothesized that the majority (more than 50%) of responses would be process-aligned.

Word count of a response was calculated using the function *wordCloudCounts* in Matlab (Mathworks; Natick, Massachusetts). This function splits the text into words, removes stop words, and combines words with a common root. Word counts were calculated for both behavioral assessment responses and intervention responses. We predicted that participants would respond to prompts with multiple words (non-yes/no). The final metric, qualitative content of responses, was determined using the categories established in the qualitative coding system. We examined whether or not a response fell into a certain category.

Descriptive statistics were calculated for the metrics of engagement. This was done in two steps: first, across responses per participant and then across participants. Each metric was summarized as a count, a proportion, or average across responses for each participant. For example, we calculated the

total number of submitted responses and the proportion of submitted responses that are non-blank for each person; each calculated separately for behavioral and intervention responses. We then calculated information about the distribution of these participant-summarized metrics of engagement: min, max, median, and 25th and 75th percentiles. Metrics of engagement are reported in text as medians across participants to represent the majority of participants, which corresponds to the 50th percentile among the tables. Majority is thus defined as over 50% of participants. To improve readability in the Results section, we will refer to a median as a value in which "over 50% of participants" had an equal or higher value. A final check was to examine whether our participant-summarized metrics of engagement were providing consistent information. To this end, we used a Pearson correlation coefficient to measure correlation between the participant-summarized metrics.

RESULTS

First-Generation College Student Cohort

Sample Characteristics

In both the exploratory and confirmatory samples, the majority of the subjects were female (88%), comprising 9 out of 10 participants and 13 out of 15 participants, respectively. A majority of participants identified as White (50% of exploratory sample; 67% of confirmatory), and a single participant in the exploratory sample identified as Hispanic. No participants in either sample reported working full-time at the time of the study. Four (40%) participants in the exploratory group and 8 (53%) of the participants in the confirmatory group were working part-time. Two exploratory participants and 1 confirmatory participant were currently using SNAP benefits ("food stamps"), and 7 (70%) participants in the confirmatory group and 9 (60%) participants in the exploratory group reported experiencing financial problems during childhood. Data on prior history of mental health treatment or therapy was not collected. The average number of behavior responses was 68.7 ($SD = 60.2$) for the exploratory group and 54.1 ($SD = 28.1$) for the confirmatory group. The average number of intervention response was 34.4 ($SD = 29.5$) for the exploratory group and 26.2 ($SD = 15.3$) for the confirmatory group. The complete sample characteristics are summarized in **Table 1**.

Metrics of Engagement

Across all participants in the confirmatory sample, 799 behavior responses were submitted. Submitted behavior responses were accompanied by an additional ACT-based intervention prompt for a total 393 times (49.1% behavior responses coinciding with the 50-50 randomization for delivering an intervention prompt). Participants submitted a response to these prompts 372 times. Four participants accounted for all blank responses submitted (21 in total), whereas the remaining 11 participants submitted a text response to every intervention prompt received. In other words, over 50% of participants always provided a non-blank response to intervention prompts. Similarly, over 50% of participants always provided a non-blank response to behavior prompts. In addition, over 50% of participants provided responses with average word counts that were at least 4.26 words in length for intervention prompts and 1.91 words for behavior prompts. The distribution of these metrics across participants are summarized in **Figure 1**.

Participant-average word count for behavior responses was positively, but not significantly, correlated with participant-average word count for intervention responses ($r = 0.44$, $p < 0.1$; **Table 2**). In turn, these two metrics were each positively, but not significantly, correlated with percent of non-blank intervention responses ($r = 0.31$ and $r = 0.26$, respectively; $p > 0.1$). Non-blank intervention responses were also significantly correlated with non-blank behavior responses ($r = 0.86$, $p < 0.001$).

Process Alignment

Over 50% of participants provided a collection of intervention responses in which over 96% were coded as process-aligned. Further, the percentage of process-aligned responses was significantly and positively correlated with percent non-blank intervention responses ($r = 0.90$, $p < 0.001$) and percent non-blank behavior responses ($r = 0.80$, $p < 0.001$). Percentage of process-aligned responses also correlated positively but not significantly with intervention word count ($r = 0.40$, $p > 0.10$).

Identifying Function of Behaviors: Values-Based or Avoidant

Behaviors, in general and those belonging to a specific behavior code, were more likely to be identified by participants as being values-based as opposed to avoidant (**Table 3**). Active behaviors, academic behaviors, exercise, social behaviors, and reading were the most likely to be categorized as values-based; specifically, at least half the participants always categorized these behaviors as values-based. At the 25th percentile, however, there was variation in behavioral function for each of those behavior types. Watching (e.g., TV) and sedentary behaviors were the least likely to be categorized as values-based behaviors.

Qualitative Themes

Over 50% of participants provided a collection of behavior responses in which at least 25% were academic related, 79% were sedentary, 10% were active, and 15% were related to self-care (**Table 4**). In the intervention responses, the dominant values were family relationships and education. The theme of time as a cost of engaging in values-based behaviors also emerged. The most mentioned types of negative affect were sadness and feeling overwhelmed or stressed. The concept of psychological flexibility also appeared in the intervention responses, with greater indications of flexibility than inflexibility.

Bipolar Cohort

Sample Characteristics

Exploratory and confirmatory groups were similar in demographics. The exploratory group consisted of a mean age of 41.3 years (SD = 10.4), 67% females, and 89% White. The confirmatory group had a mean age of 42 years (SD = 12.4), 60% of participants were female, and 73% were White. Overall, the sample represented a high-SES population. Seventy-eight percent of the exploratory group had BPI, and 22% had BPII. Within the confirmatory group, 87% had BPI and only 13% had BPII. At baseline, the average Hamilton Rating Scale for

TABLE 1 | Sample characteristics by cohort.

		Exploratory	Confirmatory
		FGCS ($n = 10$ & $n = 15$):	
Age, mean (SD)		18.7 (0.48)	18.4 (0.51)
Gender, *N* (%)	Man	1 (10%)	1 (7%)
	Woman	9 (90%)	13 (87%)
	Unknown	0 (0%)	1 (7%)
Race, *N* (%)	Caucasian	5 (50%)	10 (67%)
	Native American	0 (0%)	0 (0%)
	African American	2 (20%)	1 (7%)
	Asian/Indian	3 (30%)	3 (20%)
	Pacific Islander	0 (0%)	0 (0%)
	More than one race	0 (0%)	1 (7%)
Ethnicity, *N* (%)	Hispanic	1 (10%)	0 (0%)
Sexual orientation, *N* (%)	Heterosexual	9 (90%)	12 (80%)
	Homosexual	1 (10%)	0 (0%)
	Bisexual	0 (0%)	2 (13%)
	Pansexual	0 (0%)	1 (7%)
	Single	5 (50%)	11 (73%)
	Partnered	5 (50%)	4 (27%)
Employment, *N* (%)	Working full time	0 (0%)	0 (0%)
	Working part-time	4 (40%)	8 (53%)
	Unemployed	6 (60%)	7 (47%)
Using SNAP benefits ("food stamps") at time of study, *N* (%)	Yes	2 (20%)	1 (7%)
Experienced financial problems in childhood, *N* (%)	Yes	7 (70%)	9 (60%)
Children, *N* (%)	Yes	0 (0%)	0 (0%)
Behavior responses, mean (SD)		68.7 (60.2)	54.1 (28.1)
Intervention responses, mean (SD)		34.4 (29.5)	26.2 (15.3)
		BP ($n = 9$ & $n = 15$):	
Age, mean (SD)		41.3 (10.4)	42.0 (12.4)
Sex, *N* (%)	Female	6 (67%)	9 (60%)
	White	8 (89%)	11 (73%)
	American Indian or Alaskan Native	0 (0%)	1 (7%)
	Black or African American	1 (11%)	1 (7%)
	More than one race	0 (0%)	2 (13%)
Ethnicity, *N* (%)	Hispanic	1 (11%)	1 (7%)
Bipolar Type, *N* (%)	Type I	7 (78%)	13 (87%)
	Type II	2 (22%)	2 (13%)
Behavior responses, mean (SD)		72.2 (18.3)	67.1 (23.2)
Intervention responses, mean (SD)		33.6 (9.1)	32.6 (11.4)

Depression (HRSD) was 6.20 ($SD = 5.78$) and the average Young Mania Rating Scale (YMRS) was 1.83 ($SD = 3.29$). The average number of behavior responses were similar between groups (exploratory: $M = 72.2$, $SD = 18.3$; confirmatory: $M = 67.1$, $SD = 23.2$). The average number of intervention responses was 33.6 ($SD = 9.1$) for the exploratory group and 32.6 ($SD = 11.4$) for the confirmatory group. Complete sample characteristics are summarized in **Table 1**.

Metrics of Engagement

These data are intended to indicate the degree to which a person engages with ACT processes. Over half the participants provided a non-blank response to every intervention prompt, and all participants provided non-blank responses to every behavior prompt. In addition, over 50% of participants provided responses with average word counts that were at least 6.03 words in length for intervention prompts and 3.74 words for behavior prompts. Participants were largely categorizing their behavior responses as values-based as opposed to avoidant. A full description of these metrics per percentile can be found in **Figure 1.**

Correlations were calculated to show whether engagement metrics were related within this qualitative analysis. **Table 2** summarizes these correlations. Percent non-blank response to an intervention prompt was positively, but not significantly,

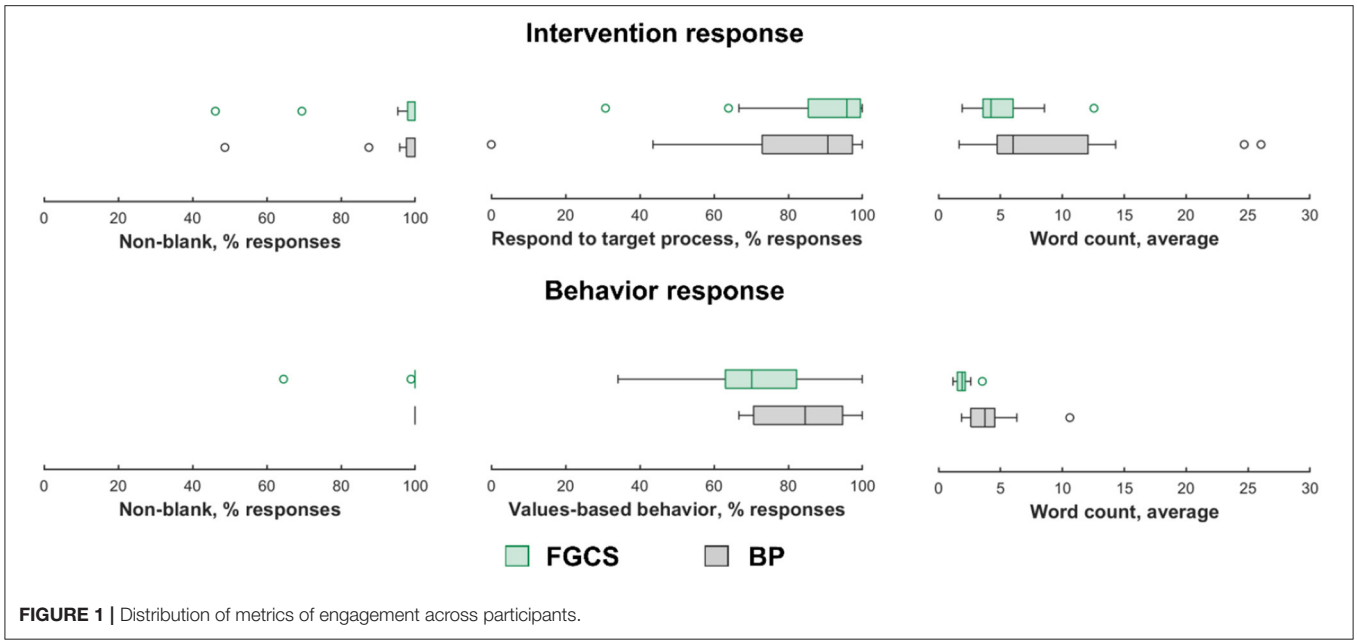


FIGURE 1 | Distribution of metrics of engagement across participants.

TABLE 2 | Correlation between metrics of engagement with ACT processes.

		Intervention		Behavior		
		Aligned with target process, %	Word count, average	Non-blank response, %	Values-based behavior, %	Word count, average
				FGCS:		
Intervention	Non-blank response, %	0.90***	0.31	0.86***	0.07	0.26
	Aligned with target process, %	-	0.40	0.80***	0.17	0.25
	Word count, average	-	-	0.31	0.12	0.44*
Behavior	Non-blank response, %	-	-	-	0.06	0.34
	Values-based behavior, %	-	-	-	-	0.19
				BP:		
Intervention	Non-blank response, %	0.41	0.37	n/a[a]	0.19	0.32
	Aligned with target process, %	-	0.20	n/a[a]	−0.21	0.16
	Word count, average	-	-		−0.01	0.83***
Behavior	Non-blank response, %	-	-	-	n/a[a]	n/a[a]
	Values-based behavior, %	-	-	-	-	0.05

*$P < 0.1$, ***$P < 0.001$.

[a]*Correlation is not defined, since all BP participants provided non-blank responses to all behavior prompts.*

correlated with percent process-aligned ($r = 0.41$, $p > 0.10$) and average word count ($r = 0.37$, $p > 0.10$). Average word count for intervention responses was significantly correlated with average word count for behavior responses ($r = 0.83$, $p < 0.001$). A weaker, and sometimes negative correlation with measures of engagement was observed with the *values-based behavior* category. This aligns with expectations that *flexibility* in categorizing behaviors as values-based is a meaningful measure of engagement as opposed to *percent* values-based behavior. We will expand on this idea below.

Process Alignment

Over 50% of participants provided a collection of intervention responses among which at least 91% of responses were processed aligned. This is lower than what we reported above in terms of 100% of intervention responses being non-blank for over 50% of participants.

Identifying Function of Behaviors: Values-Based or Avoidant

Behaviors that were always indicated as values-based by most participants included active, exercise, reading, service, and social behaviors. Unsurprisingly, behaviors that were less often categorized as values-based were watching, media, and subjective behavior. However, even behaviors largely categorized as avoidant behaviors were sometimes entered as a values-based behavior, suggesting that participants were categorizing behavior based on function in the current context. For example, not every instance of "talking with friends" is considered an avoidant

TABLE 3 | Among different behaviors, distribution of percent responses categorized as value-based across participants.

Code	Participant percentile					
	Min	25th	Median	75th	Max	*N*
			FGCS:			
Watching	0%	31%	50%	93%	100%	14
Sedentary	20%	59%	68%	78%	100%	15
Needs	46%	68%	84%	100%	100%	14
Leisure Other	0%	42%	93%	100%	100%	11
Reading	0%	55%	100%	100%	100%	9
Active	50%	80%	100%	100%	100%	15
School	50%	86%	100%	100%	100%	15
Exercise	50%	89%	100%	100%	100%	13
Social	75%	95%	100%	100%	100%	12
			BP:			
Needs	44%	77%	94%	100%	100%	15
Leisure Other	38%	86%	88%	100%	100%	15
Social	57%	85%	100%	100%	100%	15
Watching	20%	45%	69%	100%	100%	13
Work	37%	64%	95%	100%	100%	12
Media	0%	33%	56%	100%	100%	14
Reading	0%	88%	100%	100%	100%	8
Exercise	89%	100%	100%	100%	100%	11
Service	88%	98%	100%	100%	100%	8
Subjective	0%	25%	56%	70%	100%	9
Sedentary	48%	58%	81%	90%	100%	15
Active	74%	88%	100%	100%	100%	15

Each metric was first summarized as a percentage or average across responses for each participant. We then calculated information about the distribution of these participant-summarized metrics in terms of a min, max, median, and 25th and 75th percentile, which are reported in the table above.

TABLE 4 | Distribution of percent responses with a given code across participants in FGCS cohort.

Response	Metric	Participant percentile				
		Min	25th	Median	75th	Max
Behavior	School, %	8%	16%	25%	43%	58%
	Needs, %	0%	9%	15%	25%	68%
	Watching, %	0%	8%	14%	25%	49%
	Exercise, %	0%	3%	5%	9%	18%
	Social, %	0%	2%	4%	11%	48%
	Other leisure, %	0%	1%	5%	10%	34%
	Reading, %	0%	0%	1%	5%	38%
	Sedentary, %	40%	67%	79%	86%	92%
	Active, %	4%	6%	10%	19%	29%
Intervention	Flexible, %	0%	1%	5%	7%	11%
	Inflexible, %	0%	0%	3%	8%	33%
	Workability, %	0%	0%	3%	4%	11%
	Value - Education, %	0%	0%	5%	7%	23%
	Value - Family, %	0%	0%	5%	8%	13%
	Time, %	0%	3%	7%	15%	25%
	Sadness, %	0%	0%	5%	6%	33%
	Overwhelmed, %	0%	0%	5%	8%	23%
	Low positive affect, %	0%	0%	3%	10%	13%
	Physio, %	0%	1%	3%	8%	11%
	Positive affect, %	0%	2%	4%	10%	21%
	Interpersonal context, %	0%	0%	3%	11%	22%

Each metric was first summarized as a percentage or average across responses for each participant. We then calculated information about the distribution of these participant-summarized metrics in terms of a min, max, median, and 25th and 75th percentile, which are reported in the table above.

behavior. It is important to note that not all categories represent the same sample size as not all participants shared the same reported behaviors. **Table 3** provides percentiles for behavioral codes, along with specific sample sizes.

Qualitative Themes

Table 5 highlights intervention responses for the median participant as they pertain to personal aspects such as family, time, and interpersonal context. For the behavior responses, the codes demonstrated the most were sedentary, needs, and active.

DISCUSSION

The aim of this study was to investigate the degree to which individuals engaged with and learned from mobile ACT interventions in two different cohorts, hypothesizing that in both, over 50% of participants would respond to open-ended questions in a way that aligned with ACT processes. We considered evidence of clinically meaningful engagement to be both a willingness to provide responses that offer specific, personal context and a variability in participants' self-reported behavioral function, displaying a recognition that behaviors can be flexible in function based on context. Achieving such clinically meaningful engagement—without clinical support—is important, since an ideal intervention could be utilized despite individual barriers such as availability of care, lack of access to resources (time, financial), or treatment models limited to specific diagnoses.

In both cohorts, participants demonstrated an ability to independently grasp ACT concepts and apply them, as evidenced by high proportions of process-aligned responses and flexibility in the reported behavioral function. In the BP cohort, results show process alignment in 73% responses even in the 25th percentile of participants that provided any type of response; similarly, the FGCS cohort responses were process-aligned 85% of the time at the 25th percentile. Thus, even the participants who were least process-aligned were still process-aligned in the majority of responses (over 50% of responses). The findings show that engagement in digital health is not only possible—supporting previous research (1)—but also can be achieved in a clinically significant way under the conditions described in this study.

In addition, participants were able to independently recognize and identify functions (values-based or avoidant) for the same behavior type. This is the core skill participants needed to learn to reflect psychological flexibility. For example, in the BP cohort, "watching TV" was recorded as values-based behavior

TABLE 5 | Distribution of percent responses with a given code across participants in BP cohort.

Response	Metric	Participant percentile				
		Min	**25**th	**Median**	**75**th	**Max**
Behavior	Needs, %	11%	24%	33%	38%	70%
	Leisure Other, %	4%	12%	17%	21%	44%
	Social, %	6%	10%	17%	30%	75%
	Watching, %	0%	7%	17%	27%	32%
	Work, %	0%	2%	14%	30%	33%
	Media, %	0%	3%	4%	7%	29%
	Reading, %	0%	0%	3%	7%	18%
	Exercise, %	0%	0%	2%	7%	25%
	Service, %	0%	0%	1%	15%	36%
	Subjective, %	0%	0%	5%	11%	33%
	Sedentary, %	24%	41%	48%	59%	68%
	Active, %	8%	18%	29%	34%	44%
Intervention	Flexible, %	0%	3%	5%	10%	29%
	Inflexible, %	0%	0%	3%	5%	9%
	Workability, %	0%	0%	3%	3%	18%
	Health, %	0%	0%	3%	12%	21%
	Education, %	0%	0%	2%	3%	8%
	Work, %	0%	0%	3%	7%	23%
	Family, %	0%	3%	8%	12%	31%
	Friend, %	0%	0%	3%	7%	46%
	Other relationship, %	0%	0%	2%	7%	31%
	Self, %	0%	0%	3%	5%	17%
	Time, %	0%	6%	8%	16%	23%
	Sadness, %	0%	0%	3%	8%	16%
	Fear, %	0%	1%	5%	9%	19%
	Anger, %	0%	0%	3%	5%	6%
	Low positive affect, %	0%	0%	3%	5%	15%
	Physio, %	0%	3%	5%	7%	43%
	Positive affect, %	0%	3%	5%	9%	15%
	Work context, %	0%	0%	3%	8%	25%
	Interpersonal context, %	0%	4%	10%	18%	57%

Each metric was first summarized as a percentage or average across responses for each participant. We then calculated information about the distribution of these participant-summarized metrics in terms of a min, max, median, and 25th and 75th percentile, which are reported in the table above.

1 day, and an avoidant away behavior the following day. Similarly, "working" was frequently coded as both values-based and avoidant within one participant's responses. More to this point, there were differences in the likelihood of values-based vs. avoidant categorization for each behavior type. For example, exercise behaviors were almost always categorized as values-based but other types of behaviors had greater variability. Evidence of this skill is encouraging, as it suggests that users were able to grasp a major goal of ACT, which is to distinguish between form and function of a behavior. We conjecture that this process of engagement might mediate any symptom reduction from the intervention. This is even more encouraging considering that the study was only 6 weeks and some participants' only exposure to ACT might be a 20-min informational video created by the authors for this study.

We also utilized word count as a metric of engagement. Average word counts per response was small (< 7 words) for both intervention and behavior responses. Longer average word counts were found in the BP cohort vs. FGCS cohort and for intervention vs. behavior responses. Longer intervention responses are expected given that many behaviors can be expressed concisely, whereas certain intervention prompts demanded significant reflection (e.g., "What are the consequences [positive or negative] when you try to interpret your thoughts and emotions?"). For example, short behavior responses included "homework, tv", "talking with friends," or "studying" as opposed to more descriptive intervention responses such as "I am very very nervous about my midterm tonight." Although average word count was small, it was encouraging to note that some responses were lengthy, and some prompts did not require a response of any length. For example, eight intervention prompts could have been read as closed questions that could be answered with a single word, percentage value, or indicator of frequency (e.g., "Does your mind ever label you 'bad' or 'defective'?").

Coinciding with engagement, a major barrier to successfully implementing digital interventions is earning the trust of participants so that they feel safe inputting information into a digital platform. Given the sensitive nature of psychotherapy and emotions more broadly, an intervention such as the one studied here would not be possible if participants were unwilling to share their responses. We suspect that we gained the trust of some participants from both cohorts based on the word count and consistent content of responses. For example, one 89-word intervention response described a particularly difficult manic episode, and a 46-word response answered the question of how they avoid uncomfortable emotions: "I usually tend to avoid uncomfortable emotions often, as I try to avoid situations that would give me these emotions. Sometimes I cannot avoid them, and inevitably end up feeling depressed."

Most responses contained specific, personal context, displaying a willingness from participants to share their thoughts and emotions not only in an app, but with the knowledge that someone would be closely reading their responses. This allowed us to identify themes among participants' values and negative internal experiences. One of the themes in the intervention responses of both cohorts was time as a cost of or barrier to engaging in a behavior ("spending time [engaging in activity]"). The concept of time management is well-documented in research concerning college students. Effective time management has been associated with improved academic performance and reduced stress (46, 47). It has also been highlighted in a 2005 qualitative study; all participants in a small sample of 8 FGCSs noted time management as a skill important for college readiness (48). Another study found first-year college students' time management to be dynamic, changing from one semester to the next depending on their ability to meet academic goals (49). This previous work seems to establish time as an important factor in decision-making and a potential barrier or facilitator to engaging with values. Observing this theme of time in FGCS participant responses could be indicative of this importance.

The concept of time-management is less documented in research concerning BP. However, everyone is faced with the opportunity cost of spending time in one manner over another (50). Borda describes patients with BP as individuals with a broad experience of time, including the ability to be reflective, actively engaged in the moment, and thoughtful about the future (51). Examples of this reflection in our study include a preoccupation with a past event "…my [past event] was a time when I had difficulty with emotions" or optimistic thoughts about the future "…take time to relax after work." This could suggest that the ACT intervention was able to extract a commonality among BP patients—being the importance of time in their perception of self (51). More specifically, research by Rusner et al. found that with varying mood states, the concept of and connection to time may change (52). Increased mentions of time, as seen in this intervention, might provide useful insights for providers about their patients' mood states.

Academic behaviors and the values of education and family were additional dominant themes in the FGCS cohort; family was also a dominant theme in the BP cohort. Research on FGCSs' values often examines the conflict between independent social norms at academic institutions and FGCSs' experience of interdependent social norms, which place value on family (53–56).

Limitations

The coding process presented many challenges. Although coding was completed independently by two researchers, the process was nonetheless subjective as the researchers had to make choices on the meaning behind responses. For example, the research team had to discuss what types of behaviors to code as "needs" (i.e., attending to one's needs; self-care). One decision involved coding all responses that mentioned eating as "needs" regardless of the type of food described (e.g. "cake," "breakfast") thereby avoiding assumptions about what constitutes self-care for a particular individual. The behavioral coding system has its own challenge when participants listed multiple behaviors in their response. This sometimes led to responses that met for multiple codes within one broad category ("went to work, then went to class" applies to two codes in the work-related behavior category: work-related and academic); in those instances, we had to choose which code to assign. It would have been more effective to allow for multiple codes.

Coding psychiatric symptoms was particularly challenging, especially considering that the two cohorts (FGCSs and BP) are very different. For example, when participants responded with the word "upset" the researchers had to decipher what exactly was meant, i.e., did upset mean "angry" or "sad?" On that same token, a response such as "irritable" could easily be coded as anger, but in particular for the BP cohort, in terms of which mood state it was pertaining to, that remains unknown. On the flip side, a response mentioning "mania" might imply different symptomatic profiles. In other words, it was impossible to know if mania should be coded under "NEG-ANGRY" indicating the person was experiencing irritability, which is common, but not necessary, for mania. Such responses highlight the difficulty in translating these codes into their clinical significance. Future attempts at implementing ACT as a mental health app should avoid similar discrepancies by expanding the codebook to include more specific codes for what mood state a response might refer to, and potentially supplement with a form of passive data collection.

Moreover, the codes utilized with these two specific samples may not generalize to other samples, and future work might expand qualitative codes to those that are generalizable across large samples. Nevertheless, the intention of conducting parallel trials with two distinct samples was to investigate the transdiagnostic nature of the intervention and engagement with the intervention. As such, although some codes were not applicable or shifted in meaning across samples, others applied in both samples, and the similarities and differences between samples both provided important information.

The coding system is also incomplete in the sense that some responses had no codes applied to them. No codes were applied to 199 (32%) intervention responses and 8 (1%) behavior responses for the FGCS cohort, and 148 (32%) intervention responses and 11 (1%) behavior responses for the BP cohort. Several factors may have contributed: certain questions that prompted shorter responses (e.g., "yes" or "no"); no minimum word count required to submit a response; and qualitative codes removed from analysis due to low inter-coder reliability, leaving some topics unidentified, such as guilt and positive thoughts toward self. The coding system may have missed themes because of intervention prompts were too variable; intervention prompts were randomly selected from a list of 84 prompts from 3 target processes. Each target process category can be expected to elicit certain themes. Therefore, with each participant receiving a different number and assortment of prompts over their intervention period, it makes sense that high thematic frequencies were not observed within the intervention response data. By contrast, behavior responses gave us more consistent information about themes since the prompt was always the same ("What behavior are you currently engaged in?").

It is also important to note that we compared coded categories of behavior and not distinct behaviors. For example, behaviors coded as work-related included responses such as "Planning for the class I teach and chores…," "working," and "writing a cover letter for a job I was invited to apply for." Further, although it was possible for a participant's behavior response to fall under multiple behavior types ("at the movies with friends" would be 1) a watching behavior 2) a social behavior), analyses did not examine relationships between behavioral function categorization and multiple categories. Most importantly, for anyone logging a response, what the participant was actually doing at the time of response was responding to the app, and the behavior identified was presumed to be what the person was doing just prior.

Because qualitative data were collected through an app rather than in-person by a member of the research team, we were unable to clarify any response content or seek further context. A different data collection method, such as qualitative interviewing, would yield richer data than our qualitative survey items and would provide the opportunity for clarification of responses. It would also allow us to seek insight from

participants on what they found to be barriers and facilitators to engagement. In light of this, findings regarding the frequency of certain themes, such as academically oriented behaviors, have been interpreted conservatively, as has been recommended by qualitative methodologists (57).

Study design factors may have influenced the frequency of engagement observed. The compensation structure for the FGCS cohort encouraged participants to continue using the app throughout the study period. Participants were compensated on a weekly basis for each week in which they responded to the app at least 50% of the time. For the BP cohort, compensation was determined based on weeks in the study, and response rate was not a factor. Another feature of the study design – specifically, the intervention itself – was accessibility. Designed for convenience, the study app allowed participants to choose when to respond. Notification functionality provided automatic reminders to log both after waking up and before bed. Each time window in which a participant could choose to respond was 5 h long, minimizing the chance that they would receive a reminder at a time when they were otherwise engaged. The brevity of the intervention meant that participants had to expend a minimal amount of time and effort to complete a log and intervention.

Variable engagement across participants may arise from some study participants having treatment experience with mindfulness-based therapy or ACT. Participants in the BP cohort were not only older than the FGCS cohort on average (42 compared to 18.7 years old, respectively), each had a psychiatric diagnosis and a history of treatment. It is possible that participants from either cohort could have been familiar with mindfulness or ACT concepts prior to participating in this study. The generalizability of our findings is also limited by small sample sizes and a lack of demographic representativeness. BP results were not representative of the general population as the majority of participants were White women, currently not identified as employed. The participants were recruited from an established longitudinal cohort that has shown a relatively high degree of trust toward the research team. Even within the BP population, our sample consisted mainly of individuals with BPI. Lastly, we are certain that the BP cohort had been diagnosed but may or may not be engaged in psychiatric treatment, and information pertaining to current medications was not collected; diagnostic and treatment history are unknown among the FGCS cohort.

Another limitation related to the mobile app concerns the 20-min introductory video that participants were instructed to watch before using the study app. The video reviewed the central concepts to be utilized throughout the assessments (form and function of behavior, personal values, internal experiences) and interventions. The video is lengthy, and it is possible that participants watched part or none of the video before using the app. Upon setting up the app for the first time, it was possible to skip the video by indicating they had already viewed it. As a result, a lack of understanding of the ACT concepts addressed by intervention prompts could have inhibited participants' abilities to engage in a meaningful way. Moreover, our inability to confirm whether the video was watched or attended to is an important limitation.

Future Directions

Future directions with this mental health app include an evaluation of its effectiveness in promoting behavioral change and symptom reduction over time. As far as changing or improving the intervention itself, the ACT intervention prompts could be designed to elicit more person context to help limit ambiguity when coding, and a refined prompt list produced. Furthermore, if we want to elicit more consistent information between participants, the delivery of prompts could be modified: perhaps participants would benefit from an intervention in which the delivery of prompts was structured to follow a specific "lesson plan" (for example, learning about engagement with values in week 1, awareness of internal experiences in week 2, etc.). In this case, the modified study design would provide an opportunity to test a different approach to delivery and a dataset in which every participant received the same set of prompts, allowing for a more direct comparison of the content of responses. Additional information might also be collected from passive sensors, such as GPS or activity trackers, in an effort to provide more context behind individual responses. Future iterations of the intervention could also be tailored to specific study populations, expanding analysis to include diagnosis-specific outcomes.

CONCLUSION

Participants from two diagnostically and demographically distinct cohorts were able to independently learn and apply complex ACT processes in the context of their own lives, as demonstrated by participants' self-reported flexibility in the form and function of behaviors. The majority of qualitative responses were specific and personal, showing that asking participants to engage with ACT in this manner can prompt reflection and meaningful engagement. ACT holds promise as the basis of a mobile intervention that can work for transdiagnostic populations.

DATA AVAILABILITY STATEMENT

The full datasets presented in this article are not readily available to maintain patient confidentiality and participant privacy. A limited dataset that includes response codes but not individual responses and that supports the conclusions of this article will be made available by the authors, without undue reservation. Requests to access the datasets should be directed to AC; cochran4@wisc.edu.

ETHICS STATEMENT

The FGCS cohort study was approved by the Health Sciences Institutional Review Board (IRB) at the University of Wisconsin-Madison (ID 2019-0819). All participants provided electronically written informed consent. The BP cohort study was approved by University of Michigan Medical School Institutional Review

Board (IRBMED, HUM# 126732). All participants provided written informed consent to participate in the study.

AUTHOR CONTRIBUTIONS

ET, AC, SH, and ZS designed the FGCS cohort study. ET, AC, SH, and MM designed the BP cohort study. ET wrote the ACT intervention, and AC created the mobile app. SH, TSG, ET, AC, and ZS implemented the FGCS cohort study. AV, AC, and MM implemented the BP cohort study. SM, ET, and AC designed the approach to statistical analysis. SH and AV created codebook that was reviewed and edited by TSG, AC, and ET. SH and AV coded all participant responses, and TSG acted as a third coder to resolve discrepancies. TSG wrote the methods section. SH and AV drafted the remaining manuscript with ET, AC, TSG, MM, and ZS providing edits and feedback. AC completed statistical analyses and figures. All authors approved the final manuscript prior to submission.

FUNDING

The BP cohort research was supported by the Heinz C. Prechter Bipolar Research Program at the University of Michigan; the Richard Tam Foundation; and the US Department of Health and Human Services, National Institutes of Health (NIH) (K01 MH112876). Funding for the FGCS cohort study was provided by the Baldwin Wisconsin Idea Endowment, Seed Project Grant and the Clinical and Translational Science Award (CSTA) program through the NIH National Center for Advancing Translational Sciences (NCATS) (grant ULITR002373).

ACKNOWLEDGMENTS

We would like to extend our gratitude to the research participants that made this research possible with their continued generosity and dedication. We also acknowledge Rena Doyle, RN, for her vital contributions to this work.

REFERENCES

1. Van Til K, McInnis MG, Cochran A. A comparative study of engagement in mobile and wearable health monitoring for bipolar disorder. *Bipolar Disord.* (2020) 22:182–90. doi: 10.1111/bdi.12849
2. Beiwinkel T, Kindermann S, Maier A, Kerl C, Moock J, Barbian G, et al. Using smartphones to monitor bipolar disorder symptoms: a psilot study. *JMIR Ment Health.* (2016) 3:e2. doi: 10.2196/mental.4560
3. Rajagopalan A, Shah P, Zhang MW, Ho RC. Digital platforms in the assessment and monitoring of patients with bipolar disorder. *Brain Sci.* (2017) 7:11. doi: 10.3390/brainsci7110150
4. Li H, Mukherjee D, Krishnamurthy VB, Millett C, Ryan KA, Zhang L, et al. Use of ecological momentary assessment to detect variability in mood, sleep and stress in bipolar disorder. *BMC Res Notes.* (2019) 12:791. doi: 10.1186/s13104-019-4834-7
5. Rizvi SL, Hughes CD, Thomas MC. The dBT coach mobile application as an adjunct to treatment for suicidal and self-injuring individuals with borderline personality disorder: a preliminary evaluation and challenges to client utilization. *Psychol Serv.* (2016) 13:380–8. doi: 10.1037/ser0000100
6. Wright JH, Mishkind M. Computer-Assisted cBT and mobile apps for depression: assessment and integration into clinical care. *Focus (Am Psychiatr Publ).* (2020) 18:162–8. doi: 10.1176/appi.focus.20190044
7. Economides M, Ranta K, Nazander A, Hilgert O, Goldin PR, Raevuori A, et al. Long-term outcomes of a therapist-supported, smartphone-based intervention for elevated symptoms of depression and anxiety: quasiexperimental, pre-postintervention study. *JMIR Mhealth Uhealth.* (2019) 7:e14284. doi: 10.2196/14284
8. Sucala M, Cuijpers P, Muench F, Cardo? R, Soflau R, Dobrean A, et al. Anxiety: there is an app for that. a systematic review of anxiety apps. *Depress Anxiety.* (2017) 34:518–25. doi: 10.1002/da.22654
9. Miner A, Kuhn E, Hoffman JE, Owen JE, Ruzek JI, Taylor CB. Feasibility, acceptability, and potential efficacy of the PTSD coach app: a pilot randomized controlled trial with community trauma survivors. *Psychol Trauma.* (2016) 8:384. doi: 10.1037/tra0000092
10. Keen SM, Roberts N. Preliminary evidence for the use and efficacy of mobile health applications in managing posttraumatic stress disorder symptoms. *Health Systems.* (2017) 6:122–9. doi: 10.1057/hs.2016.2
11. 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) 2:e7. doi: 10.2196/mental.3935
12. Tondo L, Albert MJ, Baldessarini RJ. Suicide rates in relation to health care access in the united states: an ecological study. *J Clin Psychiatr.* (2006) 67:517–23. doi: 10.4088/JCP.v67n0402
13. Anthes E. Mental health: there's an app for that. *Nature.* (2016) 532:20. doi: 10.1038/532020a
14. Wang K, Varma DS, Prosperi M, A. systematic review of the effectiveness of mobile apps for monitoring and management of mental health symptoms or disorders. *J Psychiatr Res.* (2018) 107:73–8. doi: 10.1016/j.jpsychires.2018.10.006
15. Blanco C, Okuda M, Wright C, Hasin DS, Grant BF, Liu S-M, et al. Mental health of college students and their non–college-attending peers: results from the national epidemiologic study on alcohol and related conditions. *Arch Gen Psychiatry.* (2008) 65:1429–37. doi: 10.1001/archpsyc.65.12.1429
16. Dunbar MS, Sontag-Padilla L, Kase CA, Seelam R, Stein BD. Unmet mental health treatment need and attitudes toward online mental health services among community college students. *Psychiatr Serv.* (2018) 69:597–600. doi: 10.1176/appi.ps.201700402
17. Faurholt-Jepsen M, Frost M, Ritz C, Christensen EM, Jacoby AS, Mikkelsen RL, et al. Daily electronic self-monitoring in bipolar disorder using smartphones – the mONARCA i trial: a randomized, placebo-controlled, single-blind, parallel group trial. *Psychol Med.* (2015) 45:2691–704. doi: 10.1017/S0033291715000410
18. Romano KA, Swanbrow Becker MA, Colgary CD, Magnuson A. Helpful or harmful? the comparative value of self-weighing and calorie counting versus intuitive eating on the eating disorder symptomology of college students. *Eat Weight Disord.* (2018) 23:841–8. doi: 10.1007/s40519-018-0562-6
19. Linardon J, Fuller-Tyszkiewicz Attrition M. and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. *J Consult Clin Psychol.* (2020) 88:1–13. doi: 10.1037/ccp0000459
20. Zucchelli F, Donnelly O, Rush E, White P, Gwyther H, Williamson H, et al. An acceptance and commitment therapy prototype mobile program for individuals with a visible difference: mixed methods feasibility study. *JMIR Form Res.* (2022) 6:e33449. doi: 10.2196/33449
21. Hayes SC, Strosahl KD, Wilson K. *Acceptance and Commitment Therapy: The Process and Practice of Mindful Change*. 2 ed. New York, NY: Guilford Press (2011).
22. Hayes SC, Luoma JB, Bond FW, Masuda A, Lillis J. Acceptance and commitment therapy: model, processes and outcomes. *Behav Res Ther.* (2006) 44:1–25. doi: 10.1016/j.brat.2005.06.006
23. Gloster AT, Walder N, Levin M, Twohig M, Karekla M. The empirical status of acceptance and commitment therapy: a review of meta-analyses. *J Contextual Behav Sci.* (2020) 18:181–92. doi: 10.1016/j.jcbs.2020.09.009
24. González-Fernández S, Fernández-Rodríguez C. acceptance and commitment therapy in cancer: review of applications and findings. *Behav Med.* (2019) 45:255–69. doi: 10.1080/08964289.2018.1452713

25. Osaji J, Ojimba C, Ahmed S. The use of acceptance and commitment therapy in substance use disorders: a review of literature. *J Clin Med Res.* (2020) 12:629–33. doi: 10.14740/jocmr4311
26. Bricker JB, Watson NL, Mull KE, Sullivan BM, Heffner JL. Efficacy of smartphone applications for smoking cessation: a randomized clinical trial. *JAMA Intern Med.* (2020) 180:1472–80. doi: 10.1001/jamainternmed.2020.4055
27. Levin ME, Haeger JA, Pierce BG, Twohig MP. Web-based acceptance and commitment therapy for mental health problems in college students: a randomized controlled trial. *Behav Modif.* (2017) 41:141–62. doi: 10.1177/0145445516659645
28. Räsänen P, Lappalainen P, Muotka J, Tolvanen A, Lappalainen R. An online guided ACT intervention for enhancing the psychological wellbeing of university students: a randomized controlled clinical trial. *Behav Res Ther.* (2016) 78:30–42. doi: 10.1016/j.brat.2016.01.001
29. Lappalainen P, Langrial S, Oinas-Kukkonen H, Tolvanen A, Lappalainen R. Web-based acceptance and commitment therapy for depressive symptoms with minimal support: a randomized controlled trial. *Behav Modif.* (2015) 39:805–34. doi: 10.1177/0145445515598142
30. Räsänen P, Muotka J, Lappalainen R. Examining mediators of change in wellbeing, stress, and depression in a blended, internet-based, ACT intervention for university students. *Internet Interv.* (2020) 22:100343. doi: 10.1016/j.invent.2020.100343
31. Noel JK, Lakhan HA, Sammartino CJ, Rosenthal SDepressive R, and anxiety symptoms in first generation college students. *J Am Coll Health*. 2021:1:1–10. doi: 10.1080/07448481.2021.1950727
32. Stebleton MJ, Soria KM. Huesman Jr. RL first-generation students' sense of belonging, mental health, and use of counseling services at public research universities. *J Coll Couns.* (2014) 17:6–20. doi: 10.1002/j.2161-1882.2014.00044.x
33. NCfES. *National Postsecondary Student Aid Study (NPSAS: 12) US Department of Education*. Washington, DC: Institute of Education Sciences (2012).
34. International R. *Use of student services among freshman first-generation college students*. Washington, DC: NASPA (2019).
35. Engle J. Postsecondary access and success for first-generation college students. *American academic.* (2007) 3:25–48.
36. Bui KVT. First-generation college students at a four-year university: background characteristics, reasons for pursuing higher education, first-year experiences. *Coll.* (2002) 36:3.
37. Walpole M. Socioeconomic status and college: how SES affects college experiences and outcomes. *Rev High Ed.* (2003) 27:45–73. doi: 10.1353/rhe.2003.0044
38. Diagnostic and statistical manual of mental disorders. *DSM-5™*. 5th ed. Washington, DC: American Psychiatric Publishing, A Division of American Psychiatric Association (2013)
39. James SL, Abate D, Abate KH, Abay SM, Abbafati C, Abbasi N, et al. 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) 392:1789–58. doi: 10.1016/S0140-6736(18)32279-7
40. Baldessarini RJ, Pompili M, Tondo L. Suicide in bipolar disorder: risks and management. *CNS Spectr*. (2006) 11:465–71. doi: 10.1017/S1092852900014681
41. Dean OM, Gliddon E, Van Rheenen TE, Giorlando F, Davidson SK, Kaur M, et al. An update on adjunctive treatment options for bipolar disorder. *Bipolar Disord.* (2018) 20:87–96. doi: 10.1111/bdi.12601
42. Dagani J, Signorini G, Nielssen O, Bani M, Pastore A, Girolamo G, et al. Meta-analysis of the interval between the onset and management of bipolar disorder. *Can J Psychiatry*. (2017) 62:247–58. doi: 10.1177/0706743716656607
43. Baldessarini RJ, Tondo L, Vázquez GH. Pharmacological treatment of adult bipolar disorder. *Mol Psychiatry*. (2019) 24:198–217. doi: 10.1038/s41380-018-0044-2
44. McInnis MG, Assari S, Kamali M, Ryan K, Langenecker SA, Saunders EF, et al. Cohort profile: the heinz c.prechter longitudinal study of bipolar disorder. *Int J Epidemio*. (2018) 47:28. doi: 10.1093/ije/dyx229
45. Kroska EB, Hoel S, Victory A, Murphy SA, McInnis MG, Stowe ZN, et al. Optimizing an acceptance and commitment therapy microintervention via a mobile app with two cohorts: protocol for micro-randomized trials. *JMIR Res Protoc.* (2020) 9:e17086. doi: 10.2196/17086
46. Macan TH, Shahani C, Dipboye RL, Phillips AP. College students' time management: correlations with academic performance and stress. *J Educ Psychol.* (1990) 82:760. doi: 10.1037/0022-0663.82.4.760
47. Misra R, McKean M. College students' academic stress and its relation to their anxiety, time management, leisure satisfaction. *Am J Health Stud.* (2000) 16:41.
48. Byrd KL, MacDonald G. Defining college readiness from the inside out: first-generation college student perspectives. *Community college review.* (2005) 33:22–37. doi: 10.1177/009155210503300102
49. Thibodeaux J, Deutsch A, Kitsantas A, Winsler A. first-year college students' time use:relations with self-regulation and GPA. *J Adv Acad.* (2017) 28:5–27. doi: 10.1177/1932202X16676860
50. Palmer S, Raftery J. Economic notes: opportunity cost. *BMJ*. (1999) 318:1551–2. doi: 10.1136/bmj.318.7197.1551
51. Borda JP. Self over time: another difference between borderline personality disorder and bipolar disorder. *J Eval Clin Pract.* (2016) 22:603–7. doi: 10.1111/jep.12550
52. Rusner M, Carlsson G, Brunt D, Nyström M. Extra dimensions in all aspects of life-the meaning of life with bipolar disorder. *Int J Qual Stud Health Well-being.* (2009) 4:159–69. doi: 10.1080/17482620902864194
53. Covarrubias R, Valle I, Laiduc G, Azmitia M. "You never become fully independent": family roles and independence in first-generation college students. *J Adolesc Res.* (2019) 34:381–410. doi: 10.1177/0743558418788402
54. Tibbetts Y, Harackiewicz JM, Canning EA, Boston JS, Priniski SJ, Hyde JS. Affirming independence: exploring mechanisms underlying a values affirmation intervention for first-generation students. *J Pers Soc Psychol.* (2016) 110:635–59. doi: 10.1037/pspa0000049
55. Tibbetts Y, Priniski SJ, Hecht CA, Borman GD, Harackiewicz JM. Different institutions and different values: exploring first-generation student fit at 2-year colleges. *Front Psychol.* (2018) 9:502. doi: 10.3389/fpsyg.2018.00502
56. Phillips LT, Stephens NM, Townsend SS, Goudeau S. Access is not enough: cultural mismatch persists to limit first-generation students' opportunities for achievement throughout college. *J Pers Soc Psychol.* (2020) 119:1112. doi: 10.1037/pspi0000234
57. LaDonna KA, Taylor T, Lingard L. Why open-ended survey questions are unlikely to support rigorous qualitative insights. *Acad Med.* (2018) 93:347–9. doi: 10.1097/ACM.0000000000002088

Conflict of Interest: MM has consulted for Janssen and Otsuka Pharmaceuticals and received research support from Janssen in the past 5 years, all unrelated to the current work. ZS has received research and salary support from the National Institute of Health and the Center for Disease Control and has consulted for Sage Therapeutics.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

frontiers | Frontiers in Psychiatry

ORIGINAL RESEARCH
published: 16 May 2022
doi: 10.3389/fpsyt.2022.889602

Digital Strategies to Accelerate Help-Seeking in Youth With Psychiatric Concerns in New York State

Michael L. Birnbaum[1,2,3], Chantel Garrett[4], Amit Baumel[5], Nicole T. Germano[2], Cynthia Lee[2], Danny Sosa[1,2], Hong Ngo[6], Kira H. Fox[7], Lisa Dixon[6] and John M. Kane[1,2,3]*

[1] The Zucker Hillside Hospital, Northwell Health, Glen Oaks, NY, United States, [2] The Feinstein Institutes for Medical Research, Manhasset, NY, United States, [3] The Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY, United States, [4] Department of Health Systems and Population Health, University of Washington, Seattle, WA, United States, [5] Department of Community Mental Health, University of Haifa, Haifa, Israel, [6] Department of Psychiatry, New York State Psychiatric Institute, New York, NY, United States, [7] Department of Psychology, Barnard College, Columbia University, New York, NY, United States

Background: Mental illness in transition age youth is common and treatment initiation is often delayed. Youth overwhelmingly report utilizing the Internet to gather information while psychiatric symptoms emerge, however, most are not yet ready to receive a referral to care, forestalling the established benefit of early intervention.

Methods: A digital outreach campaign and interactive online care navigation platform was developed and deployed in New York State on October 22, 2020. The campaign offers live connection to a peer or counselor, a self-assessment mental health quiz, and educational material all designed to promote help-seeking in youth and their allies.

Results: Between October 22, 2020 and July 31, 2021, the campaign resulted in 581,981 ad impressions, 16,665 (2.9%) clicks, and 13,717 (2.4%) unique website visitors. A third (4,562, 33.2%) completed the quiz and 793 (0.1%) left contact information. Of those, 173 (21.8%) completed a virtual assessment and 155 (19.5%) resulted in a referral to care. The median age of those referred was 21 years (IQR = 11) and 40% were considered to be from low-income areas. Among quiz completers, youth endorsing symptoms of depression or anxiety were more likely to leave contact information (OR = 2.18, 95% CI [1.39, 3.41] and OR = 1.69, 95% CI [1.31, 2.19], respectively) compared to those not reporting symptoms of depression or anxiety. Youth endorsing symptoms of psychosis were less likely to report a desire to receive a referral to care (OR = 0.58, 95% CI [0.43, 0.80]) compared to those who did not endorse symptoms of psychosis.

Conclusion: Self-reported symptomatology impact trajectories to care, even at the earliest stages of help-seeking, while youth and their allies are searching for information online. An online care navigation team could serve as an important resource for individuals with emerging behavioral health concerns and help to guide the transition between online information seeking at baseline to care.

Keywords: digital advertisements, early intervention, youth mental health, help-seeking, social media, pathways to care

OPEN ACCESS

Edited by:
Giovanni De Girolamo, San Giovanni di Dio Fatebenefratelli Center (IRCCS), Italy

Reviewed by:
Marco Solmi, University of Padua, Italy
Hiran Thabrew, The University of Auckland, New Zealand

****Correspondence:***
Michael L. Birnbaum
Mbirnbaum@northwell.edu

Specialty section:
This article was submitted to Digital Mental Health, a section of the journal Frontiers in Psychiatry

Received: *04 March 2022*
Accepted: *27 April 2022*
Published: *16 May 2022*

Citation:
Birnbaum ML, Garrett C, Baumel A, Germano NT, Lee C, Sosa D, Ngo H, Fox KH, Dixon L and Kane JM (2022) Digital Strategies to Accelerate Help-Seeking in Youth With Psychiatric Concerns in New York State. Front. Psychiatry 13:889602. doi: 10.3389/fpsyt.2022.889602

INTRODUCTION

Mental health concerns in transition age youth are common and approximately 20% have a diagnosable mental illness (1–3). Symptom onset frequently occurs during formative years of adolescent and young adult development, and interferes with the establishment of healthy educational, vocational, and social foundations. While early intervention improves the likelihood recovery (4, 5), treatment initiation is often delayed (6), resulting in worse outcomes with long-lasting deleterious consequences persisting well into adulthood (7, 8).

Treatment initiation delays are multifaceted and include (i) demographic characteristics (9, 10) including age, sex, race, income, and health insurance status; (ii) systemic factors (11, 12) such as ill-defined pathways to care (iii) illness-related factors (13) such as speed of symptom onset; and (iv) environmental factors (14, 15) such as perceived stigma and level of mental health awareness within the family and the community. There is therefore no single existing strategy that would completely address this public health challenge, as it involves a constellation of factors unique to each individual. Novel, personalized, and nimble efforts are necessary to accomplish this goal.

Technological innovation, harnessing the established power of digital media, offers the prospect of facilitating treatment initiation by proactively identifying and engaging youth with behavioral health concerns online. Youth with mental health concerns overwhelmingly report utilizing the Internet first and most frequently to gather information while psychiatric symptoms emerge (16–18). Further, searching online represents one of the first proactive steps toward treatment initiation (19–21). However, the majority of youth searching for mental health related information online, including those who may be at risk for psychiatric disorders, report that they are not yet interested in receiving professional care (22), forestalling the established benefit of early intervention.

Advertisers routinely and effectively use the Internet to micro-target specific consumer segments directly beyond the capabilities of traditional media (23), however, limited efforts have focused on applying available technologies to engage help-seeking youth (and their allies) online and refer them to appropriate resources (24). Toward this goal, our team has developed a comprehensive care navigation platform (NYWell) designed exclusively to promote early intervention by identifying and engaging youth online with behavioral health concerns and to expeditiously link them to local resources. Leveraging search engine and social media-based advertisements, the project encourages participants to interact with our care navigation team online by offering peer-led support and guidance, as well as a virtual assessment conducted by a live clinician, and referral to care, if indicated.

While the project's primary research objective is to examine its effectiveness at reducing the duration of untreated psychosis for youth with first episode psychosis (FEP), we have, in the process, interacted with thousands of youths and their allies with a wide variety of behavioral health questions and concerns. The trial is currently active and the goal of this paper is to report on the pathways to care for all NYWell visitors from ad impression to receiving a referral to care based on the first 9 months of the project starting from the date of deployment (October 22, 2020) to July 31, 2021. We hypothesize that rates of engagement, assessment, and referral for both youth and their allies would vary based on self-reported psychiatric symptoms.

MATERIALS AND METHODS

Northwell Health's Early Treatment Program (ETP) and Strong 365, a non-profit dedicated to raising awareness about early psychosis intervention, collaborated on the development of a digital outreach campaign and interactive web-based platform aimed at understanding the role that digital media can play in promoting help-seeking for youth in the early stages of psychosis (**Figure 1**). This study was funded by NIMH (R34MH120790) and was approved by the Northwell Health IRB (#19-0266). Following a stepped wedge randomized controlled design, the campaign began running in select regions of New York State (NYS) on October 22, 2020, and expanded to the entire State of New York on October 25, 2021. The complete trial is due to run for a total of 18 months ending on April 22, 2022.

NYWell's content and design were created using participatory design principles (25). The project's multidisciplinary team include an advisory panel with lived experience, psychiatric researchers, peer support specialists, clinicians, social marketers, web developers, user-experience designers, graphic artists, and digital advertisers. Though our target audience is youth with early psychosis and their allies (parents, educators, primary care physicians, mental health care professionals, community/faith leaders, and friends), a key insight gained in preparatory interviews with stakeholders at the outset of prototyping was that the terms used to describe their early experiences with psychosis were broad (for example, stress, sleep, feeling different). Further, changes in mood and anxiety are often noted to be among the first symptoms to develop in emerging psychotic disorders (26). We thus strategically created an online environment that would be relevant and engaging to individuals with any question or concern pertaining to mental health. Youth and their allies also reported a desire to find information "all in one place," with simple language, easy navigation, and a clear set of potential actions, from a trusted source. Based on these insights, we aimed to construct an inviting online experience, for those who may themselves not yet be fully aware of, or ready to disclose (due to suspiciousness or paranoia, for example), psychiatric symptoms, as well as those who preferred to remain anonymous.

Campaign Overview

We developed and simultaneously deployed two separate digital media campaigns with advertising and web content targeting two audience segments. The first campaign targets youth searching online for mental health related information on behalf of themselves (Youth group). The second targets concerned caregivers and allies (Ally group). In order to provide the best opportunity to reach the intended audience, digital advertisements were placed on Google, Facebook, and

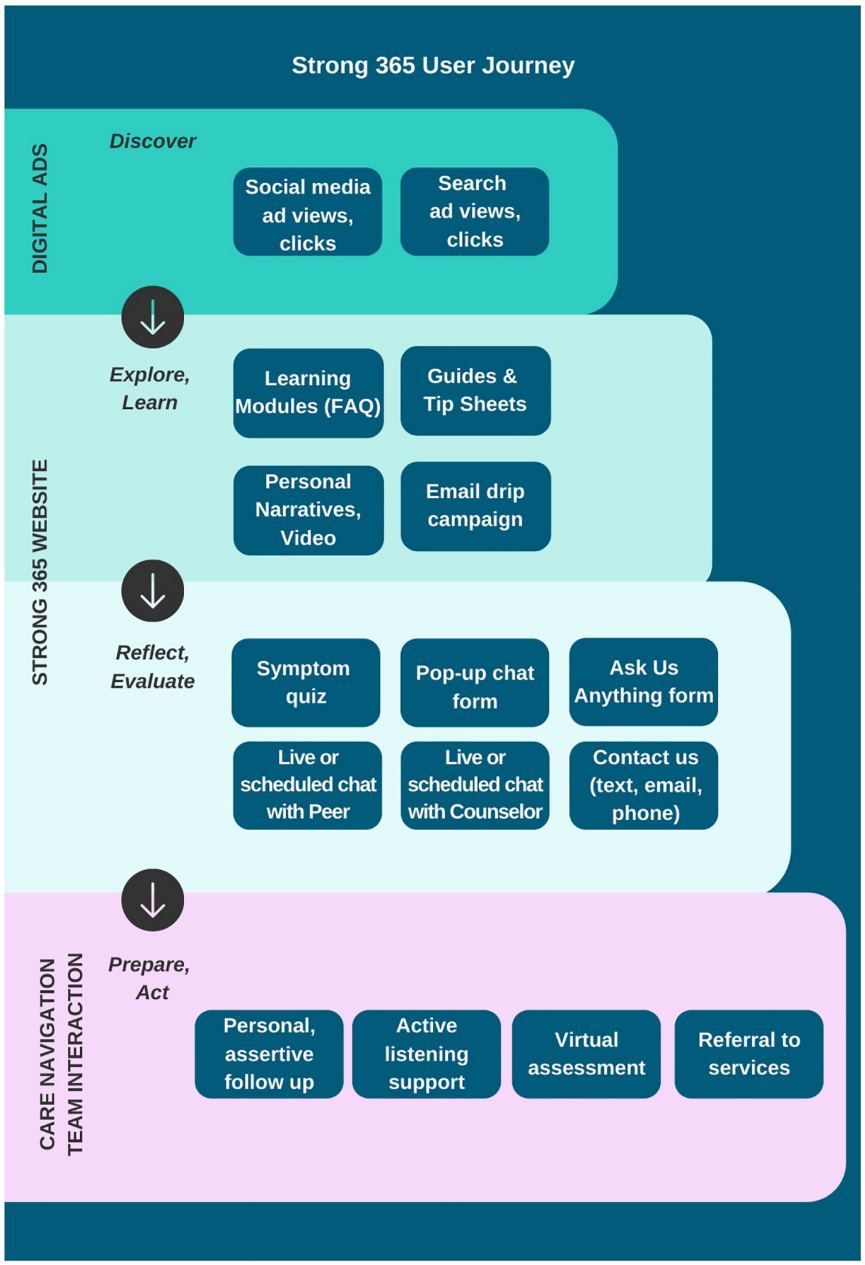


FIGURE 1 | User experience from ad to referral.

Instagram platforms. Ads are demographically, geographically, and interest-group targeted, and appear in response to online activity conducted by each user. Users click from an ad to the campaign website, where they can engage with the NYWell care navigation team via text, live/synchronous chat, asynchronous chat, email, or phone. NYWell is accessible both via desktop and mobile.

Digital Ads

More than 4,900 keywords (purchased search terms) and 240 search engine ads were created and tested, spanning numerous symptom/experience and health-behavioral categories (**Table 1**). Each ad consists of a headline, description, and call to action. Keywords and ad content were drawn from user interviews, analysis of search trends in NYS, and the performance of prior pilot campaigns (24). Additionally, 75 image and video-based ads were placed on Facebook and Instagram (**Figure 2**). Based on platform user data, a younger demographic was primarily targeted on Instagram, and an older demographic on Facebook (27). Beyond demographic and geographic targeting, the following interest groups were included: educators, parenting, medical and mental health professionals, LGBTQ + , gender

TABLE 1 | Top performing search engine ad content (by clicks).

	Top 10 Ad groups (By clicks)	Target audience	Clicks N (%)	Top performing keywords (By clicks)	Top performing headlines (By clicks)
1	Information about mental health conditions, promoting quiz as a next action	Youth	3,387 (43%)	Psychological tests Extreme anxiety Anxiety quiz Mental illness test Depression symptoms	What Do Your Symptoms Mean? Check Up on Your Mental Health Free and Confidential Quiz
2	Ally focused ads promoting quiz as a next action	Ally	1,775 (23%)	Mental health assessment Symptoms of depression Anxiety test Schizophrenia disorder Mental illness test	What Are Depression Symptoms? Understand Anxiety Symptoms Early Signs of Schizophrenia
3	Depression-focused, promoting quiz as a next action	Youth	859 (11%)	Mental depression Depression test Clinical depression Depression Help for depression	What is Major Depression? Do I Have Clinical Depression? Free and Confidential Quiz
4	Parent-focused ads promoting quiz as next action	Ally	702 (9%)	Mental illness Symptoms of depression Anxiety symptoms Depressed child Child sleep problems	Is Your Child Struggling? When To Seek Help For A Child Free Support for Parents
5	Speak with a therapist	Ally	327 (4%)	Symptoms of depression Free mental health quiz Mental health evaluation Child not sleeping Depression test	Is Your Child Struggling? Speak with Licensed Therapists Free Support for Parents
6	Free guidance and consultation	Ally	245 (3%)	NYC mental health hotline New York mental health resources Mental health symptom quiz Mental health assessment Young person mental health	Speak with Licensed Therapists Free Support for Parents Free Consult for NY Parents
7	Mental health-related experiences as described by youth	Youth	190 (2%)	Am I depressed test Am I depressed quiz Bipolar quiz Compulsive thoughts Obsessive thoughts	What Do Your Symptoms Mean? Check Up on Your Mental Health Free and Confidential Quiz
8	Information about mental health conditions, promoting peer consultation as a next action	Youth	164 (2%)	I have schizophrenia Mood disorder test Bipolar quiz How do you know if you have bipolar Psychiatry help	What Do Your Symptoms Mean? Check Up on Your Mental Health Talk to Certified Peer Mentors
9	Information about mental health conditions, promoting therapist consultation as a next action	Youth	107 (1%)	Disorder test Psychological tests Coping with stress Mental check Free therapy	Mental Health Support Online Free Mental Health Assessment Our Therapists Can Help
10	Schizophrenia-related inquiries	Ally	42 (1%)	Child schizophrenia Teenage schizophrenia teenage schizophrenia stories Symptoms of schizophrenia in teens Child schizophrenia test	Teen & Young Adult Specialists Early Signs of Schizophrenia Symptoms of Schizophrenia
Total clicks in top 10 ad groups			7,798		

queer, sports and coaching, college, high school, child protective services/foster care, juvenile justice, and homeless youth.

Campaign Website

For each campaign (Youth, Ally), an audience specific website was developed (see **Supplementary Figure 1**). To optimize visibility, the websites were designed to focus on core messages derived from prior pilot initiatives and found to be most relevant to information seeking individuals (24). These include accurately identifying and understanding early signs and symptoms of emerging psychiatric illness and finding local support (28). Based on feedback from lived-experience co-designers, the content is displayed in a simple yet informative, and engaging manner utilizing text, images, videos, and empowering personal narratives. The site is translatable into any language in recognition of the language diversity among the population

we aim to serve. The site sought to encourage anyone with psychiatric questions or concerns to engage with the care navigation team in a format that met their desire for anonymity, confidentiality, immediacy, and ease of scheduling a conversation at a convenient time. The site facilitated connection with a compassionate peer listener who could help navigate the mental health system, or a licensed mental health counselor who could offer context and suggest a set of personalized next steps.

Once on the website, psychoeducation is offered as a tool to enhance appreciation for the early warning signs and symptoms of mental illness as well as the benefits of early intervention. Educational materials include Frequently Asked Questions (FAQ), blog articles, tip sheets, videos, and short written narratives of young people sharing their personal experiences. Users are also offered the option to sign up for an email newsletter written by team members with personal lived experience, which is set up to regularly deliver wellness tips, personal stories, and educational content over the course of several months. We offer multiple paths to take action based on user preferences, including a chatbot that pops up on the site for first-time visitors asking "How we can help?," inviting them to leave their contact information for same or next-day follow-up, and an "Ask us anything" form, in which a user enters a question that is answered by the care navigation team same or next-day.

Mental Health Check-In Quiz

A mental health quiz offers the opportunity for self-evaluation. We developed and deployed a brief quiz that was designed to serve as an engagement tool, rather than a diagnostic screener, with limited clinical utility. We encouraged all quiz takers to interact with the care navigation team upon completion to learn more about their own mental health. We thus selected a broad-based approach asking about symptoms of anxiety, depression, and psychosis. The quiz contains a total of 21 items including 4 questions adapted from the PRIME screen (29), which screens for psychosis, 2 questions adapted from the Patient Health Questionnaire (PHQ-9), which screens for depression (30), and 1 question adapted from the Generalized Anxiety Disorder Questionnaire (GAD-7), which screens for anxiety (31). The Ally campaign also includes a 21-item quiz consisting of parallel questions that were adapted to be relevant to caregivers' perspectives of emerging symptoms in a loved one ("your loved one" as opposed to "I"). We additionally included optional questions regarding demographics, motivations for seeking information or support, their interest in a variety of possible next steps as a means of assessing readiness to take action, as well as how well they were able to engage in the things that are most important to them such as work, school, and relationships, on a scale from 1 to 10, with 1 representing not able and 10 representing fully able.

At the end of the quiz, users have another opportunity to immediately connect via live online chat or text with the care navigation team, schedule an appointment via an integrated online scheduling app, or leave their contact information with preferred method of outreach (email, phone, or text).

Care Navigation: Engagement, Assessment, and Referral

Once an individual's contact information is submitted via the quiz, chat, or the appointment scheduler, participants are then considered to be an "active inquiry." Staff are available generally between 9 am and 5 pm during weekdays to interact instantly with users online. They are also available intermittently on evenings and weekends and are expected to respond to users who leave contact information within 24 h. The appointment scheduling process is designed to advance help-seeking by encouraging users who leave contact information to participate in a remote clinical assessment. The team initially reaches out via the user's preferred contact method to thank them for their interest and to ask when they might be available to discuss their stated concerns. If users do not respond within 24 h, we continue to reach out systematically for the first 4 weeks, followed by enrolling users in our email drip campaign in an effort to maintain a connection. Once assessed, if clinically indicated, users are then offered a referral to local mental health resources based on their needs, location, and preferences. Referrals are enabled throughout NYS by a comprehensive third-party database of available mental healthcare and social service providers.

Informed Consent and Safety Protocols

Verbal informed consent is obtained from individuals who interact with program staff via phone, text, email, live chat, or video call. Participants who provide consent are then provided with a description of the study. Study objectives are also described in detail in the privacy policy available on the campaign website. Ongoing data security is managed by complying with industry standards including The Health Insurance Portability and Accountability Act (HIPAA) and best practices at Northwell Health, where NYWell is hosted on a secure server. Participants who interact with research staff are assessed for safety. Incoming messages from participants are also monitored by research staff daily during standard business hours, and at least once daily during weekends. Any identified safety risk is immediately reported to the clinicians on the research team including a child and adolescent psychiatrists (MB) and licensed mental health counselor (NG) and escalated as necessary, including attempts to promptly connect the participant to appropriate local services.

Data Collection and Analysis

Demographic and clinical data entered by each participant completing the quiz are collected. In addition, data is collected regarding the timing, frequency, and method (text, email, phone, live chat, video call) of each contact between the care navigation team and participants interacting with the campaign. Campaign analytics are also utilized to record digital ad and website engagement data including the amount of time spent on the website, the number of clicks while on the website, as well as user behavior while on the website such as completing the quiz, appointment scheduling, and live chat. Based on user responses to the online quiz, we identified three diagnostic risk categories (risk for psychosis, risk for depression, and risk for anxiety). In

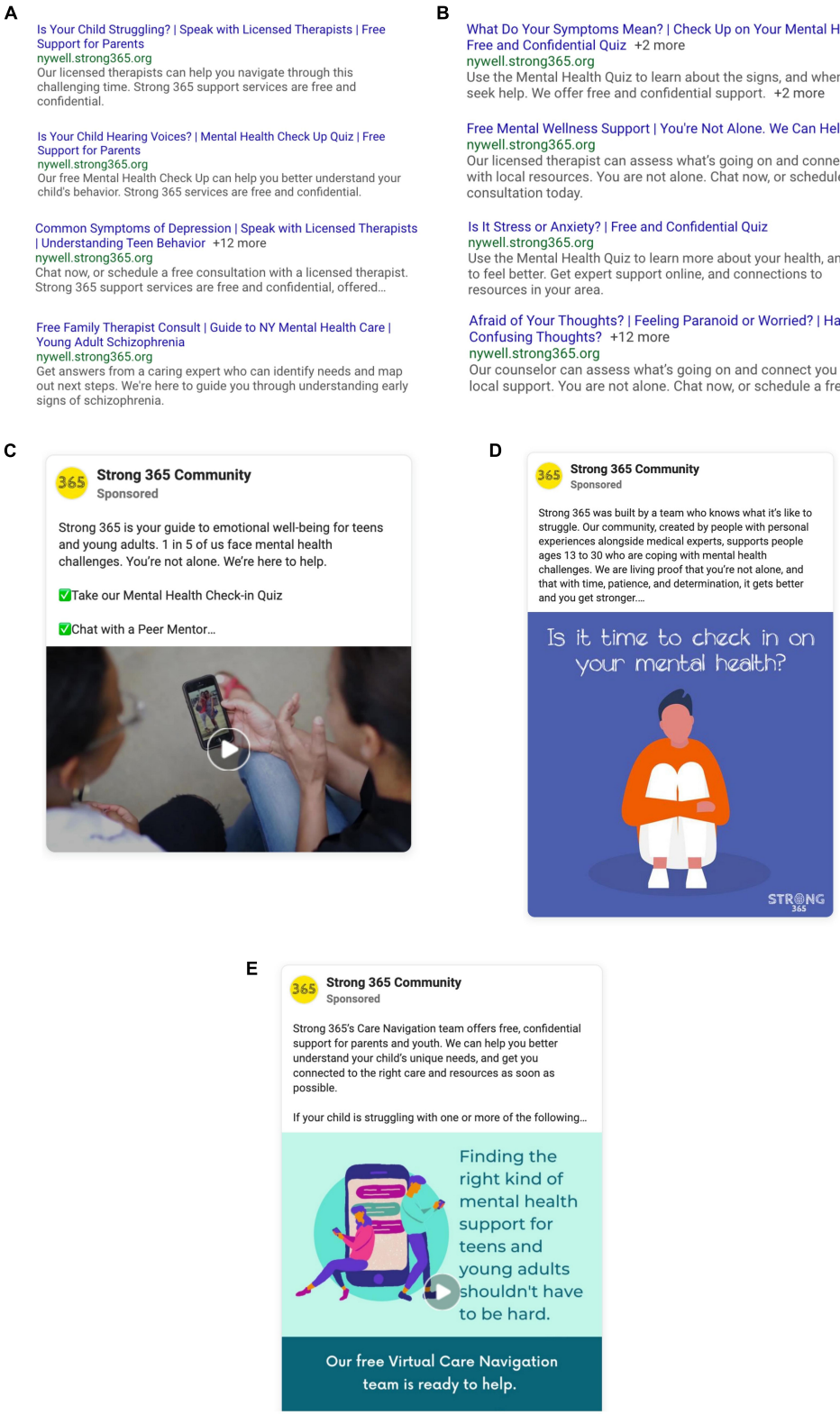


FIGURE 2 | Sample social media and search engine ads. **(A)** Example Search ads targeting allies such as parents, educators, medical professionals who are searching for mental health-related information. **(B)** Example Search ads targeting youth who are searching for mental health-related information. **(C)** Example social media video ad featuring a young person with lived experience sharing their story. **(D)** Example social media graphic ad inviting young people to take the quiz. **(E)** Example social media animation ad targeted toward parents of teens and young adults.

our analysis we selected a threshold such that reporting any of the adapted screener items pertaining to one of the three diagnostic categories suggested a risk state for that condition.

As part of the assessment process, we collect location data (city or neighborhood) whenever possible, enabling identification of local resources, and providing a measure of the socio-economic status of those referred. The American Community Survey income data available through the Economic Research Service and the City of New York was used to understand the complexion of those referred at the city and neighborhood level (32, 33). Based on these sources, low income was defined as neighborhoods or cities with a poverty rate of 20% or higher, or median family income less than 80% of median family income for the state or metropolitan area.

We used percentages, medians, and interquartile range (IQR) for descriptive statistics. Given symptom overlap between diagnostic risk categories (i.e., most participants report symptoms or experiences pertinent to more than one condition), we performed logistic regressions to analyze the relationship between each risk category (depression, anxiety, psychosis) and dichotomous variables (i.e., leaving contact information, reporting a desire to obtain a referral to care) while holding all other risk categories constant. Since we could not assume a normal distribution of contact response time, Mann-Whitney U test was performed to examine the difference in contact response time between those referred to services and those who were not.

RESULTS

User Journey From Ad Impression to Inquiry

The user journey from digital ad (impression) to active inquiry is presented in **Table 2**. Between October 22, 2020, and July 31, 2021, the campaign resulted in a total of 581,981 ad impressions, 16,665 (2.9%) ad clicks, 13,717 (2.4%) unique website visitors, and 793 (0.1%) active inquiries. Altogether, the click through rate (percentages of clicks resulting from total number of impressions) of Google ads was approximately triple the rate of social media generated clicks. Instagram outperformed Facebook ads, representing 78.1% of social media-based clicks. In total, 69.3% (403,039/581,981) of impressions targeted youth searching for information on behalf of themselves, who eventually comprised 64.1% (8,798/13,717) of the total campaign website visitors. According to campaign analytics (available from Google for only 48.4% of website visitors), 71% of Youth and 69% of Allies accessed the site via desktop, while 27% of Youth and 31% of Allies accessed via mobile device. Youth were more engaged with the platform, spending on average 158 s on the website compared to 107 s for Ally visitors. Further, 45.5% of Youth campaign website visitors completed the quiz compared to only 11.4% of Ally visitors, and 7.6% of youth left contact information for the care navigation team compared to 2.4% of Ally visitors. Among quiz takers, nearly half (49%, 1,443) were 18 and younger, 21% (610) were 19-25, 17% (517) were 26-35, 8% (232) were 36-49, and 6% (172) were 50 and older (**Figure 3**).

Quiz Completers; Responses and Desired Outcomes

Approximately one third of website visitors (4,562, 33.2%) completed the quiz. Responses, grouped by diagnostic risk categories are presented in **Table 3**. Most participants reported symptoms consistent with a risk for depression (93% youth and 84% ally). Psychosis risk was the least commonly reported diagnostic category (68% Youth and 60% Ally). Altogether (**Figure 4**), the most commonly reported desired outcome was to learn about emotional health (71% of youth and 44% of allies) and the least commonly reported goal was to receive a referral to care (12% youth and 15% ally). On average, users rated their ability to engage in work, school, relationships to be 5.35 (SD = 2.31) on a scale from 1 (unable) to 10 (fully capable). Approximately 16% of quiz-completers left contact information. Beyond the quiz, users most frequently visited the FAQ section of the website (1,050, 7.6%).

Logistic regressions were performed to explore associations between diagnostic risk categories and the likelihood of leaving contact information and reporting a desire to receive a referral to care. Regressions were performed separately for both Youth and Ally campaign groups. While holding all other predictor variables constant, Youth visitors reporting symptoms consistent with risk for depression were twice as likely to leave contact information compared to those who did not report symptoms of depression (OR = 2.18, 95% CI [1.39, 3.41]); Youth visitors reporting symptoms consistent with a risk for anxiety were 69% more likely to leave contact information compared to those who did not report symptoms of anxiety (OR = 1.69, 95% CI [1.31, 2.19]). Though not statistically significant, the percentage of Youth visitors leaving contact information was lower in those who reported symptoms consistent with psychosis risk (11%) compared with those who did not report risk for psychosis (16%). Subsequently, reporting symptoms suggestive of psychosis among Youth was associated with a significant decrease of more than 40% in the likelihood of stating an interest in a referral to care (OR = 0.58, 95% CI [0.43, 0.80]). Endorsing depression or anxiety risk did not impact the likelihood of stating an interest in a referral to care.

Among Ally visitors, no significant relationship was found between diagnostic risk categories and the predictor variables, while holding other variables constant. However, a similar pattern to the one seen among Youth visitors emerged among Ally visitors (**Table 3**). For example, Ally visitors reporting symptoms of psychosis were descriptively less likely to state an interest in a referral to care compared to those not reporting psychosis risk (22% vs. 26%, respectively); Ally visitors endorsing risk for either depression or anxiety had descriptively higher interest in receiving a referral to care (25% vs. 19% and 25% vs. 18%, respectively).

Care Navigation From Inquiries to Referrals

The trajectory from inquiry (leaving contact information) to assessment and referral, grouped by initial inquiry channel and participant (Youth vs. Ally) is presented in **Table 4**. Of the

TABLE 2 | User journey from digital ad to active inquiry.

Online funnel	Total		YOUTH		ALLY	
	N	%[a]	N	%[a]	N	%[a]
Impressions						
Google Ads impressions	136,771	24%	77,052	19%	59,719	33%
Social impressions	445,210	76%	325,987	81%	119,223	67%
Total Impressions	581,981	100%	403,039	100%	178,942	100%
Ad clicks						
Google Ad clicks	7,877	5.8%[b]	4,712	6.1%[b]	3,165	5.3%[b]
Social Ad clicks	8,788	2.0%[c]	6,865	2.1%[c]	1,863	1.6%[c]
Total Clicks	16,665	2.9%	11,577	2.9%	5,028	2.8%
Website activity						
Unique Users	13,717	2.4%	8,798	2.2%	4,919	2.8%
Average time on page (seconds)	124	N/A	158	N/A	107	N/A
Quiz takers	4,562	33.2%[d]	4,002	45.5%[d]	560	11.4%[d]
Inquiries	793	5.7%[d]	673	7.6%[d]	120	2.4%[d]

[a] *When not stated otherwise, the percentage is calculated out of the total number of impressions for the relevant column;* [b] *Percentage out of the Google Ad impressions;* [c] *Percentage out of the social-media impressions;* [d] *Percentage out of the number of unique users for the respective column.*

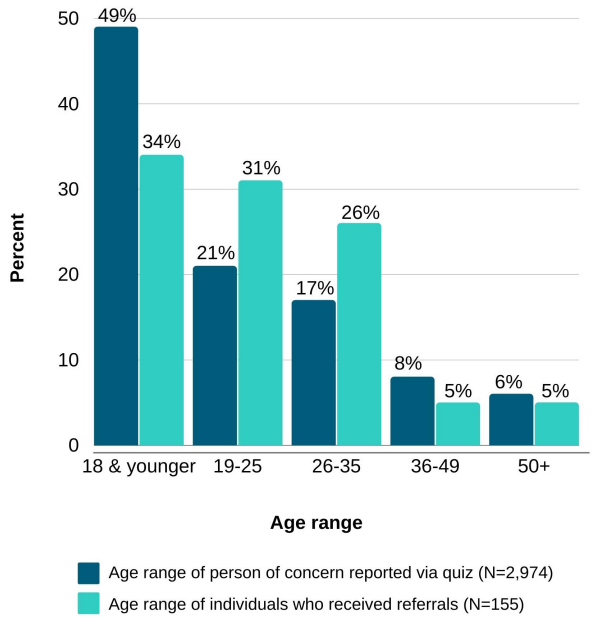


FIGURE 3 | Age range reported via quiz (n = 2,974) compared to the age range of individuals referred to care (n = 155).

793 inquiries that left contact information, 173 (21.8%) were successfully engaged to complete a virtual assessment. Most assessments resulted in a referral to care (155/173; 89.6%), indicating that those who successfully connected with the care navigation team were found to be experiencing psychiatric symptoms that warranted a referral.

Overall, the percentage of inquiries that eventually resulted in an assessment and referral was almost twice as high among Ally visitors (31.7%) compared to Youth visitors (17.4%). Of note, however, the number of visitors entering the Youth campaign was substantially larger compared to the Ally campaign, and thus most referrals to care were made via the Youth campaign. Most referrals (93.5%, 145/155) were to local general outpatient mental health services, while a minority were directed to first episode psychosis programs (3.9%, 6/155) and to clinical high-risk for psychosis programs (3.2%, 5/155). The median age of those referred was 21 years (IQR = 11). Compared to the age ranges reported in the quiz (n = 2,974), the referred population (n = 155) included a higher proportion of individuals 19–25 and 25–35, and a lower proportion of individuals 18 and under, 36–49, and 50 and older (**Figure 3**). Location data (city or neighborhood) was available for 79% (122/155) of the participants referred to care. Of those, 40% (49/122) were located in areas defined as low income.

Contact Characteristics of Those Who Received a Referral to Care

As seen in **Table 4**, inquiries resulting in a referral to care was highest among users who initially made contact using phone call (n = 3/4, 75%) or appointment scheduling (n = 11/18, 61.1%), with the lowest percentage found in users submitting first inquiries via the quiz (n = 155/793, 19.5%). Of those who were referred to care, the median number of outreach attempts initiated by the care navigation team was 6 (IQR = 4). The median time difference from first inquiry to last contact prior to receiving a referral was 14.9 days (IQR = 22.3). The team's initial response time after receiving an inquiry was significantly shorter for those who were successfully referred to care (median = 0.2 h, IQR = 9.8) compared to those who were not referred to care (median = 12.3 h, IQR = 28.7; Mann–Whitney U = 17,403, $P < 0.001$).

DISCUSSION

This manuscript aimed to characterize online help-seeking behaviors of youth and their allies interacting with a digital

TABLE 3 | User preferences grouped by diagnostic risk categories.

		Symptoms reported	Left contact information for follow up	Completed "next steps" questions (next columns)	Learn about emotional health	Learn how to open up to others; how to ask for help [a]	Learn about therapy or treatment	Obtain self-help resources	Referral to care
		N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)
Youth	**Total Youth Quizzes**	**4002 (100%)**	**624 (16%)**	**1641 (41%)**	**1171 (71%)**	**584 (36%)**	**569 (35%)**	**584 (36%)**	**199 (12%)**
	No psychosis symptoms	1284 (32%)	185 (14%)	474 (37%)	316 (67%)	115 (24%)	181 (38%)	176 (37%)	76 (16%)
	One or more psychosis symptoms	2718 (68%)	439 (16%)	1167 (43%)	855 (73%)	382 (33%)	388 (33%)	408 (35%)	123 (11%)
	No depression symptoms	296 (7%)	22 (7%)	88 (30%)	64 (73%)	12 (14%)	28 (32%)	25 (28%)	6 (7%)
	One or two depression symptoms	3706 (93%)	602 (16%)	1553 (42%)	1107 (71%)	485 (31%)	541 (35%)	551 (35%)	193 (12%)
	No anxiety symptom	758 (19%)	77 (10%)	283 (37%)	203 (72%)	72 (25%)	87 (31%)	104 (37%)	31 (11%)
	Anxiety symptom	3244 (81%)	547 (17%)	1358 (42%)	968 (71%)	425 (31%)	482 (35%)	480 (35%)	168 (12%)
Ally	**Total Ally Quizzes**	**560 (100%)**	**92 (16%)**	**253 (45%)**	**112 (44%)**	**143 (57%)**	**108 (43%)**	**76 (30%)**	**59 (23%)**
	No psychosis symptoms	226 (40%)	34 (15%)	97 (43%)	46 (47%)	54 (56%)	46 (47%)	31 (32%)	25 (26%)
	One or more psychosis symptoms	334 (60%)	58 (17%)	156 (47%)	66 (42%)	89 (57%)	62 (40%)	45 (29%)	34 (22%)
	No depression symptoms	126 (23%)	13 (10%)	58 (46%)	35 (60%)	24 (41%)	21 (36%)	14 (24%)	11 (19%)
	One or two depression symptoms	473 (84%)	79 (17%)	195 (41%)	77 (39%)	119 (61%)	87 (45%)	62 (32%)	48 (25%)
	No anxiety symptom	161 (29%)	19 (12%)	62 (39%)	25 (40%)	37 (60%)	32 (52%)	20 (32%)	11 (18%)
	Anxiety symptom	399 (71%)	73 (18%)	191 (48%)	87 (46%)	106 (55%)	76 (40%)	56 (29%)	48 (25%)

[a] *In the ally quiz, the wording of this question was altered to reflect a caregiver perspective, "Learn how to talk to or support a loved one."*

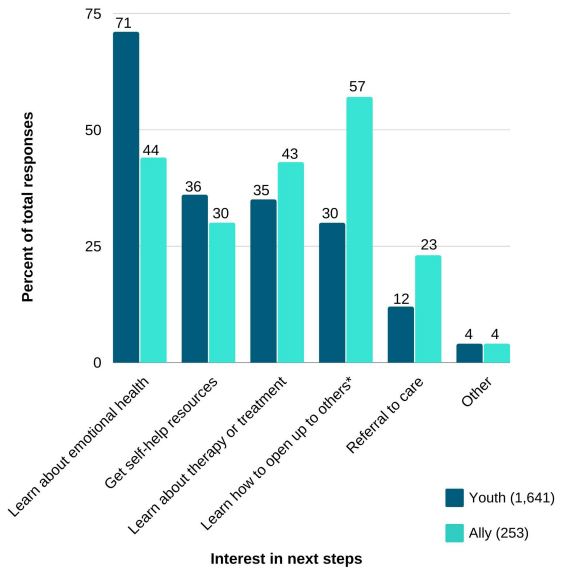


FIGURE 4 | User preferences grouped by youth and allies (n = 1,894). *Ally selection was worded, "Learn how to talk to or support a loved one."

care navigation platform in NYS. Our findings suggest that self-reported symptomatology impact trajectories to care, even at the earliest stages of help-seeking, while youth and their allies are searching for information online. An online care navigation team could serve as an important resource for individuals with emerging behavioral health concerns and help to guide and support the transition between online information seeking at baseline to care.

Prior efforts to accelerate treatment initiation (34–38) have involved broad-based strategies to screen for mental illness within the community and/or raise awareness of the benefits of early intervention. Though educational campaigns have demonstrated success, they have relied predominantly on offline mass media channels, such newspaper advertisements, transit advertising, brochures, posters, TV, movie, and radio commercials. Websites were developed, though primarily designed for healthcare professionals and/or as a source of obtaining information about the project, rather than generating community-based referrals. Mindmap, a more recent federally funded research initiative in the United States (37), included digital media channels such as Facebook, Twitter, YouTube, Reddit, and LinkedIn, in addition to traditional mass media, though this strategy was also primarily intended to contribute to education rather than generate referrals directly. Mindmap inquiries were largely made by phone rather than interactive online channels, and referral sources were predominately clinical rather than self-generated community referrals. Our findings are in line with growing literature highlighting the potential for digital technology to improve the process of identifying struggling youth and the effectiveness of outreach and engagement efforts (22, 24, 39–41). The Internet may serve as a critical resource to connect with concerned youth and their allies at the earliest possible time in the help-seeking trajectory, when individuals first begin to search for information online. While many individuals describe their first contact with psychiatry as a negative and often traumatic experience (42), online resources, staffed by peers and mental health professionals, such as the NYWell project offers an opportunity to ease the introduction to care, paving the way for greater engagement with services and thus improved outcomes for youth with behavioral health conditions.

In line with prior reports (19–22, 43), we found that younger participants were predominantly interested in obtaining behavioral health information rather than a referral to care. Their quiz responses and online behaviors reflect a desire to understand their experiences and obtain answers to questions and concerns. For example, while completing our symptom quiz was a popular online activity among youth, once completed, most youth did not leave contact information, demonstrating limited readiness to take action toward receiving support. Furthermore, while youth ages 18 and under represent nearly half of all quiz completers, most referrals to care were provided to young adults between the ages of 19–25 and 26–35. This discrepancy may be partially explained by the requirement for minors to obtain parental consent to engage in psychiatric care in NYS, representing another potential barrier to help seeking. Importantly, our data highlight that self-reported symptomatology impacts the likelihood of advancing help-seeking behaviors beyond information gathering. This pattern was apparent for both youth and their allies. Understanding the barriers and facilitators present to each diagnostic group and more granularly to each individual, will be critical to improving efforts to reduce the duration of untreated illness. Tailored strategies will likely be necessary to advance help seeking of those in need based on a thorough appreciation of their symptoms and obstacles to accessing care. For example, while substantial stigma is associated with psychosis, depression and anxiety are somewhat less stigmatized (44, 45) and may partially explain the differences in the likelihood of reporting a desire to seek care and to leave contact information. It is also plausible that symptoms associated with depression and anxiety are more readily identifiable, as exemplified by evidence supporting online assessment of these conditions (46, 47), while symptoms of psychosis are more challenging to recognize, contributing to delayed help-seeking (48). Lastly, the presence of common psychotic symptoms themselves, such as paranoia, may result in mistrust of our efforts and rapid disengagement.

Youth were highly interested in completing our self-assessment quiz, however, most did not leave contact information and were lost to follow up, highlighting both a strength and limitation to online resources. While the Internet provides a comfortable and anonymous setting for help-seeking (49), our ability to confidently ascertain who these individuals are and understand their motivation for searching online and completing a self-assessment quiz is limited. Further, we are unable to clinically confirm the presence or absence of self-reported psychiatric symptoms based on the quiz alone. While psychiatric self-screeners effectively assess for psychiatric diagnoses risk (29–31), few self-screeners have been validated in an online environment (50, 51), which may consist of a different population

TABLE 4 | Trajectory from inquiry to assessment and referral, grouped by channel and participant.

	Total		Quiz		Video		Appointment scheduling		Pop up		Phone call		Ask us anything		Text messaging	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Total sample																
Inquiries	793	100%	719	100%	23	100%	18	100%	14	100%	4	100%	14	100%	1	100%
Assessments	173	21.8%	139	19.3%	9	39.1%	11	61.1%	5	35.7%	3	75.0%	5	35.7%	1	100.0%
Referrals	155	19.5%	123	17.1%	9	39.1%	11	61.1%	4	28.6%	3	75.0%	5	35.7%	0	0.0%
Youth																
Inquiries	673	100%	633	100%	15	100%	14	100%	7	100%	0	0.0%	5	100%	0	0.0%
Assessments	130	19.3%	112	17.7%	7	50.0%	7	50.0%	2	28.6%	0	0.0%	2	40.0%	0	0.0%
Referrals	117	17.4%	100	15.8%	7	50.0%	7	50.0%	1	14.3%	0	0.0%	2	40.0%	0	0.0%
Ally																
Inquiries	120	100%	86	100%	8	100%	4	100%	7	100%	4	100%	9	100%	1	100%
Assessments	43	35.8%	27	31.4%	2	25.0%	4	100.0%	3	42.9%	3	75.0%	3	33.3%	1	100.0%
Referrals	38	31.7%	23	26.7%	2	25.0%	4	100.0%	3	57.1%	3	75.0%	3	33.3%	0	0.0%

and setting. In order to build better online resources for youth and their allies, we will likely need strategies informed by symptomatology, and further validating online self-screeners delivered over the Internet will be a critical next step.

Our findings highlight that individuals online present at very different stages along the help-seeking continuum (52) and in varying degrees of readiness for change at the time of online search query. While some were immediately prepared to take action by proceeding with a remote clinical assessment and referral, many others were reluctant to advance and either did not leave contact information or were lost to follow up. Further, motivation for help-seeking likely oscillates over time. For instance, we found that a quicker response time was more likely to result in a referral, perhaps capturing a critical window of opportunity when help-seeking motivation was higher. Future research will need to build agile online interventions and creative engagement strategies designed to meaningfully advance help-seeking behaviors based on an individual's needs and readiness for action at the moment of initial engagement. A successful care navigation platform will ultimately need to ascertain, beyond presenting symptoms, which individuals might benefit most from what kind of online support and guidance based on where they are at that moment in the help-seeking trajectory. Further, separate strategies will need to be developed and tested for allies. Our data support the critical importance of engaging allies in facilitating treatment initiation as a greater proportion were ready to take action compared to youth. Further, while youth in our dataset readily interacted with our digital ads and gravitated toward completing an online self-assessment, allies were much less likely to engage in either of these activities, preferring instead to expeditiously connect to a licensed professional. More research is necessary to understand how to effectively reach and engage allies online.

Forty percent of our referred population with known neighborhood or city data were in low-income areas. By comparison, New York State's poverty rate is 12.7%, and New York City's is 17.9% suggesting that our platform may be reaching and serving a disproportionate number of individuals with limited financial means (32, 33). Collecting a range of additional demographic data in future implementations of the platform will enable us to ascertain the potential for a youth-focused online care navigation service as a mechanism to improve access to care for populations who are underserved due to social, economic, policy, and environmental factors (53).

Limitations and Future Directions

There are several noteworthy limitations. First, we selected to develop and implement a self-assessment quiz, designed to function as an engagement tool. We did not utilize validated symptom screeners, limiting our ability to report with greater certainty on the mental health characteristics of the population we reached. Further, given substantial symptom overlap between the symptoms reported in each of the three diagnostic risk categories, future research will need to better delineate how specific diagnoses and psychiatric symptoms impact online help-seeking behaviors for both youth and allies. This may require the development of novel digital symptom self-screeners deployed in an online environment, or the incorporation of established self-screeners, followed by a remote clinical assessment to confirm diagnostic accuracy or determine the most effective threshold for accurately identifying symptoms in an online environment. Second, most participants opted to remain anonymous, and many others were lost to follow up, limiting our ability to determine what these individuals want, need, and how best to support them. Although, we can reasonably assume that these individuals are interested in obtaining mental health related information, we are unable to confirm their motivation, if they are in need a referral to traditional psychiatric care, or other forms of sub-clinical support. Many website visitors likely do not need formal psychiatric intervention and future research will need to explore novel engagement and intervention strategies for those who might benefit from support and guidance and whose symptoms do not meet the threshold necessary to warrant formal intervention from a mental health professional. Additionally, the success of online treatment programs aimed at young people (54–56) may offer innovative and effective alternatives to traditional in

person care. Fourth, the NYWell project was designed ultimately to identify individuals in the earliest stage of help-seeking and expeditiously refer them to care. This approach was successful at capturing a momentary spike in help-seeking motivation, while individuals are searching online for information, however, the campaign demonstrated limited ability to maintain help-seeking motivation over time for many in order to advance them from online information seeking at baseline to community-based care. Accordingly, this approach may have resulted in individuals being lost to follow up, who require alternate forms of engagement, or a longer timeline over which to gain the awareness, self-efficacy, and social support necessary to take action. Much more work is required between these two endpoints (identification and referral) and tailored engagement efforts will be critical to meaningfully advance help-seeking. Future interventions will need to incorporate flexible and personalized strategies based on an individual's readiness for change, needs, and expectations while incorporating lessons learned thus far such as the importance of response times and selecting the most appropriate digital platforms for each target audience. Lastly, given that the project is still active at the time of publication, data related to cost were not included. Campaign initiation is typically more expensive until algorithm optimization occurs on the digital advertising platform, which is a dynamic and ongoing process. Once the campaign is complete, a formal cost effectiveness analysis will be conducted to accurately present associated costs.

Leveraging digital technology, an online care navigation platform may prove to be a critical resource capable of refining the help-seeking process for youth and their allies. Our findings reinforce the need to further delineate how individuals progress beyond online information seeking at baseline to meaningfully taking action toward care. By better understanding motivations and barriers, we can continue to expand and implement tailored engagement strategies, designed to effectively support and guide youth and their allies with mental health questions and concerns. Moreover, we must continue to develop and test digital support tools and interventions geared toward the growing population of youth who are in need of support, but not yet ready, and may not need, formal psychiatric care delivered by a mental health clinician.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Northwell Health IRB. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

MB, CG, JK, AB, and LD conceptualized and executed the project and interpreted the results. NG, DS, CL, and KF contributed to participant recruitment and data collection. AB conducted the data analysis along with support from CG and MB. MB, AB, and CG wrote the initial draft of the manuscript. All authors contributed to the manuscript preparation and editing.

FUNDING

This work was supported through funding from the National Institutes of Mental Health (R34MH120790) as well as the One Mind Foundation.

ACKNOWLEDGMENTS

We would like to acknowledge individuals and teams who contributed to this project including: Syrena Clark, Rebecca Hurd, Amanda Lipp, Dr. Rachel Loewy, Mili Mehta, Kendra Prasad, Dr. Mark Savill, Sohil Shah, Dr. Vinod Srihari, members of the Lived Experience and Family Member Advisory Panels, LastRev, OnTrackNY, and Tenlee Media.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyt.2022.889602/full#supplementary-material

Supplementary Figure 1 | Strong 365 NYWell Youth Website.

REFERENCES

1. Merikangas KR, He JP, Burstein M, Swanson SA, Avenevoli S, Cui L, et al. Lifetime prevalence of mental disorders in U.S. adolescents: results from the National comorbidity survey replication–adolescent supplement (NCS-A). *J Am Acad Child Adolesc Psychiatry.* (2010) 49:980–9.
2. Auerbach RP, Mortier P, Bruffaerts R, Alonso J, Benjet C, Cuijpers P, et al. WHO World mental health surveys international college student project: prevalence and distribution of mental disorders. *J Abnorm Psychol.* (2018) 127:623–38. doi: 10.1037/abn0000362
3. Solmi M, Radua J, Olivola M, Croce E, Soardo L, Salazar de Pablo G, et al. Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. *Mol Psychiatry.* (2022) 27:281–95. doi: 10.1038/s41380-021-01161-7
4. Fusar-Poli P, Salazar de Pablo G, De Micheli A, Nieman DH, Correll CU, Kessing LV, et al. What is good mental health? A scoping review. *Eur*

Neuropsychopharmacol. (2020) 31:33–46. doi: 10.1016/j.euroneuro.2019.12.105
5. Salazar de Pablo G, De Micheli A, Nieman DH, Correll CU, Kessing LV, Pfennig A, et al. Universal and selective interventions to promote good mental health in young people: systematic review and meta-analysis. *Eur Neuropsychopharmacol.* (2020) 41:28–39. doi: 10.1016/j.euroneuro.2020.10.007
6. Wang PS, Angermeyer M, Borges G, Bruffaerts R, Tat Chiu W, De Girolamo G, et al. Delay and failure in treatment seeking after first onset of mental disorders in the World Health Organization's world mental health survey initiative. *World Psychiatry.* (2007) 6:177–85.
7. Dell'Osso B, Glick ID, Baldwin DS, Altamura AC. Can long-term outcomes be improved by shortening the duration of untreated illness in psychiatric disorders? A conceptual framework. *Psychopathology.* (2013) 46:14–21. doi: 10.1159/000338608
8. Membride H. Mental health: early intervention and prevention in children and young people. *Br J Nurs.* (2016) 25:552–4. doi: 10.12968/bjon.2016.25.10.552
9. Wang PS, Berglund P, Olfson M, Pincus HA, Wells KB, Kessler RC. Failure and delay in initial treatment contact after first onset of mental disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry.* (2005) 62:603–13. doi: 10.1001/archpsyc.62.6.603
10. Wang PS, Berglund PA, Olfson M, Kessler RC. Delays in initial treatment contact after first onset of a mental disorder. *Health Serv Res.* (2004) 39:393–415. doi: 10.1111/j.1475-6773.2004.00234.x
11. Anderson KK, Fuhrer R, Malla AK. The pathways to mental health care of first-episode psychosis patients: a systematic review. *Psychol Med.* (2010) 40:1585–97. doi: 10.1017/S0033291710000371
12. MacDonald K, Fainman-Adelman N, Anderson KK, Iyer SN. Pathways to mental health services for young people: a systematic review. *Soc Psychiatry Psychiatr Epidemiol.* (2018) 53:1005–38. doi: 10.1007/s00127-018-1578-y
13. MacDonald K, Ferrari M, Fainman-Adelman N, Iyer SN. Experiences of pathways to mental health services for young people and their carers: a qualitative meta-synthesis review. Soc Psychiatry Psychiatr Epidemiol. 2021 Mar;56(3):339-361. *Erratum in Soc Psychiatry Psychiatr Epidemiol.* (2020) 24:339–61. doi: 10.1007/s00127-020-01976-9
14. Schnyder N, Panczak R, Groth N, Schultze-Lutter F. Association between mental health-related stigma and active help-seeking: systematic review and meta-analysis. Br J Psychiatry. 2017 Apr;210(4):261-268. *Erratum in Br J Psychiatry.* (2017) 211:184. doi: 10.1192/bjp.bp.116.189464
15. Kaushik A, Kostaki E, Kyriakopoulos M. The stigma of mental illness in children and adolescents: a systematic review. *Psychiatry Res.* (2016) 243:469–94. doi: 10.1016/j.psychres.2016.04.042
16. Fox S. *The Social Life of Health Information*. Washington, DC: Pew Research Center (2014).
17. Berger M, Wagner TH, Baker LC. Internet use and stigmatized illness. *Soc Sci Med.* (2005) 61:1821–7. doi: 10.1016/j.socscimed.2005.03.025
18. Burns JM, Davenport TA, Durkin LA, Luscombe GM, Hickie IB. The Internet as a setting for mental health service utilisation by young people. *Med J Aust.* (2010) 192(Suppl. 11):S22–6. doi: 10.5694/j.1326-5377.2010.tb03688.x
19. Birnbaum ML, Rizvi AF, Faber K, Addington J, Correll CU, Gerber C, et al. Digital trajectories to care in first-episode psychosis. *Psychiatr Serv.* (2018) 69:1259–63. doi: 10.1176/appi.ps.201800180
20. Birnbaum ML, Rizvi AF, Confino J, Correll CU, Kane JM. Role of social media and the Internet in pathways to care for adolescents and young adults with psychotic disorders and non-psychotic mood disorders. Early Interv Psychiatry. 2017 Aug;11(4):290-295. *Erratum in Early Interv Psychiatry.* (2017) 11:539. doi: 10.1111/eip.12237
21. Van Meter AR, Birnbaum ML, Rizvi A, Kane JM. Online help-seeking prior to diagnosis: can web-based resources reduce the duration of untreated mood disorders in young people? *J Affect Disord.* (2019) 252:130–4. doi: 10.1016/j.jad.2019.04.019
22. Murphy JM, Nguyen T, Lucke C, Chiang C, Plasencia N, Jellinek M. Adolescent self-screening for mental health problems; demonstration of an internet-based approach. *Acad Pediatr.* (2018) 18:59–65. doi: 10.1016/j.acap.2017.08.013
23. Rosso M, McClelland M, Jansen B, Fleming SW. Using Google AdWords in the MBA MIS course. *J Inf Syst Educ.* (2009) 20:41–9.
24. Birnbaum ML, Garrett C, Baumel A, Scovel M, Rizvi AF, Muscat W, et al. Using digital media advertising in early psychosis intervention. *Psychiatr Serv.* (2017) 68:1144–9. doi: 10.1176/appi.ps.201600571
25. Schuler D, Namioka A(eds). *Participatory Design: Principles and Practices*. Hillsdale, NJ: Routledge, Taylor & Francis group (1993).
26. Larson MK, Walker EF, Compton MT. Early signs, diagnosis and therapeutics of the prodromal phase of schizophrenia and related psychotic disorders. *Expert Rev Neurother.* (2010) 10:1347–59. doi: 10.1586/ern.10.93
27. Auxier B, Anderson B. *Social Media Use in 2021*. Washington, DC: Pew Research Center (2021).
28. Baumel A, Faber K, Mathur N, Kane JM, Muench F. Enlight: a comprehensive quality and therapeutic potential evaluation tool for mobile and web-based eHealth interventions. *J Med Internet Res.* (2017) 19:e82. doi: 10.2196/jmir.7270
29. Miller TJ, Cicchetti D, Markovich PJ, McGlashan TH, Woods SW. The SIPS screen: a brief self-report screen to detect the schizophrenia prodrome. *Schizophr. Res.* (2004) 70:78.
30. Levis B, Sun Y, He C, Wu Y, Krishnan A, Bhandari PM, et al. Accuracy of the PHQ-2 alone and in combination with the PHQ-9 for screening to detect major depression: systematic review and meta-analysis. *JAMA.* (2020) 323:2290–300. doi: 10.1001/jama.2020.6504
31. Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. *Ann Intern Med.* (2007) 146:317–25. doi: 10.7326/0003-4819-146-5-200703060-00004
32. Economic Research Service (ERS). *U.S. Department of Agriculture. Food Environment Atlas [Internet]. Food Access Research Atlas*. (2019). Available online at: https://www.ers.usda.gov/data-products/food-environment-atlas/ (accessed February 22, 2022).
33. Nyc Mayor's Office for Economic Opportunity. *New York City Government Poverty Measure 2019 [Internet]. Poverty Measure*. (2019). Available online at: https://www1.nyc.gov/site/opportunity/poverty-in-nyc/poverty-measure.page (accessed February 25, 2022).
34. Joa I, Johannessen JO, Auestad B, Friis S, McGlashan T, Melle I, et al. The key to reducing duration of untreated first psychosis: information campaigns. *Schizophrenia Bull.* (2008) 34:466–72. doi: 10.1093/schbul/sbm095
35. Melle I, Johannesen JO, Friis S, Haahr U, Joa I, Larsen TK, et al. Early detection of the first episode of schizophrenia and suicidal behavior. *Am J Psychiatry.* (2006) 163:800–4. doi: 10.1176/ajp.2006.163.5.800
36. Chong SA, Mythily S, Verma S. Reducing the duration of untreated psychosis and changing help-seeking behaviour in Singapore. *Soc Psychiatry Psychiatr Epidemiol.* (2005) 40:619–21. doi: 10.1007/s00127-005-0948-4
37. Srihari VH, Tek C, Pollard J, Zimmet S, Keat J, Cahill JD, et al. Reducing the duration of untreated psychosis and its impact in the US: the STEP-ED study. *BMC Psychiatry.* (2014) 14:335. doi: 10.1186/s12888-014-0335-3
38. Ghio L, Gotelli S, Marcenaro M, Amore M, Natta W. Duration of untreated illness and outcomes in unipolar depression: a systematic review and meta-analysis. *J Affect Disord.* (2014) 152–154:45–51. doi: 10.1016/j.jad.2013.10.002
39. Baghaei N, Naslund JA, Hach S, Liang HN. Editorial: designing technologies for youth mental health. *Front Public Health.* (2020) 8:45. doi: 10.3389/fpubh.2020.00045
40. 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) 214:269–72. doi: 10.1192/bjp.2019.8
41. Muñoz RF, Chavira DA, Himle JA, Koerner K, Muroff J, Reynolds J, et al. Digital apothecaries: a vision for making health care interventions accessible worldwide. *Mhealth.* (2018) 4:18. doi: 10.21037/mhealth.2018.05.04
42. Xanthopoulou P, Thomas C, Dooley J. Subjective experiences of the first response to mental health crises in the community: a qualitative systematic review. *BMJ Open.* (2022) 12:e055393. doi: 10.1136/bmjopen-2021-055393
43. Thabrew H, Goodyear-Smith F. Youth version, case-finding and help assessment tool (YouthCHAT): an electronic, psychosocial screener for young people with long-term physical conditions. *J Paediatr Child Health.* (2020) 56:102–6. doi: 10.1111/jpc.14503
44. Norman RM, Windell D, Manchanda R. Examining differences in the stigma of depression and schizophrenia. *Int J Soc Psychiatry.* (2012) 58:69–78. doi: 10.1177/0020764010387062

45. Angermeyer MC, Matschinger H. Public beliefs about schizophrenia and depression: similarities and differences. *Soc Psychiatry Psychiatr Epidemiol.* (2003) 38:526–34. doi: 10.1007/s00127-003-0676-6
46. Takao Y, Figueroa E, Berna KFJ, Jo Y, Kissane LA, Yoshimura K, et al. Validation of a novel online depression symptom severity rating scale: the R8 Depression. *Health Qual Life Outcomes.* (2021) 19:163. doi: 10.1186/s12955-020-01654-z
47. Handal N, LePage J, Dayley P, Baldwin B, Roeser A, Kay J, et al. Validation, reliability, and specificity of CliniCom^{TM} psychiatric assessment software. *Psychiatry Res.* (2018) 265:334–40. doi: 10.1016/j.psychres.2018.05.029
48. Gronholm PC, Thornicroft G, Laurens KR, Evans-Lacko S. Mental health-related stigma and pathways to care for people at risk of psychotic disorders or experiencing first-episode psychosis: a systematic review. *Psychol Med.* (2017) 47:1867–79. doi: 10.1017/S0033291717000344
49. Pretorius C, Chambers D, Coyle D. Young people's online help-seeking and mental health difficulties: systematic narrative review. *J Med Internet Res.* (2019) 21:e13873. doi: 10.2196/13873
50. Alfonsson S, Maathz P, Hursti T. Interformat reliability of digital psychiatric self-report questionnaires: a systematic review. *J Med Internet Res.* (2014) 16:e268. doi: 10.2196/jmir.3395
51. van Ballegooijen W, Riper H, Cuijpers P, van Oppen P, Smit JH. Validation of online psychometric instruments for common mental health disorders: a systematic review. *BMC Psychiatry.* (2016) 16:45. doi: 10.1186/s12888-016-0735-7
52. Zhang CQ, Zhang R, Schwarzer R, Hagger MS. A meta-analysis of the health action process approach. *Health Psychol.* (2019) 38:623–37. doi: 10.1037/hea0000728
53. Compton MT, Shim RS. The social determinants of mental health. *Focus.* (2015) 13:419–25. doi: 10.1097/NCM.0000000000000518
54. Karyotaki E, Efthimiou O, Miguel C, Bermpohl FMG, Furukawa TA, Cuijpers P, et al. Internet-based cognitive behavioral therapy for depression: a systematic review and individual patient data network meta-analysis. *JAMA Psychiatry.* (2021) 78:361–71. doi: 10.1001/jamapsychiatry.2020.4364
55. Staples LG, Dear BF, Johnson B, Fogliati V, Gandy M, Fogliati R, et al. Internet-delivered treatment for young adults with anxiety and depression: evaluation in routine clinical care and comparison with research trial outcomes. *J Affect Disord.* (2019) 256:103–9. doi: 10.1016/j.jad.2019.05.058
56. van Doorn M, Nijhuis LA, Egeler MD, Daams JG, Popma A, van Amelsvoort T, et al. Online indicated preventive mental health interventions for youth: a scoping review. *Front Psychiatry.* (2021) 12:580843. doi: 10.3389/fpsyt.2021.580843

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

frontiers | Frontiers in Psychiatry

ORIGINAL RESEARCH
published: 24 June 2022
doi: 10.3389/fpsyt.2022.880972

Building Emotional Awareness and Mental Health (BEAM): A Pilot Randomized Controlled Trial of an App-Based Program for Mothers of Toddlers

Anna L. MacKinnon[1], Kaeley M. Simpson[2], Marlee R. Salisbury[3], Janelle Bobula[2], Lara Penner-Goeke[2], Lindsay Berard[2], Charlie Rioux[1,2], Gerald F. Giesbrecht[4], Ryan Giuliano[2], Catherine Lebel[5], Jennifer L. P. Protudjer[6], Kristin Reynolds[2], Shannon Sauer-Zavala[7], Melanie Soderstrom[2], Lianne M. Tomfohr-Madsen[1†] and Leslie E. Roos[2†]*

[1] Department of Psychology, University of Calgary, Calgary, AB, Canada, [2] Department of Psychology, University of Manitoba, Winnipeg, MB, Canada, [3] Department of Psychology, York University, Toronto, ON, Canada, [4] Department of Pediatrics, University of Calgary, Calgary, AB, Canada, [5] Department of Radiology, University of Calgary, Calgary, AB, Canada, [6] Department of Pediatrics and Child Health, University of Manitoba, Winnipeg, MB, Canada, [7] Department of Psychology, University of Kentucky, Lexington, KY, United States

OPEN ACCESS

Edited by:
Patricia A. Arean,
University of Washington,
United States

Reviewed by:
Pablo Martínez Díaz,
University of Chile, Chile
J. Carola Pérez,
Universidad del Desarrollo, Chile

***Correspondence:**
Leslie E. Roos
leslie.roos@umanitoba.ca

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Digital Mental Health, a section of the journal Frontiers in Psychiatry

Received: 22 February 2022
Accepted: 03 June 2022
Published: 24 June 2022

Citation:
MacKinnon AL, Simpson KM, Salisbury MR, Bobula J, Penner-Goeke L, Berard L, Rioux C, Giesbrecht GF, Giuliano R, Lebel C, Protudjer JLP, Reynolds K, Sauer-Zavala S, Soderstrom M, Tomfohr-Madsen LM and Roos LE (2022) Building Emotional Awareness and Mental Health (BEAM): A Pilot Randomized Controlled Trial of an App-Based Program for Mothers of Toddlers. Front. Psychiatry 13:880972. doi: 10.3389/fpsyt.2022.880972

Background: Families have faced unprecedented challenges during the COVID-19 pandemic, leading to increased maternal mental health problems and barriers to accessing care. Innovative programs are needed to support both maternal mental health and parenting, and to buffer the long-term impacts of stress on young children. Using a patient-oriented approach, our research team aimed to co-develop and pilot test an App-based psychoeducation and social-connection platform: Building Emotional Awareness and Mental Health (BEAM).

Methods: The co-development process involved a parent advisory board from conceptualization and design, through to direct participation in the program delivery. The BEAM program includes weekly videos and activities based on Unified Protocol therapy modules and emotion-focused parenting strategies, a weekly telehealth group review session, and access to a private online forum for support from other mothers and clinical coaches. A parallel randomized control trial was conducted across two provinces in Canada. Mothers of preschool children (aged 18–36 months old), with moderate-to-severe depression (Patient Health Questionaire-9 $\geq$ 10), were recruited online and randomized to either the 10-week BEAM intervention or treatment as usual (TAU) control group. Online surveys (ensuring researcher blinding) included questions about feasibility and acceptability of the program and pre/post self-report measures of mental health, parenting, positive coping and child behavior outcomes. The primary outcome measures were symptoms of depression and parenting stress. Data were analyzed using mixed models and an intention-to-treat approach.

Results: 65 participants were randomized, by an online allocation tool, to the BEAM (n = 33) and TAU (n = 32) groups. Engagement was relatively high at the beginning of the program, with 78.8% starting the BEAM App and 70.6% attending $\geq$1 telehealth session. Most respondents felt socially supported, satisfied with the App, and found

it easy to use. Pre-post results indicated interaction effects with greater reductions in overall mental health problems, and specifically anxiety and sleep symptoms, among BEAM vs. control participants. There were also time effects with reductions in depression symptoms across both groups. No significant treatment effects emerged for the other mental health symptoms, parenting problems, positive coping, or child behavior outcomes. Descriptive data are included to highlight possible areas of promise for future large efficacy trials. Technological difficulties and other challenges that may have led to attrition and impacted outcomes are discussed. There were no adverse events related to study participation.

Conclusions: The BEAM program has promise as a novel, feasible and acceptable intervention for improving mental health among mothers of young children.

Clinical Trial Registration: [www.ClinicalTrials.gov], identifier [NCT04772677].

Keywords: digital health (eHealth), maternal mental health, randomized controlled trial, parenting, emotion regulation

INTRODUCTION

Families have been facing unprecedented challenges during the COVID-19 pandemic, which have led to dramatic increases in maternal mental health problems. A recent meta-analysis indicated that 25–30% of mothers have experienced clinically significant symptoms of depression and anxiety (1). Additional pandemic-related stressors have been identified as key risk factors for families, including isolation, domestic conflict, and a lack of parenting support (2, 3). The absence of standard screening and a backlog at existing services means that the majority of parents will not get access to evidence-based treatments (4, 5). In Canada, only 1 in 10 mothers receive the postpartum mental health care they need (6), resulting in symptom persistence, child mental health problems, developmental impairments and enormous economic impact (7–9). Accessible and scalable programs are urgently needed to build resilience to ongoing stressors and prevent the intergenerational transmission of mental illness to children exposed to maternal mental health problems during the COVID-19 pandemic. Digital mental health interventions have potential for addressing these needs and barriers for families' care.

Children's development exhibits high environmental sensitivity in the first 3 years of life, with maternal depression linked to irritable temperament, sleep problems, and socio-emotional impairments later in childhood and adolescence (10–12). Although shared genetic lability may increase children's likelihood of experiencing mental health problems, environmental factors are understood to moderate this such that risk for long-term mental illness is particularly heightened when children are exposed to persistent maternal depression and other family stressors (13, 14). Social mechanisms through which maternal depression may impact child mental health include low maternal sensitivity (15), maternal modeling of poor emotion coping (16), and negative parenting (e.g., harsh or punitive parenting practices) (17). The parenting stress associated with maternal depression is also established to lead to low-quality interactions and harsh discipline (18), which is particularly concerning in the pandemic context in which children are spending significant time at home.

Although evidence-based treatments exist to address maternal mental illness, there are significant barriers to accessing care, particularly due to the pandemic and its associated public health measures. These include restricted access to mental health clinicians, physical distancing, high costs of individual therapy, closure of existing services, and overwhelming childcare demands (19–21). In our previous research, approximately 20% of mothers with depression during the pandemic reported accessing services (2). Digital interventions provide an accessible, low-cost way to address these barriers and needs for care. Early models of digital mental health interventions show promise for treating adult depression, however dropout rates are high and these programs rarely include parenting-related content and support (22, 23). For example, a meta-analysis of MoodGYM, a web-based program for depression and anxiety, indicated that user completion may be as low as 10% (24). A recent meta-analysis by our team suggested that digital interventions (including videoconferencing, App-based, and Web-based) targeting parent mental health, parenting skills or behavior were associated with improvements in a range of symptoms among parents of children aged 1–5 years old, with effects sizes comparable to in-person interventions (25). Other research highlights the promise of delivering mental health and parenting services through mobile applications with studies revealing App-based intervention programs can improve parent-child interactions (26), mental well-being, sleep, and resilience, and decrease anxiety (27). However, there are few digital programs that target both mental health and parenting skills. One pragmatic RCT of a 4-month nurse moderated App-based program for new mothers found no statistically significant differences compared to care as usual in depression symptoms or parenting problems, despite reporting high engagement and satisfaction (28). While early intervention is critical, the postpartum period has unique challenges and care provisions, thus treatment that extends to early childhood is needed. Addressing intergenerational mental

health concerns will require innovative program design methods to simultaneously treat maternal mental illness and address parenting risks—the theorized transmission pathways or causal mechanisms (29).

Building Emotional Awareness and Mental Health Program Development

The need for scalable online programming became clear early in the COVID-19 pandemic. We and others reported concerningly high levels of maternal mental health problems, which were associated with negative parenting practices (e.g., less responsive parenting, more harsh discipline) (2).

In May 2020, we began consulting with our parent advisory board (10 mothers with lived experience managing depression) by asking the simple question "*What would be helpful for supporting your mental health?*" Their feedback included a request for online support that would allow for connection to other mothers *via* forums or brief (1 h or less) group discussions, contact with therapists, and access to evidenced-based mental health and parenting strategies delivered in short-informational videos. These requests aligned with our knowledge synthesis work highlighting the importance of therapist contact (vs. only didactic materials) to symptom improvement in digital programs (30). Similarly, our previous qualitative research on parenting forums (1,000 posts) during the pandemic highlighted requests for strategies to support family mental health and to feel less isolated (31).

Based on this feedback, and ongoing input from our parent advisory board, we designed the Building Emotional Awareness & Mental Health (BEAM) Program to reduce mental health problems and parenting stress in a scalable, App-based format. BEAM integrates evidence-informed psychoeducation and group therapy with matched emotion-focused parenting skills. A closed-group online forum was also included to promote social support. We chose to use a transdiagnostic treatment approach (i.e., Unified Protocol) to address mental health symptoms characteristic of emotion disorders alongside emotion-focused parenting strategies.

The Unified Protocol is an effective transdiagnostic emotion-focused cognitive-behavioral therapy for emotional disorders in adults (32–34). Unified Protocol targets the underlying processes of aversive/avoidance reactions to emotions and consists of five core modules targeting emotion regulation: mindful emotion awareness; building cognitive flexibility; identifying and preventing patterns of emotion avoidance and emotion-driven behavior; awareness and tolerance of emotion-related physical sensations; and interoceptive and situation-based emotion-focused exposures (35). A recent meta-analysis indicated that Unified Protocol is associated with significant improvement in symptoms of anxiety and depression, in on-site and online formats (36).

Emotion-focused parenting approaches help parents observe and validate children's emotional reactions, increase communication, and provide support without escalating the situation (37). There are three key processes through which parents help their children learn to regulate emotions: parental reactions to child emotions, talking about emotions, and emotional expressiveness (38). Emotion socialization strategies are demonstrated to be effective for increasing children's emotional knowledge and reducing challenging emotions (39). This emotion-focused parenting approach was designed to interrupt some of the social mechanisms of parenting risk (e.g., emotional socialization, harsh parenting) through which parent mental health problems are understood to impact child mental health.

BEAM is unique from existing digital programs in that it simultaneously targets both maternal mental health and parenting skills through a therapist-led psychoeducation protocol, combined with a social connection platform. These interrelated skills are designed to reduce maternal mental health problems and synergistically increase supportive parenting behaviors in order to promote family relationships and disrupt the intergenerational transmission of mental illness.

Current Study

The pilot study had two main objectives: (1) Determine the acceptability and feasibility of version 1.0 of BEAM, and (2) Assess the initial evidence of BEAM on maternal mental health, parenting, and family function outcomes. Although BEAM is designed to target a range of emotion-based parent mental health problems, we chose to recruit based on a single diagnostic criterion (depression) in an effort to have a more homogenous clinical sample. We hypothesized that participation in BEAM would be associated with improvements in mental health and parenting outcomes including reductions in maternal depression symptomatology and parenting stress (primary outcomes) as well as a range of other mental health symptoms (anxiety, sleep problems, anger, alcohol use) and family function (i.e., parenting, coping, child behavior problems; secondary outcomes). The overall aim of this pilot randomized control trial was to use the findings and feedback to inform improvements to BEAM as an accessible and scalable program with potential for rapid, widespread dissemination.

MATERIALS AND METHODS

The current investigation represents the second arm of a larger Phase II (preliminary testing) pilot study, following the ORBIT model for developing behavioral treatments (40). See the full study protocol and trial registration on ClinicalTrials.gov (Identifier: NCT04772677). Ethics approval was obtained from the Psychology/Sociology Research Ethics Board (P2020:081) at the University of Manitoba and the Conjoint Health Research Ethics Board (REB20-1933) at the University of Calgary. All study procedures were conducted with the electronic informed consent of participants.

Trial Design

The current investigation comprised a pilot randomized controlled trial with parallel assignment to the BEAM intervention or treatment as usual (TAU) control group.

Study advertising provided a link to assess eligibility for the trial *via* online screener. Those who met inclusion criteria (listed below) were contacted *via* email with a brief description of the study and to obtain confirmation of their availability for telehealth sessions and time commitment to participate in the trial. Those who confirmed were enrolled in the trial and completed a series of self-report questionnaires before randomization (T1 = pre-intervention assessment). Participants were subsequently randomized to the intervention (BEAM App program) or control (TAU) groups in a parallel 1:1 ratio by computer generated sequencing using an online tool[1] conducted by a non-affiliate research assistant to conceal allocation. Participants randomized to the BEAM intervention group were asked to register accounts (i.e., create anonymous usernames and passwords) for the BEAM App and to complete their profile in Jane (Jane Software Inc.), a secure online platform used for charting clinical contact. All participants were asked to complete a battery of self-report questionnaires after the 10-week program ended (T2 = post-intervention assessment). Participants randomized to the BEAM App program treatment group were also asked to complete questions assessing feasibility and acceptability at T2.

Participants

Inclusion criteria were being an adult (aged 18 years or older) mother or other primary caregiver who identify as a woman (e.g., grandmother, aunt) of a child aged 18–36 months old, experiencing moderate to severe depression (Patient Health Questionnaire (PHQ-9) score $\geq$ 10), living in Alberta or Manitoba, comfortable understanding, speaking and reading English, and available for weekly telehealth sessions (*via* Zoom). Potential participants were excluded if they had significant suicidal ideation, a history of attempted suicide in the past year or self-harm in the past 6 months.

Participants were recruited in the Canadian provinces of Manitoba and Alberta, which both have access to public health care and various social services for families. Online advertisements were used including paid social media posts (e.g., Facebook, Instagram) and postings by community partner agencies *via* electronic mailing lists or public announcements.

A priori power was challenging to determine given limited research on eHealth programs for parents of young children and unknown trial aspects such as recruitment feasibility and attrition. Accordingly, we aimed to recruit a sample for one large telehealth group (including breakout rooms) and control group in a set amount of time (May–June 2021). Given participants were being randomized into a group intervention, we were obligated to have a clear start date for the program.

Intervention

The BEAM program is a novel 10-week App-based digital intervention that combines maternal mental health treatment and parenting skills training with clinician-facilitated peer support and social connection. The primary aim of the program is to improve symptoms of depression and promote a positive parent-child relationship.

There are five main components of the program (see **Figure 1**): (1) weekly expert-led psychoeducation videos (5–15 min) using (a) adapted Unified Protocol therapy modules (35), which target maternal mental health symptomology, and (b) emotion-focused parenting skills modules (37), which were designed to correspond to the Unified Protocol modules and promote maternal responsivity to children's emotions (see **Appendix Table 1** for weekly module content); (2) a monitored closed group online forum with reflection activities and open discussion to encourage social support; (3) weekly 1-h structured telehealth group sessions (*via* Zoom for Healthcare) to review program content and connect with other participants (41, 42); (4) participants are encouraged to complete weekly activities (i.e., homework) based on the mental health and parenting modules, such as worksheets, reflections, practice exercises and strategies; and (5) participants are also asked to complete a brief weekly survey measuring symptoms of depression and parenting stress. Due to budgetary and technological difficulties (discussed in section "App Interface"), interactive activities and symptom monitoring were not included within the App interface. Therefore, participants were invited to share about activities on the forum and during the telehealth group sessions, and to track their own progress (43) by recording their symptom scores from the weekly survey which was administered externally *via* RedCap link.

A mental health therapist (with a Masters or PhD in clinical psychology, and supervised by a registered clinical psychologist) and parent coach (with lived experience of participating in a similar program) facilitated the forum discussions and telehealth group sessions to increase support, and in turn accountability, consistent with evidence that therapist guidance on eHealth interventions is more effective than self-directed only programs (44). Trial therapists also contacted participants by phone or email if they requested an individual follow-up (e.g., questions about the material, partner or family conflict) or if weekly symptom surveys indicated potential suicidal ideation, for which a risk assessment was conducted and crisis services recommended if necessary. In addition to opportunities for peer support through the forum and telehealth group, mothers were encouraged to identify and engage a person to support their participation in the program (e.g., partner, friend, family member). Research suggests social support is associated with improved adherence and response to psychological interventions (42, 45–47).

The BEAM program was delivered *via* mobile application (accessible by Android and iOS devices). The research team co-developed the evidenced-based content and protocol, while working closely with a digital media company to build the BEAM App. The research team met regularly with the digital media company to determine the App design and function, coordinated extensive user acceptance testing with the parent advisory board for quality assurance, and communicated as needed regarding any necessary updates or repairs.

The control arm of the study was designed to account for the potential effects of time and regular care on change in

[1]https://www.randomlists.com/team-generator

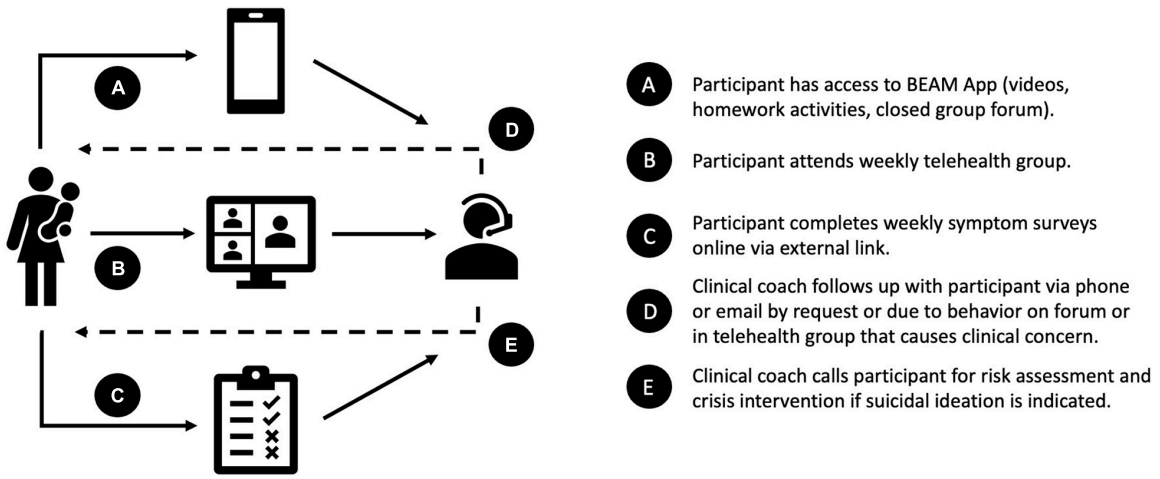


FIGURE 1 | BEAM program components.

psychological distress (40). Given ethical considerations to not withhold treatment from mothers experiencing active distress, the TAU group was encouraged to access parenting and mental health resources available in their community. The TAU group was also sent weekly parenting stress and depression symptom surveys to account for change over time; although their scores were displayed on screen immediately following completion, they did not receive instructions to note them or track progress.

All participants in the trial also received an (online) information pamphlet about local parenting and mental health resources (e.g., counseling centers, crisis lines, websites). Data were collected from all participants during the online screener regarding current psychiatric medications and mental health service use in the previous month (e.g., counseling, crisis lines, website).

Outcome Measures

Feasibility and Acceptability

Measures of recruitment, enrollment, and retention were included to assess interest in the overall program and acceptability of run-in procedures.

A narrative description of the different steps and processes involved in developing the digital BEAM program is provided, including considerations and challenges related to building the App interface, hosting platform and data storage, data tracking, and participant communication.

We intended to collect measures of App-based engagement (i.e., log-ins, time spent on app, forum activity, telehealth sessions attended) through Google Analytics and Firebase, but these data were lost due to technical challenges (see "Results" section). In addition, several questions were developed for the pilot study to measure engagement in different components of the program, including videos ("*Have you ever watched a video on the app?*" if yes, "*How many videos did you watch?*"), forum ("*Did you ever participate in the forums?*" if yes, "*How often did you use the forums?*"), and telehealth groups ("*Did you ever participate in a zoom telehealth group?*" if yes, "*How often did you participate in a zoom telehealth group?*"). Intervention group participants were also asked if "*The BEAM program was a good source of social support*" on a 6-point Likert scale from 1 (*Strongly disagree*) to 6 (*Strongly agree*).

The mHealth App Usability Questionnaire (MAUQ) was used to assess participants experience with the BEAM App (48). The MAUQ comprises three subscales that rate ease of use (5 items), interface and satisfaction (7 items), and usefulness (6 items) on a 7-point Likert scale from 1 (*Disagree)* to 7 (*Agree*), where high scores indicate better useability. The subscales demonstrated high internal consistencies in the current sample, with Cronbach α ranging from 0.89 to 0.93.

Primary

Change in Maternal Depression was measured using the Patient Health Questionnaire (PHQ-9) (49). This 9-item questionnaire assesses depression presence and severity (mild to severe), where scores $\geq$10 indicate clinically significant depression and a 5-point reduction represents clinically significant change (50). The PHQ-9 demonstrated good reliability in the current sample, with an internal consistency Cronbach α of 0.68 at T1 and 0.90 at T2.

Change in Parenting Stress was measured using the Parenting Stress Index – Short Form (PSI-SF) (51). This 36-items scale assesses the presence of difficult child behaviors and whether they were stressful for parents, where scores $\geq$90 indicate clinically significant levels. For the purposes of the current study, a 5-point reduction represents clinically significant change, as this was the approximate mean difference observed for a digital parent training intervention with children aged 2–5 years old (52). The PSI-SF demonstrated high reliability in the current sample, with internal consistency Cronbach α of 0.91 at T1 and 0.92 at T2.

Although we were interested in using the questionnaires as a weekly measure of depression and parenting stress the response rates were very low <20%, so data was not considered informative for the present trial, beyond the need to increase feasibility of mood tracking in future iterations.

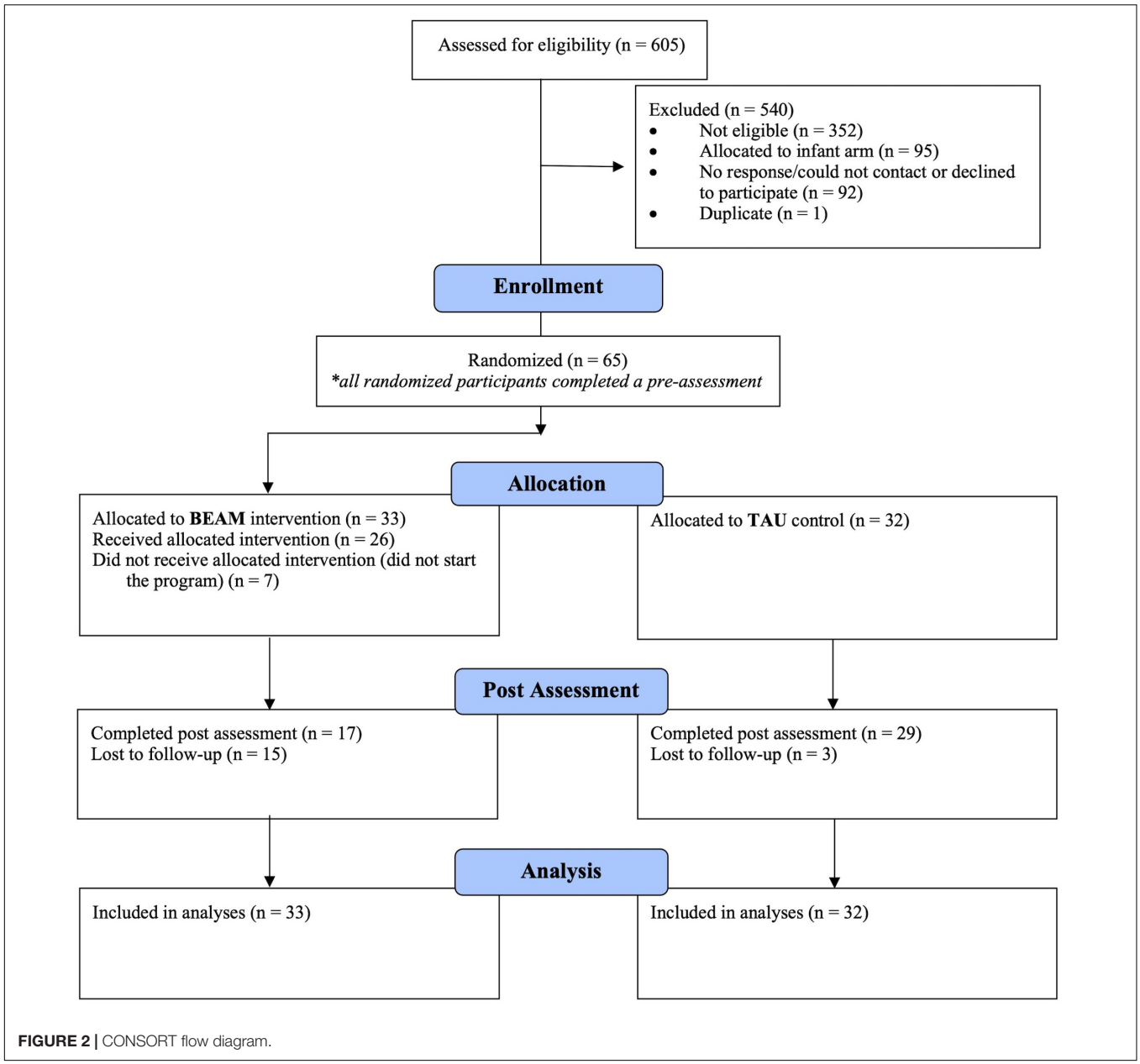


FIGURE 2 | CONSORT flow diagram.

Secondary

Change in other maternal mental health symptoms was assessed using well-validated self-report measures. Anxiety was measured using the 7-item Generalized Anxiety Disorder scale (GAD-7), where scores ≥10 indicate clinically significant symptoms (53), and a 4 point reduction represents clinically significant change (54). The GAD-7 demonstrated good reliability in the current sample, with an internal consistency Cronbach α of 0.83 at T1 and 0.89 at T2. Anger was measured using the 5-item Patient-Reported Outcomes Measurement Information System (PROMIS) Short Form v1.1-Anger 5a (55), which demonstrated good reliability in the current sample with an internal consistency Cronbach α of 0.80 at T1 and 0.91 at T2. Sleep problems were measured using the PROMIS Sleep Disturbance—Short Form 8a (56), which demonstrated good reliability in the current sample with an internal consistency Cronbach α of 0.84 at T1 and 0.87 at T2. For the PROMIS scales, T-scores of 60.0–69.9 and ≥70 indicate moderate and severe elevated problems, respectively (57), and a 2-6 point reduction in T-scores represents minimal important change (58). Substance use was measured with the 10-item Alcohol Use Disorders Identification Test (AUDIT) (59), which demonstrated good reliability in the current sample with an internal consistency Cronbach α of 0.81 at T1 and 0.83 at T2. For the purposes of the current study, scores ≥6 indicate hazardous use among women and a 5 point reduction represents clinically significant change, as these were the cut-off and approximate mean difference observed for a digital self-help intervention (60).

Change in coping abilities was assessed using the 12-item Self Compassion Scale-Short Form (SCS-SF; Cronbach α of 0.82 at T1 and 0.86 at T2) (61), the 7-item Recent Stressful Events (RSE) hopeful coping scale (Cronbach α of 0.73 at T1 and 0.75 at T2 for the 5 Likert scale items), which was developed based on recommendations from JPB Research Network on Toxic Stress at Harvard's Center on the Developing Child (62), and the 4-item Couple's satisfaction index (CSI-4; Cronbach α of 0.95 at T1 and 0.97 at T2) (63).

Change in parenting problems was measured using the 30-item Parenting Scale (PS) which includes lax, overreactive and verbosity discipline styles (64), as well as the 21-item Parenting Young Children (PARYC), which includes setting limits, proactive, and supportive behaviors (65). The PS and PARYC demonstrated good internal consistency in the current sample, with Cronbach α of 0.89 and 0.91 at T1 and 0.86 and 0.86 at T2, respectively.

Change in child internalizing and externalizing behavior problems were assessed by mother report using the 99-item Child Behavior Checklist (CBCL) for ages 1.5–5 years (66). The CBCL demonstrated excellent reliability in the current sample, with an internal consistency Cronbach α of 0.95 at T1 and 0.96 at T2.

Analytic Approach

All analyses were conducted using SPSS 26.0. Descriptive statistics were computed for the total sample and between groups for demographics and outcome measures at baseline. Longitudinal analysis of covariance, using linear mixed modeling, was conducted in order to test treatment effects on outcome measures. Treatment effects were tested by entering two-way interactions between time and group assignment. An α of 0.05 was used to determine statistical significance. Cohen's d effect sizes were derived based on recommendations for linear mixed models (67), where 0.20, 0.50, and 0.80 were interpreted as indicating small, medium, and large effects (68). An intent-to-treat (ITT) approach was used, consistent with CONSORT guidelines (i.e., all participants who were randomized to receive enrollment information for the intervention were included in analyses) (69), and missing data was handled using maximum likelihood estimation (70).

Aggregate variables were computed for mental health symptoms (PHQ-9, GAD-7, PROMIS anger and sleep scales, AUDIT), parenting problems (PSI, PS, PAYRC), and positive coping (SCS, CSI, RSE), by converting each measure to a standardized z-score then taking their average. Moderate internal consistency was observed between the standardized scores for mental health symptoms (Cronbach α = 0.68 at T1 and 0.75 at T2), for parenting problems (Cronbach α = 0.60 at T1 and 0.65 at T2), and for positive coping (Cronbach α = 0.50 at T1 and 0.60 at T2).

Exploratory analyses were also conducted to assess clinically significant change in mental health symptoms by creating binary variables for participants in each condition to code whether they achieved the minimum point reduction and/or scored below the clinical cut-offs on the primary and secondary outcome measures at T2. Fisher's exact tests were used to compare the proportion of participants who exhibited clinically significant change or reached a score below the cut-off.

RESULTS

Participant Flow and Recruitment

Participant enrolment, allocation, and retention information is provided in the CONSORT flow diagram (**Figure 2**). Recruitment ran from May 5 to June 16, 2021. Over 600 individuals completed the eligibility screener for both arms of the larger pilot study. Out of the 158 who met inclusion criteria for the preschool arm, 65 (41.1%) were enrolled and completed the T1 assessment, then subsequently randomized to the BEAM program (n = 33) or TAU (n = 32) groups. Of those participants randomized to the intervention, 78.8% started the BEAM App program (i.e., created a user accounts). The program began on July 5, 2021, with summer break from August 12 to September 8, 2021, and ended on October 13, 2021. T2 post assessment were completed from October 20 to November 22, 2021, and focus groups were conducted from November 9–30, 2022. In terms of attrition, 15 (45.5%) BEAM and 3 (9.4%) TAU participants were lost by the T2 post assessment, a statistically significant difference on Fisher's exact test (p = 0.001).

Feasibility and Acceptability

Given the significant practical and technical challenges encountered, we first describe these challenges in detail before presenting outcome-related findings. Recommended points of consideration and possible solutions for each of these feasibility and acceptability challenges described below are included in the discussion section (**Table 1**).

App Interface

Significant challenges arose in our service use agreement with a digital media company to build the BEAM App. This included challenges in communication and expectations regarding feasibly scope in our budget ($120,000) for App development and maintenance. We estimated this amount to be reasonable for developing a pilot therapy program by traditional academic standards, and based on an initial discussion with App-development. However, the features that we believed to be critical to build based on evidence-based best practices and advisory board input likely required a budget of closer to $500,000. As a result, we were not able to include multiple important features including: App-embedded mood tracking, App-embedded group telehealth videos, or any usage-related tailored push-notifications. Instead of dynamic and interactive videos which we hoped to have tech experts co-develop, we created narrated slide presentations to communicate necessary content (see **Appendix Table 1** for weekly module content). Although the forum provided a platform for participants to connect with each other and therapists anonymously (by using Bitmoji profile pictures and discrete usernames), the advised forum solution of using third party software was extremely limited because it did not integrate well into the App (e.g., separate window opens), users could not respond directly to

TABLE 1 | Lessons learned for facilitating the development of a digital mental health program.

Challenges	Recommendations
Participants were lost if they did not send a username and password for the App	Research team creates login information for participants and send to participants
Delayed start date due to technology complications	Ensure technology is in close-to-ready state prior to initiating participant recruitment
Low engagement throughout the summer	Launch in fall or winter
App functionality issues including: Videos not playing on some devices	Research team member available for tech troubleshooting (e.g., re-download App, restart device, check internet connection)
Low attendance at group telehealth sessions	Advertise telehealth groups as mandatory rather than optional, in screener have participants check off times they could commit to attend group sessions
Difficulty finding time across provinces for telehealth sessions	Offer multiple timeslots for sessions
Too many points of contact for participants	Streamline contact with participants and practice the on-ramping process
Participants missing information due to not checking emails	Send critical information and reminders *via* text and email
Engaging participants in the forum in a meaningful way	Have peer coaches facilitate organic discussions within the forum

comments or tag each other, the threads were difficult to navigate, and is was not visually appealing. We received feedback that small challenges that came up during the pilot trial, such as videos freezing, were not possible to address mid-trial, so we created *ad hoc* solutions such as sending participants YouTube-uploaded videos *via* private channel. The result of these misunderstandings in the App-development costs and process resulted in a limited product with substantial unintended participant burdens due to "friction" from poor ease of use.

Hosting Platform and Data Storage

There were lengthy time delays due to legal contract negotiations and service use agreements with the research institutions and digital technology company, in order to align agreements with university intellectual property policies and ensure the processes and product were compliant with local Personal Health Information Act (PHIA) requirements of possible personal health information disclosed on the forums. The digital media company was not able to offer PHIA-compliant server storage, so the university made an exception to standard policy to allow the app to operate within the high-security University protected data storage systems. The time delays associated with the contracting process spanned from September 2020 to March 2021 with the program testing and implementing lasting an additional 3 months, resulting in participants waiting multiple months prior to receiving services (May 2021) and the program running over the summer, which is an undesirable time to test parenting programs due to non-routine schedules. Many participant reported challenges in engaging with material and attending group during the latter half of the summer due to family vacations. After polling the telehealth groups, we made a decision to take a 4-week vacation from early-August to early-September to account for vacations and re-engage in the early fall.

Data Tracking

Another challenge was difficulty ensuring relevant usage data (e.g., App logins, time spent on App, activity completion) was being captured on the backend of the App (through Firebase and Google Analytics). Unfortunately, usable data was not collected from Firebase as custom events did not capture our desired variables and data from Google Analytics was lost due to a default *data retention period* setting. This may reflect miscommunications given terminology can differ between researchers and tech developers. Before data collection begins, it is important to determine what variables are needed (70, 71), translate them into *events* that can be tracked, and check whether they are tracked as *automatically collected events* in Google Analytics (e.g., total time spent on the App, session start) or need to be implemented as *custom events* during App development (e.g., time spent on different App screens, tracking when video content is started, paused, resumed, watched completely). This should be done as early as possible during App development as some early decisions can affect the custom events that can be collected (e.g., the choice of video player used in App development determines whether the number of minutes a video is watched can be tracked). In addition, the Google Analytics default data retention setting for storing data should be adjusted to the study period and data should be exported regularly to avoid loss. These settings should be set before study onset as they cannot be adjusted retrospectively, and any data deleted before settings are changed will be unrecoverable. We were able to collect telehealth group session attendance *via* the Jane platform and forum participation data *via* the backend of Vanilla Forums.

Participant Communication

Conducting the trial included a significant amount of contact with participants to coordinate enrollment. As the participants were first recruited *via* a public URL leading to our online screener, we emailed eligible participants to confirm their commitment to the program. If participants replied and expressed their commitment, we sent them the link to the online questionnaire and reminders to complete it if needed (i.e., if they had not started or were slow to finish). After randomization, we asked, *via* email, for participants to send us their preferred username, password, contact information, and Bitmoji profile picture for the App and Jane platforms. Many participants required multiple reminders to send us this information. After receiving this information, we manually created accounts for participants across the App, forum, and Jane platforms, and emailed participants with information on downloading and logging into the App. As we registered participants on the app,

TABLE 2 | Engagement and satisfaction measures for the BEAM App program.

Engagement	*n* (%)†	Range
Watched at least 1 video	16 (94.1)	0–16†
Participated in forum	12 (70.6)	0–15
Attended at least 1 telehealth session	12 (70.6)	0–10†
MAUQ	***M* (SD)**	**Range**
Ease of use	24.64 (7.09)	8–35
Interface and satisfaction	33.18 (11.65)	5–49
Usefulness	26.52 (9.38)	6–42

MAUQ, mHealth App Usability Questionnaire.
†*Out of 17 respondents.*

many required assistance troubleshooting technical difficulties (e.g., login issues, videos not playing) *via* email. During the program, weekly email reminders were sent with links for the symptom surveys (RedCap) and telehealth sessions (Zoom). Check-in emails were sent throughout the program to promote attendance at telehealth sessions.

Engagement and Satisfaction

Descriptive statistics for the engagement questions and MAUQ subscales are presented in **Table 2**. Of the 33 BEAM group participants, 17 (51.5%) completed the measures of feasibility and acceptability. Among these respondents, engagement was relatively high at the beginning of the program with 94.1% self-reporting watching ≥1 video and 70.6% attending ≥1 telehealth session. Video watching amongst respondents ranged from 31.3% reported watching 1–5, 12.5% 6–10, and 31.3% ≥16 of the 31 total videos. Telehealth session participation was also variable, with 25.0% attending 1–3 sessions, 33.3% attending 4–6 sessions, 16.77% attending 7–9 sessions, and 25.0% attending ≥10 sessions. 70.6% of respondents reported participating in the forum, with their use ranging from "*rarely*" to "*once or twice,*" to "*once every 2 weeks.*" Attendance tracking *via* Jane, for the 26 participants who started the BEAM program, indicated that the average number of telehealth sessions was 3.19 (ranging from 0 to 12). Backend data from Vanilla Forums indicated that 14 participants used the forum, with an average of 5.64 posts (ranging from 1 to 15).

Respondents indicated that the BEAM program was a good source of social support (58.8% rated ≥4 out of 7 agreement). MAUQ results indicated that the BEAM App had moderate ease of use (e.g., 64.7% rated ≥5 on the "*easy for me to learn*" item), interface and satisfaction (e.g., 56.3% rated ≥5 on the "*Overall, I am satisfied*" item), and usefulness (e.g., 58.8% rated ≥5 on "*useful for my health and well-being*" item).

Baseline Characteristics

Participant socio-demographic characteristics and outcome measures at baseline are presented for the full sample and by group in **Table 3**. Mothers were, on average, 33.75 years old, and had 2 children with the toddlers identified for the study being an average of 2.32 years old. Across the full sample, the majority of mothers (81.5%) had completed some form of postsecondary education, most (70.8%) identified as being from European descent, and almost half (46.2%) had an annual household income (before tax) above the Canadian median (approximately $90,000 CAD in 2019). Mean levels of depression, anxiety and anger symptoms, as well as parenting stress, were above the established clinical cut-offs at baseline. After randomization, there were more participants with male children in the TAU group ($\chi^2 = 9.600$, $p = 0.002$). The imbalance between groups in percentage of participants who reported receiving individual

TABLE 3 | Sample characteristics at baseline.

		Group	
	Total sample (*N* = 65)	**BEAM (*n* = 33)**	**TAU (*n* = 32)**
***n* (%)**			
Socio-demographics			
European Canadian	46 (70.8)	23 (69.7)	23 (71.9)
Household income > 90K	30 (46.2)	15 (45.5)	15 (46.9)
Post-secondary education	53 (81.5)	26 (78.8)	27 (84.4)
Married	50 (76.9)	27 (81.8)	23 (71.9)
Male child	40 (61.5)	14 (42.2)	26 (81.3)
***M* (SD)**			
Age of mother (years)	33.84 (5.34)	33.73 (5.19)	31.90 (5.46)
Age of child (months)	26.02 (6.38)	27.33 (7.26)	24.62 (5.04)
Parity	2.09 (1.09)	2.24 (1.25)	1.94 (0.88)
***n* (%)**			
Treatment history (past month)[b]			
Psychiatric medication(s)	27 (41.5)	15 (45.5)	12 (37.5)
Individual counseling	18 (27.7)	5 (29.4)	13 (54.2)
Group counseling	10 (15.4)	4 (25.0)	6 (37.5)
Other services[a]	6 (40.0)	2 (23.1)	4 (30.8)
***M* (SD)**			
Mental health symptoms[c]	0.000 (0.655)	0.014 (0.627)	–0.014 (0.693)
Depression (PHQ-9)	15.98 (4.02)†	16.27 (4.27)†	15.69 (3.80)†
Anxiety (GAD-7)	13.88 (4.68)†	14.40 (4.55)†	13.34 (4.82)†
Anger (PROMIS)	64.90 (6.02)†	64.79 (5.16)†	65.01 (6.88)†
Sleep disturbance (PROMIS)	58.45 (5.88)	58.71 (5.67)	58.18 (6.17)
Alcohol use (AUDIT)	4.09 (4.37)	3.38 (4.03)	4.82 (4.65)
Parenting problems[c]	–0.004 (0.736)	0.021 (0.815)	–0.030 (0.656)
Parenting stress (PSI-SF)	93.45 (20.71)†	94.00 (22.37)†	92.88 (19.19)†
Parenting discipline (PS)	3.83 (0.27)	3.82 (0.24)	3.84 (0.31)
Parenting behaviors (PARYC)[d]	75.72 (11.08)	75.19 (11.38)	76.27 (10.43)
Positive coping[c]	–0.010 (0.729)	0.032 (0.718)	–0.053 (0.750)
Self-compassion (SCS-SF)	2.25 (0.55)	2.34 (0.55)	2.16 (0.55)
Hopefulness (RSE)	12.85 (2.58)	12.97 (2.42)	12.72 (2.77)
Couple satisfaction (CSI-4)	12.91 (5.34)	12.47 (4.88)	13.42 (5.92)
Child behavior (CBCL)	39.35 (23.98)	38.31 (23.13)	40.41 (25.15)

PHQ-9, Patient Health Questionnaire; GAD-7, Generalized Anxiety Disorder; PROMIS, Patient-Reported Outcomes Measurement Information System; AUDIT, Alcohol Use Disorders Identification Test; PSI-SF, Parenting Stress Index Short Form; PS, Parenting Scale; PARYC, Parenting Young Children; SCS-SF, Self-Compassion Scale Short Form; RSE, Recent Stressful Events; CSI-4, Couples Satisfaction Index; CBCL, Child Behavior Checklist.
[a]*Other services included: App-based or online mental health programs, instant messaging mental health services, mental health crisis lines, and faith-based counseling.*
[b]*Data collected from online screener.*
[c]*Aggregate of standardized variables.*
[d]*Combined total of subscales.*
†*Mean above clinical cut-off.*

counseling in the month prior to enrolment was not statistically significant ($\chi^2 = 2.476$, $p = 0.116$).

There were no statistically significant differences in baseline characteristics between participants who completed the T2 assessment and those lost to attrition, except for more parents of male children among those retained (70.2%) than dropouts (41.2%) according to Fischer's exact tests ($p = 0.035$). There were also no statistically significant differences in mental health symptoms, parenting problems, positive coping, or children behavior between participants who completed the T2 assessment and those lost to attrition, except that dropouts reported more limit setting on PAYRC subscale ($p = 0.029$).

Treatment Effects

Results of the mixed model analyses for primary and secondary outcomes are presented in **Table 4**.[2] There were no adverse events related to study participation.

Primary Outcomes

No statistically significant treatment effects (i.e., time*group interaction) were observed for depression or parenting stress, although the effect sizes were moderate (Cohen's $d = 0.47$ and 0.33, respectively). However, results indicated a statistically significant effect of time for depression, such that symptoms decreased from pre to post assessment across both the BEAM and TAU groups.

For participants who completed T2 differences between groups regarding clinically significant change on primary outcomes was also explored. In terms of depression, 66.7% of respondents from BEAM and 48.3% from the TAU group had clinically significant change (≥5-point reduction) on the PHQ-9. Statistically, Fisher's exact test indicated this difference was not significant ($p = 0.176$). In terms of parenting stress, 47.1% of BEAM and 48.3% of TAU respondents had clinically significant change (≥5-point reduction) on the PSI-SF, which was not a statistically significant difference on Fisher's exact test ($p = 0.590$).

Secondary Outcomes

Mixed model results for aggregate variables indicated a statistically significant medium treatment effect (i.e., time * group interaction) for mental health symptoms (Cohen's $d = 0.71$), and moderate but not statistically significant treatment effects for parenting problems nor positive coping (Cohen's $d = 0.36$ and 0.41, respectively). There was also a moderate but not statistically significant treatment effect for child behavior problems (Cohen's $d = 0.41$).

Follow-up analyses of the individual scales comprising the mental health symptoms aggregate indicated statistically significant treatment effects (i.e., time * group interaction) for anxiety ($b = -2.97$, $SE = 1.37$, $p = 0.035$) and sleep problems ($b = -4.43$, $SE = 1.59$, $p = 0.007$), where a larger decrease from pre to post assessment was observed among the BEAM participants compared to the TAU group. There were not statistically significant treatment effects for anger ($b = -0.71$, $SE = 0.81$, $p = 0.387$) or alcohol use ($b = -1.08$, $SE = 0.79$, $p = 0.177$).

In terms of anxiety, 52.9% of respondents from BEAM and 34.5% from the TAU group had clinically significant change (≥4-point reduction) on the GAD-7 (Fisher's exact test; $p = 0.180$). For anger, 64.7% of BEAM and 46.4% of TAU respondents had clinically significant change (≥2-point T-score reduction) (Fisher's exact test; $p = 0.189$). For sleep, 52.9% of BEAM and 20.7% of TAU respondents had clinically significant change (≥2-point T-score reduction), (Fisher's exact test; $p = 0.028$). In terms of alcohol use, no respondents from either group had a clinically significant change (≥5-point reduction).

DISCUSSION

The current pilot randomized controlled trial investigated the development of the Building Emotional Awareness and Mental Health (BEAM) digital program, which simultaneously targeted mental health and parenting skills in mothers of young children. In terms of feasibility and acceptability there was high interest in the digital program at screening, resulting in recruitment of 65 mothers with clinically significant depressive symptoms. Several challenges to building and testing an App-based intervention were identified (see **Table 1**), which likely impacted engagement, although those who completed the program reported adequate usability and satisfaction. Analyses did not reveal significant treatment effects for symptoms of depression or parenting stress (primary outcomes), however there were greater reductions in overall maternal mental health problems for participants receiving the BEAM program. Specifically, greater reductions in anxiety symptoms and sleep problems (secondary outcomes) were observed among the BEAM vs. TAU participants. No statistically significant effects emerged for parenting problems, positive coping, or child behavior outcomes.

We followed best practices and used a guided approach (with therapists and parent coaches). In terms of acceptability, the MAUQ findings in the current study indicated slightly lower usability than those for a similar App prototype which targeted parent feeding practices with their infants and toddlers (71). Although engagement and attrition in the BEAM program was comparable to other digital interventions that target parent mental health, parenting skills, or child behavior (25), the lower rate of retention in the intervention compared to the control group may suggest program feasibility issues. Indeed, feasibility challenges in the design process and budget available (as is the case with almost any grant-based project) resulted in a non-optimal digital health platform, which may have impacted user experience, engagement, and ultimately mental health. The digital media company emphasized that we should be satisfied with a limited "minimum viable" product for this pilot trial. However, the research team ultimately felt the App was sufficiently compromised so-as to offer a marginally satisfactory test of the promise of materials and therapeutic approach. We tried to overcome the limited functionality of the App on usability and adherence (72), by ensuring learnability of the video content and adding a telehealth group with clinical contact

[2]Including child sex as a covariate did not change the pattern of findings for any of the mixed model analyses. Therefore, the results are presented here without covariates.

TABLE 4 | Pre-post mixed model results for outcome measures.

	Primary Outcomes		Secondary outcomes			
	Depression	**Parenting stress**	**Mental health**	**Parenting problems**	**Positive coping**	**Child behavior**
Estimate (SE)						
Random effects						
Intercept	12.17*** (3.65)	326.35*** (71.29)	0.301*** (0.070)	0.428*** (0.092)	0.127*** (0.026)	509.42*** (98.84)
Fixed effects						
Intercept	15.69*** (0.883)	92.88*** (3.665)	–0.014 (0.117)	–0.040 (0.132)	–0.053 (0.127)	40.41*** (4.27)
Time	**–4.33*** (0.927)**	–3.72 (2.67)	**0.203* (0.096)**	0.119 (0.095)	–0.113 (0.093)	1.16 (2.27)
Treatment group	0.585 (1.24)	1.13 (5.15)	0.024 (0.164)	0.078 (0.186)	0.085 (0.178)	–2.10 (5.60)
Time * treatment group	–2.42 (1.45)	–2.35 (4.32)	**–0.422** (0.154)**	–0.102 (0.155)	0.201 (0.151)	–3.03 (3.77)

Treatment as usual (TAU) is the reference group. Depression was measured using the PHQ-9, Parenting Stress was measured using the PSI-SF, Mental health was measured using an aggregate (PHQ-9, GAD-7, PROMIS Anger and Sleep Disturbance, AUDIT), Parenting problems was measured using the aggregate (PSI-SF, PS, PARYC), Positive coping was measured using an aggregate (SCS-SF, RSE, CSI), Child behavior was measured using the CBCL.
$^{*}p < 0.05$ $^{**}p < 0.01$ $^{***}p < 0.001$.
Bold values represent statistical significance.

and material review. On the other hand, adding components to make up for App limitations (e.g., external survey weekly link vs. a mood monitoring function, posting about homework on the forum vs. interactive activities within the App) may have inadvertently increased complexity and cognitive demand. Our team is also investigating feasibility and acceptability of the pilot in a separate project through thematic analysis of qualitative data (including open-ended survey questions, forum posts, and focus group interviews), which may provide further insight on helpful components and areas of improvement.

Small to medium treatment effects were observed for mental health symptoms and parenting problems in the current trial, but were not statistically significant due to small sample size and measurement variability. The discrepancy between the linear mixed modeling and clinical significance results for anxiety may reflect the large point reduction required for the GAD-7. Nonetheless, the pattern of change for the BEAM program aligns with meta-analyses indicating larger effects of digital interventions for anxiety than for depression or parenting stress among parents of young children (25). Together this evidence suggests that parental anxiety may be more responsive to digital health interventions, whereas the cyclical nature of depression and related anhedonia may impede improvement and motivation to engage in a treatment with limited clinical contact. The non-ideal summer start of group may also have contributed to depression reductions across groups. Longer-term follow-up will be key in future trials to assess maintenance effects.

Although experiencing parenting stress was not an inclusion criterion for the trial, the mean parenting stress scores of both intervention and control groups was above the clinically significant level. We did not observe any statistically or clinically significant change in parenting stress. The pandemic exacerbated existing gender inequalities with women experiencing more job loss, and greater home and childcare responsibilities (73, 74). Potentially, in the context of chronic pandemic parenting stress, this brief intervention was not enough to decrease it. Future trials with long-term follow up are needed to investigate whether treatment effects may be delayed until acute stressors, particularly related to the pandemic, are resolved. It was also notable that the trial began in the spring of 2021, when pandemic related restrictions were relaxed across Canada. The large reductions in depressive symptoms observed in both groups may have been associated with these easing of restrictions.

Longitudinal studies to examine the sustainability of treatment effects and extent to which reductions in parent mental health problems are linked to changes in parenting and subsequent child mental health will be critical to informing the potential of digital health program to deliver widespread impacts on family mental health. Given that meta-analyses from in-person programs indicate that addressing both mother and child parenting needs leads to ∼50% larger treatment effects, there is a clear need for developing evidence-based digital health programs for parents that prevent child mental illness and its health sequalae in the aftermath of the pandemic.

The acceptability findings and preliminary treatment effects for some of the mental health symptoms suggest that the App-based BEAM program is promising intervention for addressing family mental health and parenting needs during the pandemic and its aftermath. The BEAM App is broadly consistent with priority-setting research in parents of young children which highlights a desire for more support for families to develop healthy coping and emotion regulation (75). Parents have also indicated a need for access to evidence-based information, tailored to their needs, delivered in timely formats (75). Our team is committed to improving the BEAM program and are making multiple systematic improvements to all aspects based on pilot study results and ongoing input from the parent advisory board.

Strengths and Limitations

The current trial was strengthened by the use of a community-based, participatory action approach to co-develop an accessible, evidence-based digital mental health intervention. The incorporation of peer support and therapist contact directly responds to parents' needs (75), and follows best practices for digital interventions (42, 44). However, the findings should

be interpreted with several limitations in mind. Despite implementing strategies for participant communication (e.g., individualized check-in emails and reminders), retention proved difficult for this digital intervention. Although an intention-to-treat approach was used and dropout analyses indicated no significant differences on outcome measures at baseline, attrition (45.5%) likely reduced power to detect treatment effects. In addition, participants with a more positive perception of the program may have been more likely to complete the engagement and satisfaction measures. Larger efficacy trials are now planned with more participants in order to fully test the intervention in an adequately powered sample. There were no differences between groups in medication or mental health service use at enrollment, however data was not collected on other forms of treatment during the trial, which could potentially obscure the findings. Information on service and resource use will be gathered in future trials. The weekly depression symptom and parenting stress surveys were also sent to the TAU group, and given the benefits of self-monitoring (43, 76, 77), which has been considered as an intervention or perceived mechanism of change (e.g., increasing awareness and reflection) (78–80), this could have functioned as an unintended active control condition. To account for this, our subsequent trials have removed weekly surveys for the comparison groups. Regardless, examining weekly symptom monitoring was not feasible in the current trial due to low completion (<25%), indicating that integration and individualized feedback within the App are needed in future versions of the program. Lastly, the generalizability of the results is restricted by the demographics of the sample and should be replicated among parents with more diverse socioeconomic backgrounds, among equity seeking groups, as well as non-female identifying caregivers.

Future Directions

Investing in maternal mental health early, before problems are entrenched, is expected to yield high health and economic benefits by preventing the long-term consequences of maternal depression from becoming embedded in children's biological and behavioral development (81). Although there were no prespecified criteria for proceeding to a definitive trial given the novelty of the field, our team is using a rapid cycle iteration approach, following the IDEAS (Innovate, Design, Evaluate, Adapt, Scale) Impact Framework™ from the Harvard Center on the Developing Child, to improve our ability to affect change in the BEAM program outcomes based on feedback from focus groups with participants. These improvements include simplified content delivery, streamlined communication with participants, greater synergy between mental health and parenting content, allowing parent coaches to fully facilitate the forums, and improving the user experience in the App. Indeed, we have secured further funding to update the App and program content and will be launching several larger trials that incorporate more peer-coaching, and involvement from service providers at local agencies holding tailored expertise to community needs. Version 1.2 of the BEAM program is being tested in a larger Phase III (efficacy) randomized controlled trial and a pragmatic trial with a community organization, as well as plans for a full App re-build in an embedded longitudinal cohort study for mothers who were pregnant during the pandemic.

CONCLUSION

Results highlight both feasibility limitations in version 1.0 of the BEAM program alongside significant promise to improve family mental health through an App-based digital intervention.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Psychology/Sociology Research Ethics Board (P2020:081) at the University of Manitoba & the Conjoint Health Research Ethics Board (REB20-1933) at the University of Calgary. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

LR, LT-M, KR, RG, MS, JP, and GG contributed to the conceptualization and design of the study, as well as funding applications. SS-Z was consulted regarding adaptation of the Unified Protocol treatment program. KS, MRS, JB, and LP-G contributed to the trial coordination and data collection. AM and LB were clinical trial therapists. AM and KS conducted the data analyses and synthesis. AM, KS, MRS, CR, LP-G, LT-M, and LR contributed to the preparation of the manuscript for publication. All authors have reviewed and approved the final version.

FUNDING

This study was funded by a Research Manitoba COVID-19 Rapid Response Operating Grant. AM was supported by the Social Sciences and Humanities Research Council (SSHRC) Banting Postdoctoral Fellowship (#01353-000). CR was supported by the Postdoctoral fellowship from Research Manitoba and the Children's Hospital Foundation of Manitoba.

ACKNOWLEDGMENTS

We gratefully acknowledge the mothers who participated in the study. We would also like to thank the Parent Advisory Board members and parent coaches for their valuable contributions, Bailin Xie and Sandra Hunter for their role as clinical therapists, as well as Allyson Paton, Avaline Konkin, Cici Guo, Dana Watts, Madeline Belows, Makayla Freeman, Mateja Carevic, Jeshna Kumari, Kira Pavagadhi, Kyra Cebula, and Shayna Pierce for their research assistance in conducting the trial.

REFERENCES

1. Tomfohr-Madsen LM, Racine N, Giesbrecht GF, Lebel C, Madigan S. Depression and anxiety in pregnancy during COVID-19: a rapid review and meta-analysis. *Psychiatry Res.* (2021) 300:113912. doi: 10.1016/j.psychres.2021.113912
2. Cameron EE, Joyce KM, Delaquis CP, Reynolds K, Protudjer JLP, Roos LE. Maternal psychological distress & mental health service use during the COVID-19 pandemic. *J Affect Disord.* (2020) 276:765–74.
3. Lebel C, MacKinnon A, Bagshawe M, Tomfohr-Madsen L, Giesbrecht G. Elevated depression and anxiety symptoms among pregnant individuals during the COVID-19 pandemic. *J Affect Disord.* (2020) 277:5–13.
4. Lomonaco-Haycraft KC, Hyer J, Tibbits B, Grote J, Stainback-Tracy K, Ulrickson C, et al. Integrated perinatal mental health care: a national model of perinatal primary care in vulnerable populations. *Prim Health Care Res Dev.* (2019) 20:e77. doi: 10.1017/S1463423618000348
5. Yao H, Chen J-H, Xu Y-F. Patients with mental health disorders in the COVID-19 epidemic. *Lancet Psychiatry.* (2020) 7:e21.
6. Bowen A, Bowen R, Butt P, Rahman K, Muhajarine N. Patterns of depression and treatment in pregnant and postpartum women. *Can J Psychiatry.* (2012) 57:161–7. doi: 10.1177/070674371205700305
7. Kingston D, Tough S. Prenatal and postnatal maternal mental health and school-age child development: a systematic review. *Matern Child Health J.* (2014) 18:1728–41. doi: 10.1007/s10995-013-1418-3
8. Van Der Waerden J, Galéra C, Larroque B, Saurel-Cubizolles M-J, Sutter-Dallay A-L, Melchior M, et al. Maternal depression trajectories and children's behavior at age 5 years. *J Pediatr.* (2015) 166:1440–1448. doi: 10.1016/j.jpeds.2015.03.002
9. König H, König HH, Konnopka A. The excess costs of depression: a systematic review and meta-analysis. *Epidemiol Psychiatr Sci.* (2019) 29:e30. doi: 10.1017/S2045796019000180
10. McGrath JM, Records K, Rice M. Maternal depression and infant temperament characteristics. *Infant Behav Dev.* (2008) 31:71–80.
11. Keller PS, Kouros CD, Erath SA, Dahl RE, El-Sheikh M. Longitudinal relations between maternal depressive symptoms and child sleep problems: the role of parasympathetic nervous system reactivity. *J Child Psychol Psychiatry.* (2014) 55:172–9. doi: 10.1111/jcpp.12151
12. Park M, Brain U, Grunau RE, Diamond A, Oberlander TF. Maternal depression trajectories from pregnancy to 3 years postpartum are associated with children's behavior and executive functions at 3 and 6 years. *Arch Womens Ment Health.* (2018) 21:353–63. doi: 10.1007/s00737-017-0803-0
13. Henry J, Dionne G, Viding E, Vitaro F, Brendgen M, Tremblay RE, et al. Early warm-rewarding parenting moderates the genetic contributions to callous-unemotional traits in childhood. *J Child Psychol Psychiatry.* (2018) 59:1282–8. doi: 10.1111/jcpp.12918
14. Uher R, Zwicker A. Etiology in psychiatry: embracing the reality of poly-gene-environmental causation of mental illness. *World Psychiatry.* (2017) 16:121–9. doi: 10.1002/wps.20436
15. Bernard K, Nissim G, Vaccaro S, Harris JL, Lindhiem O. Association between maternal depression and maternal sensitivity from birth to 12 months: a meta-analysis. *Attach Hum Dev.* (2018) 20:578–99. doi: 10.1080/14616734.2018.1430839
16. Katz LF, Maliken AC, Stettler NM. Parental meta-emotion philosophy: a review of research and theoretical framework. *Child Dev Perspect.* (2012) 6:417–22.
17. Lovejoy MC, Graczyk PA, O'Hare E, Neuman G. Maternal depression and parenting behavior: a meta-analytic review. *Clin Psychol Rev.* (2000) 20:561–92. doi: 10.1016/s0272-7358(98)00100-7
18. Vostanis P, Graves A, Meltzer H, Goodman R, Jenkins R, Brugha T. Relationship between parental psychopathology, parenting strategies and child mental health. *Soc Psychiatry Psychiatr Epidemiol.* (2006) 41:509–14. doi: 10.1007/s00127-006-0061-3
19. Starace F, Ferrara M. COVID-19 disease emergency operational instructions for mental health departments issued by the Italian society of epidemiological psychiatry. *Epidemiol Psychiatr Sci.* (2020) 29:e116. doi: 10.1017/S2045796020000372
20. Moreno C, Wykes T, Galderisi S, Nordentoft M, Crossley N, Jones N, et al. How mental health care should change as a consequence of the COVID-19 pandemic. *Lancet Psychiatry.* (2020) 7:813–24. doi: 10.1016/S2215-0366(20)30307-2
21. Petts RJ, Carlson DL, Pepin JR. A gendered pandemic: childcare, homeschooling, and parents' employment during COVID-19. *Gend Work Organ.* (2021) 28:515–34. doi: 10.1371/journal.pone.0247959
22. Penner-Goeke L, Kaminski L, Sulymka J, Mitchell B, Peters S, Roos LE. A meta-analysis group-based parenting programs for preschoolers: are programs improving over time? *PsyArXiv.* (2020) [Preprint]. doi: 10.31234/osf.io/yubqs
23. Sztein DM, Koransky CE, Fegan L, Himelhoch S. Efficacy of cognitive behavioural therapy delivered over the Internet for depressive symptoms: a systematic review and meta-analysis. *J Telemed Telecare.* (2017) 24:527–39. doi: 10.1177/1357633X17717402
24. Twomey C, O'Reilly G. Effectiveness of a freely available computerised cognitive behavioural therapy programme (MoodGYM) for depression: meta-analysis. *Aust N Z J Psychiatry.* (2017) 51:260–9. doi: 10.1177/0004867416656258
25. MacKinnon AL, Silang K, Penner K, Zalewski M, Tomfohr-Madsen L, Roos LE. Promoting mental health in parents of young children using ehealth interventions: a systematic review and meta-analysis. *Clin Child Fam Psychol Rev.* (2022):1–22. doi: 10.1007/s10567-022-00385-5 [Epub ahead of print].
26. Larkin F, Oostenbroek J, Lee Y, Hayward E, Meins E. Proof of concept of a smartphone app to support delivery of an intervention to facilitate mothers' mind-mindedness. *PLoS One.* (2019) 14:e0220948. doi: 10.1371/journal.pone.0220948
27. Catuara-Solarz S, Skorulski B, Estella I, Avella-Garcia C, Shepherd S, Stott E, et al. Efficacy of 'foundations', a digital mental health app to improve mental well-being, during COVID-19: a proof-of-principle randomised controlled trial. *JMIR Mhealth Uhealth.* (2021). doi: 10.2196/30976 [Epub ahead of print].
28. Sawyer A, Kaim A, Le HN, McDonald D, Mittinty M, Lynch J, et al. The effectiveness of an app-based nurse-moderated program for new mothers with depression and parenting problems (eMums Plus): pragmatic randomized controlled trial. *J Med Internet Res.* (2019) 21:e13689.
29. Shonkoff JP, Fisher PA. Rethinking evidence-based practice and two-generation programs to create the future of early childhood policy. *Dev Psychopathol.* (2013) 25:1635. doi: 10.1017/S0954579413000813
30. Flannery JE, Penner-Goeke L, Xie EB, Prince D, Simpson KM, Callaghan B, et al. Digital parent training RCT meta-analysis and systematic review. *PsyArXiv.* (2021) [Preprint]. doi: 10.31234/OSF.IO/CPD9B
31. Pierce S, de Castro Lima HL, Jain B, Reynolds K, Tomfohr-Madsen L, Roos L. Unique role for social connection in the COVID-19 pandemic; a qualitative analysis of parenting forums. *OSF.* (2022) [Preprint]. doi: 10.31219/osf.io/g63sh
32. Farchione TJ, Fairholme CP, Ellard KK, Boisseau CL, Thompson-Hollands J, Carl JR, et al. Unified protocol for transdiagnostic treatment of emotional disorders: a randomized controlled trial. *Behav Ther.* (2012) 43:666–78.
33. Steele SJ, Farchione TJ, Cassiello-Robbins C, Ametaj A, Sbi S, Sauer-Zavala S, et al. Efficacy of the Unified Protocol for transdiagnostic treatment of comorbid psychopathology accompanying emotional disorders compared to treatments targeting single disorders. *J Psychiatr Res.* (2018) 104:211–6. doi: 10.1016/j.jpsychires.2018.08.005
34. Bullis J, Sauer-Zavala S. The unified protocol for insomnia disorder. In: Farchione TJ, Barlow DH editors. *Applications of the Unified Protocol for Transdiagnostic Treatment of Emotional Disorders*. (Oxford: Oxford University Press) (2017).
35. Barlow DH, Farchione TJ, Sauer-Zavala S, Latin HM, Ellard KK, Bullis JR, et al. *Unified Protocol for Transdiagnostic Treatment of Emotional Disorders: Therapist Guide*. Oxford: Oxford University Press (2017).
36. Sakiris N, Berle D. A systematic review and meta-analysis of the unified protocol as a transdiagnostic emotion regulation based intervention. *Clin Psychol Rev.* (2019) 72:101751. doi: 10.1016/j.cpr.2019.101751
37. Havighurst SS, Radovini A, Hao B, Kehoe CE. Emotion-focused parenting interventions for prevention and treatment of child and adolescent mental health problems: a review of recent literature. *Curr Opin Psychiatry.* (2020) 33:586–601. doi: 10.1097/YCO.0000000000000647
38. Eisenberg N, Cumberland A, Spinrad TL. Parental socialization of emotion. *Psychol Inq.* (1998) 9:241–73.

39. Havighurst SS, Wilson KR, Harley AE, Prior MR, Kehoe C. Tuning in to Kids: improving emotion socialization practices in parents of preschool children – findings from a community trial. *J Child Psychol Psychiatry*. (2010) 51:1342–50. doi: 10.1111/j.1469-7610.2010.02303.x
40. Czajkowski SM, Powell LH, Adler N, Naar-King S, Reynolds KD, Hunter CM, et al. From ideas to efficacy: the ORBIT model for developing behavioral treatments for chronic diseases. *Health Psychol.* (2015) 34:971–82. doi: 10.1037/hea0000161
41. Yardley LP, Spring BJP, Riper HP, Morrison LGP, Crane DHM, Curtis KP, et al. Understanding and promoting effective engagement with digital behavior change interventions. *Am J Prev Med.* (2016) 51:833–42. doi: 10.1016/j.amepre.2016.06.015
42. Spence SH, March S, Donovan CL. Social support as a predictor of treatment adherence and response in an open-access, self-help, internet-delivered cognitive behavior therapy program for child and adolescent anxiety. *Internet Interv.* (2019) 18:100268. doi: 10.1016/j.invent.2019.100268
43. Lambert MJ, Harmon KL. The merits of implementing routine outcome monitoring in clinical practice. *Clin Psychol Sci Pract.* (2018) 25:e12268. doi: 10.1016/j.jmwh.2006.02.008
44. Heber E, Ebert DD, Lehr D, Cuijpers P, Berking M, Nobis S, et al. The benefit of web-and computer-based interventions for stress: a systematic review and meta-analysis. *J Med Internet Res.* (2017) 19:e32. doi: 10.2196/jmir.5774
45. Ellis JG, Deary V, Troxel WM. The role of perceived partner alliance on the efficacy of CBT-I: preliminary findings from the partner alliance in insomnia research study (PAIRS). *Behav Sleep Med.* (2015) 13:64–72. doi: 10.1080/15402002.2013.838768
46. Whisman MA, Baucom DH. Intimate relationships and psychopathology. *Clin Child Fam Psychol Rev.* (2012) 15:4–13. doi: 10.1007/s10567-011-0107-2
47. Alves S, Martins A, Fonseca A, Canavarro MC, Pereira M. Preventing and treating women's postpartum depression: a qualitative systematic review on partner-inclusive interventions. *J Child Fam Stud.* (2018) 27:1–25.
48. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth app usability questionnaire (MAUQ): development and validation study. *JMIR Mhealth Uhealth.* (2019) 7:e11500. doi: 10.2196/11500
49. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med.* (2001) 16:606–13. doi: 10.1046/j.1525-1497.2001.016009606.x
50. Löwe B, Unützer J, Callahan CM, Perkins AJ, Kroenke K. Monitoring depression treatment outcomes with the patient health questionnaire-9. *Med Care.* (2004) 42:1194–201. doi: 10.1097/00005650-200412000-00006
51. Abidin R. *Parenting Stress Index*. 3rd ed. Odessa, FL: Psychological Assessment Resources (1995).
52. Breitenstein SM, Fogg L, Ocampo EV, Acosta DI, Gross D. Parent use and efficacy of a self-administered, tablet-based parent training intervention: a randomized controlled trial. *JMIR MHealth and UHealth.* (2016) 4:e36. doi: 10.2196/mhealth.5202
53. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med.* (2006) 166:1092–7. doi: 10.1001/archinte.166.10.1092
54. Toussaint A, Hüsing P, Gumz A, Wingenfeld K, Härter M, Schramm E, et al. Sensitivity to change and minimal clinically important difference of the 7-item generalized anxiety disorder questionnaire (GAD-7). *J Affect Disord.* (2020) 265:395–401. doi: 10.1016/j.jad.2020.01.032
55. Pilkonis PA, Choi SW, Reise SP, Stover AM, Riley WT, Cella D. Item banks for measuring emotional distress from the patient-reported outcomes measurement information system (PROMIS®®): depression. *Anxiety Anger Assess.* (2011) 18:263–83.
56. Yu L, Buysse DJ, Germain A, Moul DE, Stover A, Dodds NE, et al. Development of short forms from the PROMIS™™ sleep disturbance and sleep-related impairment item banks. *Behav Sleep Med.* (2012) 10:6–24. doi: 10.1080/15402002.2012.636266
57. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, et al. The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol.* (2010) 63:1179–94. doi: 10.1016/j.jclinepi.2010.04.011
58. Terwee CB, Peipert JD, Chapman R, Lai J-S, Terluin B, Cella D, et al. Minimal important change (MIC): a conceptual clarification and systematic review of MIC estimates of PROMIS measures. *Qual Life Res.* (2021) 30:2729–54. doi: 10.1007/s11136-021-02925-y
59. de Meneses-Gaya C, Zuardi AW, Loureiro SR, Crippa JAS. Alcohol use disorders identification test (AUDIT): an updated systematic review of psychometric properties. *Psychol Neurosci.* (2009) 2:83–97. doi: 10.20344/amp.15765
60. Johansson M, Sinadinovic K, Hammarberg A, Sundström C, Hermansson U, Andreasson S, et al. Web-based self-help for problematic alcohol use: a large naturalistic study. *Int J Behav Med.* (2017) 24:749–59. doi: 10.1007/s12529-016-9618-z
61. Raes F, Pommier E, Neff KD, Van Gucht D. Construction and factorial validation of a short form of the self−compassion scale. *Clin Psychol Psychother.* (2011) 18:250–5. doi: 10.1002/cpp.702
62. Center on the Developing Child. *The JPB Research Network on Toxic Stress*. Cambridge, MA: Harvard University (2021).
63. Funk JL, Rogge RD. Testing the ruler with item response theory: increasing precision of measurement for relationship satisfaction with the couples satisfaction index. *J Fam Psychol.* (2007) 21:572–83. doi: 10.1037/0893-3200.21.4.572
64. Arnold DS, O'Leary SG, Wolff LS, Acker MM. The parenting scale: a measure of dysfunctional parenting in discipline situations. *Psychol Assess.* (1993) 5:137–44.
65. McEachern AD, Dishion TJ, Weaver CM, Shaw DS, Wilson MN, Gardner F. Parenting young children (PARYC): validation of a self-report parenting measure. *J Child Fam Stud.* (2012) 21:498–511. doi: 10.1007/s10826-011-9503-y
66. Achenbach TM. Child behavior checklist. In: Kreutzer JS, DeLuca J, Caplan B editors. *Encyclopedia of Clinical Neuropsychology*. (New York, NY: Springer) (2011). p. 546–52.
67. Feingold AA. Regression framework for effect size assessments in longitudinal modeling of group differences. *Rev Gen Psychol.* (2013) 17:111–21. doi: 10.1037/a0030048
68. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: L. Erlbaum Associates (1988).
69. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *Ann Intern Med.* (2010) 152:726–32.
70. Kwok OM, Elliott TR, Yoon M, Underhill AT, Berry JW, Luo W. Analyzing longitudinal data with multilevel models: an example with individuals living with lower extremity intra-articular fractures. *Rehabil Psychol Rehabil Psychol.* (2008) 53:370–86. doi: 10.1037/a0012765
71. Mobley A, Jake-Schoffman D, Fedele D, Varela E, Zeldman J. Babybyte: usability testing of a mHealth app prototype to improve responsive feeding practices in parents of infants and toddlers. *Curr Dev Nutr.* (2021) 5(Suppl. 2):983.
72. Perez JC. Considerations in the development and use of e– and m– mental health in Latin American contexts: case analysis of an m-health intervention to prevent postnatal depression and anxiety in Chilean mothers. *ISSBD Bull.* (2021) 2:14–21.
73. Fuller S, Qian Y. Covid-19 and the gender gap in employment among parents of young children in Canada. *Gend Soc.* (2021) 35:206–17.
74. Qian Y, Fuller S. COVID-19 and the gender employment gap among parents of young children. *Can Public Policy.* (2020) 46:S89–101.
75. Brockway ML, Keys E, Bright KS, Ginn C, Conlon L, Doane S, et al. Top 10 (plus 1) research priorities for expectant families and those with children to age 24 months in Alberta, Canada: results from the family research Agenda initiative setting (FRAISE) priority setting partnership project. *BMJ Open.* (2021) 11:e047919. doi: 10.1136/bmjopen-2020-047919
76. van Os J, Verhagen S, Marsman A, Peeters F, Bak M, Marcelis M, et al. The experience sampling method as an mHealth tool to support self-monitoring, self-insight, and personalized health care in clinical practice. *Depress Anxiety.* (2017) 34:481–93. doi: 10.1002/da.22647
77. Bastiaansen JA, Meurs M, Stelwagen R, Wunderink L, Schoevers RA, Wichers M, et al. Self-monitoring and personalized feedback based on the experiencing sampling method as a tool to boost depression treatment: a protocol of

a pragmatic randomized controlled trial (ZELF-i). *BMC Psychiatry.* (2018) 18:276. doi: 10.1186/s12888-018-1847-z

78. Cerga-Pashoja A, Doukani A, Gega L, Walke J, Araya R. Added value or added burden? A qualitative investigation of blending internet self-help with face-to-face cognitive behaviour therapy for depression. *Psychother Res.* (2020) 30:998–1010. doi: 10.1080/10503307.2020.1720932
79. Darvell MJ, Kavanagh DJ, Connolly JM. A qualitative exploration of Internet-based treatment for comorbid depression and alcohol misuse. *Internet Interv.* (2015) 2:174–82.
80. Gerhards SA, Abma TA, Arntz A, de Graaf LE, Evers SM, Huibers MJ, et al. Improving adherence and effectiveness of computerised cognitive behavioural therapy without support for depression: a qualitative study on patient experiences. *J Affect Disord.* (2011) 129:117–25. doi: 10.1016/j.jad.2010.09.012
81. Bernard-Bonnin A-C, Psychosocial Paediatrics Committee, Canadian Paediatric Society. Maternal depression and child development. *Paediatr Child Health.* (2004) 9:575–98.

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX

TABLE A1 | BEAM weekly module topics.

Week	Module	Mental Health	Parenting
1	This stuff is tough	Introduction to the UP program Checking in with your mental health (self-monitoring)	Debunking parenting myths Finding your personal values
2	How do i get motivated?	Setting (SMART) goals Maintaining motivation Decisional balance exercise	Using mindfulness and self-compassion to identify moments of joy, feel less stressed, and improve relationships with children
3	The payoff to understanding emotions	The functions of emotions Three component model of an emotional response (cognitive, physiological, behavioral)	Teaching children about emotions through play and stories
4	How do my emotions work?	Putting emotions into context by identifying antecedents and consequences	Parents as co-regulators; responding with awareness and moving on together
5	Appreciating the rainbow amidst the storm	Developing and practicing mindful emotion awareness and self-compassion Meditation exercises	Loving your whole child (both positive and challenging qualities) Preparing ahead for difficult situations Responding with compassion
6	Getting unstuck	Identifying and changing negative thinking patterns and beliefs with cognitive flexibility	Motivating flexibility and adapting to children's needs to help navigate tough situations (transition toolbox)
7	Finding a better path when emotions are high	Identifying and countering emotional behaviors with alternative actions	Using alternative actions to help children manage emotional reactions
8	Feeling our feelings	Understanding physical sensations Physical exposure exercises	Coping strategies to regulate physical sensations while parenting
9	Exploring Your options	Developing an emotional exposure hierarchy (situational and imaginal)	Skills and techniques for managing tantrums
10	Where do i go from here?	Review of skills and progress Developing a practice plan	Identifying what you need to feel well and balanced

frontiers | Frontiers in Digital Health

TYPE Original Research
PUBLISHED 22 July 2022
DOI 10.3389/fdgth.2022.868970

OPEN ACCESS

EDITED BY
Brenna Renn,
University of Nevada, Las Vegas,
United States

REVIEWED BY
Michael Musker,
University of South Australia, Australia
Kevin A. Hallgren,
University of Washington,
United States

*CORRESPONDENCE
Charalampos Tsirmpas
haris@myfeel.co
Dimitrios Andrikopoulos
dimitris@myfeel.co
Panagiotis Fatouros
panagiotis@myfeel.co

†These authors have contributed equally to this work and share first authorship
‡These authors share senior authorship
§These authors share last authorship

SPECIALTY SECTION
This article was submitted to Digital Mental Health, a section of the journal Frontiers in Digital Health

RECEIVED 03 February 2022
ACCEPTED 29 June 2022
PUBLISHED 22 July 2022

CITATION
Tsirmpas C, Andrikopoulos D, Fatouros P, Eleftheriou G, Anguera JA, Kontoangelos K and Papageorgiou C (2022) Feasibility, engagement, and preliminary clinical outcomes of a digital biodata-driven intervention for anxiety and depression.
Front. Digit. Health 4:868970.
doi: 10.3389/fdgth.2022.868970

Feasibility, engagement, and preliminary clinical outcomes of a digital biodata-driven intervention for anxiety and depression

Charalampos Tsirmpas[1*†], Dimitrios Andrikopoulos[1*†], Panagiotis Fatouros[1*†], Georgios Eleftheriou[1‡], Joaquin A. Anguera[2‡], Konstantinos Kontoangelos[3,4§] and Charalabos Papageorgiou[4§]

[1]Feel Therapeutics Inc., San Francisco, CA, United States, [2]Departments of Neurology and Psychiatry, University of California, San Francisco, San Francisco, CA, United States, [3]First Department of Psychiatry, Eginition Hospital, Medical School National and Kapodistrian University of Athens, Athens, Greece, [4]Neurosciences and Precision Medicine Research Institute "Costas Stefanis", University Mental Health, Athens, Greece

Hypothesis: The main hypothesis is that a digital, biodata-driven, and personalized program would exhibit high user retention and engagement, followed by more effective management of their depressive and anxiety symptoms.

Objective: This pilot study explores the feasibility, acceptability, engagement, and potential impact on depressive and anxiety and quality of life outcomes of the 16-week Feel Program. Additionally, it examines potential correlations between engagement and impact on mental health outcomes.

Methods: This single-arm study included 48 adult participants with mild or moderate depressive or anxiety symptoms who joined the 16-week Feel Program, a remote biodata-driven mental health support program created by Feel Therapeutics. The program uses a combination of evidence-based approaches and psychophysiological data. Candidates completed an online demographics and eligibility survey before enrolment. Depressive and anxiety symptoms were measured using the Patient Health Questionnaire and Generalized Anxiety Disorder Scale, respectively. The Satisfaction with Life Scale and the Life Satisfaction Questionnaire were used to assess quality of life. User feedback surveys were employed to evaluate user experience and acceptability.

Results: In total, 31 participants completed the program with an overall retention rate of 65%. Completed participants spent 60 min in the app, completed 13 Mental Health Actions, including 5 Mental Health Exercises and 4.9 emotion logs on a weekly basis. On average, 96% of the completed participants were active and 76.8% of them were engaged with the sensor during the week. Sixty five percent of participants reported very or extremely high satisfaction, while 4 out of 5 were very likely to recommend the program to someone. Additionally, 93.5% of participants presented a decrease in at

least one of the depressive or anxiety symptoms, with 51.6 and 45% of participants showing clinically significant improvement, respectively. Finally, our findings suggest increased symptom improvement for participants with higher engagement throughout the program.

Conclusions: The findings suggest that the Feel Program may be feasible, acceptable, and valuable for adults with mild or moderate depressive and/or anxiety symptoms. However, controlled trials with bigger sample size, inclusion of a control group, and more diverse participant profiles are required in order to provide further evidence of clinical efficacy.

KEYWORDS

data-driven therapeutics, major depressive disorder, generalized anxiety disorder, psychophysiological data, emotion detection

1. Introduction

Major depressive disorder (MDD) and generalized anxiety disorders (GAD) are very common but serious mental disorders that can lead to considerable deterioration in overall health and daily functioning levels. MDD and GAD are typically manifested by adverse effects on a person's thoughts, behaviors, motivation, feelings, and sense of well-being, as well as a major disruption of a person's day-to-day life. According to the Anxiety and Depression Association of America (ADAA) (1), 18.1% of the adult population in the US is affected by anxiety disorders every year, and approximately 7% shows symptoms of MDD on a yearly basis, with MDD being the leading cause of disability among people aged 15–45. Similarly for the European Union (EU), the Organisation for Economic Co-operation and Development (OECD) (2) estimates that anxiety disorders affected an estimated 5.4% of the population in 2016, which equates to approximately 25 million people, while a recent study (3) showed that more than 6% of the EU population suffers from depression (data collected during the period 2013 — 2015). The World Health Organization (WHO) estimates that 1 in 4 people in Europe are affected by depression or anxiety each year (4). Given the high prevalence of MDD and GAD worldwide, efforts to quantify and assess their impact on a person's health and functioning leverage the Years Lived with Disability (YLD) metric (5). Using this metric, the WHO reports (6) that depressive disorders rank as the most prevalent contributor to non-fatal health loss, representing 7.5% of the total YLD worldwide, whilst also being one of the leading causes of suicide. At the same time, anxiety which is generally associated with a lower average level of disability on average (sixth largest contributor) accounts for 3.4% of the total YLD globally (6).

Apart from the direct implications on overall health and wellbeing of those affected, MDD and GAD also heavily impact the global labor market. The OECD reports (7) a 15–30% decrease in the likelihood of employment and a twice as high unemployment rate for people with mental health disorders, such as anxiety and depression. Furthermore, compared to the mentally healthy, people with mental health issues are more likely to report job strain, and are on average 25% less satisfied with their jobs, with the majority of them earning less than the overall median income as well (4, 8). Moreover, an increase in chronic absences due to sickness can be expected due to the increased burden in both their personal and professional lives. Interestingly, in the EU (9), up to half of the total sick leaves regard depression or anxiety. Consequently, all these factors combined can lead to a significant reduction in work productivity. According to a study in several OECD countries (10), it is estimated that compared to the healthy workforce, employees with mental health disorders performed three times worse. It is evident therefore that MDD and GAD may add a heavy burden to the global budget as a result of the reduced labor market participation stemming from lower employment rates and reduced productivity owed to reduced job satisfaction and unavoidable sick leaves. Combined with an increase in spending for the associated social benefits, the total indirect costs of MDD and GAD can sum up to $1 trillion annually, worldwide (9).

As can be understood, diagnosing, supporting, managing, and treating MDD and GAD is of utmost importance. Numerous reports (11–14) suggest an increasing trend in the prevalence of MDD and GAD which is further accelerated by the recent COVID-19 pandemic (15–17). Consequently, an increasing number of people are expected to develop symptoms of depression and anxiety in the following years. Although there exist tools and treatments, more than half of people suffering from MDD or GAD face barriers that prevent them from accessing mental health care resources (18–20). Among these, lack of access and utilization of mental health services is prevalent particularly in developing countries (21–23), where policies, health services, and research regarding mental health are ill-represented in the countries' budget with respect to the size of the problem. This can lead to a threefold reduction in the probability of obtaining mental health care compared

to the situation in developed countries (24, 25). Nevertheless, even in the case of more advanced countries, there exist a number of reasons that hinder access to mental health care. An indicative example regards that of the U.S. where a recent study (19) reported that almost half of the participants could not afford the cost of treatments, while approximately 17% cited reasons related to the lack of awareness of any services for reaching out. On a more individual-oriented level, more than 30% of participants raised concerns regarding social stigma, adverse effects on professional life, or unavailability for in-person treatment sessions.

Currently, traditional approaches for the management and treatment of depression and anxiety involve pharmacological treatment as well as psychotherapy sessions offered in the form of Cognitive Behavioral Therapy (CBT) or Interpersonal Therapy (IPT). Usually, for the more severe and chronic cases, a combination of both therapy and medication is followed (26). The efficacy of both approaches has been extensively investigated and has shown to be consistently high across many studies (27–29). Nevertheless, there exist additional factors that determine whether a specific approach proves beneficial for patients in the long run. The poorly established doctor-patient relationships due to lack of knowledge and insufficient training (27), the inability of patients to consistently follow psychotherapy sessions in the long run (19), along with the additional cost of medication and the adverse side effects that sometimes emerge (30), yield a risk of non-adherence and discontinuity of treatment (31, 32) which can reach values close to 60% for patients with depression or anxiety (30, 33). At the same time, subjective and patient-related factors (27) such as the severity of symptoms, comorbidities, and cultural beliefs, as well as limited response to medication due to underlying pathological conditions (27, 34) might hinder the overall efficacy of a treatment protocol, leaving an estimated 30%-40% of patients (27, 34) with minor or no symptom improvement.

Toward alleviating accessibility barriers and inequalities in mental health care, as well as potentially addressing the factors contributing to non-adherence and non-response to management and treatment protocols, digital mental health support tools have spurred during the last years and especially during the COVID-19 pandemic period (35, 36). These solutions revolve around therapeutic approaches and positive behavioral change, which can also work complementary to long-established traditional methods. Contrary to the latter, however, digital mental health offerings translate into remote and on-demand, personalized, inexpensive approaches for the treatment and management of mental health disorders. In this way, broader access to mental health care resources is achieved. The impact of such solutions has been extensively studied in the literature (37–41), with the results suggesting that an equal or greater effectiveness compared to traditional approaches can be achieved. Currently, there exist several products[1] that primarily offer remote therapy sessions which can also be complemented by tutorials, educational/training material, and exercises adapted to the specific clinical case.

These years, we are experiencing an evolution of management and treatment solutions for various physical health disorders toward precise medicine expressed by quantifiable data-driven schemes and continuous monitoring approaches. This established paradigm guides us to a potentially promising alternative for mental health support. Such approaches have been shown to significantly alter current practices and introduce considerable improvements in treatment adherence and effectiveness, with diabetes and cardiovascular diseases being two representative examples. For the former, continuous blood sugar monitoring devices along with insulin pumps are combined in a smart hybrid device that automatically regulates the delivery of insulin based on real-time readings of blood sugar levels. Similarly, for cardiovascular diseases, the pacemakers act by sending electrical signals to increase the heartbeats when their sensors pick up bradycardia conditions. In both cases, it is the introduction of objective data, as reflected by physiological measurements (i.e., glucose and heart electrical activity, respectively) that has enabled real-time, inexpensive, and unobtrusive interventions. These are usually integrated into closed-loop telehealth solutions which show increased engagement, effectiveness, and improvement of the quality of life of patients in comparison to conventional therapies (42–46). Therefore, using continuous measurement and data-driven interventions for diagnosis, management or treatment are the way forward for mental health. This introduction of continuous, passive, and objective data, in conditions such as depression and anxiety, is expected to increase engagement, and facilitate personalized treatment approaches, which are expected to be more effective.

With respect to tackling some of the factors leading to patient non-adherence and non-response to management and treatment protocols, digital mental health tools can be augmented by multimodal, digital data, *via* the utilization of mobile phones and wearable sensors. This kind of data varies in complexity, ranging from straightforwardly interpreted mobile app-based data such as user interactions, app usage, and activity tracking *via* the embedded mobile phone sensors (i.e., GPS), to much more complex data that may require dedicated devices and/or advanced collection and processing techniques such as physiological signals, voice, text, etc. The expectation is that part, or all, of the acquired data, offer a degree of objectivity and ubiquitousness and as such, it enables a greater understanding and a deeper insight into the behavior of individuals, in the

1 myStrength, AbleTo, Talkspace, Lyra, ginger, Meru Health.

context of their respective mental health conditions. At the same time, by incorporating an adaptive UX/UI such as a gamified experience, it is claimed that long-term engagement can be realized, enhancing in this way the effectiveness of the solution (47–50) and reducing both any direct and indirect costs involved (51–54). Regardless of their obvious advantages, however, digital mental health solutions need to be largely scalable in order to achieve their full potential, while maintaining user engagement and treatment effectiveness.

Acknowledging the importance of increased accessibility to mental health care and the added value of multimodal and objective data, Feel Therapeutics[2] has introduced the 16-week Feel Program (FP), which is a data-driven, digital mental health support program that uses a combination of emotion journaling, evidence-based cognitive behavioral therapy, mindfulness and positive psychology techniques augmented with the Feel Mental Health Biomarkers platform. Specifically, the FP consists of four components: (i) The Feel Emotion Sensor (FES), a wrist-worn device that provides continuous and unobtrusive monitoring of an individual's significant emotional changes. The FES continuously monitors the physiological signals of the user (i.e., Electrodermal Activity, Heart Rate Variability, and Skin Temperature) to detect changes in the activation of their Autonomic Nervous System and extract a series of other mental health metrics; (ii) The Feel mobile app, which utilizes data from the FES to provide near real-time emotion alerts and interventions, access to other mental health-related metrics and facilitate virtual sessions; (iii) Personalized weekly 15-min coaching sessions with the Feel Providers, augmented by the weekly extracted mental health metrics such as emotion-related, and self-reported data, etc.; (iv) Mental health resource center that compiles tutorials, exercises, tips and advice focusing on the development of mental health coping skills.

In order to explore and evaluate the feasibility, acceptance, and potential efficacy of the FP, a real-world data (RWD) feasibility study has been designed and conducted. In particular, focusing on mild or moderate MDD and/or GAD, Feel Therapeutics has designed and conducted a Proof-of-Concept (PoC), single-arm pilot study, aiming to (i) explore the ecological validity of the Feel emotion detection technology, (ii) validate the engagement with and (iii) evaluate the preliminary efficacy of the FP. The main hypothesis is that a remote, data-driven, and personalized program would exhibit increased levels of user retention and engagement during the 16-weeks, followed by a reduction of their depression and anxiety symptoms that would be captured by a respective decrease in associated clinical measures. In this work, we attempt to validate this hypothesis by presenting the main results of this PoC study, including (i) adoption/conversion rates; (ii) engagement levels with the program, along with the respective drop-out rates; (iii) the impact of the program on health-related quality of life; (iv) preliminary efficacy of the program; (v) validation of the emotion detection capabilities of the Feel technology measured at an in-the-wild environment.

The rest of this work is organized as follows: in Section 2, a detailed description of the materials and methods used for the PoC study is provided. Furthermore, in Section 3, we present the results of the study focusing on the feasibility and acceptability of the program (Section 3.1), the ecological validation of the Feel Emotion Detection technology (Section 3.2), the participant retention and engagement in the program (Section 3.3), as well as a preliminary assessment of the program's impact on mental health symptoms (Section 3.4), quality of life measures (Section 3.5) and participant self-assessment metrics (Section 3.6). Then, in Section 4 we continue with a thorough discussion, presenting an interpretation of the results toward supporting our hypothesis (Section 4.1), along with a few further observations (Section 4.2. Finally, the study limitations are outlined in Section 4.3 and this work is concluded in Section 4.4.

2. Materials and methods

2.1. Participant recruitment

The company developed a dedicated web page that included information about the Feel Program for depression and anxiety for participant recruitment.The target audience for the study was general public, while the study recruitment process was advertised *via*: (i) candidate referrals from the undergraduate student mental health support unit of the National and Kapodistrian University of Athens; (ii) social media ads (e.g., Facebook, Instagram, etc.) and (iii) word of mouth.

The study's inclusion and exclusion criteria were assessed based on self-reported candidate responses to the eligibility questionnaire, while they were also verified by the Feel Providers during the first introductory session. The main inclusion criteria were: (i) mild to moderate MDD ($4 < \text{PHQ-9} < 15$) and/or GAD ($4 < \text{GAD-7} < 15$); (ii) age $\geq$ 18 years old; and (iii) smartphone users/owners. On the other hand, the main exclusion criteria were: (i) severe MDD and/or GAD; (ii) personality disorders; (iii) psychotic disorders; (iv) bipolar disorder; (v) eating disorders; (vi) suicidal or self-harm thoughts; (vii) psychotropic medication; (viii) substance abuse.

2.2. Materials

2.2.1. Demographic and Eligibility Questionnaire

This questionnaire was completed by candidates at the screening stage, in order to assess their eligibility and included demographic information (e.g., gender, age, location, etc.), presence of any of the exclusion criteria, along with wrist measurements, in order to determine the appropriate sensor

2 myFeel.

size, in case of a positive eligibility assessment. The Demographic and Eligibility Questionnaire was embedded at the recruitment web page and completed at baseline.

2.2.2. Patient Health Questionnaire 9-item (PHQ-9)

The PHQ-9 is a self-administered questionnaire for assessing the severity of depressive symptoms (55). The questionnaire is composed of 9 items, each one scoring the frequency of occurrence of the 9 DSM-IV criteria on a scale from 0 (not at all) to 3 (nearly every day). The total score is the sum of the scores of the individual items. The threshold scores for classifying the severity of the depressive symptoms as mild, moderate, moderately severe, and severe depression are 5, 10, 15, and 20, respectively. The PHQ-9 was completed at baseline (embedded in the recruitment web page) and at weeks 8 (mid-program) and 16 (end-of-program) in the Feel app.

2.2.3. Generalized Anxiety Disorder-7 (GAD-7)

The GAD-7 is a self-administered questionnaire that serves as a brief clinical measure for assessing the severity of GAD (56). The questionnaire is composed of 7 items regarding DSM-IV criteria, each one scoring the frequency of occurrence of symptoms on a scale from 0 (not at all) to 3 (nearly every day). Based on the total score, which is the sum of the individual scores from the 7 items, the severity of the anxiety symptoms is assessed. More specifically, for the classification of the symptoms as mild, moderate, or severe anxiety, threshold scores of 5, 10, and 15 have been used, respectively. The GAD-7 was completed at baseline (embedded in the recruitment web page) and at weeks 8 (mid-program) and 16 (end-of-program) in the Feel app.

2.2.4. Life Satisfaction Questionnaire (LISAT-11)

The LISAT-11 questionnaire is a self-administered tool for measuring Life Satisfaction (57). It is comprised of 11 items; 1 global item evaluating life as a whole, and 10 domain-specific items including vocational situation, financial situation, leisure, contact friends, sexual life, activities of daily living, family life, partnership relationship, physical health, and psychological health. Each item is scored on a range from 1 to 6, with the total score being the mean of the individual scores. Higher scores indicate a greater level of perceived life satisfaction. The LISAT-11 was completed at baseline (embedded in the recruitment web page) and at weeks 8 (mid-program) and 16 (end-of-program) in the Feel app.

2.2.5. Satisfaction With Life Scale (SWLS)

The SWLS is a self-administered subjective well-being questionnaire that measures global life satisfaction (58). It consists of 5 items, each one scored from 1 (strongly disagree) to 7 (strongly agree) with the cutoff scoring values being 10, 15, 20, 21, 26, and 31, for life satisfaction levels ranging from extremely dissatisfied to extremely satisfied. The SWLS was completed at baseline (embedded in the recruitment web page) and at weeks 8 (mid-program) and 16 (end-of-program) in the Feel app.

2.2.6. Self-assessment questionnaire

The self-assessment questionnaire is a custom tool, aiming to capture the participants' perception regarding their accomplishments and progress throughout the 16 weeks of the program. Indicative questions of this survey are: "My concerns that brought me to the program have improved as a result of the services provided" and "I learned to think more clearly/accurately to reduce distressing emotions or behaviors," among others. A 5-point Likert-type scale was used with participant responses ranging from 1 (strongly disagree) to 5 (strongly agree). The self-assessment questionnaire was completed at week 16 (end-of-program) in the Feel app.

2.2.7. Mobile app interaction metrics

A wide range of mobile-app-related metrics was collected during the 16 weeks of the program, including participants' responses to emotion notifications, number of times each component of the FP was accessed, exercises completed, time spent in the Feel app, number of weekly sessions with the Feel provider attended, etc.

2.2.8. User feedback survey

After the completion of the program (i.e., at 16 weeks), the participants were asked to complete a feedback survey (administered in the Feel app), in order to identify and measure various aspects of their experience throughout the program. The first part of the survey addressed the overall level of satisfaction with the program. Then, a group of questions helped to assess the ease of use of the different program components (e.g., FES, Feel app, etc.) and the responsiveness of Customer Support. Next, questions on the importance and value of each program component followed. Finally, the participants were given the option to provide open-ended comments on aspects and features of the FP they particularly liked or considered useful, as well as recommendations for improvements. A 5-point Likert-type scale was used with participant responses ranging from 1 (not at all) to 5 (extremely).

2.3. Methods

2.3.1. Participant screening and onboarding flow

The study design was single-arm and was conducted in accordance with the Declaration of Helsinki and was approved

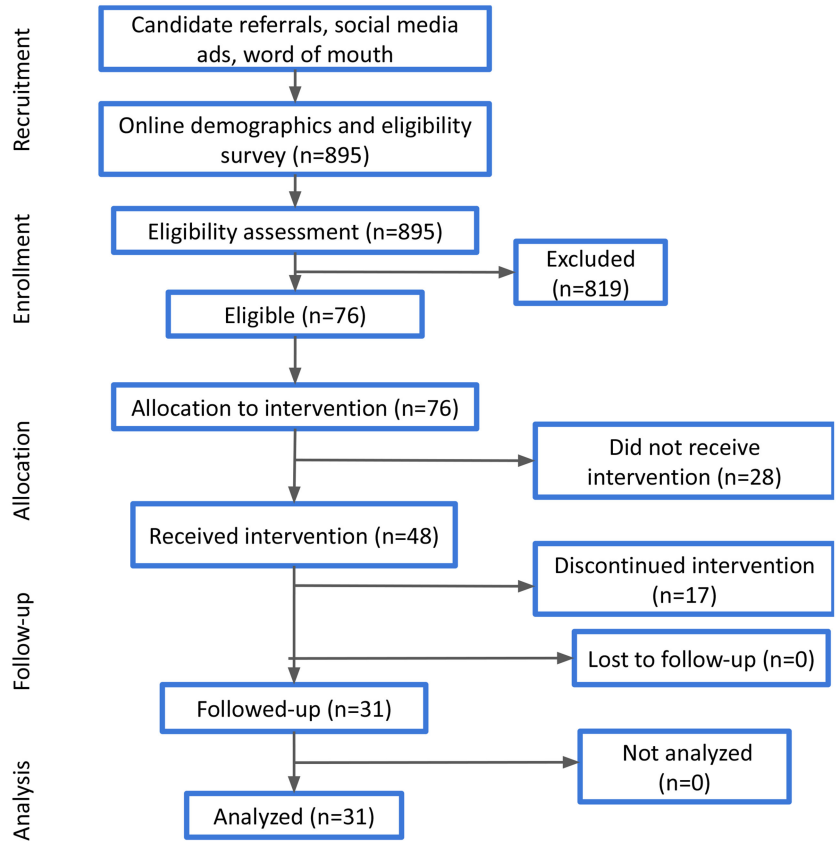


FIGURE 1
Overview of the participant screening and onboarding flow.

by the Ethics Committee of the National and Kapodistrian University of Athens, 1st Department of Psychiatry, Eginition Hospital. The study took place between January 2020 and October 2020. Firstly, individuals interested in participating in the study should complete an online demographic and eligibility questionnaire, as well as the PHQ-9 and GAD-7. The applicants' eligibility was evaluated considering their self-reported survey responses, along with the study's inclusion and exclusion criteria (see Section 2.2). Candidates who fulfilled any of the exclusion criteria were immediately disqualified and received proper communication accordingly. Prior to the start of the program, eligible participants were provided a comprehensive description of the Program and study's scope and components, and scheduled an orientation session with a member of our customer success team that would introduce and onboard them on the program components.

Participants that have accepted the invitation to join the study were provided with a Feel Emotion Sensor, downloaded the Feel app, and attended the orientation session, where they were familiarized with the basic components of the program and the key functionalities of the sensor. As a next step, they had to register to the Feel app and complete the onboarding quests that aim to walk them through the different parts of the program. Among these, participants could check the providers' availability, select their preferred one and book their 16 weekly sessions. All participants who completed these steps finally joined the study. Their progress was monitored at the middle (i.e., after the 8th session) and at the end of the study (i.e., after the 16th session). Participant demographics (e.g., age, gender, etc.) that were acquired at the completion of the online demographic survey were also verified during the orientation session. A high-level presentation of the experimental process is illustrated in Figure 1.

2.3.2. Feel Program

The Feel Program (FP) is a fully-remote mental health support program created by Feel Therapeutics, in an effort to bring objective data and precise therapeutics in the management and treatment of mental health conditions and address the ever-increasing need of people to enhance emotional awareness and self-regulating skills. The program uses a combination of emotion journaling and evidence-based approaches, such as cognitive behavioral therapy (CBT), mindfulness and Positive Psychology (PP) techniques. Furthermore, it is augmented by

FIGURE 2
The Feel Emotion Sensor (FES).

physiological data that reflect the activation of the individual's Autonomic Nervous System and capture mental health and emotion-related information. The program is available as a patient support system in the USA and Europe and is available for download *via* the App Store (iOS) or Play Store (Android). For the present study, the FP focuses on people suffering from depression and/or anxiety and expands over 16 weeks and consists of the following components:

- **Feel Emotion Sensor**: The Feel Emotion Sensor (FES) is a wrist-worn electronic device designed and manufactured by Feel Therapeutics (Figure 2). The device consists of three main sensors for ubiquitous and unobtrusive monitoring of physiological signals: (i) a custom proprietary electrodermal activity (EDA) sensor measuring changes in skin conductance, (ii) an off-the-shelf photoplethysmogram (PPG) sensor for measuring Heart Rate and Heart Rate Variability (HRV) and (iii) an off-the-shelf temperature sensor for measuring skin temperature (ST). In addition, the FES contains a 9-axis Inertial Measurement Unit (IMU) serving as an accelerometer, gyroscope, and magnetometer. It also contains standard equipment adhering to the Bluetooth (BLE) protocol for pairing with a mobile device using the Feel App. Finally, the FES has obtained the following certifications: CE-RED (EN55032, EN61000, EN55035, EN301489, EN62368, EN300328, EN62479), IEC62133, WEEE, RoHs 2.0, FCC ID, FCC sDoc, CEC, US CA Prop 65, BQB/Bluetooth SIG, UN38.3 and MSDS.
- **Feel Mobile App**: The Feel Mobile App is a mobile application available in Android and iOS. It connects to the FES, collects data, and transfers it from the FES to the Feel cloud-based processing infrastructure. The Feel Mobile App helps the participants to onboard the Feel Program, guiding them on how to connect and use the FES, providing information on the program and the theory behind emotion journaling, and facilitating the scheduling of weekly sessions with their providers. Furthermore, the app responds in real-time to changes in the emotional state of the participants and helps them journal the emotions they experience. Access to weekly educational material that explains the evidence-based practices used in the program is offered, as well as default and personalized exercises to practice the various concepts. Self-guided tools such as mood boosters that increase engagement and further help the participants to reach their goals are integrated into the app. Moreover, the clear and coherent program structure and the progress bar within the app ensure that the participants can effortlessly track their progress in the program anytime.
- **Feel Mental Health Biomarkers platform**: The Feel Mental Health Biomarkers platform focuses on the discovery, extraction, and validation of mental health-related biomarkers and metrics for various mental and physical (where comorbid mental health conditions emerge) health use cases. In the context of this study, the Feel Emotion Detection (FED) has been the main functionality that has been deployed. The FED is the backbone of the Feel Program and is based on affective computing technology principles, translating physiological signals (i.e., EDA, HRV, and ST) to emotional events. The platform brings together many different data processing and insights extraction components, including data curation, artifact detection, signal processing and denoising, dynamic segmentation, feature extraction, personalization, and decision models.The FED infrastructure was developed and extensively tested by Feel Therapeutics, is proprietary and protected by U.S. patent (59).
- **Personalized weekly sessions**: The weekly sessions are administered remotely *via* teleconference, by the Feel Providers and have a duration of 15 min (apart from the 1st introductory session that lasts 45 min), augmented by the data provided by the Feel Emotion Sensor and the Feel app. There are a total of 16 sessions with weekly educational material and exercises. The first 3 weeks of the program are foundational and build upon the participants' knowledge. The following 5 weeks focus on CBT, biopsychology and positive psychology and the final weeks build upon the Thoughts, Feelings, Behavior cycle with developing skills and resiliency. The Feel Provider utilizes the Feel Dashboard to access the data to identify themes, key words and behavior patterns to prepare for the session in order to connect-the-dots with the weekly material and exercises.

During the session, the provider is able to target feedback and personalize interventions based on each participant's data. The provider also discusses the participant's next steps after graduation and how they can continue applying what they learned to their daily lives.

- **Mental health resource center**: This component is directly integrated into the Feel mobile app and compiles tutorials, exercises, and tips that focus on the challenges present when dealing with depression and anxiety. The information is designed to engage the participant on their journey and provide the scientific basis for the program interventions and tools. The exercises compliment the material and are customized by the provider based on the participant's data. The participant establishes motivation to engage in the program, set program goals and enhance the knowledge with theoretical frameworks and evidenced-based practices. The educational material and exercises assigned promote mind-body awareness and self-management to further improve the quality of life according to the goals of the participant. The goal is to help participants develop mental health coping skills. All of the material in the mental health resource center has been created by Feel Therapeutics utilizing evidence-based techniques, including CBT, Mindfulness, Biopsychology and Positive Psychology, and is available to the participants anytime *via* the Feel mobile app.

2.3.3. Emotion journaling

The emotion journaling aspect constitutes one of the core components of the data-driven nature of the FP. Thus, an engaging and intuitive journaling user experience has been designed and implemented in the Feel app. This journaling process integrates all the information required both for personalized interventions and for empowering the providers with emotion-related insights and patterns that could be leveraged during the weekly sessions. Additionally, it is used for algorithmic validation purposes, as well as improvements and enhancements of the Feel Biomarkers platform. The emotion journaling process can be either triggered by an emotion notification received by the FES (i.e., FES-triggered) or be manually logged by the participant. For each emotional event detected by the platform, the participant receives a notification in the Feel app to register their response to the detected event. The participant is presented with two high-confidence answers, "accept" or "reject" and with two low-confidence ones, "skip" and "not sure." In case of an accurate detection (i.e., accepted event), the participant is also asked to specify the perceived intensity of the emotional event, ranging from 1 to 10, in the next mandatory step. Furthermore, participants are encouraged to input the triggers, thoughts, behaviors, and physical sensations associated with the specific emotional event in free-form text or voice recording in the following non-mandatory steps. It should be noted that for the emotional events manually registered by the participants (i.e., participant has not received an emotion notification), the journaling procedure is exactly the same, except for the first step requiring the response to a detected event. The complete structure for the accurately detected events involving all user input (both mandatory and not) is referred to as an emotion journal, while the sole completion of the mandatory steps constitutes an emotion log. All emotion logs can be accessed *via* the Feel app anytime throughout the program.

2.3.4. Mid- and end-of-program assessment

Upon the completion of the 8th week of the program, participants were asked to complete a set of questionnaires assessing their depressive and anxiety symptoms, as well as their life satisfaction levels, *via* the Feel app. Finally, at the end of the program participants responded to the same questionnaires, along with the self-assessment and the user feedback survey. No monetary incentive was provided to the participants.

2.3.5. Statistical analysis

Regarding the FP feasibility assessment, the overall onboarding process is summarized and presented, followed by the participant responses to the eligibility survey. Considering that this is a PoC single-arm study that serves as a preliminary evaluation of the participant engagement with the intervention, followed by a preliminary assessment of the efficacy of the intervention, we have followed a per-protocol analysis approach. As stated by the FDA "the use of the per protocol set may maximize the opportunity for a new treatment to show additional efficacy in the analysis, and most closely reflects the scientific model underlying the protocol" (60). Therefore, adopting such an approach provides an opportunity to assess the preliminary effect of receiving the assigned intervention (61), which is the primary focus of this study. In the context, when evaluating engagement or preliminary efficacy aspects, only participants who have completed the study have been considered (i.e., 31 participants). We define a completed participant in our analysis as a participant who has not explicitly requested to withdraw from the study, regardless of their engagement level with the various study components (e.g., assessment surveys, sensor, weekly sessions, etc.).

For the evaluation of engagement metrics, the average values and percentages of participants engaging with the different program components are presented to evaluate engagement aspects. For such metrics, we present the aggregate values over a specified period of time (e.g., day, week, etc.), averaged over the number of participants. As previously discussed, for this analysis only the participants that completed the study have been considered. Finally, aggregate values of the participant

responses to the user feedback survey have been employed to assess program acceptability.

With respect to the preliminary assessment of the intervention effect on depressive and anxiety symptoms, as well as on the participants' quality of life and life satisfaction levels, average values at baseline, mid-program and end-of-program have been used. The statistical significance of the results has been validated using the Wilcoxon signed-rank test and matched pairs rank-biserial correlation (r) for the effect sizes. Minimal clinically important differences (MCID) are defined as at least five points for the PHQ-9 scores (62) and at least four points for the GAD-7 scores (63). According to the per-protocol analysis methodology followed, only participants who have completed the program have been included in the analysis. The missing values accounted for 6.5% of the total assessment survey values that were used in the analysis and were imputed by utilizing the multivariate feature imputation available in the open source scikit-learn python package (64).

Finally, the different types of data collected from the participants during the eligibility, onboarding and assessment process, their physiological data, as well as data from the Feel app are stored and processed in our secure cloud-based infrastructure in Europe. For privacy reasons and in order to adhere to GDPR regulations, all data has been pseudonymized before any processing and insights extraction.

3. Results

3.1. Recruitment, feasibility, and acceptability

During the recruitment and onboarding process, 895 candidates answered the online demographic and eligibility survey, out of which 76 were eligible and considered to receive the intervention, while finally 48 participants actually joined the program. Candidate demographics, as well as responses to the various eligibility questions are presented in Table 1, along with the respective percentages. It should be noted that the overall eligibility ratio is not derived as the combination of the different percentages of the non-eligible responses, as many candidates may have had more than one condition (e.g., severe MDD and eating disorder, psychotic disorder and psychotropic medication). Thus, the overall eligibility ratio was 8.5%.

Out of the 48 participants who were invited to and joined the Program, all of them downloaded the Feel Mobile App, attended the orientation session, scheduled their weekly sessions with the Feel Provider and joined at least one session. Additionally, their mean age was 37.67 (SD = 10.11), with almost 70% of them being 18–40 years old, while their gender distribution was 62.5% females-37.5% males (Table 2). The average baseline PHQ-9 score was 9.1, where 50% of participants had mild depressive symptoms, 39.6% moderate and the rest minimal. Similarly, the average baseline GAD-7 score was 7.7, where 52.1% of

TABLE 1 Candidate responses to the demographic and eligibility questionnaire.

Candidate responses	
Age, *n* (%)	
< 18	17 (1.9%)
18 or older	878 (98.1%)
Gender, *n* (%)	
Female	682 (76.2%)
Male	213 (23.8%)
Income level, *n* (%)	
< 5,000	150 (16.8%)
5,001 – 15,000	243 (27.2%)
15,001 – 25,000	158 (17.6%)
25,001 – 50,000	134 (15%)
> 50,000	38 (4.2%)
No response	172 (19.2%)
Education, *n* (%)	
No schooling completed	23 (2.6%)
High school graduate	260 (29%)
Bachelor's degree	170 (19%)
Master's degree	132 (14.7%)
Professional degree	254 (28.4%)
Doctorate degree	14 (1.6%)
No response	42 (4.7%)
Exclusion disorders*, *n* (%)	
Yes	220 (24.6%)
No	675 (75.4%)
Psychotropic medication, *n* (%)	
Yes	324 (36.2%)
No	571 (63.8%)
Suicidal or self-harm thoughts, *n* (%)	
Yes	128 (14.3%)
No	767 (85.7%)
Substance abuse, *n* (%)	
Yes	63 (7%)
No	832 (93%)
MDD severity, *n* (%)	
Minimal	105 (11.7%)
Mild-Moderate	464 (51.8%)
Moderately Severe-Severe	278 (31.1%)
No response	48 (5.4%)
GAD severity, *n* (%)	
Minimal	123 (13.8%)
Mild-Moderate	517 (57.8%)
Severe	200 (22.3%)
No response	55 (6.1%)

*Exclusion disorders were: personality disorders, psychotic disorders, bipolar disorder, eating disorders.

TABLE 2 Participant demographic characteristics and baseline assessment scores.

Participant characteristics	
Age, *n* (%)	
18 – 30	13/48 (27.1%)
31 – 40	20/48 (41.7%)
41 – 50	9/48 (18.7%)
Older than 50	6/48 (12.5%)
Gender, *n* (%)	
Female	30/48 (62.5%)
Male	18/48 (37.5%)
MDD symptom severity, *n* (%)	
Minimal	5 (10.4%)
Mild	24 (50%)
Moderate	19 (39.6%)
GAD symptom severity, *n* (%)	
Minimal	9 (18.7%)
Mild	25 (52.1%)
Moderate	14 (29.2%)

participants had mild anxiety symptoms, 29.2% moderate and the remaining minimal. Moreover, 31 participants completed the Program, while 17 discontinued for various reasons, ranging from technical challenges when using the FES or the app to unwillingness and limited time availability to commit to the program.

Participant responses to the user feedback survey are presented in Table 3, with a 5-point Likert-type scale being used, where it becomes evident that the overall participant satisfaction levels are quite high with an average of 4 out of 5. More than 65% of participants are reporting at least very high satisfaction levels. Similarly, the average Net Promoter Score is also 4 out of 5, as it would be highly or extremely likely for close to 80% of participants to recommend the FP to someone they know. Regarding the usability of the different program components, 70% of participants found it easy to use the Feel Mobile app, while only 15% of participants seem to have had faced technical difficulties engaging with the FES. The level of responsiveness to questions or concerns about the FP (i.e., Customer Support) has received a remarkable 4.5 out 5 average rating with close to 90% of participants being very or extremely satisfied. Furthermore, regarding the importance of the different program components, the personalized data-driven sessions stand out with over 80% of participants perceiving them as very important, while 70% of participants identify the FES as a very important program component. Finally, participants identified 1) the emotion journaling flow, 2) the integration of their physiological data into the Program guideline and 3) the in-the moment interventions as the most important features of the program. On the other hand, the option to visualize the collected data (e.g., heart rate), the sensor ergonomics and battery life, along with the enhancement of the interactive material (e.g., weekly exercises) have emerged as the features that could be improved or added to the program.

3.2. Ecological validity of the Feel Emotion Detection

As previously discussed, among the primary aims of this study was to validate in-the-wild the performance of the FED that focuses on the identification of the significant emotional moments participants have been experiencing. In this context, an analysis of the FES-triggered emotion logs follows, considering the participants' responses (positive or negative) on the notifications sent, as well as the perceived intensity input by the participants during the emotion logging flow. Overall, the average precision levels, that reflect the ratio of total accepted notifications of the study to the sum of the accepted and rejected, were 87%. On an individual level, the mean precision among the participants was 88% with a standard deviation of 0.2. Moreover, when weighing each participant's precision with the total number of notifications, the obtained (weighted) average precision was 86%. Furthermore, it should be noted that 75% of the participants had a precision of at least 85% (25^{th} percentile). The combination of the above supports the capability of FED to correctly identify the participants' emotional events in the wild.

Furthermore, a balanced distribution between positive and negative emotional events was observed (54 vs. 46%, respectively) indicating minimum bias toward the detection of emotions of a particular valence. More specifically, when referring to the positive valence emotions, study participants logged 32% "content" and 22% "joyous" emotions, while for the negative valence ones, they logged 18% "sad" and 28% "distressed" (Figure 3). Meanwhile, regarding the participant-perceived emotion intensity, the mean intensity level logged was 6. Considering that participants may label their emotion intensity at a scale of 1 to 10, an intensity threshold at the midpoint of the scale (i.e., 5) has been selected, with emotional events rated 6 or higher perceived as high intensity, while the ones logged with 1 to 5 as low intensity. During the study, high intensity emotion logs constituted more than 60% of the total FES-triggered emotion logs. Furthermore, the mean intensity level of the accurately detected events for the majority ($>$ 70%) of the participants was at least 6. The last two observations support that the FES captures the higher intensity (and perhaps more meaningful events, at least as experienced by each participant) with a great degree of uniformity across multiple individuals.

TABLE 3 Participant responses to the user feedback survey.

Survey questions	Participant responses, *n* (%)
Participant satisfaction	
Overall, how satisfied are you with the Feel Program?	Extremely: 11 (36.7%) Very: 9 (30%) Neutral: 9 (30%) Slightly: 1 (3.3%) Not at all: 0 (-)
How likely are you to recommend the Feel Program to someone?	Extremely: 10 (33.3%) Very: 14 (46.7%) Neutral: 3 (10%) Slightly: 3 (10%) Not at all: 0 (-)
Program components usability	
How easy is it to navigate the Feel app?	Extremely: 9 (30%) Very: 12 (40%) Neutral: 3 (10%) Slightly: 5 (16.7%) Not at all: 1 (3.3%)
How easy is it to use the Feel emotion sensor?	Extremely: 6 (20%) Very: 7 (23.3%) Neutral: 12 (30%) Slightly: 3 (10%) Not at all: 2 (6.7%)
Customer support	
How responsive have we been to your questions or concerns about the Feel program?	Extremely: 20 (66.7%) Very: 6 (20%) Neutral: 4 (13.3%) Slightly: 0 (-) Not at all: 0 (-)
Program components importance	
Feel Emotion Sensor	Extremely: 14 (46.7%) Very: 7 (23.3%) Neutral: 6 (20%) Slightly: 1 (3.3%) Not at all: 2 (6.7&)
Feel Mobile App	Extremely: 14 (46.7%) Very: 7 (23.3%) Neutral: 3 (10%) Slightly: 6 (20%) Not at all: 0 (-)
Mental Health Resource Center	Extremely: 14 (46.7%) Very: 13 (43.3%) Neutral: 3 (10%) Slightly: 0 (-) Not at all: 0 (-)
Personalized Data-driven Sessions	Extremely: 22 (73.3%) Very: 3 (10%) Neutral: 5 (16.7%) Slightly: 0(-) Not at all: 0 (-)
Open-ended feedback*	
Features participants particularly liked	Emotion journaling flow: 15 (68.2%) Data integration: 12 (54.5%) In-the-moment interventions: 8 (36.4%)
Features to improve/add	Physiological data visualization: 8 (36.4%) Sensor ergonomics or battery life: 7 (31.8%) Enhance interactive material: 5 (22.7%)

*Participants could select up to 3 features, so percentages do not add up to 100%.

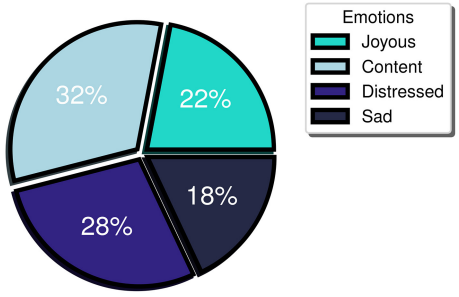


FIGURE 3
Distribution of emotions logged by type of emotion.

3.3. Participant retention and engagement in the FP

In order to measure and extract meaningful insights associated with the participant retention levels in the FP, we divide the FP into four parts (modules), each one consisting of four personalized weekly sessions spread across a time period of 1 month. The monthly retention rate is illustrated in Figure 4. About 25% of the onboarded participants discontinued the program during the first month, while approximately 15% of the ones who went through the 1st month discontinued during the second one. Then, 100% of participants who completed the first 8 weekly sessions (i.e., the second monthly module) continued toward full completion of the FP. In other words, zero dropouts were observed after the second month. The overall retention throughout the study was 65%. The metrics presented in the remainder of this section refer to the participants that completed the study ($n = 31$, 65% of the total participants).

Regarding participant engagement, we gauge it by metrics that capture all the different aspects of the FP. We define as "active," a completed participant that shows engagement with

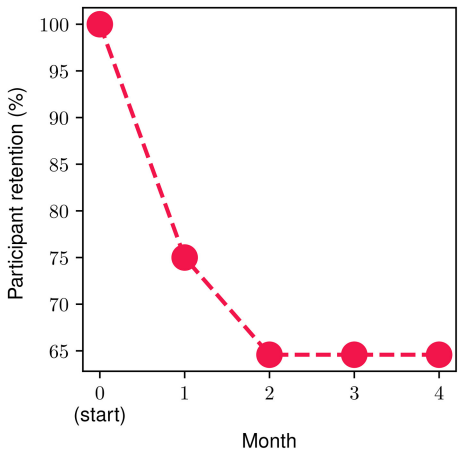


FIGURE 4
Participant monthly retention during the program.

one or more of the study components (e.g., FES, Mental Health resources, etc.). With respect to the overall activity in the FP during the study, it was observed that participants engaged with any component of the FP on average for 3.1 days per week, with an average of 96% of them being active on a weekly basis. The participants spent close to 60 min per week (*SD* = 10.3) in the Feel Mobile app and accessed it on average for 16.5 times throughout the week which translates to roughly 2.5 times per day. Moreover, regarding the weekly data-driven sessions, we observed that the session compliance ratio was 96.9%, while on average, the participants completed 5 mental health exercises/educational material provided *via* the mental health resource center of the mobile app per week. Aiming to capture the overall participant engagement and their progress/efforts toward improving their mental health, an aggregate engagement metric, the weekly Mental Health Actions (MHA), is introduced. This metric includes the number of emotion logs, completed mental health exercises/educational material, sessions with the Provider attended, as well as whether the participant has engaged with the sensor or not and is calculated on a weekly basis. During this study, participants' MHA reached an average of 13 actions per week.

With respect to the FES, we found out that on average, participants engaged with the FES for 74.5% of their time throughout the FP. In more detail, Figure 5 depicts the percentage of participants that engaged with the FES on a weekly basis throughout the FP. At the beginning of the FP (i.e., week 1), it can be noticed that almost all of the participants (more than 95%) engaged with the FES. Then, a slight decline can be observed during the next 2 weeks resulting in approximately 80% of the participants engaging with the FES by the end of Week 4. During the next 2 months of the FP (i.e., week 5-week 12), the proportion of participants attains an average value of 75.4% with a corresponding standard deviation of only 3.2%. The relatively high engagement ratio of 75.4% slightly drops to an average of 68.5% during the last month of the FP which can be attributed to a relatively reduced engagement level, as the study completion time point approaches. Overall, the mean weekly participant engagement with the FES was 76.8% throughout the study.

Regarding the emotion journaling activity throughout the program, we investigated the average number of total emotion logs (both FES-triggered and manually logged) registered by the participants. More specifically, in Figure 6 we present the number of total emotion logs for each of the 16 weeks of the FP, averaged over the number of the study participants during each week. The results indicate that during the FP, the average number of total emotion logs ranged from 3.65 to 6.4, with a mean value of 4.92 emotion logs per week, while 91.3% of them are also journalled (see Emotion Journaling in Section 2.3). In total, close to 2000 emotional moments have been logged during the 16 weeks of the study, with only 13.7% corresponding to days that participants did not engage with the FES.

3.4. Preliminary assessment of impact on mental health symptoms

Figure 7 showcases the mean scores from the mental-health-related questionnaires at baseline, mid-program (week 8) and end-of-program (week 16) evaluations. For both PHQ-9 (Figure 7A) and GAD-7 (Figure 7B) questionnaires, a decrease in mean participant scores can be observed, suggesting an important improvement in depressive and anxiety symptoms. More specifically, for the PHQ-9 assessment, the mean baseline score was 8.23 (*SD* = 3.44), while for the mid-program evaluation the obtained mean reduces to 6.23 (*SD* = 3.15), and reaches an average of 3.76 (*SD* = 2.48) at the end of the program. Similarly, for the GAD-7 questionnaire, at baseline the mean score was 7.30 (*SD* = 3.68), reducing to an average of 4.26 (*SD* = 2.08) at the mid-program assessment, before reaching a mean value of 3.3 (*SD* = 2.18) at the end of the program. These results suggest that the overall average mental health symptom reduction throughout the FP was 54.3 % for the depressive symptoms and 54.8% for the anxiety ones. The reduction was statistically significant in both cases (Wilcoxon signed-rank test with $p < 10^{-4}$ and effect size r equals to 0.93 and 0.98 for the PHQ-9 and GAD-7 scores, respectively). More specifically, 87% of participants exhibited a reduction in depressive symptoms, while 83.8% had a reduction in anxiety symptoms. Additionally, 93.5% of participants presented a decrease in at least one the two symptom categories (i.e., PHQ-9 or GAD-7), while 77.4% of them showed a decrease in both of them. Referring to participants exhibiting clinically significant symptom improvement (62, 63), 51.6%

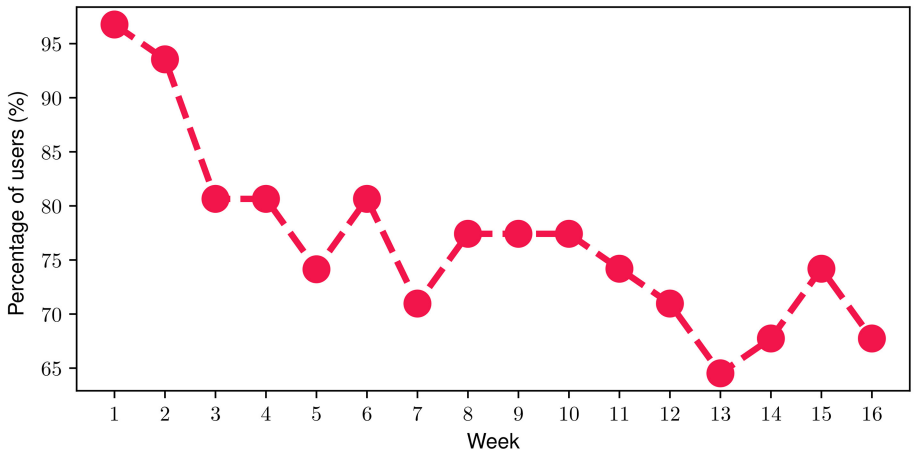


FIGURE 5
Percentage of participants engaging with the FES during each week of the FP.

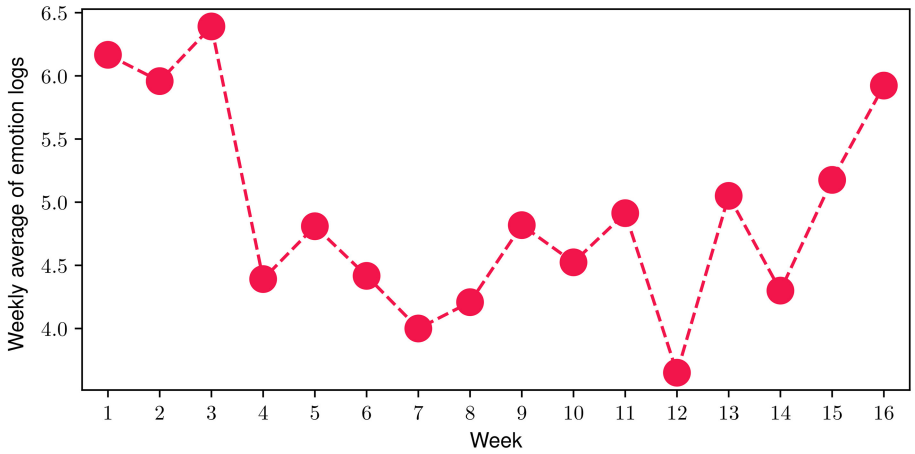


FIGURE 6
Weekly average of FES-triggered and manual emotion logs during the FP.

of them showed improvement in depressive symptoms and 45% in anxiety symptoms. Finally, 74.2% of participants had improved by at least one severity level (e.g., changed from moderate to mild) in depressive symptoms and 71% in anxiety symptoms.

3.5. Preliminary assessment of impact on quality of life results

In Figure 8, we present the mean scores for the SWLS (Figure 8A) and LISAT-11 (Figure 8B) questionnaires regarding quality of life aspects at the baseline, mid-program and end-of-program evaluations. An increase in both scores can be observed throughout the FP, with a 24% and 15% overall improvement for SWLS and LISAT-11 mean scores, accordingly. The reduction was statistically significant in both cases (Wilcoxon signed-rank test with $p < 10^{-4}$ and effect size r equals 0.95 and 0.96 for the LISAT-11 and *SWLS* scores, respectively). More specifically, at the baseline evaluation, the mean SWLS score was 19.9 (*SD* = 5.42), at mid-program it increased to 23.8 (*SD* = 5.76) while at the end-of-program evaluation, it reached 24.7 (*SD* = 4.71). At the same time, the corresponding scores for the LISAT-11 questionnaire were at baseline 3.86 (*SD* = 0.59), at mid-program 4.24 (*SD* = 0.68) and at the end-of-program evaluation 4.45 (*SD* = 0.54). Finally, 80.7% of participants demonstrated an increase at the SWLS scores throughout the program, while 71.4% showed an increase at the LISAT-11 scores.

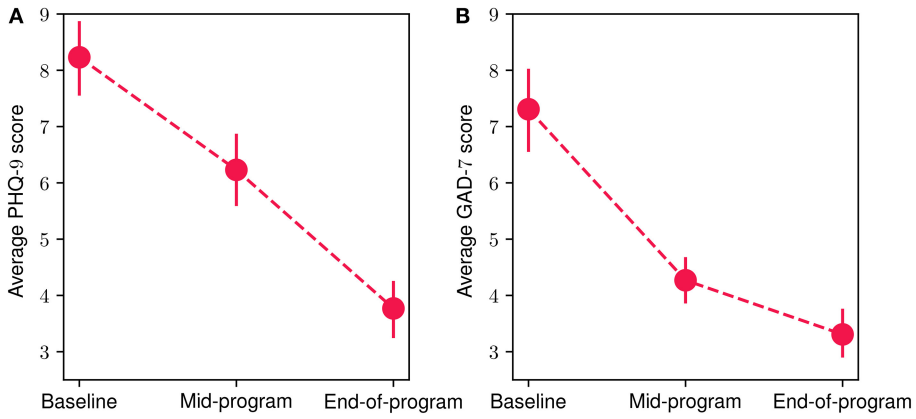


FIGURE 7
Mean participant PHQ-9 **(A)** and GAD-7 **(B)** scores at baseline, mid-program and end-of-program evaluations. The vertical bars represent the standard error.

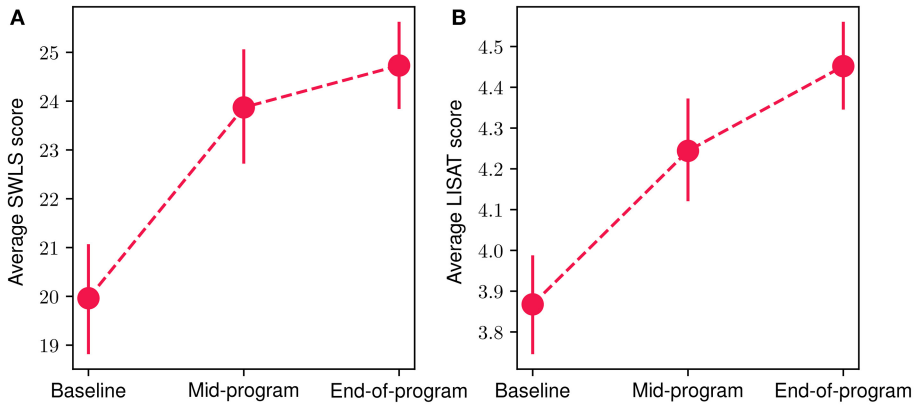


FIGURE 8
Mean participant SWLS **(A)** and LISAT-11 **(B)** scores at baseline, mid-program and end-of-program evaluations. The vertical bars represent the standard error.

3.6. Preliminary assessment of participant self-assessment results

An additional tool used to assess the impact participants perceive that the FP has had, along with their accomplishments during the program, was the participant self-assessment questionnaire. Participant responses, following a Likert-type scale, are presented in Table 4. Overall, it is evident that participants anticipate that the program has had an important impact on them, as 100% of respondents (30 participants) state that the concerns that led them to the program have improved during their participation. Additionally, almost 97% of participants feel that they have made progress toward the goal they had set at the beginning of the program, while the same percentage expressed that their everyday lives have improved. Finally, all participants responded that they learnt to think more clearly to reduce distressing emotions/behaviors and more than 85% of them increased their ability to recognize, name, and/or appropriately express their emotions.

4. Discussion

4.1. Key findings

In this study, we have presented the results of a RWD feasibility study that serves as a PoC for a digital data-driven mental health program (i.e., the Feel Program) for people suffering from MDD and/or GAD. In this context, an experimental protocol involving the deployment of the 16-week Feel Program at a population with mild or moderate MDD/GAD, has been designed and executed. Our first aim was to explore the feasibility of such a program, as captured by the

TABLE 4 Participant responses to the self-assessment questionnaire.

Survey questions	Participant responses, *n* (%)
My concerns that brought me to the program have improved as a result of the services provided.	Strongly agree: 17 (56.7%) Agree: 13 (43.3%) Neither disagree nor agree: 0 (-) Disagree: 0 (-) Strongly disagree: 0 (-)
I feel I made progress toward my set goal.	Strongly agree: 18 (60%) Agree: 11 (36.7%) Neither disagree nor agree: 1 (3.3%) Disagree: 0 (-) Strongly disagree: 0 (-)
My everyday life has improved.	Strongly agree: 15 (50%) Agree: 14 (46.7%) Neither disagree nor agree: 1 (3.3%) Disagree: 0 (-) Strongly disagree: 0 (-)
I learned to think more clearly/accurately to reduce distressing emotions or behaviors.	Strongly agree: 22 (73.3%) Agree: 8 (26.7%) Neither disagree nor agree: 0 (-) Disagree: 0 (-) Strongly disagree: 0 (-)
I increased my ability to recognize, name, and/or appropriately express my emotions.	Strongly agree: 19 (63.3%) Agree: 7 (23.4%) Neither disagree nor agree: 3 (10%) Disagree: 1 (3.3%) Strongly disagree: 0 (-)

responses of potential participants' to the recruitment campaign. We, therefore, implemented a relatively broad campaign, with the aim of exploring the various personas, mental health conditions and demographic profiles of individuals that were more interested in the program and, therefore, in joining the study. At the same time, the study protocol did include a set of strict inclusion and exclusion criteria (e.g., exclusion of specific mental health disorders, no psychotropic medication, etc.), ensuring that only participant profiles aligned with the study scope were finally enrolled. Therefore, the recruitment campaign attracted a high number of respondents to the demographics and eligibility questionnaire, indicating that there is a great need for mental health support resources for numerous conditions. The combination of the high number of responses to the broad recruitment campaign and the presented exclusion and inclusion criteria led to an eligibility ratio of 8.5%. In particular, a high prevalence of the exclusion disorders (i.e., personality disorders, psychotic disorders, bipolar disorder, eating disorders) was observed, with 3 out of 4 applicants reporting the existence of at least one of the above. Additionally, we noticed that the vast majority of respondents were females (76.2%), frequently low-to-mid income level (44% with less than 15,000 annual income) and often with a basic level of education (32% high-school graduates or no schooling completed). This is in line with findings from previous studies (18, 65) that report that income and education levels can limit access to mental health services. Finally, it should be highlighted that 1 out of 3 respondents reported moderately severe or severe depressive symptoms and 1 out 4 severe anxiety symptoms. The gender distribution of the eligible participants that joined the study was more balanced, with 62.5% female participants and 37.5% males, and the $31-40$ age group has the highest proportion of participants (41.7%). Overall, 70% of participants were aged 40 years or younger, while interestingly 12.5% were older than 50, suggesting that technology-driven programs could be attractive to older adults too. Finally, the distribution of baseline symptom severity was well balanced with regards to MDD with 50 and 40% of participants exhibiting mild and moderate depressive symptoms at the beginning of the study respectively, and 50 and 30% reporting mild and moderate anxiety symptoms, respectively.

The second aim of the present study, after its feasibility was established, was to explore the main hypothesis which revolved around the fact that a remote, data-driven and personalized program would boost the participant engagement, while a significant improvement of their depressive and anxiety symptoms would be observed. In order to support our hypothesis, our analysis focuses on three main sections: (i) in-the-wild Feel technology validation; (ii) acceptability and participant engagement and (iii) preliminary assessment of impact on depressive and anxiety symptoms, as well as on quality of life aspects. In the following, we discuss in more detail the key findings of our investigation with regards to these aspects. In the context of technology validation, we aimed to demonstrate that the enabler of the data-driven nature of the FP, the Feel Emotion Detection technology, can be deployed in real-world settings maintaining similar performance levels compared to our testing environments. Successful deployment of technology is dependent on; (i) system-design-specific factors impacting performance (e.g., high precision and minimal bias toward specific emotions or individuals), and (ii) successfully withstanding challenges associated with real-world applications, such as unknown emotional stimuli, noise signal artifacts and stochastic participant behavior. The results of this study show a very high average precision level (87%) in identifying emotional events for this group of individuals suffering from MDD and/or GAD. Moreover, positive and negative emotional events were detected in a balanced (54 and 46% respectively), suggesting that there was a minimal bias toward either valence category. Finally, we have shown that, for the majority of the participants ($> 70\%$), their FES-triggered events were registered on average as high-intensity (i.e., with a participant-perceived intensity level ≥ 6). These findings indicate that the Feel Emotion Detection technology performs well enough to capture significant emotional events from any participant.

The next step toward validating our initial hypothesis includes assessing the acceptability of and participant engagement with the FP. Regarding the former, participant responses to the user feedback survey collected after the completion of the 16-week study suggest that participant satisfaction levels were considerably high with over 65% (20 participants) reporting very or extremely high satisfaction levels. Additionally, 4 out of 5 participants were at least very likely to recommend the program to someone. These findings imply that the FP sufficiently addressed their needs and met their expectations. Furthermore, the data-driven nature of the program—in the form of the data-driven weekly sessions with the provider and the FES—was recognized in the survey responses as one of the most important features of the program. Specifically, more than 90% of participants identified the weekly data-driven sessions as the most important feature of the program, while 70% selected the FES. Participants also reported particularly liking the data integration to the program and the in-the-moment interventions, powered by the FES-triggered emotion logs.

Regarding participant engagement, we focused on participant retention levels throughout the program, as well as the degree of engagement with the different components. An overall 65% retention rate was observed with most discontinuations from the program occurring during its 1st month while no participants dropped out after the midpoint of the program. This could be related to the initial introductory period required for participants to acquaint themselves with the various program components, and the level of commitment, effort, and time resources required. Equally, this could possibly be attributed to the structure of the FP and its use of the theoretical framework of the Stages of Change (i.e., Contemplation, Preparation, Action) (66). During the first introductory session, the participant sets their program goals and then when they are faced with taking action, they either regress into Contemplation stage and drop out of the program or progress into Action stage and increase their engagement in the program.

Turning our attention toward participant engagement, a wide range of extracted metrics referring to all program components indicate high engagement levels. Overall, completed participants spent on average 60 min per week in the Feel app, accessing it on average 2.5 times per day. This equated to an interaction with at least one of the FP components on average for approximately 3.1 days per week. Additionally, the term weekly Mental Health Actions was introduced as a metric to capture participants' weekly activity during the week which reached an average of 13 actions per week for the completed ones. The highest amount of activity was the on average 5 mental health exercises accessed *via* the mental health resource center per week, followed by the emotion journaling feature, with completed participants registering an average of 4.92 emotional events, either FES-triggered or manually input, per week. Finally, a very high rate of attendance to the weekly sessions with the provider (96.9%) was also observed.

With respect to the main data-driver in the FP, the FES, completed participants engaged with it for 74.5% of the time while they were on the program, with an average of 76.8% of them using the sensor on a weekly basis. An interesting observation was that increased engagement with the FES boosted overall completed participant engagement in terms of both increased FES-triggered emotion logs, and also increased manually registered emotional events. The former seems quite intuitive, as the more physiological data are available, the more events can be detected. However, what is particularly interesting is that increased FES engagement was also highly associated with a higher number of the manual logs. More specifically, we observed that on the days that completed participants engaged with the FES, the number of manual emotion logs more than doubled, compared to the days without any FES engagement. The contribution of the FES to emotion journaling can be quantified by an approximate 5.5-fold increase in the probability of an emotion log occurring on days with FES engagement. The FES is considered a driving factor for overall participant engagement with the program, since it is strongly associated with the majority of the emotion logs, which are in turn linked with the various other components of the FP (e.g., data-driven weekly sessions, app usage, weekly exercises, etc.).

Having discussed the in-the-wild performance of the technology and the fact that the data-driven nature of the program significantly enhances engagement, we lastly focused on depressive and anxiety symptoms, as well as on quality of life levels. For both MDD and GAD symptoms—reflected by PHQ-9 and GAD-7 scores, respectively—a continuing decrease in average scores was observed from baseline to mid-program, and all the way to the end-of-program evaluations, with 9 out of 10 participants reporting symptom reduction for at least one symptom category. Overall, the average reduction in both scales during the 16-week program was over 50%. For almost 3 out of 4 participants the severity of their symptoms decreased by at least one severity level, while a clinically significant improvement was characterized for almost half of the participants. This symptom improvement was followed by a subsequent increase of 24% on average in quality of life metrics. It should be noted that this study did not include participants with severe depression or anxiety symptoms, in order to exclude the effect of pharmacotherapy on symptom improvement. We anticipate that the inclusion of more severe cases will significantly enhance the total average reduction of mental health symptoms in a 16-week program like FP (67).

In summary, the study results support that very high levels of participant engagement with a 16-week personalized data-driven digital mental health program, with its data-driven nature—coming from the use of a wearable sensor—being a key engagement factor. Moreover, the increased participant engagement levels and the data enhancement of the provided

TABLE 5 Highlights of the study.

Key outcomes	
Engagement metrics	
Users active during the week	96%
Users engaging with the FES during the week	76.8%
Weekly time in app	60 min
Weekly mental health actions	13
Weekly mental health exercises	5
Weekly emotion logs	4.92
Average notification precision	87%
Session with the provider compliance	96.9%
Impact on mental health outcomes	
Average improvement of depressive/anxiety symptoms	54.3%/54.8%
Participants with clinically significant depressive/anxiety symptom improvement	51.6%/45%
Participants with improvement in at least one the two symptom categories	93.5%
Participants with improvement in both symptom categories	77.4%
Participants that improved by at least one severity level in depressive/anxiety symptoms	74.2%/71%

mental health support, may be linked to a significant reduction of depressive and anxiety symptoms, as well as an improvement in quality of life. Table 5 summarizes the highlights of this study.

4.2. Further observations

Having discussed our initial hypothesis which evolved around the feasibility, acceptability, engagement and potential impact on mental health symptoms of the FP, a few interesting observations are presented. Considering that the emotion journaling aspect constitutes a core component of the program, we introduced an additional metric—the mood index—to capture the progression of participants' emotional patterns throughout the program. This new metric is derived from a combination of the different positive and negative emotions logged by the participants, where each emotion valence category is mapped accordingly to a positive (+1) or negative (−1) value, accordingly. The daily value is then derived from the average value of all emotions logged during each particular day. Figure 9 illustrates the 30-day centered rolling average of the mood index for the duration of the study. A very interesting observation is that the average values of the mood index reach very low negative values when participants join the study, indicating that the vast majority of emotions experienced at the start of the study are negative. However, a steadily increasing trend is observed from the beginning of the program, which aligns with the above-mentioned improvement in depressive and anxiety symptoms and quality of life. Close to the 40 day time point, the index transitions into a positive value, where it stabilizes for the rest of the program.

As an additional step toward exploring the importance of the data-driven component of the FP, we explored the level of engagement of different participant cohorts based on their baseline symptom severity. In this context, the average number of emotion logs per day was selected as a measure of engagement, as it constitutes one of the most important engagement metrics. In Figure 10, the average engagement levels throughout the study for participants with mild and moderate symptom severity at baseline are presented. We should highlight that for the depressive symptoms cohort, a clear difference between the two groups can be noticed, with the moderate severity group exhibiting a more than 50% higher engagement compared to the mild severity (0.83 vs. 0.54). When grouping participants according to their anxiety symptom severity however, no significant differences in engagement between the two groups were observed (0.63 vs. 0.66). These observations may serve as an indicator of greater participant engagement when symptom severity is higher, especially for those with depressive symptoms.

Finally, we explored the effects of differing engagement levels (i.e., low and high engagement) on the program's impact on symptom severity. By utilizing the aforementioned index of engagement, we assigned participants to the high (or low) engagement group, if they had more (or less) than 1 emotion log every two days. This threshold was selected based on the average weekly rate of emotion logs for all participants. In Figure 11, we present the percentage reduction in depression and anxiety symptom severity for the two engagement cohorts. It is notable that the high engagement cohort shows a more than 70% greater improvement in depressive symptoms compared to the low engagement cohort (60.8 vs. 35.4%). The outcomes are more balanced for the case of anxiety symptoms, where the high engagement cohort presents a 10% higher improvement compared to the low engagement (54 vs. 49.2%). Altogether, these findings serve as an indication that increased engagement with the program, as reflected by participants' emotion logging activity, seems to be associated with greater improvement of depressive and anxiety symptoms.

4.3. Limitations

When interpreting the results presented in such a paper, study limitations should always be considered. The most important among these is the lack of a control group, which could provide further evidence as to whether the overall symptom and quality of life improvement can be attributed to the intervention or not. Secondly, the sample size of the study was relatively small with 31 participants actually completing the program. Additionally, the different cohorts include a small

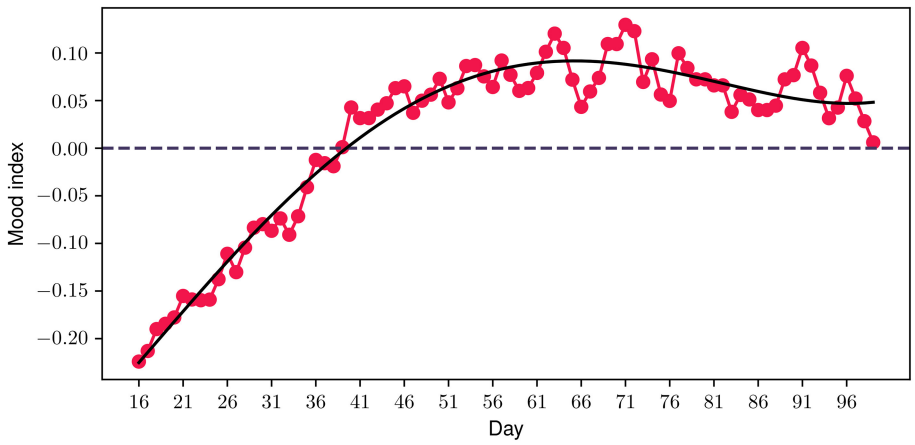


FIGURE 9
Participant mood index progression during the program. The solid black line represents a polynomial fit.

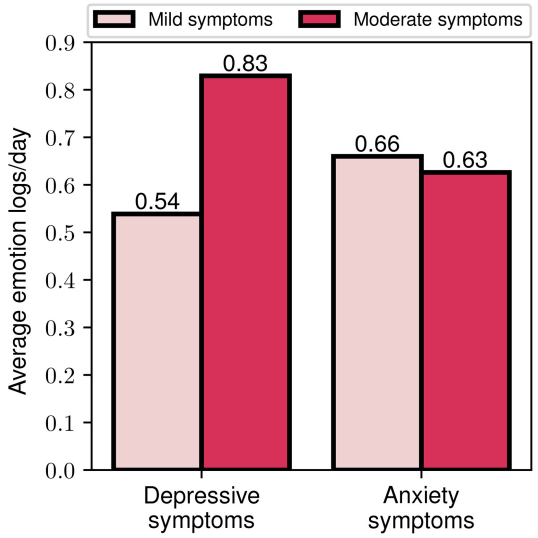


FIGURE 10
Average emotion logs per day for mild and moderate depressive and anxiety symptoms baseline assessment.

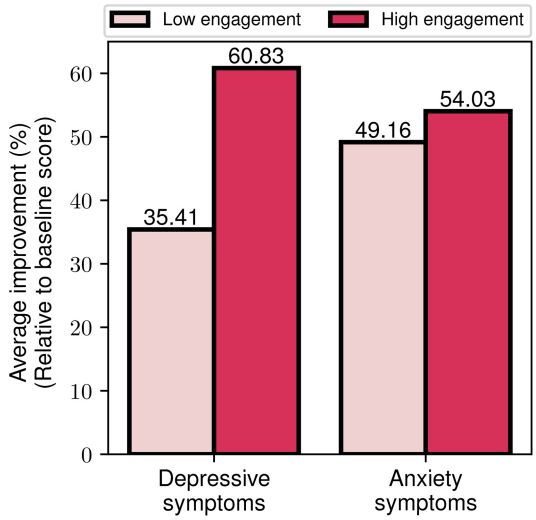


FIGURE 11
Average depressive and anxiety symptom improvement for low and high engagement participant groups.

number of participants, so the respective insights should be treated with caution. Thirdly, a more diverse demographic group (e.g., expanded age groups, balanced gender groups, racially diverse groups, etc.) of participants could be included in future studies, in order to generalize the findings to a general demographic audience. Additionally, the study includes only participants with mild and moderate depressive and anxiety symptoms severity. Examining the impact of the program on severe cases in a follow-up study would be valuable. Finally, the level of mobile device proficiency was not monitored and its impact on participants' engagement was not assessed.

4.4. Conclusions

In summary, this study aimed to explore the hypothesis that a digital, biodata-driven mental health program, the Feel Program, would introduce multiple benefits for people suffering from mild or moderate MDD and/or GAD. Therefore, we first focused on providing evidence supporting the feasibility of this program. Feedback from the participants revealed very high satisfaction scores across the different components of the program, and highlighted the significance of their data-driven nature. Moreover, overall completed participant engagement remained at high levels throughout the program and was

significantly boosted by the use of the sensor. Additionally, we have shown that the key differentiating program component—the Feel Emotion Sensor and the Feel Emotion Detection—can be successfully deployed in-the-wild and can accurately detect significant emotional events. Overall, participants exhibited significant improvement on depressive and anxiety symptoms, with almost all of them showing symptom reduction at the end of the study. However, a controlled trial will be required to demonstrate further evidence of the clinical efficacy of the program. Finally, our indicators regarding the value of using personalized data - mainly collected by the FES—suggest enhanced engagement for groups with more severe depressive symptoms, as well as greater symptom improvement in participants with higher engagement throughout the program.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding authors.

Ethics statement

The studies involving human participants were reviewed and approved by the Ethics Committee of the National Kapodistrian University of Athens, First Department of Psychiatry, Eginition Hospital. The patients/participants provided their written informed consent to participate in this study.

Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

Conflict of interest

Authors CT and GE are employed by Feel Therapeutics Inc., receive a salary and own a large share of the company stocks. DA and PF are employed by Feel Therapeutics Inc., receive a salary and own options of the company. JA is an advisor to and holds options of the company.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

5. Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Anxiety and Depression Association of America (ADAA). *Understand Anxiety and Depression: Facts and Statistics* (2022). Available online at: https://adaa.org/understanding-anxiety/facts-statisticshere (accessed on January 22, 2022).

2. OECD. Promoting mental health in Europe: Why and how?. In: *Health at a Glance: Europe 2018: State of Health in the EU Cycle*. Paris: OECD Publishing. (2018). doi: 10.1787/health_glance_eur-2018-4-en

3. Jorge A, Gemma V, Amy R, Antoni SB, Vicente M, Michele P, et al. Prevalence and variability of current depressive disorder in 27 European countries: a population-based study. *Lancet Public Health*. (2021) 60:e729–38. doi: 10.1016/S2468-2667(21)00047-5

4. World Health Organization. *Depression in Europe: Facts and Figures* (2022). Available online at: https://www.euro.who.int/en/health-topics/noncommunicable-diseases/mental-health/news/news/2012/10/depression-in-europe/depression-in-europe-facts-and-figureshere; https://www.euro.who.int/en/health-topics/noncommunicable-diseases/mental-health/news/news/2012/10/depression-in-europe/depression-in-europe-facts-and-figures

5. World Health Organization. *Disability-Adjusted Life Years (DALYs)* (2013). Available online at: https://www.who.int/data/gho/indicator-metadata-registry/imr-details/158here (accessed January 16, 2022).

6. World Health Organization. *Depression and Other Common Mental Disorders: Global Health Estimates* (2017). Available online at: https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdhere (accessed on January 16, 2022).

7. OECD. Assessment and conclusions - The OECD mental health and work policy framework. In: *FitMind, Fit Job: From Evidence to Practice inMental Health and Work*. Paris: OECD Publishing (2015). doi: 10.1787/9789264228283-4-en

8. OECD. *Fitter Minds, Fitter Jobs* (2021). Available online at: https://www.oecdilibrary.org/content/publication/a0815d0f-en

9. World Health Organization. *Mental Health and Substance Use* (2022). Available online at: https://www.who.int/teams/mental-health-and-substance-use/promotion-prevention/mental-health-in-the-workplacehere (accessed on January 16, 2022).

10. OECD. *Mental Health and Work* (2021). Available online at: https://www.oecd.org/health/mental-health-and-work.htmlhere (accessed on January 12, 2022).

11. Weinberger AH, Gbedemah M, Martinez AM, Nash D, Galea S, Goodwin RD. Trends in depression prevalence in the USA from 2005 to 2015: widening disparities in vulnerable groups. *Psychol Med*. (2018) 48:1308–15. doi: 10.1017/S0033291717002781

12. Mojtabai R, Olfson M, Han B. National trends in the prevalence and treatment of depression in adolescents and young adults. *Pediatrics*. (2016) 138:e20161878. doi: 10.1542/peds.2016-1878

13. Fulton S, Decarie-Spain L, Fioramonti X, Guiard B, Nakajima S. The menace of obesity to depression and anxiety prevalence. *Trends Endocrinol Metab*. (2021) 48:18–35. doi: 10.1016/j.tem.2021.10.005

14. NHS. *Adult Psychiatric Morbidity Survey: Survey of Mental Health and Wellbeing* (2022). Available online at: https://webarchive.nationalarchives.gov.uk/ukgwa/20180328140249; http://digital.nhs.uk/catalogue/PUB21748here (accessed on January 21, 2022).

15. Jia H, Guerin RJ, Barile JP, Okun AH, McKnight-Eily L, Blumberg SJ, et al. National and state trends in anxiety and depression severity scores among adults during the COVID-19 Pandemic – United States, 2020–2021.

MMWR Morb Mortal Wkly Rep. (2021) 70:1427–32. doi: 10.15585/mmwr.mm7040e3

16. Hou WK, Li TW, Liang L, Liu H, Ettman CK, Hobfoll SE, et al. Trends of depression and anxiety during massive civil unrest and COVID-19 in Hong Kong, 2019–2020. *J Psychiatr Res*. (2022) 145:77–84. doi: 10.1016/j.jpsychires.2021.11.037

17. Santomauro DF, Mantilla Herrera AM, Shadid J, Zheng P, Ashbaugh C, Pigott DM, et al. Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic. *Lancet*. (2021) 398:P1700–12. doi: 10.1016/S0140-6736(21)02143-7

18. Evans-Lacko S, Aguilar-Gaxiola S, Al-Hamzawi A, Alonso J, Benjet C, Bruffaerts R, et al. Socio-economic variations in the mental health treatment gap for people with anxiety, mood, and substance use disorders: results from the WHO World Mental Health (WMH) surveys. *Psychol Med*. (2018) 48:1560–71. doi: 10.1017/S0033291717003336

19. Chekroud AM, Foster D, Zheutlin AB, Gerhard DM, Roy B, Koutsouleris N, et al. Predicting barriers to treatment for depression in a U.S. National sample: a cross-sectional, proof-of-concept study. *Psychiatr Serv*. (2018) 69:927–34. doi: 10.1176/appi.ps.201800094

20. World Health Organization. *Depression: Fact sheets* (2022). Available online at: https://www.who.int/news-room/fact-sheets/detail/depressionhere (accessed on January 16, 2022).

21. World Mental Health Survey Consortium TW. Prevalence, severity, and unmet need for treatment of mental disorders in the world health organization world mental health surveys. *JAMA*. (2004) 291:2581–90. doi: 10.1001/jama.291.21.2581

22. Izibeloko J Omi, Leana U. Barriers to mental health services utilization in the Niger Delta region of Nigeria: service users perspectives. *Pan Afr Med J*. (2013) 14:1970. doi: 10.11604/pamj.2013.14.159.1970

23. Ali SH, Agyapong VIO. Barriers to mental health service utilisation in Sudan - perspectives of carers and psychiatrists. *BMC Health Serv Res*. (2015) 16:31. doi: 10.1186/s12913-016-1280-2

24. Galderisi S, Heinz A, Kastrup M, Beezhold J, Sartorius N. Toward a new definition of mental health. *World Psychiatry*. (2015) 14:231–3. doi: 10.1002/wps.20231

25. Abdelgadir EKM. *Exploring Barriers to the Utilization of Mental Health Services at the Policy and Facility Levels in Khartoum State, Sudan*. Seattle, WA: Queensland University of Technology (2012).

26. Anxiety and Depression Association of America (ADAA). *Understand Anxiety and Depression: Treatment & Management* (2021). Available online at: https://adaa.org/understanding-anxiety/depression/treatment-managementhere (accessed on January 30, 2022).

27. Bystritsky A. Treatment-resistant anxiety disorders. *Mol Psychiatry*. (2006) 11:805–14. doi: 10.1038/sj.mp.4001852

28. Munder T, Flückiger C, Leichsenring F, Abbass AA, Hilsenroth MJ, Luyten P, et al. Is psychotherapy effective? A re-analysis of treatments for depression. *Epidemiol Psychiatr Sci*. (2019) 28:268–74. doi: 10.1017/S2045796018000355

29. Cipriani A, Furukawa T, Salanti G, Chaimani A, Atkinson L, Ogawa Y, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet*. (2018) 391:1357–66. doi: 10.1016/S0140-6736(17)32802-7

30. Marasine NR, Sankhi S. Factors associated with antidepressant medication non-adherence. *Turkish J Pharmaceut Sci*. (2021) 18:242–9. doi: 10.4274/tjps.galenos.2020.49799

31. Hung C. Factors predicting adherence to antidepressant treatment. *Curr Opin Psychiatry*. (2014) 27:86. doi: 10.1097/YCO.0000000000000086

32. Fortney JC, Pyne JM, Edlund MJ, Stecker T, Mittal D, Robinson DE, et al. Reasons for antidepressant nonadherence among veterans treated in primary care clinics. *J Clin Psychiatry*. (2011) 72:827–34. doi: 10.4088/JCP.09m05528blu

33. Bautista LE, Vera-Cala LM, Colombo C, Smith P. Symptoms of depression and anxiety and adherence to antihypertensive medication. *Am J Hypertens*. (2012) 25:256. doi: 10.1038/ajh.2011.256

34. Al-Harbi KS. Treatment-resistant depression: therapeutic trends, challenges, and future directions. *Patient Prefer Adherence*. (2012) 6:369–88. doi: 10.2147/PPA.S29716

35. Fierce Healthcare. *Demand for Virtual Mental Health Care Is Soaring. Here Are Key Trends on Who Is Using it and Why* (2022). Available online at: https://www.fiercehealthcare.com/tech/demand-for-virtual-mental-health-soaring-here-are-notable-trends-who-using-it-and-whyhere (accessed on January 16, 2022).

36. Fierce Healthcare. *Digital Behavioral Health Startups Scored $588M in Funding Amid COVID-19 Pandemic* (2022). Available online at: https://www.fiercehealthcare.com/tech/funding-for-digital-behavioral-health-startups-surged-amid-covid-19-pandemichere (accessed on January 16, 2022).

37. Sharry J, Davidson R, McLoughlin O, Doherty G. A service-based evaluation of a therapist-supported online cognitive behavioral therapy program for depression. *J Med Internet Res*. (2013) 15:2248. doi: 10.2196/jmir.2248

38. Derek R, Thomas R. Computer-based psychological treatments for depression: a systematic review and meta-analysis. *Clin Psychol Rev*. (2012) 32:4. doi: 10.1016/j.cpr.2012.02.004

39. David K, Glyn L, Surinder K, Nicola W, Michael K, Scott W, et al. Therapist-delivered internet psychotherapy for depression in primary care: a randomised controlled trial. *Lancet*. (2009) 374:9690. doi: 10.1016/S0140-6736(09)61257-5

40. Wade AG. Use of the internet to assist in the treatment of depression and anxiety: a systematic review. *Prim Care Companion J Clin Psychiatry*. (2010) 12:PCC.09r00876. doi: 10.4088/PCC.09r00876blu

41. de Graaf L, Gerhards SAH, Arntz A, Riper H, Metsemakers JFM, Evers SMAA, et al. Clinical effectiveness of online computerised cognitive-behavioural therapy without support for depression in primary care: randomised trial. *Br J Psychiatry*. (2009) 195:73–80. doi: 10.1192/bjp.bp.108.054429

42. Hommel E, Olsen B, Battelino T, Conget I, Schütz-Fuhrmann I, Hoogma R, et al. Impact of continuous glucose monitoring on quality of life, treatment satisfaction, and use of medical care resources: analyses from the SWITCH study. *Acta Diabetol*. (2014) 51:845–51. doi: 10.1007/s00592-014-0598-7

43. Thivolet C, Gaudilliere M, Villar Fimbel S, Bendelac N, Mestre B, Protsenko N, et al. Hybrid closed Loop improved glucose control compared to sensor-augmented pumps in outpatients with type 1 diabetes in real-life conditions with telehealth monitoring. *Acta Diabetol*. (2021) 59:395–401. doi: 10.1007/s00592-021-01820-9

44. Lind M, Polonsky W, Hirsch IB, Heise T, Bolinder J, Dahlqvist S, et al. Continuous glucose monitoring vs. conventional therapy for glycemic control in adults with type 1 diabetes treated with multiple daily insulin injections: the GOLD randomized clinical trial. *JAMA*. (2017) 317:19976. doi: 10.1001/jama.2016.19976

45. López-Liria R, López-Villegas A, Leal-Costa C, Peiró S, Robles-Musso E, Bautista-Mesa R, et al. Effectiveness and safety in remote monitoring of patients with pacemakers five years after an implant: the poniente study. *Int J Environ Res Public Health*. (2020) 17:1431. doi: 10.3390/ijerph17041431

46. Udo E, van Hemel N, Zuithoff N, Nijboer H, Taks W, et al. Long term quality-of-life in patients with bradycardia pacemaker implantation. *Int J Cardiol*. (2013) 168:2159–63. doi: 10.1016/j.ijcard.2013.01.253

47. Anderson EM, Lambert MJ. A survival analysis of clinically significant change in outpatient psychotherapy. *J Clin Psychol*. (2001) 57:1056. doi: 10.1002/jclp.1056

48. Fulmer R, Joerin A, Gentile B, Lakerink L, Rauws M. Using psychological artificial intelligence (tess) to relieve symptoms of depression and anxiety: randomized controlled trial. *JMIR Ment Health*. (2018) 5:9782. doi: 10.2196/mental.9782

49. Economides M, Lehrer P, Ranta K, Nazander A, Hilgert O, Raevuori A, et al. Feasibility and efficacy of the addition of heart rate variability biofeedback to a remote digital health intervention for depression. *Appl Psychophysiol Biofeedback*. (2020) 45:75–86. doi: 10.31234/osf.io/8wpq7

50. Valerie LFH, Benjamin WN, Kristian R, Albert N, Outi H, Joao D. Significant reduction in depressive symptoms among patients with moderately-severe to severe depressive symptoms after participation in a therapist-supported, evidence-based mobile health program delivered *via* a smartphone app. *Internet Intervent*. (2021) 25:100408. doi: 10.1016/j.invent.2021.100408

51. Hull TD, Mahan K. A study of asynchronous mobile-enabled sms text psychotherapy. *Telemed e-Health*. (2017) 23:240–7. doi: 10.1089/tmj.2016.0114

52. Reena LP, Morris M, Peters A, Spettell CM, Feifer R, Gillis W. Leveraging remote behavioral health interventions to improve medical outcomes and reduce costs. *Am J Manag Care*. (2015) 21:e141–51.

53. AbleTo. *Our Impact in Virtual Therapy and Counseling* (2022). Available online at: https://www.ableto.com/our-impact/here (accessed on January 15, 2022).

54. Richards D, Enrique A, Eilert N, Franklin M, Palacios J, Duffy D, et al. A pragmatic randomized waitlist-controlled effectiveness and cost-effectiveness trial of digital interventions for depression and anxiety. *npj Digit Med*. (2020) 3:85. doi: 10.1038/s41746-020-0293-8

55. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. (2001) 16:606–13. doi: 10.1046/j.1525-1497.2001.016009606.x

56. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. *Arch Intern Med*. (2006) 166:1092. doi: 10.1001/archinte.166.10.1092

57. Fugl-Meyer AR, Bränholm IB, Fugl-Meyer KS. Happiness and domain-specific life satisfaction in adult northern Swedes. *Clin Rehabil*. (1991) 5:25–33. doi: 10.1037/t12710-000

58. Diener E, Emmons R, Larsen R, Griffin S. The satisfaction with life scale. *J Pers Assess*. (1985) 49:71–5.

59. Eleftheriou G, Fatouros P, Tsirmpas C. *METHOD FOR DETECTING AND RECOGNIZING AN EMOTIONAL STATE OF A USER*. (2020). (20200075039). U.S. Patent No. 20200075039. Available online at: https://www.freepatentsonline.com/y2020/0075039.html

60. U S Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. *Guidance for Industry: E9 Statistical Principle for Clinical Trials*. International Conference on Harmonisation of Technical Requirements for Registration for Human Use (ICH) (1998).

61. Tripepi G, Chesnaye NC, Dekker FW, Zoccali C, Jager KJ. Intention to treat and per protocol analysis in clinical trials. *Nephrology*. (2020) 25:513–17. doi: 10.1111/nep.13709

62. Kroenke K. Enhancing the clinical utility of depression screening. *CMAJ*. (2012) 184:281–2. doi: 10.1503/cmaj.112004

63. Toussaint A, Hüsing P, Gumz A, Wingenfeld K, Härter M, Schramm E, et al. Sensitivity to change and minimal clinically important difference of the 7-item generalized anxiety disorder questionnaire (GAD-7). *J Affect Disord*. (2020) 265:32. doi: 10.1016/j.jad.2020.01.032

64. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. *J Mach Learn Res*. (2011) 12:2825–30. Available online at: https://www.jmlr.org/papers/v12/pedregosa11a.html

65. Rowan K, McAlpine DD, Blewett LA. Access and cost barriers to mental health care, by insurance status, 1999–2010. *Health Aff*. (2013) 32:133. doi: 10.1377/hlthaff.2013.0133

66. Prochaska J, DiClemente C. Stages and processes of self-change of smoking: toward an integrative model of change. *J Consult Clin Psychol*. (1983) 51:390. doi: 10.1037/0022-006X.51.3.390

67. Shapiro DA, Barkham M, Rees A, Hardy GE, Reynolds S, Startup M. Effects of treatment duration and severity of depression on the effectiveness of cognitive-behavioral and psychodynamic-interpersonal psychotherapy. *J Consult Clin Psychol*. (1994) 62:522. doi: 10.1037/0022-006X.62.3.522

frontiers | Frontiers in Psychiatry

TYPE Opinion
PUBLISHED 05 August 2022
DOI 10.3389/fpsyt.2022.900615

OPEN ACCESS

EDITED BY
Benjamin Nelson,
Meru Health Inc., United States

REVIEWED BY
Alicia Salamanca-Sanabria,
ETH Zürich, Singapore
Akira Monji,
Saga University, Japan
Robert Blumenfeld,
California State Polytechnic University,
Pomona, United States

*CORRESPONDENCE
P. Murali Doraiswamy
murali.doraiswamy@duke.edu

SPECIALTY SECTION
This article was submitted to
Digital Mental Health,
a section of the journal
Frontiers in Psychiatry

RECEIVED 21 March 2022
ACCEPTED 29 June 2022
PUBLISHED 05 August 2022

CITATION
Nwosu A, Boardman S, Husain MM and Doraiswamy PM (2022) Digital therapeutics for mental health: Is attrition the Achilles heel?
Front. Psychiatry 13:900615.
doi: 10.3389/fpsyt.2022.900615

Digital therapeutics for mental health: Is attrition the Achilles heel?

Adaora Nwosu[1], Samantha Boardman[2], Mustafa M. Husain[3] and P. Murali Doraiswamy[1,4]*

[1]Department of Psychiatry and Behavioral Sciences, Duke University School of Medicine, Durham, NC, United States, [2]Department of Psychiatry, Weill Cornell Medical College, New York, NY, United States, [3]Departments of Psychiatry, Neurology and Biomedical Engineering, Southwestern Medical Center, The University of Texas, Austin, TX, United States, [4]Duke Institute for Brain Sciences, Duke University, Durham, NC, United States

KEYWORDS

ADHD, depression, anxiety, Alzheimer's disease, PTSD, autism, schizophrenia, addiction

Introduction

Large unmet needs in mental health combined with the stress caused by pandemic mitigation measures have accelerated the use of digital mental health apps and software-based solutions (1). Global investor funding for virtual behavioral services and mental health apps in 2021 exceeded $5.5 billion, a 139% jump from 2020, according to CB Insights (2). While there are thousands of apps claiming to improve various aspects of mental wellbeing, many of them have never gone through clinical trials or regulatory scrutiny. The term "digital therapeutic" is used in the literature to distinguish high quality evidence-based software programs from wellness apps (3). Regulators use the term "software as a medical device" (SaMD) or "software in a medical device" (SiMD) to refer to software that functions as a medical device and is promoted to treat a specific condition. When a SaMD or SiMD is deployed on a phone it is referred to as a mobile medical app (MMA) (4, 5). The International Medical Device Regulators Forum (IMDRF), a voluntary group of medical device regulators from around the world has developed detailed guidance on definitions, framework for risk categorization, quality management, and the clinical evaluation of such devices (6–8). Non-traditional approaches, outside of RCTs, to evaluate efficacy for such tools has also been discussed elsewhere (9).

To date, only a few clinically tested software devices have been authorized by the U.S. Food and Drug Administration for treating specific mental health disorders (excluding devices marketed under pandemic-related emergency use authorization). These include reSET for substance abuse disorder (10), reSET-O for opioid use disorder (11), Somryst for chronic insomnia (12, 13) and EndeavorRx for pediatric attention deficit hyperactivity disorder (14, 15). SaMDs and MMAs for treating mild cognitive impairment, Alzheimer's disease, schizophrenia, autism, depression, social anxiety disorder, phobias and PTSD are in clinical trials (1, 3, 5, 16, 17) and may also come to market soon. The state of efficacy for non-regulated, wellness apps (e.g., for mindfulness or stress management) is beyond the scope of this article, and readers are referred elsewhere for information on these apps (16).

High attrition and low engagement

While digital therapeutics and apps undoubtedly hold promise, relatively little attention has focused on attrition rates. Even effective apps will have limited impact if they are not highly engaging and result in high attrition (18, 19). Attrition, the loss of a randomized subject(s) from a study sample, is a very common issue in clinical trials and results from several causes such as refusal to participate after randomization, an early dropout from the study, and loss of subject's study data. Attrition can substantially bias estimates of efficacy and reduce generalizability (20). Traditionally, regulatory trials of psychopharmacological agents have used the last observation carried forward (LOCF) statistical method to accommodate attrition–but this has been increasingly replaced with mixed-effects models, and pattern-mixture and propensity adjustments (20). Compliance in trials of psychopharmacological agents is traditionally measured *via* pill counts. However, in virtual platform trials of digital therapeutics, compliance cannot simply be measured by the number of times a subject logs on to an app and it is important to also measure and report how engaged users were with the app (21). Currently, there is no standard way to define what constitutes meaningful engagement and how to compare engagement across different digital therapeutic devices (21). There is also no consensus as to how to deal with users who are non-engaged but stay in the study.

As patients typically use apps on their own time, they must be intrinsically motivated to do so and must perceive the benefits from the app as meaningful (18). Such intrinsic motivation may be low for psychiatric patients with depression, anhedonia, or cognitive difficulties. For example, in one study of internet-based cognitive behavioral therapy for depression, the highest engagers comprised just 10.6% of the sample (22). This is further highlighted by a 2020 meta-analysis of 18 randomized trials of (non-FDA cleared) mobile apps for treating depression (trial duration ranging from 10 days to 6 months), in which the pooled dropout rate was 26.2% and rose to 47.8% when adjusted for publication bias (19). The authors concluded that this raises concern over whether efficacy was overstated in these studies. Real-world attrition rates for non-FDA cleared mental wellness apps are not readily available for direct comparison. But one study of 93 non-FDA approved Android apps (median installs 100,000), targeting mental wellbeing, found the medians of app 15 and 30-day retention rates were very low at 3.9% and 3.3% (23). In that study, the daily active user rate (median open rate) was only 4.0%. (23) These data highlight that the number of app installs has very little correlation with daily long-term usage.

Attrition and engagement rates (self-defined by study sponsors) in the pivotal studies for four FDA-authorized neuropsychiatric digital therapeutics are shown in Table 1. The studies reported significant benefits for the digital therapeutic vs. a control condition (Table 1), (10–15). Sample sizes were adequate, ranging from 170 to 1,149 participants (Table 1). Active intervention durations were relatively short ranging from 4 to 12-weeks (Table 1). Trial design, nature of therapy, incentives, and diagnosis influenced attrition. The Somyrst trials additionally reported 6 and 12-month follow up data. Attrition was lowest and compliance was highest in the pivotal study (14) of EndeavorRx for pediatric ADHD (Table 1) – this was a short 4-week trial of an interactive videogame where compliance was monitored electronically and there was close parental supervision. However, in their open 12-week study (15), the average missions engaged (with the videogame therapeutic) dropped by 34% at week 4 and by 50% at week 12 (Table 1). In the Somryst study for chronic insomnia (12), only 60% of subjects completed all 6 core modules of CBT and frequency of subject logins varied from 0 to 142 times (median of 25). While efficacy was sustained even at 12-month follow up, the decrease in insomnia score was greater in subjects who completed all 6 modules vs. those who did not. In the Somryst study for subclinical depression with insomnia (13), attrition rate was 58% at 6 weeks and on average only 3.5 of the 6 modules were completed. Patients completing <4 modules had no significant overall benefits vs. the control condition and were not different from the control condition at 6 months. In the reSET study for substance use disorder (10), the drop-out rate was low (12%) – this was likely because subjects were seen twice a week in the clinic, supervised by therapists, paid prizes ranging from "thank you" notes to up to $100 cash for compliance, and on average, earned $277 in prizes over 12 weeks. In their long-term follow-up (10), when this contingency incentive ended, the superiority of the digital therapeutic over the control condition also ended.

Closing the attrition-efficacy gap

Mental health conditions, like major depressive disorder, ADHD, and PTSD, require sustained treatment. Because the field of digital therapeutics is still in its early stages, currently, there is little long-term efficacy data. If even well-designed, gamified, digital therapeutics have a 50% drop in engagement in 3 months then the outlook for long-term efficacy is grim. While drop-out rates in clinical trials can be kept low through frequent clinician contact, gamification, feedback, and cash incentives, this is not practical in the real world and hence attrition rates will be far higher. Finally, if the costs of increasing engagement and compliance equals that of a getting live psychiatric care, then digital therapies would become less attractive as a scalable low-cost solution.

Scientific gaps identified

Our scrutiny of published data also reveals several scientific gaps. First is the lack of standardized definitions of attrition and engagement in the field of digital therapeutics. Second is

TABLE 1 Engagement and Attrition rates in the pivotal studies of FDA-cleared Digital Therapeutics for menta health.

Reference number	N	Indication	Study intervention and dose	Trial design	Trial duration	Engagement	Attrition rate for active intervention arm
(14)	348	Pediatric ADHD	RCT of Endeavor Rx vs. control video game 25 min per day, 5 days per week, for 4 weeks.	Hybrid	4 weeks	83%	6%
(12)	303	Chronic Insomnia	SHUTi (Somryst) six sequential modules completed within 9 weeks	Remote	9 weeks, with 12- month follow-up.	60.3%	9.2% 17.5%[a]
(15)	206	Pediatric ADHD	Open Label study of Endeavor Rx +/- pharmacotherapy 4 weeks (25 mins/day, 5 days/week), followed by a 4 week pause, and then another 4-week use of therapeutic.	Hybrid	12 weeks	68% (pharmacotherapy) 58% (no pharmacotherapy)[b]	12% (pharmacotherapy) 12% (no pharmacotherapy)
(11)	170	Opioid Use Disorder	Therapeutic Education System (reSET-O) TES modules 3 times per week for 12 weeks.	In-Clinic	12 weeks	All sessions were supervised by a live therapist in the clinic.	20%
(10)	507	Substance Use Disorder	Therapeutic Education System (reSET) 4 TES modules per week for 12 weeks.	Hybrid	12 weeks	76.3%[c]	12%
(13)	1149	Subclinical depression and insomnia	SHUTi (Somryst) six sequential modules completed within 6 weeks	Remote	6 weeks, with 6 month follow up	58%	57% 61%[d]

[a] Attrition at 12 month follow up. Studies used variable definitions and often did ot break down reasons.
[b] Compliance was the percentage of total possible recommended sessions. Engagement metrics were not reported in a standardized manner.
[c] 36.6 modules out of recommended 48 (range 0–72).
[d] Attrition at 6 month follow up.

the lack of standardized reporting requirements by journals. A single digital therapeutic session can generate a dozen or more different metrics of how a user may interact with the app. Even widely used clinical trials reporting checklists, such as CONSORT, have not yet required the reporting of all such engagement metrics in digital trials. This makes it hard to extract such data from published reports and compare metrics across trials and products. Third, is the lack of a standardized definition of compliance. Fourth is the lack of standardized statistical methods, such as mixed models or last observation carried forward, to account for attrition and engagement biases in digital trials.

Emerging solutions

Fortunately, several constructs are emerging as promising features to increase engagement – both related to external factors of motivation and UX design (24). Several factors such as ease of use, gamification, ability to personalize app, in-app symptom monitoring, numerical feedback, ability to chart progress, socialization within the app, and integration with clinical services, have been reported to increase engagement (17, 18). A machine learning analysis of 54,604 adult patients with depression and anxiety identified 5 distinct engagement patterns for digital cognitive behavior therapy over 14-weeks: low engagers [36.5% of sample], late engagers [21.4%], high engagers with rapid disengagement [25.5%], high engagers with moderate decrease [6.0%], and high persistent engagers [10.6%]. Depression improvement rates were lowest for the low engagers (22). This study suggested machine learning algorithms may be useful to tailor interventions and a human touch for each of these five groups. Kaiser Permanente found that integrating digital mental health solutions – provided *via* clinician referral – into their health care delivery system was able to successfully enhance engagement (25). Fears around privacy and data security for mental health data may be a factor in engagement and attrition for some participants and this should be addressed upfront. While we do not have all the solutions,

encouraging the availability of raw data from clinical trials through trial registries, analyzing long-term real-world data on patient reported outcomes, user experience (engagement and compliance) and product reliability (18) will be important to enhance their utility.

Digital therapeutics for mental health are here to stay. As the pivotal studies demonstrate, they benefit a substantial number of patients. However, the gap between intention and real-world efficacy for digital therapeutics remains large. There is an urgent need to recognize this gap and for stakeholders–regulators, technology developers, clinicians, patients–to come together to close this gap and ensure that this form of treatment is useful to clinical populations.

Author contributions

AN and PMD drafted the article. SB and MH provided critical edits. All authors helped with data interpretation of cited studies, contributed to the article, and approved the submitted version.

Funding

The study was funded by a grant from the National Institute of Aging to PMD.

Conflict of interest

MH has received grants from NIMH, NIA, Brain Initiative and Stanley Foundation for other projects. For other projects, PMD has received grants from NIA, DARPA, DOD, ONR, Salix, Avanir, Avid, Cure Alzheimer's Fund, Karen L. Wrenn Trust, Steve Aoki Foundation, and advisory fees from Apollo, Brain Forum, Clearview, Lumos, Neuroglee, Otsuka, Verily, Vitakey, Sermo, Lilly, Nutricia, and Transposon. PD is a co-inventor on patents for diagnosis or treatment of neuropsychiatric conditions and owns shares in biotechnology companies.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

1. Doraiswamy PM, London E, Varnum P, Harvey B, Saxena S, Tottman S, et al. Empowering 8 billion minds: Enabling better mental health for all via the ethical adoption of technologies. NAM Perspect. (2019). doi: 10.31478/201910b. Available online at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8406599/ (accessed May 9, 2022).

2. CB Insights. *State of Mental Health Tech 2021 report. CB Insights Research*. (2022). Available online at: https://www.cbinsights.com/research/report/mental-health-tech-trends-2021/ (accessed May 9, 2022).

3. Wehrwein P. Digital therapeutics shaping the future of care. *Managed Healthc Exec.* (2021) 31.

4. U.S. Food and Drug Administration. Policy for device software functions and mobile medical applications. (2019). Available online at: https://www.fda.gov/media/80958/download (accessed May 9, 2022).

5. Shuren J, Doraiswamy PM. Digital Therapeutics for MCI and Alzheimer's disease: A Regulatory Perspective – Highlights From The Clinical Trials on Alzheimer's Disease conference (CTAD). *J Prev Alzheimers Dis.* (2022) 9:236. doi: 10.14283/jpad.2022.28

6. International Medical Devices Regulator Forum. Software as a medical device (SaMD): key definitions. (2013). Available online at: https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-131209-samd-key-definitions-140901.pdf (accessed May 9, 2022).

7. International Devices Regulator Forum. "Sofware as a medical device": possible framework for risk categorization and corresponding considerations. (2014). Available online at: https://www.imdrf.org/sites/default/files/docs/imdrf/final/technical/imdrf-tech-140918-samd-framework-risk-categorization-141013.pdf

8. U.S. Food and Drug Administration. Software as a medical device (SaMD) clinical evaluation. Guidance for Industry and Food and Drug Administration staff. (2017). Available online at: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/software-medical-device-samd-clinical-evaluation/ (accessed May 9, 2022).

9. Guo C, Ashrafian H, Ghafur S, Fontana G, Gardner C, Prime M. Challenges for the evaluation of digital health solutions – a call for innovative evidence generation approaches. *NPJ Digit Med.* (2020) 3:110–110. doi: 10.1038/s41746-020-00314-2

10. Campbell ANC, Nunes EV, Matthews AG, Stitzer M, Miele GM, Polsky D, et al. Internet-delivered treatment for substance abuse: a multisite randomized controlled trial. *Am J Psychiatry*. (2014) 171:683–90. doi: 10.1176/appi.ajp.2014.13081055

11. Christensen DR, Landes RD, Jackson L, Marsch LA, Mancino MJ, Chopra MP, et al. Adding an internet-delivered treatment to an efficacious treatment package for opioid dependence. *J Consult Clin Psychol*. (2014) 82:964–72. doi: 10.1037/a0037496

12. Ritterband LM, Thorndike FP, Ingersoll KS, Lord HR, Gonder-Frederick L, Frederick C, et al. Effect of a web-based cognitive behavior therapy for insomnia intervention with 1-year follow-up: a randomized clinical trial. *JAMA Psychiatry*. (2016) 74:68–75. doi: 10.1001/jamapsychiatry.2016.3249

13. Christensen H, Batterham PJ, Gosling JA, Ritterband LM, Griffiths KM, Thorndike FP, et al. Effectiveness of an online insomnia program (SHUTi) for prevention of depressive episodes (the GoodNight study): a randomised controlled trial. *Lancet Psychiatry*. (2016) 3:333–41. doi: 10.1016/S2215-0366(15)00536-2

14. Kollins SH, DeLoss DJ, Cañadas E, Lutz J, Findling RL, Keefe RSE, et al. A novel digital intervention for actively reducing severity of paediatric ADHD (STARS-ADHD): a randomised controlled trial. *Lancet Digit Health*. (2020) 2:e168–78. doi: 10.1016/S2589-7500(20)30017-0

15. Kollins SH, Childress A, Heusser AC, Lutz J. Effectiveness of a digital therapeutic as adjunct to treatment with medication in pediatric ADHD. *NPJ Digit Med.* (2021) 4:58–58. doi: 10.1038/s41746-021-00429-0

16. Lau N, O'Daffer A, Colt S, Yi-Frazier JP, Palermo TM, McCauley E, et al. Android and iPhone mobile apps for psychosocial wellness and stress management: systematic search in app stores and literature review. *JMIR Mhealth Uhealth*. (2020) 8:e17798. doi: 10.2196/17798

17. Bodner KA, Goldberg TE, Devanand DP, Doraiswamy PM. Advancing computerized cognitive training for MCI and Alzheimer's disease in a pandemic and post-pandemic world. *Front Psychiatry*. (2020) 11:557571. doi: 10.3389/fpsyt.2020.557571

18. Boardman S. Creating wellness apps with high patient engagement to close the intention-action gap. *Int PsychGeriatr*. (2021) 33:551–2. doi: 10.1017/S1041610220003385

19. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. *J Affect Disord*. (2020) 263:413–9. doi: 10.1016/j.jad.2019.11.167

20. Leon AC, Mallinckrodt CH, Chuang-Stein C, Archibald DG, Archer GE, Chartier K. Attrition in randomized controlled clinical trials: methodological issues in psychopharmacology. *Biol Psychiatry*. (2006) 59:1001–5. doi: 10.1016/j.biopsych.2005.10.020

21. Torous J, Michalak EE, O'Brien HL. Digital health and engagement - looking behind the measures and methods. *JAMA Net Open*. (2020) 3:e2010918. doi: 10.1001/jamanetworkopen.2020.10918

22. Chien I, Enrique A, Palacios J, Regan T, Keegan D, Carter D, et al. A machine learning approach to understanding patterns of engagement with internet-delivered mental health interventions. *JAMA Netw Open*. (2020) 3:e2010791. doi: 10.1001/jamanetworkopen.2020.10791

23. 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) 21:e14567. doi: 10.2196/14567

24. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. *J Med Internet Res*. (2020) 22:e17572. doi: 10.2196/17572

25. Mordecai D, Histon T, Neuwirth E, Heisler S, Kraft A, Bang Y, et al. How Kaiser permanente created a mental health and wellness digital ecosystem. *NEJM Catalyst Innovations Care Deliv*. (2021) 2. doi: 10.1056/CAT.20.0295

frontiers | Frontiers in Psychiatry

TYPE Original Research
PUBLISHED 11 August 2022
DOI 10.3389/fpsyt.2022.807116

OPEN ACCESS

EDITED BY
Patricia A. Arean,
University of Washington, United States

REVIEWED BY
Long-Biao Cui,
Fourth Military Medical University, China
Gerard Anmella,
Hospital Clinic of Barcelona, Spain
Yuezhou Zhang,
King's College London, United Kingdom

*CORRESPONDENCE
George D. Price
George.price.gr@dartmouth.edu

SPECIALTY SECTION
This article was submitted to Digital Mental Health, a section of the journal Frontiers in Psychiatry

RECEIVED 01 November 2021
ACCEPTED 27 July 2022
PUBLISHED 11 August 2022

CITATION
Price GD, Heinz MV, Nemesure MD, McFadden J and Jacobson NC (2022) Predicting symptom response and engagement in a digital intervention among individuals with schizophrenia and related psychoses. *Front. Psychiatry* 13:807116. doi: 10.3389/fpsyt.2022.807116

Predicting symptom response and engagement in a digital intervention among individuals with schizophrenia and related psychoses

George D. Price[1,2]*, Michael V. Heinz[1,3], Matthew D. Nemesure[1,2], Jason McFadden[4] and Nicholas C. Jacobson[1,2,3,5]

[1]Center for Technology and Behavioral Health, Geisel School of Medicine, Dartmouth College, Lebanon, NH, United States, [2]Quantitative Biomedical Sciences Program, Geisel School of Medicine, Dartmouth College, Lebanon, NH, United States, [3]Department of Psychiatry, Geisel School of Medicine, Dartmouth College, Lebanon, NH, United States, [4]Dartmouth College, Lebanon, NH, United States, [5]Department of Biomedical Data Science, Geisel School of Medicine, Dartmouth College, Lebanon, NH, United States

Introduction: Despite existing work examining the effectiveness of smartphone digital interventions for schizophrenia at the group level, response to digital treatments is highly variable and requires more research to determine which persons are most likely to benefit from a digital intervention.

Materials and methods: The current work utilized data from an open trial of patients with psychosis (N = 38), primarily schizophrenia spectrum disorders, who were treated with a psychosocial intervention using a smartphone app over a one-month period. Using an ensemble of machine learning models, pre-intervention data, app use data, and semi-structured interview data were utilized to predict response to change in symptom scores, engagement patterns, and qualitative impressions of the app.

Results: Machine learning models were capable of moderately (r = 0.32–0.39, R^2 = 0.10–0.16, MAE_{norm} = 0.13–0.29) predicting interaction and experience with the app, as well as changes in psychosis-related psychopathology.

Conclusion: The results suggest that individual smartphone digital intervention engagement is heterogeneous, and symptom-specific baseline data may be predictive of increased engagement and positive qualitative impressions of digital intervention in patients with psychosis. Taken together, interrogating individual response to and engagement with digital-based intervention with machine learning provides increased insight to otherwise ignored nuances of treatment response.

KEYWORDS

schizophrenia, digital intervention, machine learning, intervention engagement, qualitative impressions, app use, Brief Symptom Inventory

Introduction

Schizophrenia, a primary psychotic disorder, is a serious mental illness characterized by debilitating symptoms including delusions, hallucinations, disorganized speech and behavior, and diminished emotional expression (1, 2). Current studies estimate that schizophrenia affects up to 0.64% of the United States population (3–5). However, despite schizophrenia's comparatively low incidence (6), it had an economic burden of $155.7 billion in 2013 (7), and remains a major contributor to the global burden of disease with two-thirds of affected individuals experiencing persistent symptoms following treatment (1, 8). Further, schizophrenia is often comorbid with anxiety disorders and depression (9, 10). Thus, illness burden is of particular concern, with symptom severity shown to be negatively correlated with physical health, psychological health, and relationships, and one in ten completing suicide (11, 12). Despite recent improvements in diagnostic accuracy and treatment efficacy, schizophrenia continues to negatively impact patients' overall quality of life (11), highlighting the necessity for additional efforts to address treatment response at the individual level.

Currently, the mainstay treatment for schizophrenia includes the use of antipsychotics, often coupled with regular psychosocial interventions (13, 14). Psychosocial interventions can be offered as individual (supportive counseling, personal therapy, social skills therapy, vocational rehabilitation therapy); cognitive-behavioral; and group (interactive/social therapy) interventions (14). Despite the relative efficacy of these combined approaches, the side-effects of antipsychotic treatment can prove debilitating to schizophrenia patients. Antipsychotic side-effects may include physical symptoms, such as movement disturbances, metabolic derangements and weight gain, sedation, and drooling (15, 16, 17–19), as well as emotional and cognitive blunting (20). As a result, the integration of alternative interventions may prove useful in reducing the side-effect burden of antipsychotics.

Research has shown promise in the use of scalable digital therapeutics in patients living with serious mental disorders, such as schizophrenia and bipolar disorder (21). Although few studies have directly analyzed the efficacy of digital interventions for schizophrenia, existing research has suggested that this method of intervention may be efficacious for schizophrenia treatment or management. For example, PRIME, or Personalized Real-time Intervention for Motivational Enhancement, is a mobile application designed to supplement antipsychotic and psychotherapeutic treatments for schizophrenia. The intervention was found to be effective in improving the mood and motivation of young patients with schizophrenia. Further, PRIME users experienced improvements in depression, self-efficacy, and reward learning (22, 23). Similarly, App4Independence (A4i), a community-centric app designed for schizophrenia patients, offers forums, appointment scheduling, and text-based functions aimed at improving illness self-management (24). Like PRIME, A4i showed modest reductions in symptoms of depression, further pronounced after controlling for gender, age, and other baseline symptoms (24).

Treatment of schizophrenia remains a challenge, partially owing to its highly heterogeneous nature (25–29), with little known of personalized prognostic and treatment factors. With growing numbers of digital treatment options for mental health in the context of this heterogeneity, it is important to understand not only the group efficacy of such interventions, but also the profile of those individuals most likely to benefit. With such an understanding, individuals could be effectively "matched" to the specific interventions most likely to improve their symptoms.

Machine learning, coupled with highly dimensional datasets, like A4i, is uniquely positioned to address these challenges, having shown promise in constructing personalized models, with a capacity to accurately predict individual-level treatment response (30, 31). Specific to individuals with schizophrenia, an extensive corpus of research has shown machine learning efficacious in classifying schizophrenia and psychosis-related symptoms from neuroimaging data (32–34), qualitative social media information (35, 36), and passively collected smartphone data. Notably, virtual communication was positively associated with increased negative affect measures (37), highlighting the necessity to interrogate the driving factors in engagement with digital interventions in individuals with schizophrenia. Furthermore, machine learning has the capacity to model complex relationships in large datasets, with model introspection made possible by high power computing methods, such as Shapley Additive Explanations (SHAP) (38). Methods, like SHAP, address the "black box" problem of machine learning by offering a manner of visually representing the directionality and magnitude of those features most important in the model's outcome prediction. This allows for the development of downstream digital biomarkers and phenotypes of psychiatric disorders.

In the current study, we utilize data from an open trial delivering the A4i intervention to persons with schizophrenia, schizoaffective disorder, and related psychoses that found modest symptom improvement at the group level (24), to better understand personalized markers of digital intervention engagement and response. We hypothesized (1) *unique baseline patient* characteristics paired with machine learning would moderately predict ($r > 0.3$) individual symptom response to, engagement with, and sentiment toward A4i (an app-delivered, digital intervention) (2) specifically, we hypothesized that people with higher affective symptoms (e.g., depression) and lower psychotic symptoms would have the most robust use and response, resulting in more positive sentiment toward the app; this hypothesis can be contextualized in research which suggests better overall prognosis and treatment response for patients with higher affective symptoms (39). Further, (3) we

hypothesized that persons with higher interpersonal sensitivity would have the strongest response to the intervention (change in composite BSI score), given the built-in peer-engagement application feature, which was reported in the original study as a user-reported strength of the A4i intervention (24). To evaluate these hypotheses, we developed three machine learning pipelines aimed at modeling the relationship between baseline patient characteristics and (i) response to the digital intervention, (ii) level of engagement with the intervention, and (iii) user sentiment toward the intervention.

Materials and methods

Participants

Participants (N = 38, 2.6% transgender, 71.1% men, 26.3% women, age_{mean} = 31.42 ± 8.60) were included in the final study population based on previously described inclusion and exclusion criteria (24). The initial study design consisted of a 3–4 week engagement with the A4i app with pre-post assessments. The study was reviewed and approved by an institutional Research Ethics Board; participants provided written consent, and the protocol was registered with clinicaltrials.gov (NCT03649815) (24).

Intervention

The A4i functionality included personalized prompts, activity scheduling, connections to social engagement resources, evidenced-based content tailored to management of psychotic symptoms, a peer engagement network, daily wellness check-ins, and passively collected phone-use information, used as a proxy for sleep and activity (24).

Data collection and outcome metrics: Quantitative data set

Participants provided demographic information, mobile technology use information, and completed quantitative pre-post symptom and intervention engagement metrics (24). The quantitative metrics included: (a) the Brief Adherence Rating Scale (BARS) to examine implications of A4i for medication use (40), (b) the Person Recovery Outcome Measure (PROM) to assess degree of engagement in the recovery process (41), and (c) the Brief Symptom Inventory (BSI) to assess psychiatric symptoms pre and post-intervention. The BSI comprises domains measuring: psychoticism, somatization, depression, hostility, phobia, obsessive-compulsive, anxiety, paranoia, and interpersonal sensitivity (42). Additionally, A4i usage data from each participant during the trial was passively collected, including: (i) a count of participants' total active interaction with the app, (ii) the number of days each participant engaged with the app, (iii) the participants' average interaction with the app per day, and (iv) app usage categorized as "low" or "high." Participant BSI total and subdomain scores were included in the present analysis; BARS scores, PROM scores and demographic information were not.

Post-intervention semi-structured interview

A semi-structured interview was completed at the post-A4i use assessment, which included a series of seven questions providing qualitative feedback from the participants on the functionality, effectiveness, and overall experience interacting with the app (e.g., "What were your favorite features of the app?") (24). The complete semi-structured interview is provided in the Supporting information section of the original publication (24).

Semi-structured interview response sentiment quantification

We extracted participants' overall response sentiment from the semi-structured interviews. Individual responses to all questions were concatenated using Python (v3.8.3) for uniformity across participants (43). Overall response sentiment was derived from the concatenated qualitative data, using the Python package *TextBlob* (Version 0.16.0) to assess polarity (i.e., the valence of the participants responses on a –1 to 1 scale, with a lower score reflecting a more negative statement and a higher score reflecting a more positive statement) (44).

Data preprocessing

Baseline BSI total score, baseline BSI subcategory scores, and passively collected A4i use metric features included in modeling were individually standardized resulting in a $\mu = 0$ and $\sigma = 1$ within that feature. Feature standardization has been shown to increase model efficiency and accuracy in machine learning approaches (45), and provides a consistent value range for features when considering their relative influence on a model's predictions.

Theoretical machine learning modeling framework

We implemented a hypothesis driven framework *via* the utilization of three separate ensemble machine learning models

to interrogate individual-level factors driving (1) app efficacy, (2) app engagement, and (3) qualitative app impressions (see Table 1). The machine learning models implemented in this study were as follows:

1. *Symptom Severity Change*: Fourteen features, including baseline BSI total score, subcategory scores, and passively collected A4i use metrics were used to predict change in BSI total score (Table 1, Model 1). Change in BSI total score was measured as the difference between baseline BSI total score and post-intervention BSI total score (e.g., a negative change reflects an overall decrease in reported symptoms).
2. *A4i Engagement*: Ten features, including baseline BSI total score and subcategory scores were used to predict a participants' overall interaction with the A4i app (Table 1, Model 2).
3. *Intervention Impressions*: Ten features, including baseline BSI total score and subdomain scores were used to predict an individual participant's sentiment toward the intervention (Table 1, Model 3). A participants' sentiment was represented by the polarity score of their concatenated semi-structured interview responses.

TABLE 1 Machine learning model corresponding hypotheses, features and outcomes.

Modeling approach	Model features	Model outcome
Model 1: Symptom Severity Change	Baseline BSI Composite Score (Overall Symptoms), Baseline BSI Anxiety, Baseline BSI Depression, Baseline BSI Hostility, Baseline BSI Interpersonal Sensitivity, Baseline BSI Obsession-Compulsion, Baseline BSI Paranoid Ideation, Baseline BSI Phobic Anxiety, Baseline BSI Psychoticism, Baseline BSI Somatization, Total A4i Interaction, Total Days of A4i Interaction, Binary A4i Use (High/Low), Average A4i Use Per Day	Change in Composite BSI Score
Model 2: A4i Engagement	Baseline BSI Composite Score (Overall Symptoms), Baseline BSI Anxiety, Baseline BSI Depression, Baseline BSI Hostility, Baseline BSI Interpersonal Sensitivity, Baseline BSI Obsession-Compulsion, Baseline BSI Paranoid Ideation, Baseline BSI Phobic Anxiety, Baseline BSI Psychoticism, Baseline BSI Somatization	Total Interaction with A4i
Model 3: Intervention Impressions	Baseline BSI Composite Score (Overall Symptoms), Baseline BSI Anxiety, Baseline BSI Depression, Baseline BSI Hostility, Baseline BSI Interpersonal Sensitivity, Baseline BSI Obsession-Compulsion, Baseline BSI Paranoid Ideation, Baseline BSI Phobic Anxiety, Baseline BSI Psychoticism, Baseline BSI Somatization	Semi-structured Interview Response Sentiment (Polarity)

Input features and predicted outcomes for the three interrogated ensemble models.

Practical machine learning model framework

All machine learning modeling followed the same nested leave-one-out (LOO) cross-validation framework (46). A nested cross-validation framework in machine learning is efficacious in allowing for unbiased performance estimates, regardless of sample size (47). In this process, one subject was completely held out, while the rest of the subjects were used as part of a simple LOO cross-validation approach to tune the hyperparameters of the model. This process was repeated N times so each subject was held out at least once. We used an ensemble approach, whereby distinct machine learning models (i.e., linear models, tree based models, and multilayer perceptrons) were individually trained on the data; an approach which has been shown to consistently outperform base algorithms in mental health disorder related outcomes (48). Predictions from these models were used as inputs to a final "deciding" model, which returned a consensus score. The specific modeling architecture and hyperparameters of the Symptom Severity Change, A4i Engagement, and Intervention Impressions models are provided in Supplementary Table 1. The ensemble pipeline's predictions were evaluated against the observed values to determine the correlative strength (r), the proportion of the variance in the outcome explained by the model's predictions (R^2), and the normalized mean absolute error (MAE_{norm}) of the respective ensemble model. The normalized mean absolute error was calculated by dividing the mean absolute error by the range of the observed outcome, offering an outcome-agnostic representation of the model's mean absolute error.

Model explainability

We implemented Shapley Additive Explanations (SHAP) to aid in model interpretability by evaluating the top five most influential features in each of the three models. Intuitively, SHAP allows for model introspection by iteratively perturbing the input data and assessing how this affects the output (38). In this way, the process can determine feature importance,

as well as the marginal contribution of each independent variable to the predicted outcome at the patient level, represented by an individual values positioning on the x-axis of Figure 2. Using this information, SHAP can estimate relative feature importance, directional relationships between predictors and outcomes, as well as different order interactions between variables.

Results

Predictive performance and interpretability: Symptom severity change

Participants' A4i use metrics and baseline BSI scores (Table 1, Model 1) were capable of moderately predicting change in symptom severity (e.g., Pre-Post BSI score difference, where a negative change corresponds to a decrease in BSI score, and thus reduced overall symptoms) (r = 0.32, R^2 = 0.10, MAE_{norm} = 0.29) (Figure 1A). Model introspection *via* SHAP suggested that the most influential feature for predicting change in BSI score was *interpersonal sensitivity* on the baseline BSI (Figure 2A), where participants with high baseline interpersonal sensitivity were predicted to have a greater reduction in psychiatric symptomatology. Furthermore, lower psychotic and obsessive compulsive traits were predictive of a reduction in psychiatric symptomatology across the intervention. These findings directly address study hypothesis (1), that unique patient characteristics will moderately predict app response, and more specifically study hypothesis (3), which suggested interpersonal sensitivity as a positive prognostic marker for A4i response.

Predictive performance and interpretability: App4Independence engagement

Baseline symptom severity (measured by pre-intervention BSI scores) moderately predicted participant engagement across the A4i intervention (r = 0.39, R^2 = 0.16, MAE_{norm} = 0.16) (Figure 1B). The BSI subdomain *depression* was the most influential feature, where participants with high baseline depression interacted with the app more during the intervention; however, overall BSI score showed an inverse relationship, where participants with lower overall total symptoms were predicted to have greater app interaction. These findings directly address study hypothesis (1), that that unique patient characteristics will moderately predict person-level app engagement, and more specifically hypothesis (2) that higher affective symptoms would drive A4i response.

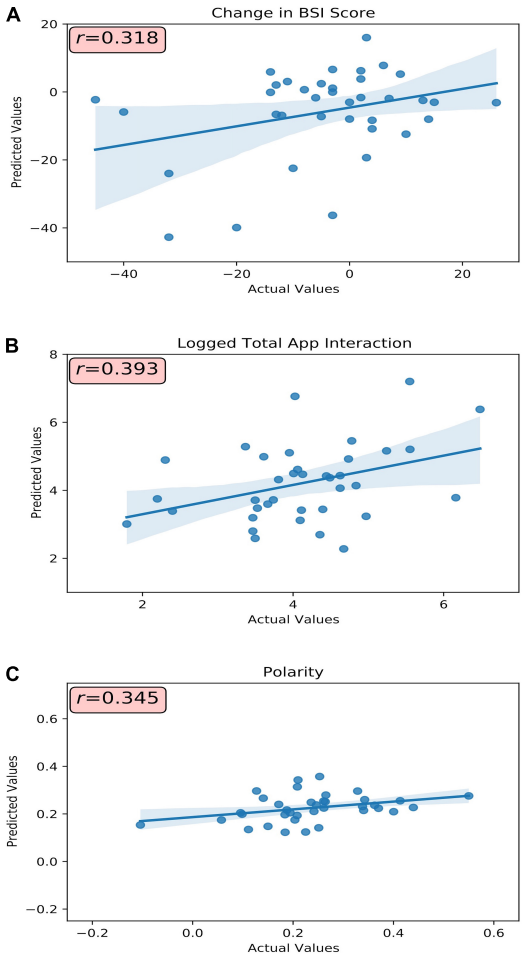


FIGURE 1
Model(s) actual versus predicted values plotted with respective correlative strength. **(A)** Baseline BSI total score, subcategory scores, and passively collected A4i use metrics were used to predict change in BSI total score. **(B)** Baseline BSI total score and subcategory scores were used to predict a participants' overall interaction (visualized as the log-transformation) with the A4i app. **(C)** Baseline BSI total score and subdomain scores were used to predict an individual participant's sentiment toward the intervention. r, Pearson's correlation coefficient.

Predictive performance and interpretability: Intervention impressions

Baseline BSI scores moderately predicted valence of interview responses (r = 0.34, R^2 = 0.12, MAE_{norm} = 0.14) (Figure 1C), with BSI depression scores having the greatest importance. Specifically, participants with high baseline *depression* were more positive when discussing the app in the semi-structured interview. Interestingly, similar to the A4i engagement model, overall BSI score showed an inverse relationship to the BSI *depression* subdomain, with lower overall BSI being predictive of more positive qualitative

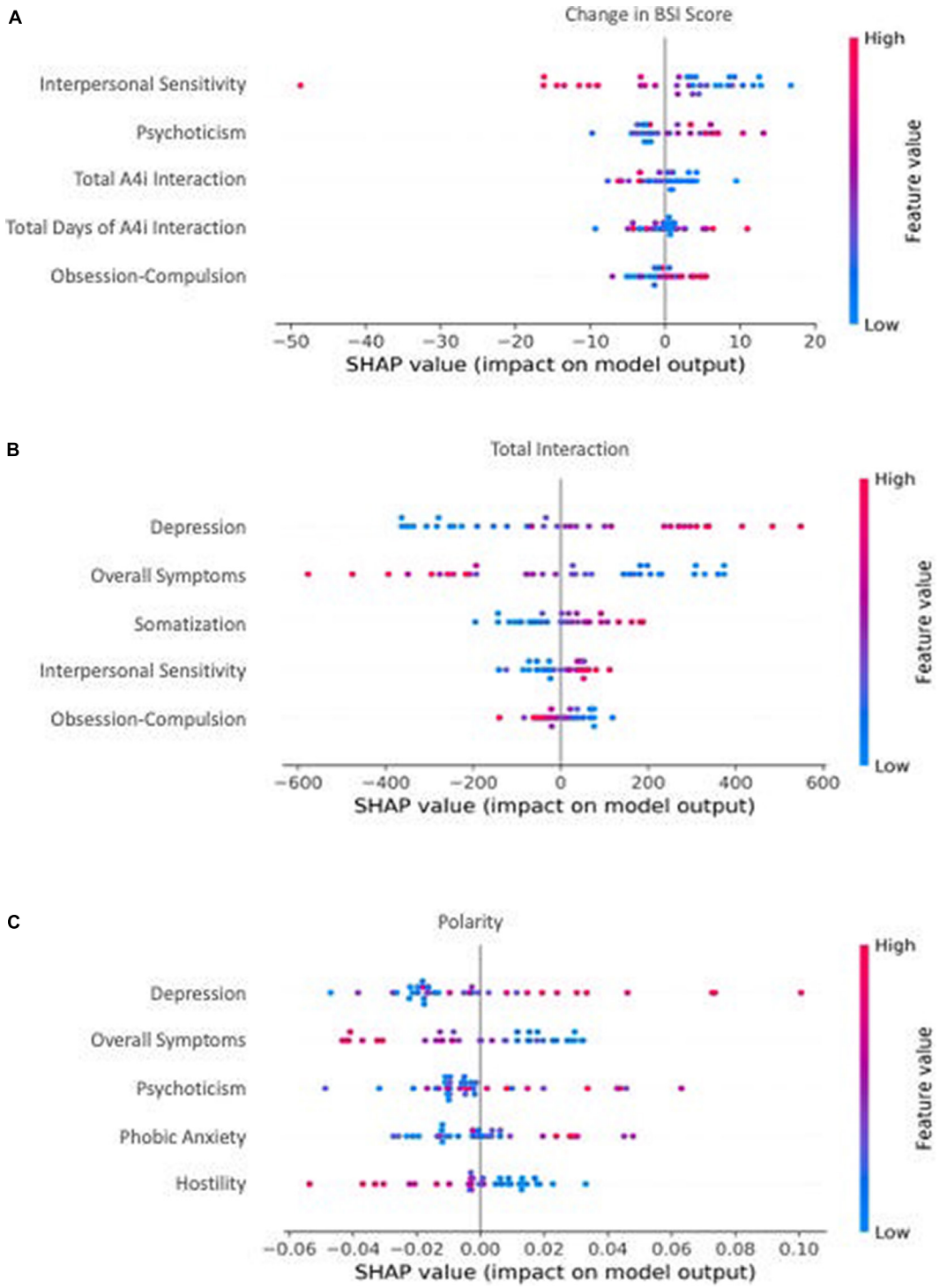


FIGURE 2
The top 5 most influential features by model. Individual dot color corresponds to the value of the feature, and location on the x-axis corresponds to that point's relative impact on the model output [e.g., a high-feature value (red) with a corresponding high x-axis value (SHAP value) represents a point that strongly, positively influences the model's outcome prediction]. **(A)** The most influential features from baseline BSI total score, subcategory scores, and passively collected A4i use metrics for predicting change in BSI total score. A positive x-axis value (SHAP value) corresponds to an increase in overall symptoms. **(B)** The most influential features from baseline BSI total score and subcategory scores for predicting a participants' overall interaction) with the A4i app. A positive x-axis value (SHAP value) corresponds to an increased interaction with A4i. **(C)** The most influential features from baseline BSI total score and subdomain scores for predicting an individual participant's sentiment toward the intervention. A positive x-axis value (SHAP value) corresponds to an increase in qualitative A4i impressions.

impressions of the intervention. These findings directly address study hypothesis (1), that unique patient characteristics will moderately predict overall app sentiment and, as in section "Predictive performance and interpretability: A4i engagement" – hypothesis (2) – that higher affective symptoms would drive A4i response.

Discussion

This study demonstrates the capacity of unique patient-level factors to predict response to a digital treatment among patients with schizophrenia, schizoaffective disorder, and related psychoses. Important factors included interpersonal sensitivity, psychotic traits, depressive traits, and overall symptom severity (as determined by baseline BSI), as well as digital intervention interaction metrics passively collected during the study. Subsequent analysis established factors associated with participant interaction, engagement and general attitudes toward the digital therapeutic intervention. As a whole, this work aimed to investigate unique patient markers of digital treatment response, as well as highlight those factors most important in predicting high engagement among persons with psychosis. Our results contribute to ongoing development and implementation of mental health digital interventions by identifying unique patient markers to suggest intervention response as well as engagement.

Higher interpersonal sensitivity

The *symptom severity change* model (**Table 1**, Model 1) predicted that participants with high baseline interpersonal sensitivity would benefit more from the A4i intervention. This finding may suggest that high interpersonal sensitivity corresponds to increased participant responsiveness to the community-centric platform of the intervention. The role of patient interaction is highlighted by the users during the semi-structured interviews, with one patient responding that the function of the A4i app is to serve as "a safe space online app based community platform where psychosis can be met with care and empathy." (24). Thus, in line with findings that lack of interpersonal relationships is known to be a significant contributor to reduced quality of life among individuals with schizophrenia (49), patients that actively interacted with the A4i community responded better to the intervention. Further, these findings are congruent with the A4i intervention goals which sought to target interpersonal aspects of psychotic disorders (24).

A second hypothesis for higher emotional sensitivity predicting better A4i response involves a potential association between interpersonal-affective sensitivity and psychotic symptom disorder severity in persons with psychosis. Interpersonal hypersensitivity has been shown characteristic of prodromal psychosis in clinically high risk patients (50). By contrast, patients at later illness stages show a generalized deficit in affect recognition, not characteristic of their earlier-illness-stage counterparts (51). Moreover, fine recognition of sad and neutral affective states have been inversely correlated with measures of disorganization in schizophrenia (51). Taken together these findings suggest (1) interpersonal and affective sensitivity as a potential surrogate marker of earlier-, or prodromal-, stage illness, which is likely to be more responsive to intervention (52) and (2) higher interpersonal sensitivity as potential marker of lower disorganization in psychotic illness, implying a greater capacity to participate and benefit from psychosocial treatments, such as A4i.

Lower psychotic and obsessive compulsive traits

Conversely, participants with *lower* baseline psychotic traits and/or *lower* baseline obsessive compulsive traits were also predicted to benefit from the A4i intervention. Psychosis is often characterized by marked perceptual disturbances and disorganization, which may suggest that individuals who were more organized were able to engage more effectively and consistently with the intervention. Notably, individuals at risk of psychosis are found to experience difficulties with interpersonal relationships, manifesting as an inability to communicate distressing psychological experiences to others (53). These troubles with communication and experiential expression likely also affect patients with high levels of psychotic traits resulting in difficulties engaging with other users of A4i, and thus the intervention overall. Similarly, intervention and community engagement may have proven difficult for patients' with obsessive compulsive traits, as these may involve intrusive, and distressing thoughts. Patients with OCD often have difficulties recognizing affective social cues, regulating emotion (54), and communicating (55). These difficulties likely also exist in individuals with primary or secondary psychoses who have obsessive compulsive traits, preventing effective engagement with socially dependent interventions, such as A4i.

App interaction and feedback

Notably, higher baseline *depression* predicted both higher A4i engagement and more satisfaction with the app (as measured by post-intervention feedback polarity, see **Figures 2B,C**). This finding is consistent with research to date demonstrating the prognostic value of affective symptoms (e.g., depression) in schizophrenia (39, 56, 57). In particular, comorbid depressive symptoms have been associated with more positive outcomes, including fewer hospitalizations and fewer illness relapses in schizophrenia (39). Considered in the context of the present study results, we hypothesize depression

to likewise represent a marker of ability to engage with a multi-feature digital intervention. Intuitively, it follows that individuals who had greater app engagement also had greater symptom improvement (see **Figure 2A**) and therefore would have greater satisfaction with the app (see **Figure 2C**).

In contrast to *depression*, higher *overall symptom severity* at baseline predicted lower A4i engagement and less favorable sentiment toward the app. We hypothesize individuals with greater symptom severity had more disorganization and executive functioning impairment, making it difficult to fully engage with a multi-feature digital intervention, like A4i. It follows that patients who interacted less with the app would have less symptom improvement (see **Figure 2A**) and therefore would have less favorable impressions (see **Figure 2C**). These findings are important in light of evidence suggesting overall lower digital app engagement among populations with schizophrenia, likely skewed by small subgroups of heavily engaged participants (58). Understanding the individual-level characteristics that drive engagement and sentiment toward digital technologies is essential for future mental health app development and implementation.

Limitations

Despite the strengths of leveraging a LOO cross-validation machine learning framework to investigate unique patient markers of digital treatment response for individuals with schizophrenia, there are a number of limitations concerning the study sample and design that should be addressed. (1) The original study used a 38-person sample drawn from urban Canadian residents, limiting generalizability of the reported (2) This study did not conduct long-term patient follow-up, an important aspect of comprehensive treatment analyses, thus it is unknown whether the observed improvements in schizophrenia psychopathology will persist for these patients. As such, the long-term efficacy of A4i, and the long-term importance of the identified unique patient markers, cannot be evaluated. (3) Due to the method of analysis, the present results only reflect predictive capacity, not causality. (4) While the BSI captures positive and negative symptoms associated with psychotic disorders, structured interviews specific to psychosis and schizophrenia (e.g., PSYRATS-D) were not included. (5) The present analyses did not incorporate demographic information or lifestyle-related information, and thus cannot account for the manner in which demographic features or living and work environments influenced participant engagement with the A4i intervention.

Conclusion

The present study sought to interrogate the A4i app as a digital intervention for schizophrenia patients *via* an ensemble LOO machine learning approach, allowing insight into the most influential, unique patient characteristics for predicting intervention response and engagement. Notably, high interpersonal sensitivity was predictive of total symptom reduction across the digital intervention, and high depression was predictive of increased digital intervention engagement and positive qualitative impressions. Taken together, these findings highlight the necessity for patient-level interrogation of treatment efficacy in mental health, particularly for schizophrenia where clinical presentation is heterogeneous. Future work should build upon the present findings to consider how individual demographic characteristics (e.g., gender, race) may influence differential engagement with a digital intervention, particularly in individuals with psychosis. Additionally, future work should aim to extend the current methodology to more traditional interventions for individuals with schizophrenia (e.g., psychosocial interventions combined with antipsychotic use) to identify unique patient characteristics predictive of intervention response, and thus tailor treatment type based on an individual's clinical presentation.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0219491. All relevant data are within the supporting information material.

Ethics statement

The studies involving human participants were reviewed and approved by Research Ethics Board, Centre for Addiction and Mental Health. The patients/participants provided their written informed consent to participate in this study.

Author contributions

GP, MH, and NJ contributed to conceptualization and design of the study. GP and MN implemented the modeling and analyses. GP, MH, MN, JM, and NJ wrote sections of the manuscript. All authors contributed to manuscript revision, read, and approved the submitted version.

Funding

This work was funded by the National Institute of Mental Health and the National Institute of General Medical Sciences under grant 1 R01 MH123482-01.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyt.2022.807116/full#supplementary-material

References

1. Saha S, Chant D, Welham J, McGrath J. A systematic review of the prevalence of schizophrenia. *PLoS Med.* (2005) 2:e141. doi: 10.1371/journal.pmed.0020141

2. Tandon R, Gaebel W, Barch DM, Bustillo J, Gur RE, Heckers S, et al. Definition and description of schizophrenia in the DSM-5. *Schizophr Res.* (2013) 150:3–10. doi: 10.1016/j.schres.2013.05.028

3. Desai PR, Lawson KA, Barner JC, Rascati KL. Estimating the direct and indirect costs for community-dwelling patients with schizophrenia: schizophrenia-related costs for community-dwellers. *J Pharmaceut Health Serv Res.* (2013) 4:187–94. doi: 10.1111/jphs.12027

4. Kessler RC, Birnbaum H, Demler O, Falloon IRH, Gagnon E, Guyer M, et al. The prevalence and correlates of nonaffective psychosis in the National Comorbidity Survey Replication (NCS-R). *Biol Psychiatry.* (2005) 58:668–76. doi: 10.1016/j.biopsych.2005.04.034

5. Wu EQ, Shi L, Birnbaum H, Hudson T, Kessler R. Annual prevalence of diagnosed schizophrenia in the USA: a claims data analysis approach. *Psychol Med.* (2006) 36:1535–40. doi: 10.1017/S0033291706008191

6. McGrath J, Saha S, Chant D, Welham J. Schizophrenia: a concise overview of incidence, prevalence, and mortality. *Epidemiol Rev.* (2008) 30:67–76. doi: 10.1093/epirev/mxn001

7. Cloutier M, Aigbogun MS, Guerin A, Nitulescu R, Ramanakumar AV, Kamat SA, et al. The economic burden of schizophrenia in the United States in 2013. *J Clin Psychiatry.* (2016) 77:764–71. doi: 10.4088/JCP.15m10278

8. Bell CC. DSM-IV: diagnostic and statistical manual of mental disorders. *JAMA.* (1994) 272:828. doi: 10.1001/jama.1994.03520100096046

9. Achim AM, Maziade M, Raymond E, Olivier D, Merette C, Roy M-A. How prevalent are anxiety disorders in Schizophrenia? A meta-analysis and critical review on a significant association. *Schizophr Bull.* (2011) 37:811–21. doi: 10.1093/schbul/sbp148

10. Häfner H, Maurer K, Trendler G, an der Heiden W, Schmidt M, Könnecke R. Schizophrenia and depression: challenging the paradigm of two separate diseases—A controlled study of schizophrenia, depression and healthy controls. *Schizophr Res.* (2005) 77:11–24. doi: 10.1016/j.schres.2005.01.004

11. Singh P, Midha A, Chugh K, Solanki R. Schizophrenia: impact on quality of life. *Indian J Psychiatry.* (2008) 50:181. doi: 10.4103/0019-5545.43632

12. Siris SG. Suicide and schizophrenia. *J Psychopharmacol.* (2001) 15:127–35. doi: 10.1177/026988110101500209

13. Kane JM, Marder SR. Psychopharmacologic treatment of schizophrenia. *Schizophr Bull.* (1993) 19:287–302. doi: 10.1093/schbul/19.2.287

14. Patel KR, Cherian J, Gohil K, Atkinson D. Schizophrenia: overview and treatment options. *P T.* (2014) 39:638–45.

15. Burke RE, Fahn S, Jankovic J, Marsden CD, Lang AE, Gollomp S, et al. Tardive dystonia: Late-onset and persistent dystonia caused by antipsychotic drugs. *Neurology.* (1982) 32:1335–1335. doi: 10.1212/WNL.32.12.1335

16. Caroff SN, Mann SC, Campbell EC, Sullivan KA. Movement disorders associated with atypical antipsychotic drugs. *J Clin Psychiatry.* (2002) 63(Suppl 4):12–9.

17. De Fazio P, Gaetano R, Caroleo M, Cerminara G, Maida F, Bruno A, et al. Rare and very rare adverse effects of clozapine. *Neuropsychiatr Dis Treat.* (2015) 11:1995–2003. doi: 10.2147/NDT.S83989

18. Lieberman JA, Stroup TS, McEvoy JP, Swartz MS, Rosenheck RA, Perkins DO, et al. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *New Engl J Med.* (2005) 353:1209–23. doi: 10.1056/NEJMoa051688

19. van Harten PN, Hoek HW, Kahn RS. Acute dystonia induced by drug treatment. *BMJ.* (1999) 319:623–6.

20. Moritz S, Andreou C, Klingberg S, Thoering T, Peters MJV. Assessment of subjective cognitive and emotional effects of antipsychotic drugs. Effect by defect? *Neuropharmacology.* (2013) 72:179–86. doi: 10.1016/j.neuropharm.2013.04.039

21. Biagianti B, Hidalgo-Mazzei D, Meyer N. Developing digital interventions for people living with serious mental illness: Perspectives from three mHealth studies. *Evid Based Ment Health.* (2017) 20:98–101. doi: 10.1136/eb-2017-102765

22. Schlosser D, Campellone T, Kim D, Truong B, Vergani S, Ward C, et al. Feasibility of PRIME: a cognitive neuroscience-informed mobile app intervention to enhance motivated behavior and improve quality of life in recent onset schizophrenia. *JMIR Res Protocols.* (2016) 5:e77. doi: 10.2196/resprot.5450

23. Schlosser DA, Campellone TR, Truong B, Etter K, Vergani S, Komaiko K, et al. Efficacy of PRIME, a mobile app intervention designed to improve motivation in young people with schizophrenia. *Schizophr Bull.* (2018) 44:1010–20. doi: 10.1093/schbul/sby078

24. Kidd SA, Feldcamp L, Adler A, Kaleis L, Wang W, Vichnevetski K, et al. Feasibility and outcomes of a multi-function mobile health approach for the schizophrenia spectrum: App4Independence (A4i). *PLoS One.* (2019) 14:e0219491. doi: 10.1371/journal.pone.0219491

25. Buchanan RW, Carpenter WT. Domains of psychopathology: an approach to the reduction of heterogeneity in schizophrenia. *J Nerv Mental Dis.* (1994) 182:193–204.

26. Buchsbaum MS, Haier RJ. Biological homogeneity, symptom heterogeneity, and the diagnosis of schizophrenia*. *Schizophr Bull.* (1978) 4:473–5. doi: 10.1093/schbul/4.4.473

27. Carpenter WT, Kirkpatrick B. The heterogeneity of the long-term course of schizophrenia. *Schizophr Bull.* (1988) 14:645–52. doi: 10.1093/schbul/14.4.645

28. Lang FU, Kösters M, Lang S, Becker T, Jäger M. Psychopathological long-term outcome of schizophrenia—A review. *Acta Psychiatr Scand.* (2013) 127:173–82. doi: 10.1111/a.12030

29. Wolfers T, Doan NT, Kaufmann T, Alnæs D, Moberget T, Agartz I, et al. Mapping the heterogeneous phenotype of schizophrenia and bipolar disorder using normative models. *JAMA Psychiatry.* (2018) 75:1146. doi: 10.1001/jamapsychiatry.2018.2467

30. Meinlschmidt G, Tegethoff M, Belardi A, Stalujanis E, Oh M, Jung EK, et al. Personalized prediction of smartphone-based psychotherapeutic micro-intervention success using machine learning. *J Affect Disord.* (2020) 264:430–7. doi: 10.1016/j.jad.2019.11.071

31. Nemesure MD, Heinz MV, Klein R, McFadden J, Jacobson NC. Predictive modeling approach to evaluate individual response to a physical activity digital intervention for subjects with major depressive disorder. *PsyArXiv [Preprint].* (2021). Available online at: https://psyarxiv.com/3kjyh/ (accessed on June 06, 2022). doi: 10.31234/osf.io/3kjyh

32. Chand GB, Dwyer DB, Erus G, Sotiras A, Varol E, Srinivasan D, et al. Two distinct neuroanatomical subtypes of schizophrenia revealed using machine learning. *Brain.* (2020) 143:1027–38. doi: 10.1093/brain/awaa025

33. Schnack HG, Nieuwenhuis M, van Haren NEM, Abramovic L, Scheewe TW, Brouwer RM, et al. Can structural MRI aid in clinical classification? A machine learning study in two independent samples of patients with schizophrenia, bipolar disorder and healthy subjects. *Neuroimage.* (2014) 84:299–306. doi: 10.1016/j.neuroimage.2013.08.053

34. Veronese E, Castellani U, Peruzzo D, Bellani M, Brambilla P. Machine learning approaches: from theory to application in schizophrenia. *Computat Math Methods Med.* (2013) 2013:867924 . doi: 10.1155/2013/867924

35. Bae YJ, Shim M, Lee WH. Schizophrenia detection using machine learning approach from social media content. *Sensors.* (2021) 21:5924. doi: 10.3390/s21175924

36. Birnbaum ML, Ernala SK, Rizvi AF, De Choudhury M, Kane JM. A collaborative approach to identifying social media markers of schizophrenia by employing machine learning and clinical appraisals. *J Med Internet Res.* (2017) 19:e289. doi: 10.2196/jmir.7956

37. Wang R, Wang W, Obuchi M, Scherer E, Brian R, Ben-Zeev D, et al. On predicting relapse in schizophrenia using mobile sensing in a randomized control trial. *Proceedings of the 2020 IEEE International Conference on Pervasive Computing and Communications (PerCom).* Austin, TX: IEEE (2020). p. 1–8. doi: 10.1109/PerCom45495.2020.9127365

38. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Red Hook, NY: Curran Associates Inc (2017). p. 4765–74.

39. Arrasate M, González-Ortega I, García-Alocén A, Alberich S, Zorrilla I, González-Pinto A. Prognostic value of affective symptoms in first-admission psychotic patients. *Int J Mol Sci.* (2016) 17:1039. doi: 10.3390/ijms17071039

40. Byerly MJ, Nakonezny PA, Rush AJ. The Brief Adherence Rating Scale (BARS) validated against electronic monitoring in assessing the antipsychotic medication adherence of outpatients with schizophrenia and schizoaffective disorder. *Schizophr Res.* (2008) 100:60–9. doi: 10.1016/j.schres.2007.12.470

41. Barbic SP, Kidd SA, Durisko ZT, Yachouh R, Rathitharan G, McKenzie K. What are the personal recovery needs of community-dwelling individuals with mental illness? Preliminary findings from the Canadian personal recovery outcome measurement (C-PROM) Study. *Can J Commun Ment Health.* (2018) 37:29–47. doi: 10.7870/cjcmh-2018-005

42. Derogatis LR, Melisaratos N. The brief symptom inventory: an introductory report. *Psychol Med.* (1983) 13:595–605.

43. Rossum G, van Drake FL, Van Rossum G. *The Python Language Reference (Release 3.0.1 [Repr.])*. Wilmington: Python Software Foundation (2010).

44. Loria S. *Textblob Documentation*. (2018). Available online at: https://buildmedia.readthedocs.org/media/pdf/textblob/latest/textblob.pdf (accessed April 26, 2020).

45. Thara DK, PremaSudha BG, Xiong F. Auto-detection of epileptic seizure events using deep neural network with different feature scaling techniques. *Pattern Recogn Lett.* (2019) 128:544–50. doi: 10.1016/j.patrec.2019.10.029

46. Webb GI, Sammut C, Perlich C, Horváth T, Wrobel S, Korb KB, et al. Leave-one-out cross-validation. In: Sammut C, Webb GI editors. *Encyclopedia of Machine Learning*. Berlin: Springer (2011). p. 600–1. doi: 10.1007/978-0-387-30164-8_469

47. Varma S, Simon R. Bias in error estimation when using cross-validation for model selection. *BMC Bioinformatics.* (2006) 7:91. doi: 10.1186/1471-2105-7-91

48. Tao X, Chi O, Delaney PJ, Li L, Huang J. Detecting depression using an ensemble classifier based on Quality of Life scales. *Brain Informat.* (2021) 8:2. doi: 10.1186/s40708-021-00125-5

49. Stevens AK, McNichol J, Magalhaes L. Social relationships in schizophrenia: a review. *Pers Ment Health.* (2009) 3:203–16. doi: 10.1002/pmh.82

50. Masillo A, Day F, Laing J, Howes O, Fusar-Poli P, Byrne M, et al. Interpersonal sensitivity in the at-risk mental state for psychosis. *Psychol Med.* (2012) 42:1835–45. doi: 10.1017/S0033291711002996

51. Romero-Ferreiro MV, Aguado L, Rodriguez-Torresano J, Palomo T, Rodriguez-Jimenez R, Pedreira-Massa JL. Facial affect recognition in early and late-stage schizophrenia patients. *Schizophr Res.* (2016) 172:177–83. doi: 10.1016/j.schres.2016.02.010

52. McGlashan TH, Johannessen JO. Early detection and intervention with schizophrenia: rationale. *Schizophr Bull.* (1996) 22:201–22. doi: 10.1093/schbul/22.2.201

53. Byrne R, Morrison AP. Young people at risk of psychosis: a user-led exploration of interpersonal relationships and communication of psychological difficulties: young people at risk of psychosis. *Early Intervent Psychiatry.* (2010) 4:162–8. doi: 10.1111/j.1751-7893.2010.00171.x

54. Jansen M, Overgaauw S, De Bruijn ERA. Social cognition and obsessive-compulsive disorder: a review of subdomains of social functioning. *Front Psychiatry.* (2020) 11:118. doi: 10.3389/fpsyt.2020.00118

55. Cullen B, Samuels J, Grados M, Landa R, Bienvenu OJ, Liang K-Y, et al. Social and communication difficulties and obsessive-compulsive disorder. *Psychopathology.* (2008) 41:194–200. doi: 10.1159/000120988

56. McCabe MS, Fowler RC, Cadoret RJ, Winokur G. Symptom differences in schizophrenia with good and poor prognosis. *Am J Psychiatry.* (1972) 128:1239–43. doi: 10.1176/ajp.128.10.1239

57. Morrissette DA, Stahl SM. Affective symptoms in schizophrenia. *Drug Discovery Today.* (2011) 8:3–9. doi: 10.1016/j.ddstr.2011.10.005

58. Torous J, Staples P, Slaters L, Adams J, Sandoval L, Onnela J, et al. Characterizing smartphone engagement for schizophrenia: results of a naturalist mobile health study. *Clin Schizophr Relat Psychoses.* (2017) [Online ahead of print]. doi: 10.3371/CSRP.JTPS.071317.

frontiers | Frontiers in Digital Health

TYPE Perspective/Commentary
PUBLISHED 11 August 2022
DOI 10.3389/fdgth.2022.890081

OPEN ACCESS

EDITED BY
Patricia A. Arean,
University of Washington, Seattle, United States

REVIEWED BY
Johannes Schobel,
Neu-Ulm University of Applied Sciences, Germany

*CORRESPONDENCE
Jessica E. Flannery
flannery@limbix.com

SPECIALTY SECTION
This article was submitted to Digital Mental Health, a section of the journal Frontiers in Digital Health

RECEIVED 05 March 2022
ACCEPTED 26 July 2022
PUBLISHED 11 August 2022

CITATION
Strauss G, Flannery JE, Vierra E, Koepsell X, Berglund E, Miller I and Lake JI (2022) Meaningful engagement: A crossfunctional framework for digital therapeutics.
Front. Digit. Health 4:890081.
doi: 10.3389/fdgth.2022.890081

Meaningful engagement: A crossfunctional framework for digital therapeutics

Gabriel Strauss, Jessica E. Flannery*, Elise Vierra, Xin Koepsell, Emily Berglund, Ian Miller and Jessica I. Lake

Limbix Health, San Francisco, California, United States

Digital mental health interventions, or digital therapeutics, have the potential to transform the field of mental health. They provide the opportunity for increased accessibility, reduced stigma, and daily integration with patient's lives. However, as the burgeoning field continues to expand, there is a growing concern regarding the level and type of engagement users have with these technologies. Unlike many traditional technology products that have optimized their user experience to maximize the amount of time users spend within the product, such engagement within a digital therapeutic is not sufficient if users are not experiencing an improvement in clinical outcomes. In fact, a primary challenge within digital therapeutics is user engagement. Digital therapeutics are only effective if users sufficiently engage with them and, we argue, only if users meaningfully engage with the product. Therefore, we propose a 4-step framework to assess meaningful engagement within digital therapeutics: (1) Define the measure of value (2) Operationalize meaningful engagement for your digital therapeutic (3) Implement solutions to increase meaningful engagement (4) Iteratively evaluate the solution's impact on meaningful engagement and clinical outcomes. We provide recommendations to the common challenges associated with each step. We specifically emphasize a cross-functional approach to assessing meaningful engagement and use an adolescent-focused example throughout to further highlight developmental considerations one should consider depending on their target users.

KEYWORDS

digital therapeutics, engagement, meaningful engagement, clinical outcomes, adolescence

Introduction

Digital therapeutics are evidence-based mental health therapeutic interventions driven by software programs to prevent, manage, or treat a medical disorder or disease (1). The demand for digital therapeutics is steadily increasing as the demand for mental health services increases, but existing face-to-face services remain limited. Digital therapeutics have the opportunity to transform the field of mental health– improving access to quality mental health services and the mental health of populations previously neglected in treatment options. Several factors can further limit an individual's ability or desire to seek face-to-face treatment, such as stigma, cultural acceptance, embarrassment, access to services, financial constraints, or a preference for self-reliance (2, 3). For developmental

populations, access can be further limited by transportation, child-care for other siblings, or caregiver alignment with the need for therapy (2, 3). While digital therapeutics offer a promising solution to barriers in access to care, user engagement within digital therapeutics is a primary challenge the industry faces (4). Without sufficient user engagement, the success and promise of digital therapeutics is limited.

To encourage sufficient engagement, we believe it's critical to integrate cross-functional perspectives from clinical science, product management, product design, content, and user experience research teams to assess common challenges and recommendations for engagement within digital therapeutics. Within this context, we discuss the various definitions of engagement and advocate for alignment with a focus on *measures of meaningful engagement*. Drawing from work in the consumer and software as a service (SaaS) industries, we propose a 4-step framework to address common challenges and recommendations for identifying, measuring, and driving meaningful engagement in digital therapeutics: (1) Define the measure of value (2) Operationalize meaningful engagement for your digital therapeutic (3) Implement solutions to increase meaningful engagement (4) Iteratively evaluate the solution's impact on meaningful engagement and clinical outcomes (See **Figure 1**.) This process uses the build-measure-learn cycle, emphasizing theory, user-centered design, and quantitative and qualitative feedback. We focus on implementing our four-step process in digital therapeutic development targeting adolescents. We also provide recommendations to overcome the general challenges digital therapeutics face in assessing and encouraging meaningful engagement in treatment within and outside of the app.

Defining engagement

While there is broad agreement that product engagement is critical for digital therapeutics to be impactful, the precise definition of engagement and measurement processes are vague (5–7). Increasingly, digital health technologists and researchers agree that assessing "generic" measures of engagement (e.g., number of sessions, weekly active usage, or program completion) may not be sufficient if they are not strong mediators of outcomes (8–10). Consequently, there is a growing call to follow a clinically informed and data-driven approach to identify specific engagement metrics that uniquely predict the long-term value for a digital therapeutic, referred to as *measures of meaningful engagement* (4, 9, 11, 12).

Driving meaningful engagement

To drive meaningful engagement in a digital therapeutic we built upon two well-known frameworks by adapting them for the unique challenges of digital therapeutic development: The Build-Measure-Learn framework popularized by The Lean Startup and the Design Thinking framework popularized by IDEO (13, 14). Both processes have significant similarities (15). They each involve building prototypes, testing and measuring the success of those prototypes through qualitative user feedback and quantitative experiment, generating insights and applying them to subsequent iterations of prototyping, testing, and learning. A key tenet of the Lean Startup framework is to define an appropriate metric for the success of a product (13). As has already been mentioned, the digital health industry needs to move away from generic measures of engagement, and identify metrics that predict clinical outcomes. We therefore elevated this into dedicated steps in our process. A key tenet of design thinking is to "understand" and "observe" users through the synthesis of existing research and by directly engaging with users as part of the design process (14). In line with this, our process emphasizes drawing upon clinical science and theory, and involving users as active partners in the design process.

We distilled these elements into a 4-step process-oriented framework for driving meaningful engagement in digital therapeutics that accounts for the unique challenges faced by developers of digital therapeutics.

1: Define the measure of value

Defining the measure of value is an area in which there is a fundamental difference between digital therapeutics and most consumer or SaaS products. For example, in a consumer product like TikTok, users are looking to be entertained, and therefore retention (consistently coming back to the app) is an excellent measure of the value of that product. Alternatively, most digital therapeutics' primary measure of value is clinical outcomes. Users have health needs and digital therapeutics must address those needs. If users come back to the app every day for months (strong retention), but their symptoms do not improve, they have not received the primary intended value from the product.

Recommendation: Optimize for clinical outcomes. As outlined above, for most digital therapeutics, the primary goal is to improve patients' clinical outcomes. Take this example of a cognitive behavioral therapy (CBT)-based digital therapeutic for adolescent depression. Patients and their parents, providers, and payors all care about improving patients' depressive symptoms. We analyze the decrease in the patient health questionnaire score (PHQ score; a measure of depressive symptom severity; (16). As such, we evaluate all efforts to improve engagement against their impact on reducing PHQ scores. We continue with this example below.

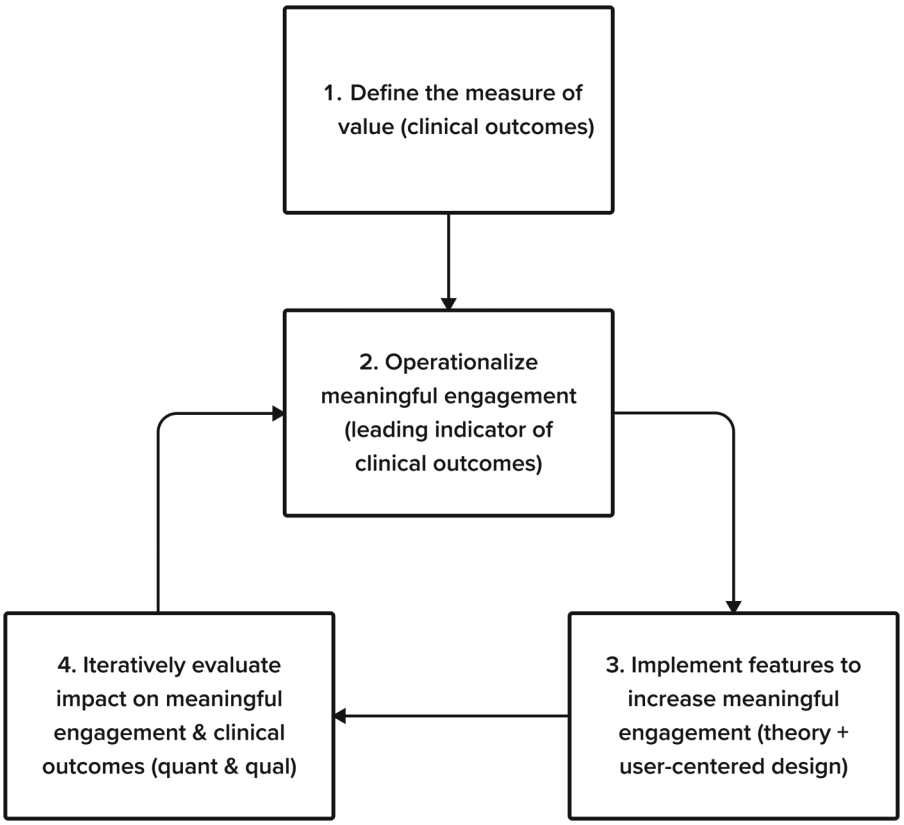


FIGURE 1
Our 4-step framework to address common challenges and recommendations for identifying, measuring, and driving meaningful engagement in digital therapeutics.

2: Operationalize meaningful engagement for your digital therapeutic

After identifying the measure of value, a common challenge is selecting the engagement behaviors that are the strongest leading indicators of that value. In product development, there are often limited resources; therefore, efforts must be focused where they will have the greatest clinical return on investment.

In our example digital therapeutic, we seek the *best* leading indicator of a drop in PHQ score. Such engagement metrics usually correspond to interactions with the "active ingredients" in a digital therapeutic. Just as traditional medicine can be fractionated into a delivery mechanism and the active ingredients (e.g., a pill and the drug compound), so too can digital therapeutics be fractionated into delivery mechanisms and their active ingredients (e.g., screen views and scheduling an in-app behavioral activation). However, identifying these active ingredients can prove difficult, as this fractionated process is still under development for face-to-face interventions.

Recommendation: Investigate predictors of positive clinical outcomes and map out possible digital analogs. We recommend taking a theory-driven approach to identifying clinical outcomes and then validating this with product data. Many digital health interventions are digitized versions of face-to-face interventions; therefore, it can be common to examine theoretical predictors of clinical outcomes in face-to-face interventions and then look for digital analogs (i.e., the digital version of what happens face-to-face) of those factors. As noted above, while these theoretical predictors are still debated within face-to-face interventions, they provide a foundation to start hypothesis testing, particularly for new digital therapeutics.

Coming back to our example to select our measure of meaningful engagement, we first investigate what predicts favorable clinical outcomes in face-to-face CBT for adolescents and identify multiple predictors. Next, we map the theoretical indicators to specific, measurable interactions within the digital intervention.

For example, we may first identify the following predictors of positive clinical outcomes in face-to-face CBT:

1. Showing up for weekly appointments
2. Doing assigned homework (e.g., completing mood-activity logs and behavioral activations) (16)
3. Having a positive therapeutic alliance with the therapist
4. Parent involvement (e.g., showing up to session, supporting teen in homework)

Which may correspond to the following digital analogs:

1. Weekly active usage (a generic engagement measure)
2. Completing specific, in-app therapeutic exercises (e.g., logging behavioral activations)
3. A questionnaire that measures therapeutic alliance as part of weekly symptom check-ins
4. Completion of parent assigned tasks, or adolescent and parent reported ratings of parental support in homework

Based on the above, we might choose completing specific in-app exercises, such as behavioral activation logs, as the most promising engagement metric, because it is the in-app action most closely related to the hypothesized active ingredients within behavioral activation therapy. Despite the heterogeneity of engagement metrics used to evaluate digital therapeutics, in the literature, adherence to the recommend usage has been shown to be strong indicator of positive outcomes (17). This behavior (adherence) is also understood to be influenced by a users' developmental stage (e.g., age).

For new digital therapeutic programs or those with limited data, a large part of this challenge is that there is little to no prior evidence of the leading indicators of clinical outcomes within this modality, which may force one to rely on theory alone. For products with existing data, however, exploratory analyses can be conducted to refine the theory-based hypotheses.

3. Implement solutions to increase meaningful engagement

3a: Hypothesize theory-driven solutions that will drive meaningful engagement in the digital therapeutic

This can be difficult within the new space of digital therapeutics, when prior solutions to increase engagement were largely based on generic metrics of engagement or consumer based products (e.g., retention for retention's sake; (11).

Recommendation: Identify engagement techniques based on developmental, behavioral, and clinical science theory and research. To design solutions that improve clinical outcomes by effectively driving meaningful engagement, we advocate for leveraging an understanding of behavioral change techniques. For a focus on adolescent depression, it is further important to leverage behavioral change theory from a developmental lens. There are many well-established techniques to improve user engagement that are beyond the scope of the article–such as usability, visual design, narratives, goal-setting, self-monitoring, professional support, reminders, interactivity, narrative, user control accountability, personalization, social support, digital therapeutic alliance, credibility, and treatment expectancy (11). Here, we specifically focus on an example with a developmental lens (2, 18–20).

Since the early days of behaviorism, researchers have long established that rewards are one of the most effective ways to influence behavior (21, 22). Reward systems within digital therapeutics generally incentivize target behaviors by providing extrinsic rewards, such as badges, points, or level progression (23). Rewards can be provided for the target behavior itself (e.g., going for a run), effort towards the target behavior (e.g., scheduling a run), or approximations to the target behavior (e.g., going for a walk). Rewards are often grouped into a larger category of gamification elements, which are designed to provide extrinsic motivation to engage with the intervention. Though gamification elements can improve engagement in digital health interventions, it's worth noting there is some debate (24, 25), and there are very few studies examining the precise impact of rewards by comparing the same intervention with and without rewards or other gamification elements. There are other components of gamification that can influence users' motivation beyond reward, however, such as motivation by purpose, autonomy, relatedness, or competence (26). For an adolescent focused intervention, there are unique developmental considerations that influence motivation, reward, and punishment (27–29). Those additional aspects of gamification (e.g., autonomy and relatedness) also have particular developmental relevance to adolescents. For example, during adolescence, young people are gaining more autonomy from their parents, exploring their self-identity, and are neurobiologically more sensitive to social rewards, making them more likely to take a riskier (unknown) option for the opportunity to learn, than choose a known reward (30). A successful adolescent-focused reward structure should integrate those considerations. For example, is there a way to offer a menu of tailored reward options, ensure rewards are salient to your target population, provide frequent and different sizes of rewards toward incremental progress, or add an element of choice regarding when they cash in rewards?

3b: Tailor engagement technique for maximum impact

Understanding the theory behind behavioral change techniques is necessary but insufficient to drive engagement. The same behavioral change technique can lead to very different effect sizes (31). A major reason for this is that a

behavioral change technique can be implemented in a variety of contexts (32). It is critical to tailor the implementation of the technique to the specific characteristics of your users.

Recommendation: **Implement user-centered design processes to fine-tune the implementation of the engagement techniques.** User-centered design is an approach that focuses on users and their needs in every step of the design process (33). To this end, we recommend employing a range of techniques for engaging end-users as creative partners in the design process.

The reason for incorporating a user-centered design process is simple: product developers are not often their users. Due to differences across age, culture, life experience, and cognition, what is engaging to the developer may be different than what is engaging to their users. In building a digital intervention for adolescent depression, even if developers may remember (or think they remember!) what it was like to be an adolescent, there is no substitute for incorporating the voices of end users. Figure 2 illustrates an example product development process paired with steps for user-centered design.

There is a wide range of tools available, and it is essential to know which tool to use at each stage of the product development cycle. The first step of this process is to create a product requirement document (PRD). We can use **problem interviews** at this early stage to identify user needs and how users are currently addressing those needs.

After creating the PRD, we generate initial solution concepts using **co-design.** Co-design involves working with participants to generate potential solutions to a design problem. This often involves a process in which participants sketch out potential solutions, share them with each other, and then iterate [For more details, see (34)].

We then create a low-fidelity prototype, a rudimentary solution abstraction. We gain feedback from users with **group critiques and brainstorming**—sometimes referred to as a solution interview. We brief participants on the design goal, show them early designs, gather feedback, and then ask them to brainstorm improvements as a group.

Next, we create a high-fidelity prototype. We run **usability tests,** which involve users completing defined tasks within a prototype while they think aloud. The goal is to assess how easy the intervention is to use and understand.

After incorporating feedback, the product development team may prepare a limited release for **diary studies**– which involve asking users to document their experiences and thoughts while completing the intervention. Diary studies enable us to get in-the-moment feedback that more accurately represents users' experiences.

After implementing the solution in a larger release (e.g., in a clinical trial), we gather more feedback through **user interviews.** This involves collecting feedback from users after they complete the intervention through a series of structured questions about their experience. Moving from this stage to an even larger public release brings up new challenges and considerations, discussed below.

4: Iteratively evaluate the solution's impact on meaningful engagement and clinical outcomes

4a: Test product changes

Developing a feature with qualitative feedback (user-centered design process) also requires quantitative evaluation with users to first, evaluate the effect on the meaningful engagement metric and, secondly, to determine if that metric was associated with a change in clinical outcomes. In the wider technology industry, running A/B tests– continuously

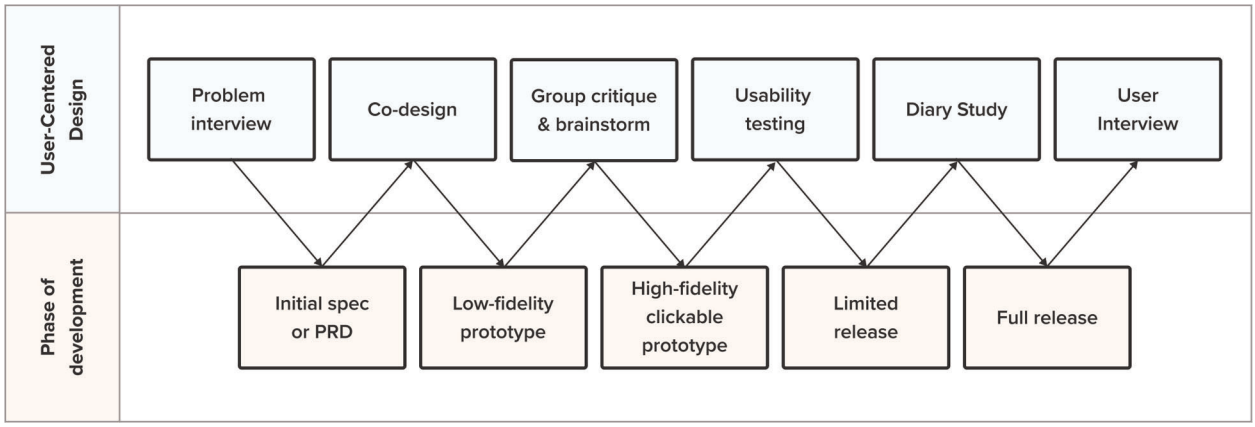


FIGURE 2
An example product development process paired with steps for user-centered design.

releasing minor changes to a subset of users and comparing outcomes– is common. However, for digital therapeutics, the impact of this type of testing could have major impacts on users' health and wellbeing, so A/B testing may not be feasible from a safety or regulatory perspective (e.g., if your product is FDA-regulated).

Recommendation: Run a series of small-scale studies before releasing public changes. Coming back to our example, we run a series of small-scale Institutional Review Board (IRB)-regulated studies to test specific hypotheses around meaningful engagement. Thus, many digital therapeutic developers build their own internal research infrastructure to run small-scale clinical trials much more quickly and inexpensively than would otherwise be possible by partnering with third-party research organizations. For example, in one study, we can test if breaking down behavioral activations into smaller chunks, with more frequent rewards for incremental completion, increases the number completed (a hypothesized indicator of clinical outcomes).

Once we confirm the feature is safe for adolescents and leads to similar or better clinical outcomes than the existing product version, we can release it to the entire user base.

4b: Establishing mechanisms of action

After running each of the above-mentioned trials, it can still be difficult to determine causality and empirically validate whether the hypothesized meaningful engagement metric contributes to clinical outcomes. Psychological processes are complex and require massive data sets to untangle the many competing factors contributing to outcomes. Even in face-to-face interventions, which have undergone decades of research and clinical trials, researchers are still attempting to pinpoint the therapeutic "active ingredients" that contribute to clinical outcomes (35, 36). Furthermore, it is also important to be cautious about mining the data to find correlations [p-hacking or hypothesizing after the fact; (37)], which can lead to spurious conclusions about these mechanisms of action and ultimately irreproducible effects.

Recommendation: Rely on theory where appropriate and be cognizant of limitations. There are no easy solutions to this challenge. As digital therapeutics scale, there is real potential to gain the critical mass of data necessary to identify reliable effect sizes. This is one of the major advantages of digital therapeutics over traditional therapies.

In the meantime, we recommend understanding the theoretical mechanisms underlying clinical outcomes and taking a cross-functional approach to triangulating the "why" of an outcome. For example, after we see quantitative support for our hypothesis, we can bring back in user experience research to interview a representative cohort of study participants to better understand the qualitative "why" behind any quantitative patterns we observed. Furthermore, finding a sustained correlation between the use of the leading indicator and the clinical outcomes may be a good indication that we have found a meaningful engagement metric.

Regardless of sample size, to avoid the trap of data mining for correlations that leads to spurious findings, we recommend following transparent and reproducible study design and analysis pipelines (e.g., pre-registration, open code, clearly labeled exploratory findings in studies, and heavier reliance on effect size than *p*-value (38, 39).

Discussion

We proposed a 4-step framework to tackle common challenges to creating digital therapeutics with meaningful engagement. As the field evolves and more data are available, however, there are additional challenges and opportunities to consider, such as blending multiple metrics, segmentation, measuring behavior outside of the app, and determining the minimum effective dose. For example, there is rarely only one metric of meaningful engagement in an intervention. Instead, there may be multiple metrics, in which case they may be combined into a hybrid measure of engagement or you might categorize someone as meaningfully engaged if they do any two out of a list of five leading indicators within the program in a given week (40, 41). Meaningful engagement is also likely to differ across users. For example, users with more severe symptoms might benefit from a different style of engagement than users with mild-to-moderate symptoms or differ across users of different socioeconomic, geographic, or racial backgrounds. Optimal engagement style may even change for the same user as they progress through the intervention or recovery, or change based on the user's starting motivation types, as detailed in the Hexard Scale for gamification (26). To this end, SilverCloud and Microsoft recently published an article that outlined their use of machine learning to identify different engagement styles (8). It is also worth noting that constraining meaningful engagement metrics to objective in-app measures may limit the ability to detect real-world clinical outcome improvement. Assessing digital biomarkers (objective and passive user data), such as wearable devices or smartphone interaction patterns may afford a better opportunity to detect real-world indicators of clinical outcomes (42). With the proliferation of digital health apps, standardized frameworks, e.g., the Mobile App Rating Scale (43), will be increasingly useful for evaluating the

quality of a mobile app on a number of dimensions, including engagement. To ensure digital therapeutics meet a high quality bar for engagement, it will be prudent to adopt a cross-functional framework grounded in theoretical, user-centered, and rigorous approaches to design and interpretation to optimally determine meaningful engagement, and ultimately improve clinical outcomes for the intended users.

Author contributions

All authors contributed to the ideas, framework, editing, and revising of this manuscript. Authorship is in order of level of that contribution. Each author also uniquely contributed based on their cross-functional expertise: GS is director of product, JF is head of clinical science, EV is director of content, XK is director of design, EB is a marketing content writer, IM is a senior project analyst, JIL is chief science officer. All authors contributed to the article and approved the submitted version.

Funding

Product development work that contributed to the theory of this paper was funded by NIH grant 1R44MH125636.

Conflict of interest

All authors are employees at Limbix Health, Inc., a digital therapeutics company for adolescent mental health.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Understanding DTx. Digital Therapeutics Alliance. Available at: https://dtxalliance.org/understanding-dtx/ (cited 2022 Mar 3).

2. Liverpool S, Mota CP, Sales CMD, Čuš A, Carletto S, Hancheva C, et al. Engaging children and young people in digital mental health interventions: Systematic review of modes of delivery, facilitators, and barriers. *J Med Internet Res.* (2020) 22(6):e16317. doi: 10.2196/16317

3. Gulliver A, Griffiths KM, Christensen H. Perceived barriers and facilitators to mental health help-seeking in young people: A systematic review. *BMC Psychiatry.* (2010) 10(1):113. doi: 10.1186/1471-244X-10-113

4. Torous J, Michalak EE, O'Brien HL. Digital health and engagement—looking behind the measures and methods. *JAMA Netw Open.* (2020) 3(7):e2010918. doi: 10.1001/jamanetworkopen.2020.10918

5. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. *Am J Prev Med.* (2016) 51(5):833–42. doi: 10.1016/j.amepre.2016.06.015

6. Luderer HF, Campbell ANC, Nunes EV, Enman NM, Xiong X, Gerwien R, et al. Engagement patterns with a digital therapeutic for substance use disorders: Correlations with abstinence outcomes. *J Subst Abuse Treat.* (2022) 132:108585. doi: 10.1016/j.jsat.2021.108585

7. Kim JP, Hwang J, Tregarthen J. Engagement patterns of users and providers: A study of messaging on app usage in a smartphone app for the treatment of eating disorders. *J Technol Behav Sci.* (2021) 6(4):572–7. doi: 10.1007/s41347-021-00220-8

8. Chien I, Enrique A, Palacios J, Regan T, Keegan D, Carter D, et al. A machine learning approach to understanding patterns of engagement with internet-delivered mental health interventions. *JAMA Netw Open.* (2020) 3(7):e2010791. doi: 10.1001/jamanetworkopen.2020.10791

9. Ng MM, Firth J, Minen M, Torous J. User engagement in mental health apps: a review of measurement, reporting, and validity. *Psychiatr Serv Wash DC.* (2019) 70(7):538–44. doi: 10.1176/appi.ps.201800519

10. Donkin L, Hickie IB, Christensen H, Naismith SL, Neal B, Cockayne NL, et al. Rethinking the dose-response relationship between usage and outcome in an online intervention for depression: randomized controlled trial. *J Med Internet Res.* (2013) 15(10):e2771. doi: 10.2196/jmir.2771

11. Yeager CM, Benight CC. If we build it, will they come? Issues of engagement with digital health interventions for trauma recovery. *mHealth.* (2018) 4:37–37. doi: 10.21037/mhealth.2018.08.04

12. Birnbaum F, Lewis D, Rosen RK, Ranney ML. Patient engagement and the design of digital health. *Acad Emerg Med Off J Soc Acad Emerg Med.* (2015) 22 (6):754–6. doi: 10.1111/acem.12692

13. Reis E. The lean startup. New York: Crown Business (2011). p. 2016–20.

14. Plattner H, Meinel C, Weinberg U. *Design Thinking.* Munich: mi-wirtschaftsbuch (2009).

15. Mueller RM, Thoring K. Design thinking vs. lean startup: A Comparison of two user-driven innovation strategies (2012).

16. McCauley E, Schloredt KA, Gudmundsen GR, Martell CR. *Dimidjian S. behavioral activation with adolescents: A clinician's guide.* New York, NY, US: Guilford Press (2016). xiv, p. 226 (Behavioral activation with adolescents: A clinician's guide.).

17. Fuhr K, Schröder J, Berger T, Moritz S, Meyer B, Lutz W, et al. The association between adherence and outcome in an internet intervention for depression. *J Affect Disord.* (2018) 229:443–9. doi: 10.1016/j.jad.2017.12.028

18. Szinay D, Jones A, Chadborn T, Brown J, Naughton F. Influences on the uptake of and engagement with health and well-being smartphone apps: systematic review. *J Med Internet Res.* (2020) 22(5):e17572. doi: 10.2196/17572

19. Jeffrey B, Bagala M, Creighton A, Leavey T, Nicholls S, Wood C, et al. Mobile phone applications and their use in the self-management of type 2 diabetes Mellitus: a qualitative study among app users and non-app users. *Diabetol Metab Syndr.* (2019) 11:84.

20. Carman KL, Dardess P, Maurer M, Sofaer S, Adams K, Bechtel C, et al. Patient and family engagement: A framework for understanding the elements and developing interventions and policies. *Health Aff (Millwood).* (2013) 32 (2):223–31. doi: 10.1377/hlthaff.2012.1133

21. Schultz W. Predictive reward signal of dopamine neurons. *J Neurophysiol.* (1998) 80(1):1–27. doi: 10.1152/jn.1998.80.1.1

22. Galván A. Adolescent development of the reward system. *Front Hum Neurosci.* (2010) 4. Available at: https://www.frontiersin.org/article/10.3389/neuro.09.006.2010, (cited 2022 Feb 25).

23. Cugelman B. Gamification: What it is and why it matters to digital health behavior change developers. *JMIR Serious Games.* (2013) 1(1):e3. doi: 10.2196/games.3139

24. Looyestyn J, Kernot J, Boshoff K, Ryan J, Edney S, Maher C. Does gamification increase engagement with online programs? A systematic review. *Plos One*. (2017) 12(3):e0173403. doi: 10.1371/journal.pone.0173403

25. Schmidt-Kraepelin M, Thiebes S, Stepanovic S, Mettler T, Sunyaev A. Gamification in health behavior change support systems - A synthesis of unintended side effects (2019).

26. Tondello GF, Wehbe RR, Diamond L, Busch M, Marczewski A, Nacke LE. *The gamification user types hexad scale*. In: *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play*; Austin, Texas, USA, (2016). ACM. p. 229–43 (cited 2022 Jul 10). Available at: https://dl.acm.org/doi/10.1145/2967934.2968082

27. Fareri DS, Martin LN, Delgado MR. Reward-related processing in the human brain: Developmental considerations. *Dev Psychopathol*. (2008) 20 (4):1191–211. doi: 10.1017/S0954579408000576

28. Cuartas J, Ward KP, Ma J, Grogan-Kaylor A. Physical punishment and Colombian children and adolescents' cognitive and behavioral outcomes. *J Appl Dev Psychol*. (2020) 68:101140. doi: 10.1016/j.appdev.2020.101140

29. Simons RL, Wu CI, Lin KH, Gordon L, Conger RD. A cross-cultural examination of the link between corporal punishment and adolescent antisocial behavior. *Criminology*. (2000) 38(1):47–80. doi: 10.1111/j.1745-9125.2000.tb00883.x

30. Sanders RA. Adolescent psychosocial, social, and cognitive development. *Pediatr Rev*. (2013) 34(8):354–8; quiz 358–9. doi: 10.1542/pir.34.8.354

31. Peters GJY, de Bruin M, Crutzen R. Everything should be as simple as possible, but no simpler: towards a protocol for accumulating evidence regarding the active content of health behaviour change interventions. *Health Psychol Rev*. (2015) 9(1):1–14. doi: 10.1080/17437199.2013.848409

32. Moller AC, Merchant G, Conroy DE, West R, Hekler E, Kugler KC, et al. Applying and advancing behavior change theories and techniques in the context of a digital health revolution: proposals for more effectively realizing untapped potential. *J Behav Med*. (2017) 40(1):85–98. doi: 10.1007/s10865-016-9818-7

33. Abras C, Maloney-Krichmar D, Preece J. *User-centered Design*. *Bainbridge W Encycl Hum-Comput Interact*. (2004) 37(4):445–56.

34. Eyles H, Jull A, Dobson R, Firestone R, Whittaker R, Te Morenga L, et al. Co-design of mHealth delivered interventions: A systematic review to assess key methods and processes. *Curr Nutr Rep*. (2016) 5(3):160–7. doi: 10.1007/s13668-016-0165-7

35. Kazantzis N, Luong HK, Usatoff AS, Impala T, Yew RY, Hofmann SG. The processes of cognitive behavioral therapy: A review of meta-analyses. *Cogn Ther Res*. (2018) 42(4):349–57. doi: 10.1007/s10608-018-9920-y

36. David D, Cristea I, Hofmann SG. Why cognitive behavioral therapy is the current gold standard of psychotherapy. *Front Psychiatry*. (2018) 9. doi: 10.3389/fpsyt.2018.00004

37. Kerr NL. HARKing: hypothesizing after the results are known. *Personal Soc Psychol Rev*. (1998) 2(3):196–217. doi: 10.1207/s15327957pspr0203_4

38. Flournoy JC, Vijayakumar N, Cheng TW, Cosme D, Flannery JE, Pfeifer JH. Improving practices and inferences in developmental cognitive neuroscience. *Dev Cogn Neurosci*. (2020) 45:100807. doi: 10.1016/j.dcn.2020.100807

39. Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, du Sert NP, et al. A manifesto for reproducible science. *Nat Hum Behav*. (2017) 1 (1):0021. doi: 10.1038/s41562-016-0021

40. Steegen S, Tuerlinckx F, Gelman A, Vanpaemel W. Increasing transparency through a multiverse analysis. *Perspect Psychol Sci*. (2016) 11(5):702–12. doi: 10.1177/1745691616658637

41. Vélez JI. Machine learning based psychology: Advocating for a data-driven approach. *Int J Psychol Res*. (2021) 14(1):6–11. doi: 10.21500/20112084.5365

42. Nelson BW, Low CA, Jacobson N, Areán P, Torous J, Allen NB. Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research. *Npj Digit Med*. (2020) 3(1):90. doi: 10.1038/s41746-020-0297-4

43. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: A new tool for assessing the quality of health mobile apps. *JMIR MHealth UHealth*. (2015) 3(1):e27. doi: 10.2196/mhealth.3422

frontiers | Frontiers in Psychiatry

TYPE Original Research
PUBLISHED 09 September 2022
DOI 10.3389/fpsyt.2022.893073

OPEN ACCESS

EDITED BY
Rosa M. Baños,
University of Valencia, Spain

REVIEWED BY
Giulia Corno,
University of Quebec in Outaouais, Canada
Carolyn Ponting,
University of California, Los Angeles, United States

*CORRESPONDENCE
Carmen Gonzalez
cmgonzal@uw.edu

SPECIALTY SECTION
This article was submitted to Digital Mental Health, a section of the journal Frontiers in Psychiatry

RECEIVED 09 March 2022
ACCEPTED 17 August 2022
PUBLISHED 09 September 2022

CITATION
Gonzalez C, Ramirez M, Mata-Greve F, Diaz A, Duran MC, Johnson M, Grote N and Areán PA (2022) Acceptability of virtual therapy for postpartum women during COVID-19: A national mixed methods study. *Front. Psychiatry* 13:893073. doi: 10.3389/fpsyt.2022.893073

Acceptability of virtual therapy for postpartum women during COVID-19: A national mixed methods study

Carmen Gonzalez[1,2,3,4]*, Magaly Ramirez[1,2,3,4], Felicia Mata-Greve[1,2,3,4], Autumn Diaz[1,2,3,4], Miriana C. Duran[1,2,3,4], Morgan Johnson[1,2,3,4], Nancy Grote[1,2,3,4] and Patricia A. Areán[1,2,3,4]

[1]Department of Communication, University of Washington, Seattle, WA, United States, [2]Department of Health Services, School of Public Health, University of Washington, Seattle, WA, United States, [3]Department of Psychiatry and Behavioral Sciences, School of Medicine, University of Washington, Seattle, WA, United States, [4]School of Social Work, University of Washington, Seattle, WA, United States

Background: Postpartum depression (PPD) affects one in eight women in the U.S., with rates increasing due to the COVID-19 pandemic. Given the unique circumstances of COVID-19, virtual therapy might be a unique way to overcome barriers to mental health services. The study sought to explore the acceptability of virtual therapy among women in the postpartum period.

Methods: Using an online recruitment mixed methods approach, we collected data from a U.S. national cross-sectional sample of women ($N = 479$) who gave birth in the last 12 months.

Findings: Results show that 66% of women endorsed items consistent with possible depression during the COVID-19 pandemic. Only 27% accessed therapy services during the postpartum period. While 88% were open to engaging in virtual therapy services, 12% identified several major concerns with virtual therapy, namely: (1) preference for in-person therapy (2) no perceived need for therapy (3) uncomfortable with virtual therapy, and (4) lack of privacy. Of note, 36% more Latinas reported dissatisfaction with quality of care received during virtual therapy compared to non-Latina participants. Despite a major shift to virtual care with COVID-19, future work is needed to make virtual mental health services more accessible for women with PPD.

KEYWORDS

postpartum depression, virtual therapy, COVID-19, telehealth, digital health

Introduction

Postpartum depression (PPD) is the most common pregnancy complication and affects approximately one in eight women in the U.S (1). Women of color and women from low-income backgrounds report even higher rates of PPD (2). With changes for many individuals due to mandated stay-at-home orders during the COVID-19

pandemic, rates of PPD have *doubled* compared to pre-pandemic rate (3). With the unique circumstances of the pandemic, digital mental health, or services delivered through technology platforms, might be a unique way to overcome barriers to care and effectively provide mental health support in the postpartum period. However, little is known about the acceptability of this form of care among postpartum women.

While postpartum mood disorders increase, there are still several barriers to postpartum mental health care as well as a shortage of providers that offer culturally appropriate services (4). For example, among 500 postpartum women from diverse backgrounds, only 4% of the sample indicated that they had no barriers to accessing mental health care (5). While many women report that common barriers to mental health care include lack of time, childcare, knowledge about PPD, stigma, and transportation concerns (4, 5), systemic inequities further affect treatment options available to women from low-income backgrounds (6), women of color (7), and women with limited English proficiency (4). Latina women often report being under- or uninsured as well as challenges with childcare, transportation, and leave (6, 8).

COVID-19 has underlined these disparities by increasing financial stress and social isolation for new parents. Home environments have also shifted during the pandemic, with mothers often taking on more care responsibilities as schools, businesses, and social services moved online (9). For postpartum women, the experience of giving birth and taking care of a newborn during a pandemic can lead to compounded trauma.

Digital mental health interventions have been proposed as a novel way to overcome barriers and shortages of postpartum mental health care. As a subset of digital mental health services, self-guided digital mental health apps have been developed and hold potential for reaching women with untreated PPD (10) as well as individuals negatively impacted by the COVID-19 pandemic (11). In fact, a recent systematic review of 21 experimental and randomized controlled trials found that perinatal women in high-income countries that used self-guided digital mental health apps and coach-guided digital mental health apps reported significantly decreased depression from pre- to post-intervention (12). These self-guided digital mental health apps might also contribute to the reduction of barriers related to transportation, childcare, time constraints, and language fluency for Latina women (13).

Fewer digital mental health services have used the feature of *virtual therapy*, or one-on-one contact between a patient and mental health provider in real-time (e.g., video conferencing, telepsychotherapy, and message-based care). Nair et al. (14) conducted a systematic review of 10 randomized controlled trials published between 2000 and 2018 to assess the effectiveness of virtual therapy, which included phone calls, emails, coach-guided mental health apps, and coach-guided websites to treat PPD. A majority of these studies reported significant improvement in depression scores post intervention, signaling promise for an understudied use of telemedicine. Virtual therapy can reduce barriers for postpartum mental health care, especially for women from underserved backgrounds, who are more likely to report difficulties accessing in-person therapy due to transportation, scheduling, and childcare concerns (4).

There are few studies that explore whether women in the postpartum period are *receptive* to virtual interventions (15). Outside of the postpartum literature, a recent study found that patients are open to virtual therapy and mental health apps with understandable apprehension surrounding privacy and quality of mental health care (16). Videoconferencing psychotherapy (VCP) specifically has been found to have similar outcomes to in person psychotherapy and is associated with positive user satisfaction (17). Thus, research is needed to understand the specific concerns of virtual therapy among women in the postpartum period, particularly during the COVID-19 pandemic and the rapid shift to telehealth.

The objectives of this mixed-methods cross-sectional study included: (a) to describe experiences with mental health services among women in the postpartum period during the COVID-19 pandemic, and (b) to explore the specific concerns of virtual therapy among women in the postpartum period. Because overwhelming evidence has demonstrated how COVID-19 has disproportionately affected Latinx populations in the U.S., we also examined how these experiences and concerns varied between Latina and non-Latina women (1). The findings from this study will inform how to address specific concerns of virtual therapy in future development of virtual mental health support.

Methods

Procedures

Participants completed the online cross-sectional survey between 7/16/20 and 8/28/20 through Prolific.co., an online platform that allows users to complete surveys and experiments (18). During the study period various states were still under stay at home, shelter in place, or social distancing orders that aimed to slow the spread of COVID-19. Survey participants included those that are: female, U.S. residents, with at least one child, 18–50 YO. In addition, participants completed an eligibility questionnaire to determine whether they had given birth in the last 12 months. Survey items were piloted and refined during five practice interviews. This study was reviewed and approved by an Institutional Review Board.

The survey took an average of 10 min to complete. Survey responses were captured using Research Electronic Data Capture (REDCap) (19, 20), a secure web application developed to capture data for research that complies with HIPAA standards. Prolific.co ID numbers were stored with the survey to avoid duplicate responses. Personal

data were not collected with the exception of women who voluntarily provided their contact information (which was stored in a password protected computer at the home institution) for a follow-up interview. Participants received $5 compensation for their participation in the survey.

Measures

Demographics

Participants completed a demographic questionnaire, which assessed age, assigned sex at birth, employment status, marital status, race, ethnicity, and English proficiency.

Experiences and beliefs about mental health services and virtual therapy

Survey questions were developed by the authors (experts in mental health, health services, health information and communication technology) to understand the experiences of women seeking mental health services in the postpartum period and their concerns with using virtual therapy during the COVID-19 pandemic. These questions asked about prior experiences with therapy, their willingness to consult a therapist, their preference for in-person or virtual therapy, and potential challenges to talking to a therapist online. Some items were open-ended (e.g., "Can you tell us a bit about your experience with receiving emotional support from a professional during your pregnancy?") and some provided response options [e.g., "To what extent would you agree or disagree with the following statement: I am satisfied with the quality of care I received during virtual (therapy) interactions (strongly agree, agree, disagree, or strongly disagree)"].

Depression symptoms

Participants completed the Edinburgh Postnatal Depression Scale (EPDS) (21, 22), a 10-item questionnaire that screens for PPD. Participants indicate whether they have experienced symptoms of depression (e.g., "I have blamed myself unnecessarily when things went wrong"). Summary scores are calculated (range from 0 to 30). We categorized respondents into four categories based on scoring guidelines outlined by Cox et al. (21): depression not likely (<8), depression possible (9–11), fairly high possibility of depression (12–13), and probable depression (14 and higher). We grouped women into depression not likely (score of 8 or below) and depression possible (score of 9 or above). In the current study, internal consistency reliability was excellent (Cronbach $\alpha = 0.90$; omega = 0.90). Participants who reported suicidality were sent a follow up email with a list of postpartum and suicide prevention resources.

Analysis

Quantitative data analysis

Means and frequencies were used to document demographic information, postpartum mental health outcomes, experiences with mental health services in the postpartum period, and concerns with virtual therapy. Percentages reported in the results do not reflect missingness ($N = 479$). To further explore how COVID-19 has disproportionately affected the Latinx population in the U.S., we assessed for differences in experiences and concerns between Latina and non-Latina participants using Fisher's exact chi-square tests, chi-square tests, and independent *t*-tests. We only present results when there is a statistically significant difference between Latina and non-Latina participants. All quantitative statistical analyses were performed with SAS version 9.4.

Qualitative data analysis

We conducted a thematic analysis of the responses to an open-ended question about reasons why participants would not be open to consulting a therapist virtually. Textual responses to the open-ended questions varied from a few words to a few sentences, with participants writing an average of 50 words. For theme identification, we first used an inductive approach that involved two authors pile-sorting each participants' response into categories based on affinity (i.e., the reason for why they were not interested in virtual therapy). Affinity diagramming, a user-centered design approach, is an exploratory qualitative analysis method that facilitates the creation of a codebook; a pile sort is a grouping of similar thoughts, perspectives, or statements (23). This process identified four salient categories: (1) preference for in-person therapy, (2) no perceived need for therapy, (3) not comfortable speaking to a therapist, and (4) lack of privacy at home. To identify subthemes, the two authors further pile-sorted responses within each of the four categories. We provide a description of each category and include representative quotes from participants' responses. All qualitative analyses were performed using Dedoose coding software.

Results

Participant characteristics

See Table 1 for characteristics of study participants. There were 2416 individuals who were interested in completing the survey. In total, 479 women were screened eligible and completed the survey. The average age was 30.3 years. Most (72%) of the sample identified as White, and ∼7% of the sample identified as Latina. Most participants (97%) indicated that they

spoke English "very well." Participants reported residing in state across the U.S., with larger states such as California, Florida, and New York more highly represented. Approximately two-thirds of the sample were in the "depression possible" range for PPD.

Experiences with mental healthcare services

Of the 479 participants, 27% (126/474) reported having spoken to a counselor, therapist, or other mental health professional since giving birth. About half (64/126) of participants reported that the encounter was virtual. On average, it had been 67 ($SD = 82.5$) days between the time they took the survey and the last time that they spoke with their therapist *via* virtual encounter. Virtual encounters happened *via* video (42/64; 66%), telephone (29/64; 45%), and text message (8/64; 13%). Among all the participants who had received postpartum mental health care ($n = 126$), the degree of satisfaction with virtual mental healthcare encounters in general (regardless of the nature of their postpartum care) was significantly lower among Latinas ($n = 11$) compared to non-Latina participants ($n = 115$; Table 2). In fact, 36% more Latinas disagreed or strongly disagreed with the statement about being satisfied with the quality of care received during virtual encounters compared to non-Latina participants.

Concerns regarding virtual mental healthcare

In the total sample, 88% of participants reported being open to consulting a mental health professional in the future. Among those open to future mental health services, 91% (382/418) reported they would consider virtual mental health services. The top concerns for accessing virtual mental healthcare were privacy (226/479; 47.2%), cost (222/479; 46.3%), time (130/479; 27.1%), and trust (128/479; 26.7%). Participants who were not interested in virtual mental healthcare encounters provided open-ended responses to further elaborate on their lack of interest. The following themes emerged from these qualitative responses: (1) preference for in-person therapy, (2) no perceived need for therapy, (3) not comfortable speaking to a therapist, and (4) lack of privacy at home.

Theme 1: Preference for in-person therapy

Participants stated a preference for in-person therapy. Participants stated that they would feel uncomfortable discussing sensitive topics in a virtual environment (Box 1, extracts 1–2). These participants perceived that face-to-face conversations would be the most appropriate way to share their

TABLE 1 Sample characteristics.

Characteristic	Total ($N = 479$)
Age	
N	474
Mean (*SD*)	30.3 (5.1)
Race, *n* (%)	
Black or African American	60 (12.5%)
Hispanic or latino/x	22 (4.6%)
White or caucasian	344 (71.8%)
Asian or pacific islander	24 (5.0%)
Multiracial	24 (5.0%)
Unlisted	5 (1.0%)
Hispanic or Latina[a], *n* (%)	32 (6.7%)
English fluency, *n* (%)	
Very well	460 (97.0%)
Well	14 (3.0%)
Missing	5
Speaks a language other than English, *n* (%)	63 (13.3%)
Marital status, *n* (%)	
Currently married	367 (77.4%)
Widowed	2 (0.4%)
Divorced	11 (2.3%)
Separated	11 (2.3%)
Never married	83 (17.5%)
Missing	5
Education attainment, *n* (%)	
Less than high school diploma	2 (0.4%)
High school diploma or GED	49 (10.3%)
Some college, but no degree	84 (17.7%)
Associates degree	32 (6.8%)
Bachelor's degree	197 (41.6%)
Post-graduate degree	110 (23.2%)
Missing	5
Employment status, *n* (%)	
Employed full time	191 (40.3%)
Employed part time	90 (19.0%)
Self-employed	21 (4.4%)
Out of work and seeking opportunities	20 (4.2%)
Homemaker or stay at home parent	149 (31.4%)
Unable to work	3 (0.6%)
Missing	5
Has childcare during COVID-19 pandemic, *n* (%)	187 (39.5%)
Household size	
N	474
Mean (SD)	4.0 (1.2)
EDPS total score	
N	479
Mean (SD)	11.1 (6.1)
EDPS interpretation, *n* (%)	
Depression not likely (range: 0–8)	165 (34.4%)
Depression possible (range: 9–30)	314 (65.6%)

[a]Identifies Latino/a participants classified as multiracial.

TABLE 2 Participant satisfaction with virtual mental healthcare encounters, by ethnicity.

	Latina participants ($N = 11$)	Non-Latina participants ($N = 115$)	Group difference (Latina—Non)	*P*-value
I am satisfied with the quality of care I received during these virtual interactions, *n* (%)				0.01[a]
Strongly agree	3 (27.3%)	41 (35.7%)	−8.4%	
Agree	3 (27.3%)	63 (54.8%)	−27.5%	
Disagree	3 (27.3%)	7 (6.1%)	21.2%	
Strongly disagree	2 (18.2%)	4 (3.5%)	14.7%	

[a]Fisher exact *p*-value.

BOX 1 Theme 1—Preference for in-person therapy.

1. 32-year-old White participant: "I have comfort issues regarding virtual sessions of that nature and would much rather speak face to face with a counselor or therapist."
2. 40-year-old Asian participant: "I just prefer being in person while talking about intimate things."
3. 23-year-old African American participant: "I like for it (therapy) to feel more personal and the only way to do that is in person for me."
4. 27-year-old White participant: "It (virtual therapy) doesn't work for me."
5. 32-year-old White participant: "I feel like virtually would not feel as effective for me."

emotions. There was also aversion to virtual therapy because of a perception that it would not be tailored to their unique situation. Participants doubted that virtual therapy could be delivered in a way that made it feel as "personal" as a face-to-face encounter (Box 1, extract 3). Finally, participants believed that virtual therapy would not be as effective as in-person treatment. For some, this perception was influenced by negative past experiences with virtual therapy (Box 1, extracts 4–5).

Theme 2: No perceived need for therapy

Participants reported that they were not interested in virtual therapy because they currently did not have a need. Participants stated that they were not experiencing symptoms of depression and anxiety or that they felt well-emotionally (Box 2, extracts 1–3). However, the EPDS scores of 6/19 participants stating no perceived need for therapy indicated possible depression. Some participants explained that if there were days when they were feeling down (which could be triggered by the COVID-19 pandemic), they already had coping strategies, which included reaching out to a spouse/partner and other family members, engaging in physical activity, eating healthy, meditating, and getting adequate sleep (Box 2, extracts 4–5).

Theme 3: Not comfortable speaking to a therapist

Participants stated a disinterest in virtual therapy because they would feel uncomfortable. A few participants described experiencing shyness and social anxiety, regardless of whether interactions with others were in-person or virtual (Box 3, extracts 1–2). Other participants explained that they would be uncomfortable speaking about sensitive topics with someone they never met before (Box 3, extracts 3–5). They expressed a desire to meet the mental health professional face-to-face before transitioning to a virtual environment.

Theme 4: Lack of privacy at home

Participants explained that they were not interested in virtual therapy because they did not have privacy at home. Some participants did not want others to find out they were speaking to a therapist (Box 4, extract 1). Other participants were worried about members of their household listening during phone calls and reading text message conversations with a therapist (Box 4, extract 2–3).

Discussion

The mixed methods study supports that even though two-thirds of the U.S. sample of women reported possible PPD during the COVID-19 pandemic, only 27% had consulted a mental health professional. Of note, Latina women reported lower satisfaction with virtual therapy compared to non-Latina women. This finding is concerning given that the U.S. Latinx

BOX 2 Theme 2—No perceived need for therapy.

1. 29-year-old White participant (EPDS score = 8; depression not likely): "I'm not feeling down or hopeless."
2. 32-year-old White participant (EPDS score = 3; depression not likely): "I'm not suffering from baby blues or any other depression."
3. 22-year-old Asian participant (EPDS score = 11; possible depression): "I am sure that I am feeling okay mentally and spiritually."
4. 32-year-old Asian participant (EPDS score = 11; possible depression): "I have a few random days here and there of anxiety, but that is related to COVID and are rare. I just do some meditation and go to sleep and feel back to normal the next day."
5. 27-year-old White participant (EPDS score = 4; depression not likely): "I am happy and know how to navigate and pull myself back up if I happen to fall into a slump. We eat well, regularly exercise, I take my children hiking 3 days a week and do yoga 5 days a week too. I have a strong family support system and love my in-laws."

BOX 3 Theme 3—Not comfortable speaking to a therapist.

1. 18-year-old Latina participant: "I wouldn't feel comfortable. I'm a very shy person."
2. 39-year-old participant*: "I have pretty bad social anxiety so even if I do go to talk to someone ill most likely shut down at the time."
3. 31-year-old White participant: "I would want to meet with someone in person first. I feel it is a better way of getting to know someone."
4. 24-year-old White participant: "Not comfortable sharing sensitive information to someone I haven't seen."
5. 23-year-old White participant: "It's hard to feel comfortable with someone you haven't met in real life."

*No race or ethnicity reported.

BOX 4 Theme 4—Lack of privacy at home.

1. 24-year-old Latina participant: "I would not sure want anyone to know I am seeing a therapist."
2. 27-year-old White participant: "I would be afraid I would get interrupted, other people could hear me."
3. 35-year-old White participant: "Too worried about people breaking into that to get my info or chats."

population is disproportionately affected by both COVID-19 (24) and PPD (25) compared to their White counterparts. The results support that even though U.S. residents are in the midst of a pandemic causing many healthcare services to shift to telehealth, one-quarter of the sample still reported hesitancy with virtual therapy. For those not interested in virtual therapy, primary reasons identified for apprehension included a preference for in-person therapy, no perceived need, uncomfortable with the idea of virtual therapy, and lack of privacy at home.

Postpartum depression and therapy use

This study reported higher rates of possible PPD compared to previous U.S. studies (3). Through the self-reported Pregnancy Risk Assessment Monitoring System, the CDC (1) reports that 1 in 8 women experience PPD, whereas this study's results were closer to 6 in 8 women. Our increased rates might be explained by diagnostic clarity. Many cases of PPD typically go undiagnosed due to structural barriers as well as a lack of proper assessment (8, 21, 22). We used the EPDS, a scale that specifically assesses depression symptoms during the postpartum period. Another explanation is the stress associated with the COVID-19 pandemic, which has included exacerbated fear (26), isolation (27), and financial strain (28).

Implications for practice

Only 27% of women indicated consulting a mental health professional during the postpartum period. This discrepancy in need and use necessitates an urgent solution to increase accessibility to mental health services as well as proper education of symptoms related to PPD. Of those that did receive mental health services, approximately half were done virtually *via* video, telephone or text message platforms. Since COVID-19 has shifted many services to virtual platforms, it is important to understand patient concerns. Consistent with pre-pandemic research (16), some participants identify a preference for in-person therapy. This preference seems to stem from perceived increased authenticity in person (16) and worries that virtual therapy is less effective (16, 29).

Participants identified that they would not engage with virtual therapy because they did not perceive a need for therapy. Interestingly, of the 19 women that listed this concern, six endorsed elevated levels consistent with possible depression on the EPDS. Previous studies have highlighted stigma and beliefs about motherhood as barriers to treatment (30). While the

American College of Obstetricians and Gynecologists and the US Preventive Services Task Force have strongly recommended universal screening, there may also be an increased need for education of PPD and its treatment (31).

Participants stated they would not be comfortable with virtual therapy, and that they had concerns with a lack of privacy at home. The COVID-19 pandemic has changed home environments for many with an increase in adults working from home (32) and school-age children engaging in remote learning (33). There is evidence to support that women are taking on more care and responsibilities during the COVID-19 pandemic (9), which may be driving comfort and privacy concerns regarding virtual therapy compared to pre-pandemic times.

Limitations and future directions

This study had several limitations. Because our sample was recruited through an online platform, it is likely that participants were more comfortable with technology and may have different views about data security and privacy. This study was cross-sectional, thus, opinions about virtual therapy may change over time as the population has more access to telehealth during the pandemic. Such limitations do not detract from the importance of the findings, which highlight the low utilization and trust in virtual mental health care by women who are currently using technology. Likely because of our recruitment methods, our sampling of women of color was relatively low (e.g., 7% of this sample was Latina v. 17% of the U.S. population is Latinx). The recruitment reach of the platform used for this study makes it difficult to target specific demographics; a longer study period could help to oversample specific groups (including postpartum women). Survey instruments were also available in English only, further limiting the generalizability of our findings to English-speaking women who are likely more acculturated. We recommend a future study comparing remotely recruited samples to locally recruited (and demographically targeted) samples in their acceptance of virtual mental health care. We also did not collect information about the presence of a diagnosed mental disorder; future studies could benefit from such contextual information.

Despite these limitations, the data from this study have important implications about the concept that virtual mental health care is a viable alternative to traditionally delivered care. To improve the approachability of virtual therapy, researchers and clinicians may employ user-centered design alongside patients and better understand how women intend to use virtual therapy for support (13).

Conclusion

These results further support that rates of psychological distress, such as PPD, are elevated during the COVID-19 pandemic. Despite most of our sample reporting symptoms consistent with possible depression symptoms, only one-quarter have used psychotherapy services. While many participants were open to engaging in virtual therapy, primary concerns included preference for in-person services, lack of perceived need, discomfort with virtual therapy, and lack of privacy. We recommend that researchers and clinicians continue to find ways to make virtual therapy more approachable to the needs of women experiencing PPD.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving human participants were reviewed and approved by University of Washington Institutional Review Board. The patients/participants provided their written informed consent to participate in this study.

Author contributions

CG and MR: guarantors. CG, MR, PA, and NG: study concept and design. CG, MR, MD, MJ, and FM-G: acquisition, analysis, or interpretation of data. CG, FM-G, MR, MD, and AD: drafting of the manuscript. CG, MR, PA, and NG: obtaining funding. All authors critical revision of the manuscript, had full access to the study data, take responsibility for the integrity of the complete work, and the final decision to submit the manuscript.

Funding

This work was supported by the National Institutes of Mental Health (grant numbers P50MH115837, T32MH020021). The content is solely the responsibilities of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

1. Centers for Disease Control and Prevention. *Reproductive Health: Depression Among Women*. Centers for Disease Control and Prevention (2020). Available online at: https://www.cdc.gov/reproductivehealth/depression/index.htm (accessed February 15, 2022).

2. Howell EA, Mora PA, Horowitz CR, Leventhal H. Racial and ethnic differences in factors associated with early postpartum depressive symptoms. *Obstet Gynecol.* (2005) 105:1442–50. doi: 10.1097/01.AOG.0000164050.34126.37

3. Davenport MH, Meyer S, Meah VL, Strynadka MC, Khurana R. Moms are not ok: COVID-19 and maternal mental health. *Front Global Women Health.* (2020) 1:1. doi: 10.3389/fgwh.2020.00001

4. Bina R. Predictors of postpartum depression service use: a theory-informed, integrative systematic review. *Women Birth.* (2020) 33:e24–32. doi: 10.1016/j.wombi.2019.01.006

5. Goodman JH. Women's attitudes, preferences, and perceived barriers to treatment for perinatal depression. *Birth.* (2009) 36:60–9. doi: 10.1111/j.1523-536X.2008.00296.x

6. Kozhimannil KB, Trinacty CM, Busch AB, Huskamp HA, Adams AS. Racial and ethnic disparities in postpartum depression care among low-income women. *Psychiatric Serv.* (2011) 62:619–25. doi: 10.1176/ps.62.6.pss6206_0619

7. Hansotte E, Payne SI, Babich SM. Positive postpartum depression screening practices and subsequent mental health treatment for low-income women in Western countries: a systematic literature review. *Public Health Rev.* (2017) 38:3. doi: 10.1186/s40985-017-0050-y

8. Abrams LS, Dornig K, Curran L. Barriers to service use for postpartum depression symptoms among low-income ethnic minority mothers in the United States. *Qual Health Res.* (2009) 19:535–51. doi: 10.1177/1049732309332794

9. Power K. The COVID-19 pandemic has increased the care burden of women and families. *Sustainabil Sci Pract Policy.* (2020) 16:67–73. doi: 10.1080/15487733.2020.1776561

10. Zhou C, Hu H, Wang C, Zhu Z, Feng G, Xue J, et al. The effectiveness of mHealth interventions on postpartum depression: a systematic review and meta-analysis. *J Telemed Telecare.* (2020) 28:83–95. doi: 10.1177/1357633X20917816

11. Zhou X, Snoswell CL, Harding LE, Bambling M, Edirippulige S, Bai, X., et al. The role of telehealth in reducing the mental health burden from COVID-19. *Telemed e-Health.* (2020) 26:377–9. doi: 10.1089/tmj.2020.0068

12. Dol J, Richardson B, Murphy GT, Aston M, McMillan D, Campbell-Yeo M. Impact of mobile health interventions during the perinatal period on maternal psychosocial outcomes: a systematic review. *JBI Evid Synth.* (2020) 18:30–55. doi: 10.11124/JBISRIR-D-19-00191

13. Fortuna LR, Tolou-Shams M, Robles-Ramamurthy B, Porche MV. Inequity and the disproportionate impact of COVID-19 on communities of color in the United States: the need for a trauma-informed social justice response. *Psychol Trauma Theory Res Pract Policy*. (2020) 12:443–5. doi: 10.1037/tra0000889

14. Nair U, Armfield NR, Chatfield MD, Edirippulige S. The effectiveness of telemedicine interventions to address maternal depression: a systematic review and meta-analysis. *J Telemed Telecare.* (2018) 24:639–50. doi: 10.1177/1357633X18794332

15. Barrera AZ, Wickham RE, Muñoz RF. Online prevention of postpartum depression for Spanish-and English-speaking pregnant women: a pilot randomized controlled trial. *Int Int*. (2015) 2:257–65. doi: 10.1016/j.invent.2015.06.002

16. Renn BN, Hoeft TJ, Lee HS, Bauer AM, Areán PA. Preference for in-person psychotherapy versus digital psychotherapy options for depression: survey of adults in the US. *NPJ Dig Med.* (2019) 2:1–7. doi: 10.1038/s41746-019-0077-1

17. Backhaus A, Agha Z, Maglione ML, Repp A, Ross B, Zuest D, et al. Videoconferencing psychotherapy: a systematic review. *Psychol Serv.* (2012) 9:111. doi: 10.1037/a0027924

18. Prolific. Oxford: Retrieved from Prolific.co (2019).

19. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. A metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform.* (2009) 42:377–81. doi: 10.1016/j.jbi.2008.08.010

20. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, et al. The REDCap consortium: building an international community of software platform partners. *J Biomed Inform.* (2019) 95:103208. doi: 10.1016/j.jbi.2019.103208

21. Cox JL, Holden JM, Sagovsky R. Detection of postnatal depression: development of the 10-item Edinburgh postnatal depression scale. *Brit J Psychiatry.* (1987) 150:782–6. doi: 10.1192/bjp.150.6.782

22. Wisner KL, Parry BL, Piontek CM. Postpartum depression. *N Engl J Med.* (2002) 347:194–9. doi: 10.1056/NEJMcp011542

23. Haskins Lisle A, Merenda C, Gabbard J. Using affinity diagramming to generate a codebook: a case study on young military veterans and community reintegration. *Qualit Res.* (2020) 20:396–413. doi: 10.1177/1468794119851324

24. Oppel RA, Gebeloff R, Lai KKR, Wright W, Smith R. *The Fullest Look Yet at the Racial Inequity of Coronavirus*. New York Times (2020).

25. Lucero NB, Beckstrand RL, Callister LC, Sanchez Birkhead AC. Prevalence of postpartum depression among Hispanic immigrant women. *J Am Acad Nurse Pract.* (2012) 24:726–34. doi: 10.1111/j.1745-7599.2012.00744.x

26. Fitzpatrick KM, Harris C, Drawve G. Fear of COVID-19 and the mental health consequences in America. *Psychol Trauma Theory Res Pract Policy*. (2020) S1:S17–21. doi: 10.1037/tra0000924

27. Benke C, Autenrieth L, Asselmann E, Pané-Farré CA. Lockdown, quarantine measures, and social distancing: associations with depression, anxiety, and distress at the beginning of the COVID-19 pandemic among adults from Germany. *Psychiatry Res.* (2020) 293:113462. doi: 10.1016/j.psychres.2020.113462

28. Mimoun E, Ben Ari A, Margalit D. Psychological aspects of employment instability during the COVID-19 pandemic. *Psychol Trauma Theory Res Pract Policy*. (2020) 12(Suppl. 1):S183–5. doi: 10.1037/tra0000769

29. Linardon J, Shatte A, Tepper H, Fuller-Tyszkiewicz M. A survey study of attitudes toward, and preferences for, e-therapy interventions for eating disorder psychopathology. *Int J Eating Dis.* (2020) 53:23268. doi: 10.1002/eat.23268

30. Callister LC, Beckstrand RL, Corbett C. Postpartum depression and help-seeking behaviors in immigrant Hispanic women. *J Obst Gynecol Neonatal Nursing.* (2011) 40:440–9. doi: 10.1111/j.1552-6909.2011.01254.x

31. Baker-Ericzén MJ, Connelly CD, Hazen AL, Dueñas C, Landsverk JA, Horwitz SM. A collaborative care telemedicine intervention to overcome treatment barriers for Latina women with depression during the perinatal period. *Famil Syst Health.* (2012) 30:224–40. doi: 10.1037/a0028750

32. Bick K, Blandin A, Mertens A. Work from home after the COVID-19 outbreak. *Monet Eco Fluctuat*. (2020). doi: 10.2139/ssrn.3786142

33. Garbe A, Ogurlu U, Logan N, Cook P. COVID-19 and remote learning: Experiences of parents with children during the pandemic. *Am J Qualit Res.* (2020) 4:45–65. doi: 10.29333/ajqr/8471

frontiers | Frontiers in Digital Health

TYPE Original Research
PUBLISHED 22 September 2022
DOI 10.3389/fdgth.2022.872404

OPEN ACCESS

EDITED BY
Patricia A. Arean,
University of Washington, United States

REVIEWED BY
Rüdiger Christoph Pryss,
Julius Maximilian University of Würzburg, Germany
André Luiz Monezi Andrade,
Pontifical Catholic University of Campinas, Brazil

*CORRESPONDENCE
Lily Mainstone-Cotton
lilymainstone.cotton@gmail.com
Santiago de Ossorno Garcia
sossorno@kooth.com

SPECIALTY SECTION
This article was submitted to Digital Mental Health, a section of the journal Frontiers in Digital Health

RECEIVED 09 February 2022
ACCEPTED 31 August 2022
PUBLISHED 22 September 2022

CITATION
Mindel C, Mainstone-Cotton L
de Ossorno Garcia S, Sefi A
Sugarman G, Salhi L, Brick H, Jackson K and Hanley T (2022) The design and development of an experience measure for a peer community moderated forum in a digital mental health service.
Front. Digit. Health 4:872404.
doi: 10.3389/fdgth.2022.872404

The design and development of an experience measure for a peer community moderated forum in a digital mental health service

Charlotte Mindel[1], Lily Mainstone-Cotton[1]*, Santiago de Ossorno Garcia[1]*, Aaron Sefi[1,2], Georgia Sugarman[1], Louisa Salhi[1,3], Holly Brick[1], Katherine Jackson[1] and Terry Hanley[4]

[1]Kooth Plc, London, United Kingdom, [2]Department of Psychology, University of Exeter, Devon, United Kingdom, [3]School of Psychology, University of Kent, Canterbury, United Kingdom, [4]School of Environment, Education and Development, The University of Manchester, Manchester, United Kingdom

Online digital mental health communities can contribute to users' mental health positively and negatively. Yet the measurement of experience, outcomes and impact mechanisms relating to digital mental health communities is difficult to capture. In this paper we demonstrate the development of an online experience measure for a specific children and young people's community forum inside a digital mental health service. The development of the Peer Online Community Experience Measure (POCEM) is informed by a multi-phased design: (i) item reduction through Estimate-Talk-Estimate modified Delphi methods, (ii) user testing with think-aloud protocols and (iii) a pilot study within the digital service community to explore observational data within the platform. Experts in the field were consulted to help reduce the items in the pool and to check their theoretical coherence. User testing workshops helped to inform the usability appearance, wording, and purpose of the measure. Finally, the pilot results highlight completion rates, differences in scores for age and roles and "relate to others", as the most frequent domain mechanism of support for this community. Outcomes frequently selected show the importance of certain aspects of the community, such as safety, connection, and non-judgment previously highlighted in the literature. Experience measures like this one could be used as indicators of active therapeutic engagement within the forum community and its content but further research is required to ascertain its acceptability and validity. Multi-phased approaches involving stakeholders and user-centred design activities enhances the development of digitally enabled measurement tools.

KEYWORDS

digital mental health, online community, experience measures, multi-phased design, moderated forum

1. Introduction

Online peer communities can provide a platform of social interaction for young people. Children and young people are considered digital natives (1), most have lived with relative ease of access to the internet since childhood. This influences children and young people's attitude towards the internet and how they seek support, with most considering the internet as the first option for seeking information, advice, or emotional support for mental health problems (2, 3). When paired with the importance of social peers' attitudes, beliefs, and behaviour during adolescence (4) online peer communities can offer an important form of support for young people seeking information or struggling with mental health. Those online communities can be formed through instant communication tools, social media networks, and asynchronous forums, where people share content with the intention of being seen by peers.

The importance of online peer communities in supporting adolescent mental health is shown by a strong but complex relationship between online social networks, mental health and wellbeing (5, 6). When online social networks are used to seek support, reports of depressed moods are correspondingly minimised or maximised depending on whether users were passive or active in their online use (7, 8). However, some studies have reported low-quality connections, depression and "comparison effects" for specific social media platforms (9, 10). Liu and colleagues' (6) meta-analysis highlights how different digital communication tools and media usage affects wellbeing depending on the intimacy and activity type in the medium. More widely, the harm of some online social network platforms has been explored in more detail and found to predict an increase in body dissatisfaction (11). Therefore, it is important to recognise that the nature and characteristics of an online community will influence whether the impact on the mental health of users is positive or negative. For example, visiting pro-anorexia sites was negatively associated with perception of appearance, drive for thinness and perfectionism (12), whilst online social support has shown to act as both protective and risk factor mediating how web content is internalised (13). Conversely, online mental health communities can be seen as the analogue to traditional mental health face-to-face support groups, especially for a subset of the population seeking advice or to express emotions online (14, 15). The anonymity and social connectedness in these spaces can help people to overcome stigma and make positive disclosures of experiences and problems (16). Nevertheless, others have demonstrated how dependency on these communities can hinder recovery from stigma, especially in online spaces without moderation or supervision (17).

It is when online mental health communities have appropriate characteristics (e.g., moderation, anonymity, facilitation) that they can help individuals, and maximize support when experiencing mental health difficulties (18, 19). The potential negative impact of online mental health communities on young people can be mitigated using moderation of content, creating safety, preserving anonymity and other mechanisms to create a boundary environment which in turn is designed to avoid judgement and promote wellbeing. Observations of unmoderated platforms commonly identified signs of self-harm normalisation and increase of suicidal ideation (20, 21). Comparatively, users of moderated mental health forums report a reduction in frequency and severity of self-harm behaviour (22). Given the mixed impacts of online mental health communities, it is important to examine and attempt to measure experiences within these kinds of online communities, especially those designed to provide peer support, reduce risk of harm, preserve safety, and enhance wellbeing of online mental health experiences.

1.1. Measurement in online digital mental health communities

Determining how to evaluate the impact and effectiveness of online community mental health support should be a key focus for platforms providing online peer support services. The indirect and asynchronous nature of a community forum support presents challenges using standardised measures for its evaluation. This is particularly the case when the community is not directed to a target population, user-led, and not focused on a specific mental health concern leading to a specific outcome or mental health difficulty such as anxiety, eating difficulties or depression.

When standardised measures have been used in online communities' research, there have been mixed results. One online peer support group for young people found users improved in anxiety scores but did not show any changes in depression (23). Others found a non-significant reduction in depression of forum users or no differences in body dissatisfaction between forums users and the control group (24).

A clearer picture on the benefits of online mental health communities is found when qualitative and mixed methods are used. Horgan and colleagues (25) used thematic analysis on forum posts, alongside standardised outcomes for depression and anxiety. Young people using the forum frequently discussed the immediate benefits of sharing their feelings on the forum and described a sense of not being alone. Forum posts also mentioned the benefits of individuals comparing themselves to others, and consequently believing that their situation was less bad than previously thought. In regard to a self-harm community investigated, young people reported that they felt they learnt more about mental health from other users, compared to information sites, and they felt it easier to disclose information online, in part because they

were less likely to be judged than in real life (26). More recent investigations have shown how self-efficacy and access to further support can increase thanks to the use of these online communities (27). They can also provide a sense of belonging, tackling feelings of loneliness regarding mental health experiences (28). Qualitative studies also reveal how social modelling allows encouragement between peers to use pro-social behaviour and receive support within and outside the community (29).

Qualitative methods do, however, have limitations in measuring outcomes for online communities at scale. They are time intensive and cannot be used repeatedly to track user experience and satisfaction, nor be used as a method to routinely collect information about the community. However, they provide an in-depth understanding of why young people use online mental health communities and what outcomes are achieved. The findings can be used as the functional theory to develop an experience measure for an online community.

Online peer mental health communities aiming to support users require understanding on how their resources and content help or hinder the user's wellbeing. Measurements can be collected and may be routinely aggregated to personalise a community experience in the platform or create automated recommendations of community resources likely to contribute to the recovery or support of the individual. Developing a self-reported measure for this endeavour should aid understanding of how helpful the content is, what kind of help the content can provide, and how different users may benefit from it. Ultimately, an experience measure will provide an indicator of therapeutic active engagement with the forum and community content, using a parameter of engagement that goes beyond the forum analytics and often reported digital contexts (e.g., Views, clicks, time, popularity). Measuring the helpfulness of community content may provide insights on how resources are consumed and contribute to a positive, negative, or neutral experience. The measure should also understand the mechanisms that lead to the helpful or unhelpful experiences in the community, and what types of outcomes users are achieving in relationship with their engagement with the forum content.

1.2. The peer online community forum within a Digital Mental Health Service

The Kooth.com (referred to from here onwards as Kooth) online community is a user-led forum inside a multi-component digital mental health service where the content revolves around the changing needs and experiences of the young people in its platform. The forum promotes a wide variety of professionally and non-professionally curated content about mental health and wider wellbeing topics, aiming to reduce stigma and contribute to meaningful conversations about people's mental health experiences. The content of the community consists of three core types of posts: (1) a co-created magazine with a combination of psycho-educational, creative, and informative content written by the service users and practitioners, (2) discussion forums authored by users, providing direct interactions between peers but still moderated by professionals (with mental health backgrounds), and (3) mini-activities, a specific type of content created with the intention of helping users build life skills and promote planned action. All user submitted content is moderated before being published on the platform to safeguard, categorise, and age-restrict content where necessary. The forum is part of a wider UK online service that provides with direct synchronous support with professionals, which is anonymous and free for users. When designing a measure for a specific online community and its characteristics, a framework to measure quality-of-care is required. These frameworks will be especially useful when the programme theory and mechanisms of change for the online support community have been previously investigated, so both can be combined to develop a specific and relevant measurement.

1.3. The Quality-of-Care measurement framework to develop a Peer Online Community Experience Measure

Donabedian's (30) quality of care framework recommends measuring care through assessing structure, process, and outcomes. For an online community forum, the quality of care is reflected in the structural elements of the community (e.g., Content, relationships, posts), the process or mechanisms of accessing support through peers and consuming content within the community, and the outcomes of the community that can be achieved when meaningfully engaged with it. Each of the three components have a bidirectional relationship. The structure of the community will influence the process of peer and community interactions, which will then impact the outcomes that are achieved. Positive or negative outcomes may change the process of peer interactions, and potentially change the type of content available within the structure. By using Donabedian's framework as the foundation for an experience measure, we seek to capture information into the forum community helpfulness (structure), peer and content interactions (process) and relevant reported outcomes for the individual.

The design, and structure of the Peer Online Community Experience Measure (POCEM) was divided into three parts, each representing one of the domains of Donabedian's measurement framework. The items in the measure were initially identified through the Kooth Theory of Change research where mechanisms of support and outcomes of peer support for the service were previously examined (31). Taking

the framework approach adapted to the context of care relevant to the peer support community forum the measure is set up to assess the following:

a) **Structure assessment:** Assesses the quality of the community structure, focusing on "helpfulness" of magazine articles and forum discussions using a "emoji"-based Likert scale in response to the question "Did you find this part of Kooth helpful".
b) **Process assessment:** Assesses why online community users found a specific structure (community resources) helpful or unhelpful, depending on their response to the helpfulness Likert rating. Respondents selected one of four support processes reflecting the possible interactions they were looking for in the community (1: Emotional interpersonal; 2: Emotional intra-personal; 3: Informational inter-personal; 4: Informational intra-personal).
c) **Outcome assessment:** Explores what outcomes are achieved, specific to the structures and processes considered helpful in reference to the area of support received. This means that a different subset of outcomes may be achieved depending on the mechanisms or processes of helpfulness that the user has previously identified in the assessments. Furthermore, within the context of Kooth there are two avenues of engagement for a community member, through active contribution in generating community content by writing, or through the consumption of content posted and available within the community by reading. These types of engagement are likely to be associated with different outcomes, depending on their role in the forum, whether the user is creating (contributor) or consuming community content (viewers). Therefore, the assessment of outcomes within the measure should be able differentiate depending on the user's role to inform the experience of the community as a whole.

A measurement that covers the three layers of assessment of quality of care should help to understand how peer support in an online community relates into a quality-of-care framework for the intended context, and how feasible is to measure the experience through an instrument tailored for a digital service context and program theory.

1.4. The present study

The present study describes the (i) development, (ii) user testing, and (iii) pilot results of the measure implemented in a dynamic and multifaceted digital mental health service. This study involved different key stakeholders and participants that influenced each phase iteratively. The ethos of Donabedian's framework (30, 32) was applied to the initial development of the Peer Online Community Experience Measure (POCEM) so the assessment domains of quality of care were included within the measure. In the (i) development subject matter experts and previous literature on the service were used aiming to answer the following question: *Which Items generated by subject matter experts best represent the high-level domain of support (process)?*

To ensure POCEM is a meaningful experience measure for the people using it, participatory think aloud protocols were conducted to guide the development of items and appearance of the measure with high-fidelity and clickable prototypes. Quasi-realistic simulations and (ii) user testing can help to discern the face validity of the instrument, and were used in this development process to address following research questions: *Does the respondent understand what is being measured? Do people understand the measure within the forum platform?*

Finally, to understand its feasibility as a measure, further observations through a (iii) pilot study to examine usage, completion, and item selection from the measure were collected within the online forum platform and mental health service, to answer the questions: *How acceptable by response rates is the phased measure within digital community? Do response rates influence scores for the instrument? And What are the most frequently selected processes domains and outcomes during the pilot?*

An iterative multi-phased design process aims to integrate evidence collected from practitioners, researchers, design experts, and young people. The design of the POCEM and its development provides an opportunity to collect data on the peer support community and assess structure, processes, and outcomes within the wider service for users. This study aims to provide a foundational design structure and outline a systematic development process for an online community measure, so others can be guided in the process to develop their own community experience measures, that are relevant and context specific, implications and lessons learned from each phase of the study are discussed. The study describes the mixed-method development of the POCEM divided by three phases including the implementation of the measure in a natural environment.

2. Methods

A multi-phased design process was used involving iterative development, reflective decision making and real-world application of the findings (33, 34). It is an iterative process that aims to integrate the perspectives of key stakeholders into the phases of measurement development into the digital domain. The development involved a group of practitioners, researchers, user experience designers as experts, young people from schools in which the service operates, and users accessing the digital forum community in the service. Their

involvement as stakeholders and participants was iterative following three key phases of design, representing each study:

- *Phase 1, Item generation and reduction*: A three-part measure developed with digital product experts and designers. The content of the measure and items were created by combining qualitative thematic indicators of outcomes and mechanisms. Delphi rounds were used to reduce items and explore the content for the measure.
- *Phase 2, User testing*: To directly explore, using think-aloud workshops, the face validity of the measure with young people. The focus was to verify the appropriateness of language and how design of the measure was experienced on the platform as a prototype.
- *Phase 3, Pilot study*: A 10-week pilot of the measure within the digital online community. Exploring completion rates, average scores, item frequency selection and correlations between items scores.

Each phase provides iterative results that inform the overall improvement and design and development of the measure, from theoretical foundation to practical design thinking decision-making. Multiple stakeholders and phases of development required a mixed-method approach with qualitative and quantitative data collection activities and incremental and iterative changes to the development of the POCEM. A break-down of stages, procedure, and results of the multiphase design is illustrated in **Figure 1**.

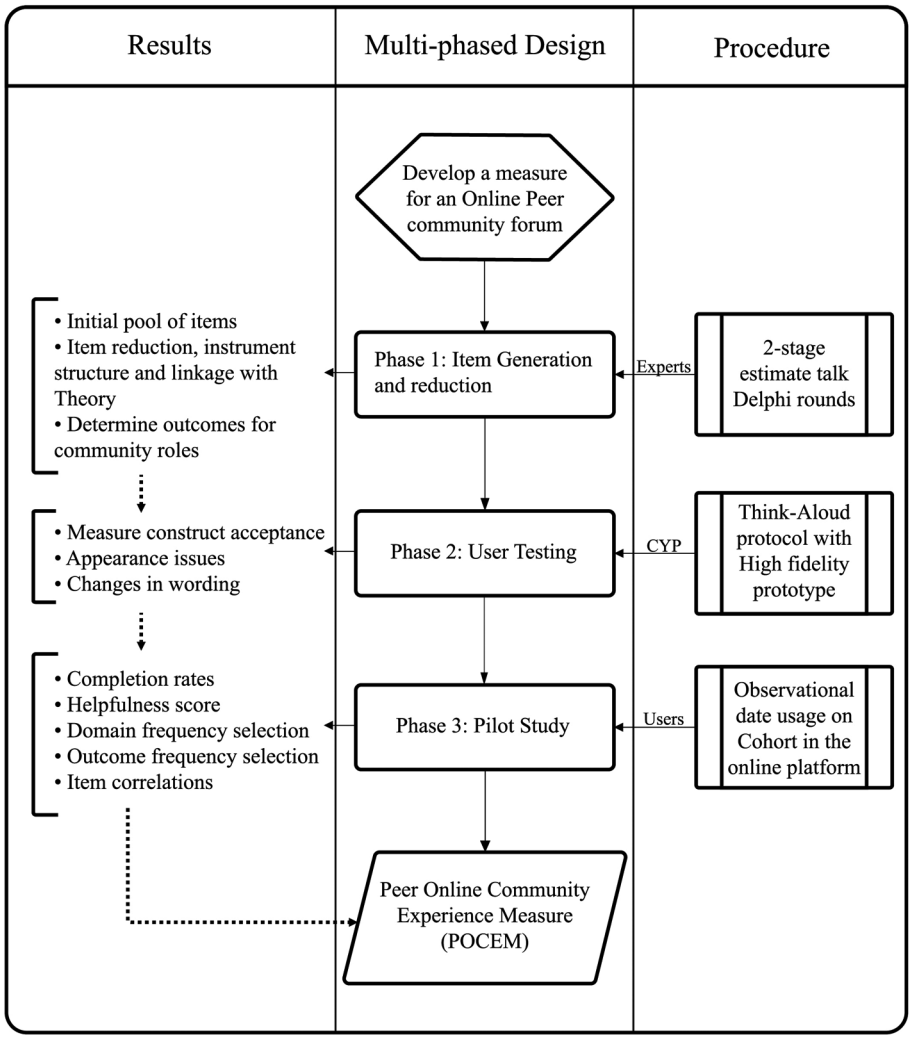


FIGURE 1
Multi-phased design with stakeholders and iterative results in the development of POCEM.

2.1. Phase 1: item generation and reduction

The process of item generation and reduction for the first phase of POCEM development was carried out using the Estimate-Talk-Estimate Delphi technique (35). The technique is used to achieve expert consensus through multiple discussion sessions between a panel of experts. The Estimate-Talk-Estimate method differs from the standard Delphi technique by then allowing for verbal interaction between panel members. The Delphi technique is frequently used in healthcare research and has previously been used to modify a social responsiveness scale (36), while the Estimate-Talk-Estimate method variation has been used in developing a framework for mental health apps (37).

To develop a measure of peer community experiences that reflected both young people's views and expert opinions, a two-stage Estimate-Talk-Estimate Delphi process was used. The first stage involved panel members with mental health practice expertise to compose an initial pool of items based on previous theory known about the service (31). The second stage of the Delphi process involved discussion between researchers to assess the items generated, identify links between the generated items and the constructs from the theory, and reduce initial pool of items using an inter-rater agreement approach to reach the final round.

2.1.1. Participants

Two panel groups were recruited, with of a total of six expert panels for the Estimate-Talks-Estimate workshops. Most experts belonged to the service and one to a university institution. Emails advertising the research participation opportunity were sent to the service employees involved in research or in providing support to service users within the community, and to external researchers in the service's existing network. Experts registered interest in the project *via* email and specified whether they were interested in participating in (a) the service's Theory of Change thematic analysis, (b) the item generation stage of the online community measure, or (c) the item reduction stage of the online community measure. Experts could volunteer for multiple parts of the project. All panel members had extensive experience researching or providing support and moderation in the digital mental health platform.

Panel members were recruited to participate in two projects concurrently: the development of the service's Theory of Change (31), and the generation of the content for POCEM. One panel member was involved in both the item generation stage and the later item reduction stage. The continued involvement of one panel member was used to ensure continuity between the two stages (**Table 1**).

TABLE 1 Expert panel members for item generation.

Panel member	Organisation role	Institution	Project involvement
PM1	Mental health practitioner	Kooth	Thematic analysis; Item generation
PM2	Mental health community practitioner	Kooth[1]	Thematic analysis; Item generation
PM3	Lecturer in counselling psychology	University of Manchester	Thematic Analysis; Item generation; Item reduction
PM4	Lead researcher	Kooth	Item reduction
PM5	Chief research officer	Kooth	Item reduction
PM6	Lead experience designer	Kooth	Item reduction

2.1.2. Procedure

Four rounds of panel workshops were carried out with the expert participants. Workshops were held face-to-face and videoconference. Asynchronous communications through email were used to prepare the experts, and anonymous questionnaires for voting were provided in each round. The rounds had different aims regarding the content relevance and structure, reduction of proposed items and changes in wording of items to improve quality. All rounds were documented through field notes supervised by the research lead (TH).

2.1.2.1. Initial round

The first stage of the process was a face-to-face meeting group with the expert panel, wherein a broad list of initial items was generated. The items were generated deductively using service programme theory (31). The experts involved in this initial item generation were concurrently involved in the Theory of Change research, allowing for a deeper understanding on the transcript's findings and theoretical foundation of the Kooth online community outcomes and mechanisms. The process of generating the initial pool of items utilized a thematic analysis approach, consistent with Braun and Clarke (38) analysis in psychology research. The thematic analysis investigated the factors influencing positive behaviour change for children and young people accessing an online peer support intervention and described the desirable outcomes for the online community. The items were generated by each panel member independently and decided in group through a panel discussion process. Items were generated using the framework for at least each of the desirable outcomes and mechanisms or processes for positive change in the online community identified in the thematic analysis.

2.1.2.2. Item pool development rounds

This round with panel members focused on mapping the initial pool of items to the domains of support. The domains of

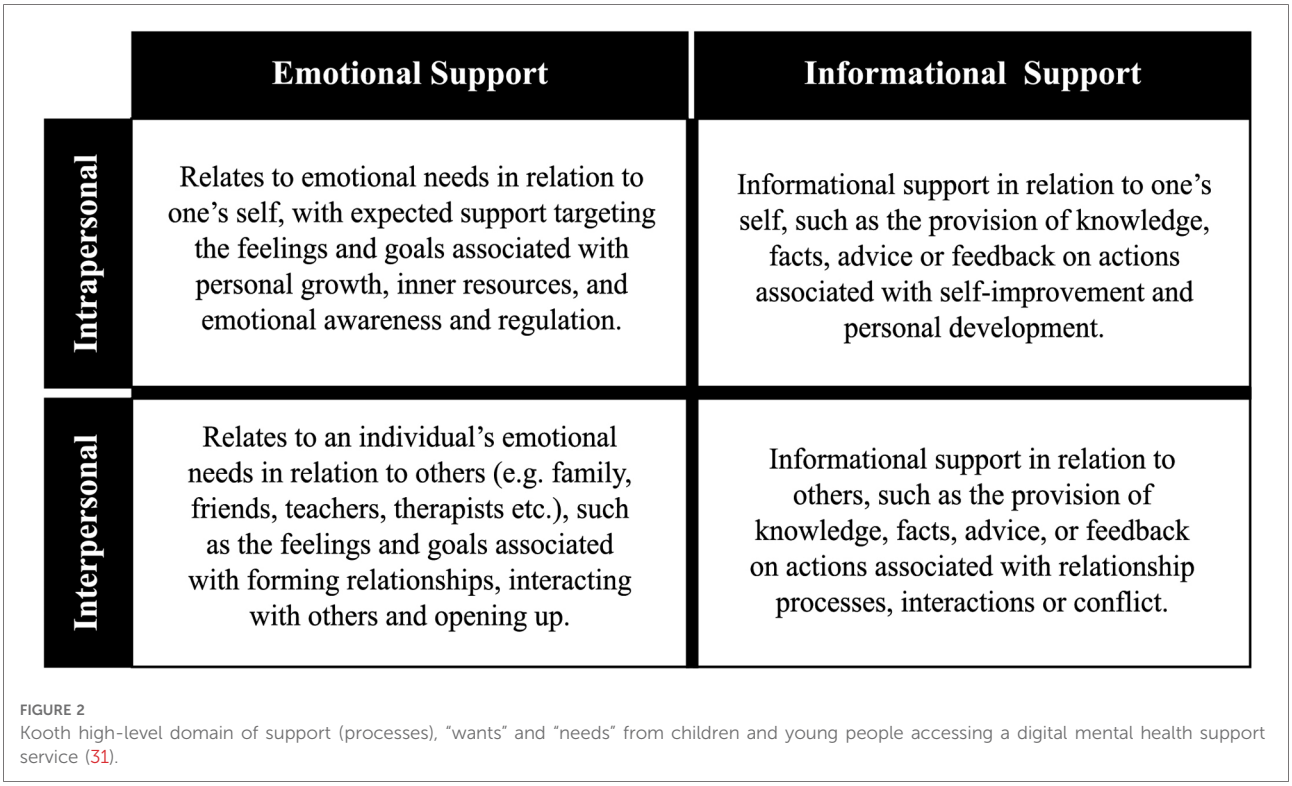


FIGURE 2
Kooth high-level domain of support (processes), "wants" and "needs" from children and young people accessing a digital mental health support service (31).

support were mechanisms formulated in earlier research, and they are intended to represent the high-level types of help from children and young people asking for mental health support within an online digital service. These domains of support are covered in the process assessment part of the POCEM, as the mechanism that leads to that online community resource outcome being helpful or unhelpful (Figure 2). After mapping domains for each item in the initial pool, panel members were asked to select the items more likely to be selected by two different types of online community members or roles: (i) those contributing; or (ii) those consuming (accessing by reading) the community resources. The workshop with experts focused on which items were more relevant to each type of community member and provided rationale on their decisions. The panel members voted on items' relevance to identify those to be discarded. Panel members were given two weeks to make their evaluations independently. A workshop was carried out to discuss the relevance items findings, and the relative ratings of the different members. The discussion focused on whether the items were repetitive, reflected the support domains as processes, and were representative of the outcomes from the thematic tree.

2.1.2.3. Final round

In the final round, panel members were presented with those items selected through voting. Panel members were given three weeks to present their review. The workshop focused on editing the wording of the items, and further reducing the items down due to similarity, or coherence with the theory used to develop it. The discussion considered the independent comments made prior to the workshop to each item and comments recorded throughout all rounds.

2.1.3. Analysis

Descriptive statistics were used to describe participants' characteristics and frequency in votes and selection was recorded for each expert. The field notes outputs from each round of the item generation and reduction phase were discussed sequentially, influencing the materials taken to the panel of experts in each round. In round three, when panel members independently and anonymously rated their preference of the items, an intraclass correlation coefficient (ICC) was calculated to understand agreement between panel members on their decisions for item selection.

The process of item generation and reduction was done iteratively over four Delphi rounds. A flow chart of the outcomes from each round of the item generation and reduction process are shown in Figure 3.

2.1.4. Results

2.1.4.1. Initial round

The items initially generated for the online community measure were produced based on the panel members' understanding of previous literature on online peer support communities and

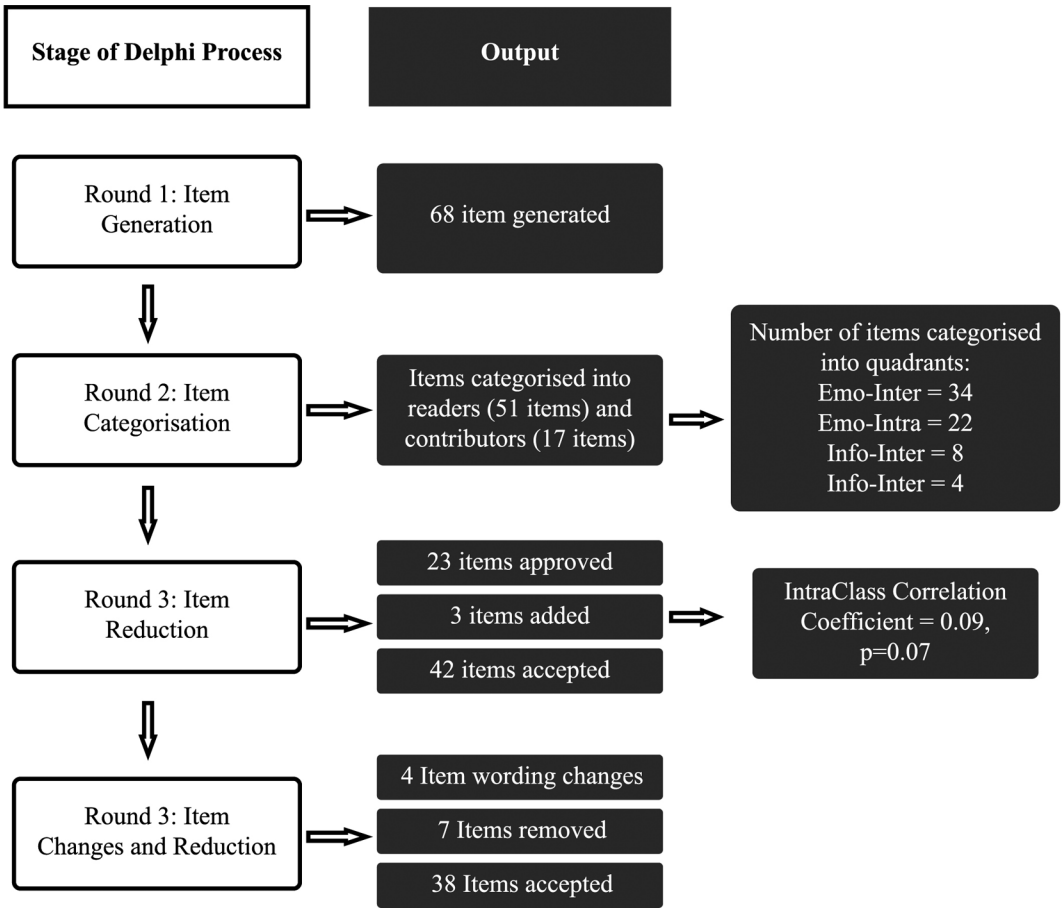


FIGURE 3
A flow chart of the results from each stage of the Delphi process for the generation and reduction of items.

service's Theory of Change (31) describing a high-level domain matrix of the type of support that children and young people "Want" and "Need" from the digital service ecosystem.

The thematic analysis revealed desirable outcomes for positive behaviour change in the online community, the peer ecosystem: (i) Relatedness and Self-Expression; (ii) Hope and Help Seeking; (iii) Building a Safe Community; (iv) Digital Altruism; (v) Hope and Help Receiving; (vi) Making Change. The identified themes as outcomes were used to generate the initial pool of 68 items based in these desirable outcomes and categorised through their mechanisms or domains of support (see **Supplementary Table SA1**).

2.1.4.2. Item pool development rounds

The first round aimed to divide items based on two criteria. First panel members categorized each item based on the type of community member that will find the item useful. Most of the items (75%; $n = 51$), were classified as relevant to users reading or consuming resources in the community, whilst the remaining 17 (25%) were relevant for users contributing with content to the community.

The second criteria was categorized into domains of support and it was used to inform the second part of POCEM. The emotional domains were more frequently used to categorize the items in the pool compared with informational domains (**Table 3**). Through a discussion process, participants agreed on a three-part structure to the measure, with respondents only shown items relevant to the selected process.

Each panel member voted on the items that they believed should be kept for the measure. An Intraclass correlation (ICC) estimate with 95% confidence intervals was calculated based on three judges, absolute agreement, with a 2-way mixed-effects model. The inter-rater reliability between panel members was poor at this stage of the item reduction process (ICC = 0.09, $p = 0.07$). In the following workshop the ratings were discussed amongst the panel. The outcome of the workshop was the removal of 23 items, and the addition of three further items. All the items that were selected by at least two ratters (12 items) were kept.

TABLE 2 User testing—affinity diagram output from workshops with POCEM interactive prototype at Kooth.com.

Magazine	Discussion boards	General community experience	Measure Insights
Clearly seeing who authored content can build or break trust, impacting engagement levels	Comments are a powerful tool for support (if positive)	Seeing content that is positive, distraction based, or more generic life advice was well received and unexpected to first time users	Selecting multiple options would allow YP to explain more about why something was helpful
Relying on users to read and process large chunks of text is both off-putting and risky	YP will more likely respond accurately to content that relates to them or content that keeps them engaged	Peer support or "community" are important alternatives to counselling	Free text fields allow YP to add their own voice/explain themselves, which is important to feeling understood
YP want the ability to reflect on or re-engage with an activity or content that has previously helped or inspired them	Being able to explain themselves or detail the 'why' behind how they feel or feedback is important to YP	Moderation is an important safety net for YP and Kooth's policy is not made clear enough	YP may be more likely to respond to content that was helpful to them
		YP relate digital "social" styled interactions to social media	The measure was interpreted as being related ti the specific content or category
		There is some expectation that user feedback will result in more personalized site activity	It is helpful for YP to know who will see the results on what they interact with, as this can impact if they engage
		YP who may not be directly struggling are empathetic to others who may have nowhere to turn	

TABLE 3 Frequency of items categorization into high-level domains of support after round 1.

Wording of quadrant	Domain of high-level support	Frequency of items
"It helped me to relate to others"	Emotional-Interpersonal	34
"I feel better about myself"	Emotional-Intrapersonal	22
"I felt it was important to me"	Informational-Intrapersonal	8
"I learned some skills I can try with others"	Informational-Interpersonal	4

TABLE 4 Changes made to the wording of items for the third part of the measure during the final round.

Original wording	Change in wording
I benefited from feedback from others	I benefited from comments from others
I feel optimistic for the future	I now feel more hopeful
I have implemented a suggestion from someone else	I have done something positive after a suggestion from someone else
I felt comfortable seeking support from my peers	I felt comfortable to share with others
I didn't feel judged by others	I feel good about not being judged
I felt connected	I felt connected to someone

2.1.4.3. Final round

Following independent and asynchronous evaluation of item wordings, nine suggestions were made regarding wording changes to the items. Six items were changed after the workshop where experts discussed ICC scores and disagreements between ratings and interpretation of the item (Table 4). The fourth and final round also resulted in the removal of seven further items. The output of the last round was a composed set of statements of three-part measure aiming to capture 38 different outcomes from four different processes on the helpfulness of community content and resource, this instrument content and structure was taken to prototype generation for the next User Testing phase.

2.2. Phase 2: user testing

Once the development brought the content and design of the measure to a prototype, this was designed and integrated as a high-fidelity prototype using the current experience of the community forum. A further layer of validity is required when testing in digital environments, a human-computer interaction understanding to improve its face validity, but also to understand the overall performance and understanding of the measure by the target population, this should also improve construct and content validity with richer findings.

2.2.1. Participants

A voluntary purposive sample of 11 young people was recruited amongst four primary and secondary schools in Manchester, UK. The sample was used to conduct user testing workshops. The 11 young people aged 12–17 (7 female, 4 male) expressed their interest for participating in the workshops. Young people aged below 10 or older than 18 were excluded from the workshops, as well as those at risk of safeguarding concerns or also not deemed appropriate by school staff to participate (e.g., lack of capacity or competency). The study was advertised through teaching staff,

participants had no previous experience as users within the digital service (Kooth.com), parental and individual consent was sought for each participant and an incentive of £10 was given to participants to attend the 60-minutes workshops. Participants could drop-out of the workshop at any point. Two researchers, one with participatory research expertise and a user experience designer conducted and analysed the user testing sessions.

2.2.2. Instruments and materials

2.2.2.1. Kooth prototype: clickable high-fidelity

A high-fidelity prototype is a smartphone-based interactive representation of the product, the digital service, with strong similarities to the final design in terms of details and functionality. The high-fidelity prototype was developed with the vector graphic editor Sketch software (39) and included the POCEM inside the online peer support community allowing the users to click around and interact with the whole platform. In the context of measure usability, a high-fidelity prototype allows exploration of wording, structure, relevance, and comprehensiveness of the measurement and its functionality in interaction with the whole platform.

2.2.2.2. Peer online community experience measure (POCEM)

The POCEM is an online community measure designed and build on theory specific to the digital service (Kooth.com). Its aim is to measure areas of care around satisfaction and quality that an online community forum has in relationship with the individual in the context of the digital mental health service. The measure automatically differentiates between roles

FIGURE 4
High-fidelity clickable prototype of the three stages for POCEM.

on the forum community by contributors and readers. The measure is divided in three stages and contains some logic based on the score responses and selection (see stages in **Figure 4**):

- The first stage contains single item question ("*Did you find this part of Kooth helpful?*") scored with a 1–5 Likert scale (1: No; 2: Not really; 3: Don't Know; 4: A bit 5: Loads!; all scores aided with emojis) to assess the helpfulness of the online community resource. The Likert scale scores determines the helpfulness as a benchmark and branch the measurement into stage two.
- Depending on the scoring in stage one, a new single-item question will prompt "*What were you hoping for?*" for 1–2 scores and "*Why did you find it helpful?* for 4–5 scores for the respondents to select between four quadrants representing high-level domains of support from the service. Respondents who select score 3 "*Don't Know*" do not complete more steps in the measure.
- The last stage is displayed only to those users who completed stage one and two (scoring 4–5 in step one and selecting the domain in step two), A single item (multiple response) question ("*What type of things have you learned?*") will ask to select from a group of outcomes found at phase 1, readers can select between 23 outcomes and contributors can select 14 outcomes, the outcomes displayed will depend on the domain selected in the previous step, the outcomes available to select were generated in Phase 1 by experts.

2.2.2.3. Lookback.io: screen recording & audio

Lookback.io (40) is a user testing software for mobile UX user recording tool. It allows recording of screen interactions alongside voice audio recordings when conducting supervised sessions of remote user testing (41). This software allows secure storage and organization of your user testing sessions for qualitative analysis. This tool allowed the recording of both screen behavior and audio from participants attending the user testing sessions.

2.2.3. Procedure

The user testing was structured in one-to-one sessions delivered at each of the four schools. Participants were provided with a smartphone which had a loaded a high-fidelity prototype of the measure within the platform. Sessions were facilitated by a user experience expert researcher and observed by another researcher to safeguard the session and take notes. The sessions were voice recorded and screen recorded, for further transcript and analysis.

The facilitator encouraged young people to verbalise their thoughts and perceptions using the think-aloud protocol (42) as they navigated their way through the platform while following the facilitator instructions with the prototype. Instructions followed a protocol of specific tasks within the prototype measure that aimed to identify any issues with the interface, allowing facilitators to observed participant specific behaviour in relationship with the task. Facilitation tasks included asking about expectations in relationship to the next event that the interface showed during the session, and whether there were any issues with the wording clarity and relevance for the measure.

2.2.4. Analysis

Affinity diagrams or KJ methods are adopted in user testing for prototype interaction (43). They are a good technique tool to synthetize and organize large amounts of qualitative data post-task, the user testing sessions were synthetized in affinity cards representing each participant's observations and quotes. Such cards are later jointly analysed to create a diagram. Affinity cards for issues more frequently raised, and for higher severity reported by participants tend to take more priority to address as changes in the prototype.

We followed the adapted four stage (creating notes, clustering notes, walking the wall, and documentation) process from Lucero (44). Researchers worked using rows to represent participants and columns for each affinity note. A total of 236 affinity cards were collected from field notes, screen recordings and audio recordings from each session. Rounds of clustering by researcher identified two main clusters in reference to the measure, and to the prototype and task performed (58.48% Measure affinity notes and 41.52% Prototype and tasks affinity notes). Twelve clustering issues were collected across clusters, some directly related with the measure such as including an "other" personalized option, and issues with the platform and prototype such as difficulties in navigation. The affinity diagram was then created, walking the wall exercises with other researchers and experts at the service ($n = 3$) provided with synthesis and identification of priority changes in the measure and prototype interaction by looking at frequency, feedback notes and quotes presented in the affinity diagram. Documentation on the output from the affinity diagram discussed by experts is provided in **Table 2**.

2.2.5. Results

The affinity diagram findings identified issues and recommendations for the POCEM. Many participants perceived the measure to be linked with the type of content consumed or accessed-at-the-time by the user inside of the community, being mainly forum posts and subsequent comments.

This is well illustrated by one of the participants quotes when prompted in the session to explain what the measure is intending to do [Participant 3]: "*How? in my experience was just reading the person and the comments under it*".

Most young people reported that they will be more likely to complete the measure if the content of the forum post was helpful for them.

The feedback suggests there may be an agreement bias effect deterring users from providing negative feedback to a peer within the online community, or encouraging users to ignore the measure when the content is not perceived helpful; [Participant 9] said "*If it is related to what I am doing, I would fill in the measure, to see similar articles*" and [Participant 1] stated that connection with peers will be a key motivation to complete the measure: "*… if I had trouble making friends I would say loads (of motivation)*".

Most of participants found it normal that the measure will appear in piece of content within the community. Although, some participants expressed difficulty in finding the measure in the platform interface without prompts. This provided evidence regarding the measure appearance suggested that changes in design may affect measure completion. The comment from participant 9 indicated that some service users may believe that completing the measure contributes to content recommendation within the service. At the time of testing, content recommendation was not an intended outcome for the measure but highlights users' expectations and assumptions. It was identified that one emoji under the scale had a mismatched emotion. [Participant 10] explained: "The "No" just looks like they're about to cry or something". Despite the majority appeal to use emojis within the scale, for instance [Participant 1] stated his preference: "The emojis are more neutral not grumpy or red as might give wrong impression to others". Findings also reveal difficulties from participants understanding who will see their responses. Four participants demonstrated doubts about the information being publicly available for peers to see in the community. For instance [Participant 8] showed: "I thought it would instil confidence in the author to write more". User experience findings around physical appearance of the measure and its instructions led to changes for version of the measure taken pilot phase.

Finally, user testing workshops provided a scenario to review all item wordings based on experiences of participants interacting with the prototype during the exercises. Some statements changes are presented based on the rationale given by participants extracted from the affinity diagram. Wording review steered two changes on the process assessment part, and four wording changes in the last part of the measure (outcome assessment) (**Table 5**).

Overall, user testing allowed identification of appearance issues, validated the focus of construct measurement (it measures the specific community resource), and allowed changes on wording by the intended population. These findings allowed to implement and administer the POCEM measure at Kooth.com, providing the results of this implementation in the next face of the study.

2.3. Phase 3: pilot study

In contrast to content validity which is more concerned with having the breadth and accuracy of items to measure a construct, face validity assesses the degree of respondents judging that the instrument and its items are appropriate for the targeted assessment (45). For an experience measure to provide useful and valuable information, it must first be considered acceptable by service users within the context it is implemented. We used completion rates to assess how acceptable the measurement was within the online community and compared drop-out effects at each stage of the three-part assessment measure. The demographic differences were analysed between the assessment of structure scores, and the assessment of structure scores were compared between different outcomes and mechanisms. During a 10-week pilot we collected qualitative and quantitative data from the users in the online community completing the measure inside the platform (11–25-year-old service users).

2.3.1. Participants

The measure was iteratively released onto the service's platform. Online service users who either contributed to a forum or submitted an article during the testing period were

TABLE 5 User testing affinity diagram findings and changes in wording of the measure.

Original statement	POCEM	Rationale [Participant]	Changed statement
It feels important to me	Part two	"It means what he said is important to you because you relate to it" [N]	I learned something important to me
I understand myself more	Part two	"Maybe you don't realise how you feel because maybe sometimes you can be sad but you don't know why then you read all this it can make sense to you"[N]	I feel better about myself
I feel safe in this community	Part Three	"Does it mean Kooth as a whole or the outside community?"[N]	I feel safe in the Kooth community
I feel good about now being judged	Part Three	"At first I was like what does it mean?"[N]; "Obviously when someone judges you its negative, but in the app you can feel more better about not being judged"[N]	It feels good not to be judged
I have done something positive after a suggestion of somebody else	Part Three	"I don't think you're gonna do this—I don't get how you're gonna do something positive? I don't think you're gonna come back and do the quiz"	I have learnt enough to make a positive change
I feel excited to support other people with my new found knowledge	Part Three	"It is kind of the same as I've done something positive (Item) …"	Item removed

presented with the contributors' measure after submission. Service users who read an article or forum were presented with the readers measure at the end of the post. Users who did not provide research consent during sign-up to the service were excluded from analysis. Data was collected between the 13 November 2019 and the 22 January 2020.

2.3.2. Procedure

The clickable prototype of the POCEM was implemented as a feature for service improvement in the online community at the service's platform, changes from the previous Phase 2 were included in the measure for pilot. For a period of 10 weeks the measure was tested within the platform and data was collected on users engaging with the community at the digital mental health service. Routinely collected monitoring information was used alongside peer support data to investigate the measure performance. All users accessing the platform community were able to complete and see the measure during the 10-week period.

2.3.3. Analysis

Frequencies and descriptive analysis were carried out on completion rates for users who accessed the online community and those who completed the measure. Descriptive statistics and frequency of selection on the three steps of the measure were calculated to understand if items were being selected sufficiently. As POCEM measurement is divided in three assessments that interrelate, different analytical approaches were taken for each section of the measure. For the assessment of structure, the helpfulness score, Kruskal-Wallis non-parametric test alongside Dwass-Steel-Critchlow-Fligner pairwise comparisons post-hoc tests were used to ascertain differences in demographic variables (age, gender, ethnicity) on the score. Further analysis then explored the type of community interactions (whether the respondent was a reader or a contributor), using a two-sample *t*-test.

For the POCEM process assessment, we explored the effect of the domain selection on the score through Kruskal-Wallis, post-hoc analysis using Dwass-Steel-Critchlow-Fligner pairwise comparisons were performed looking at the average helpfulness scores for each four domains of support, and the average score for respondents who dropped-out at this step.

The POCEM outcome assessment was explored looking at the differences in scores based on the outcome selected in the measure. The aim of the pilot analysis was to see POCEM acceptability by service users using completion rates and whether the phased design resulted in a drop-off of respondents. We also explored the outcomes and processes more frequently selected by users of the measure while in the community.

2.3.4. Results

2.3.4.1. Completion rates

The measure was tested between the 11 of December 2019 and the 20 of January 2020, with 2,140 unique service users completing a total of 4,897 administrations POCEM. There was a total of 68,439 views of community content on the site by service users who gave research consent during this time, and a total of 2,425 contributions in the form of article or discussion posts in the online forum community. Completions rates were divided between readers and contributors to better understand overall completion of the instrument across community members (**Table 6**).

2.3.4.2. Participant demographic characteristics

The respondents ages ranged from 10 to 25, the range allowed in the community and the service. However, five respondents reported an age over 25 and were removed from the dataset, as these will be outliers of the service. The remaining respondents ages ranged from 10 to 25, with a mean of 13.47 (SD = 2.09). Most service users completing the POCEM were female, white, and aged between 10 and 14 years (**Table 7**).

2.3.4.3. POCEM structure assessment: helpfulness of peer community

The most frequently selected helpfulness score was 5: "Loads!", indicating that the content helped the service user considerably. The frequency with the rating of 1: "Not really" was selected the least frequently. The mean helpfulness score was 3.77 (SD = 1.14).

Demographic differences were analysed to investigate whether the POCEM showed different experiences of structure between service users.

There were no significant differences between different genders [$H(3) = 2.4$, $p = .40$], or between different ethnicities [$H(4) = 8.4$, $p = .07$]. The age had a small significant effect on the perceived helpfulness scores, with a Kruskal-Wallis test showing a significant effect of age on helpfulness score [$H(2) = 7.89$, $p = .02$]. Post-hoc tests using Dwass-Steel-Critchlow-Fligner pairwise comparisons were carried out for three groups

TABLE 6 The unique users, POCEM completions, and proportion of completions within the community.

Engagement type	Participants	Frequency of measure views	POCEM total completions	Community completion rates
Readers	2,083	68,439	4,685	6.85%
Contributors	57	2,425	212	8.74%
All	2,140	70,864	4,897	6.91%

TABLE 7 Demographic characteristics and frequencies of unique users completing POCEMs.

Demographic	Frequency	Relative frequency
Gender		
Female	1562	73%
Male	474	22.2%
Gender Fluid	71	3.3%
Agender	33	1.5%
Age		
10–14	1143	72.6%
15–19	557	26.0%
20–25	46	1.4%
Ethnicity		
Asian	128	6%
Black	81	3.8%
Mixed	109	5.1%
White	1753	81.9%
Other	69	3.2%

and showed service users aged 10–14 gave a significantly ($p = .03$) higher helpfulness score ($M = 3.8$, $SD = 1.13$) compared to service users aged 15–19 ($M = 3.7$, $SD = 1.5$). There was no difference ($p = 0.4$) between service users aged 10–14 and aged 20–25 ($M = 3.4$, $SD = 1.43$), or between service users aged 15–19 ($M = 3.68$, $SD = 1.15$) and aged 20–15 ($p = .80$).

For the role as a member of the community, T-test frequency comparisons showed a statistically significant difference in the mean helpfulness score [$t(247) = 8.8$, $p < .001$] between readers and contributors. Service users who completed the POCEM after contributing to the community content selected the helpfulness score of 5:'Loads!' substantially more frequently than service users who read the community content (Figure 5).

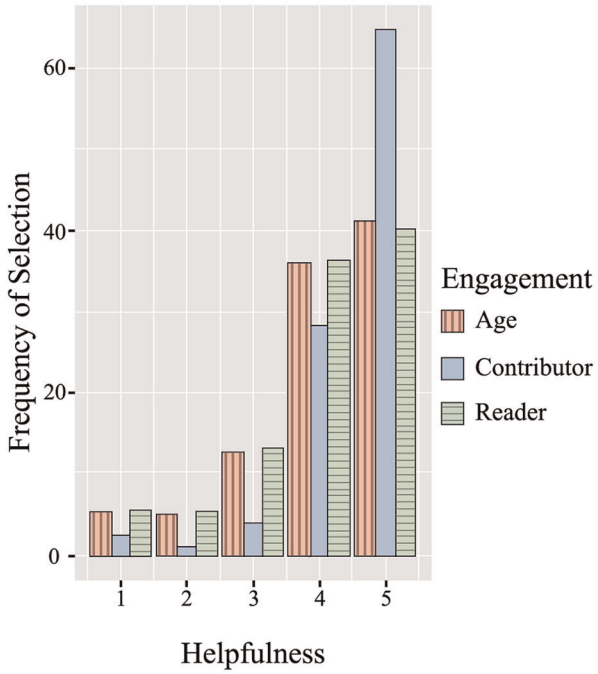


FIGURE 5
Frequency of selection across the five helpfulness scores for each type of engagement.

2.3.4.4. POCEM process assessment: high-level domain of support selection

Out of the 4,897 completions of the measure, 14.2% of responses gave 3: "Don't know" as the helpfulness score. For this score response, the rest of the measure was not shown, and responses ($n = 619$) were removed.

As seen in Figure 6, the most frequently selected high-level domain of support was "Help me relate to others", with 55.1% of respondents selecting the option. Across respondents who gave positive feedback more than half (58.2%) selected the domain from the process assessment "Help me relate to others". When looking at respondents who gave negative feedback, 32.3% selected the "Help me relate to others" (Emotional-Interpersonal) domain of support.

Out of the 4,278 responses with a score positive or negative score (1,2,4, or 5), 10.05% of respondents stopped answering the measure after providing a helpfulness score. When splitting the responses by positive or negative feedback, 34.5% of those giving negative scores did not answer next process assessment part of the measure and dropped out. Comparatively, out of all respondents who gave a positive response, only 6.8% dropped out of the measure without selecting a process domain.

An analysis of the helpfulness scores for process support domains gives the same message as the frequency findings, with respondents who dropped out of the measure giving a lower average score. A Kruskal-Wallis test showed a significant effect of process domain selection on helpfulness score [$H(4) = 207.45$, $p < .001$]. Post-hoc tests reveal that there was no significant difference between the helpfulness scores on the domains of POCEM process assessment. The post-hoc analysis revealed more about this difference and showed service users who did not give a response ("No response") gave a significantly lower helpfulness score ($m = 3.28$), compared to those who selected the other domains "Important to me" ($M = 4.29$, $p < .001$), "Learn skills" ($M = 4.09$, $p < .001$), "Relate to others" ($M = 4.31$ $p < .001$), and "Understand myself" ($M = 4.28$, $p < .001$) (Table 8).

2.3.4.5. POCEM outcome assessment: outcome based selection

The analysis was run after removing cases where respondents did not answer the process domain selection phase ($n = 430$)

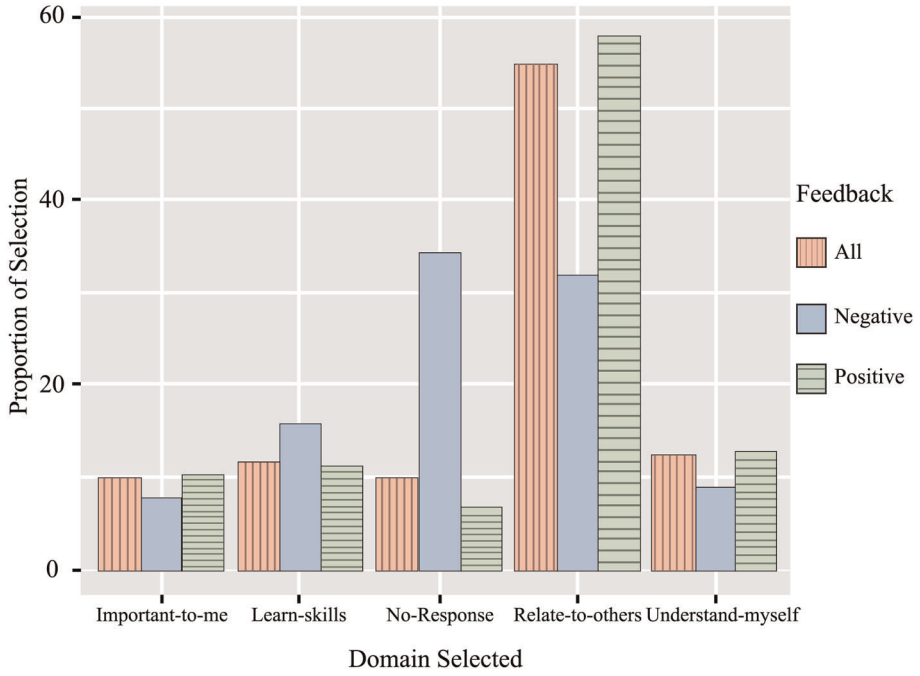


FIGURE 6
Frequency of selection for high-level support domains in process assessment of the measure for each type of feedback.

TABLE 8 Dwass-Steel-Critchlow-Fligner pairwise comparisons for domain helpfulness scores.

Item 1	Item 2	W	*p*
Relate-to-others	No Response	−19.784	<.001
Relate-to-others	Learn-skills	−1.855	.684
Relate-to-others	Understand-myself	1.991	.623
Relate-to-others	Important-to-me	2.278	.491
No Response	Learn-skills	12.733	<.001
No Response	Understand-myself	15.671	<.001
No Response	Important-to-me	15.011	<.001
Learn-skills	Understand-myself	2.683	.319
Learn-skills	Important-to-me	2.889	.246
Understand-myself	Important-to-me	0.397	.999

as drop-out no scores were recorded in the administration. The frequency of outcome selection was analysed for each of the process domains, as the outcome items shown to respondents was dependent on the earlier selection. For all but "Relate to others", the most frequent action from respondents was to drop-out of the measure, making up 20% or more of the responses in each domain (see **Figures 7**, graphs A–D). When looking across all the outcome responses, dropping out of the measure after the process question accounted for 25.38% of the sample who reached this stage of the POCEM. This is a higher drop-out rate compared to the 10.05% of respondents who dropped out at the previous stage.

For the domain "Important to me" the outcome item selected most frequently was "Others have the same experience" (18.9%), for the domain "Understand myself" the item "Felt accepted" was most selected (19.5%), and for the "Learn Skills" domain the most selected item was "Skills to help others" (17.9%). The process domain "Related to others" had the item "Felt connection" selected the most frequently (20.7%) (**Figure 7**, graph B). The item "I now feel able to ask for support outside of Kooth" was selected the least frequently out of the total items, with only 4.27% of respondents selecting the "Emotional-Intrapersonal" domain choosing the outcome (**Figure 7**, graph C).

The pattern of lower helpfulness scores for respondents who dropped out of the measure before completion continued for the outcome item stage. The Kruskal-Wallis test showed a statistically significant difference in helpfulness score between outcomes [$H(22) = 407$, $p < .001$], and the post-hoc test showed a significant difference in the helpfulness score when respondents dropped out before answering the outcome assessment stage, compared to those selecting an outcome in the instrument. There were no significant differences between other selection of outcomes (**Supplementary Table SA2**).

3. Discussion

This study outlines and discusses a novel multi-phased design method for developing an experience measure for

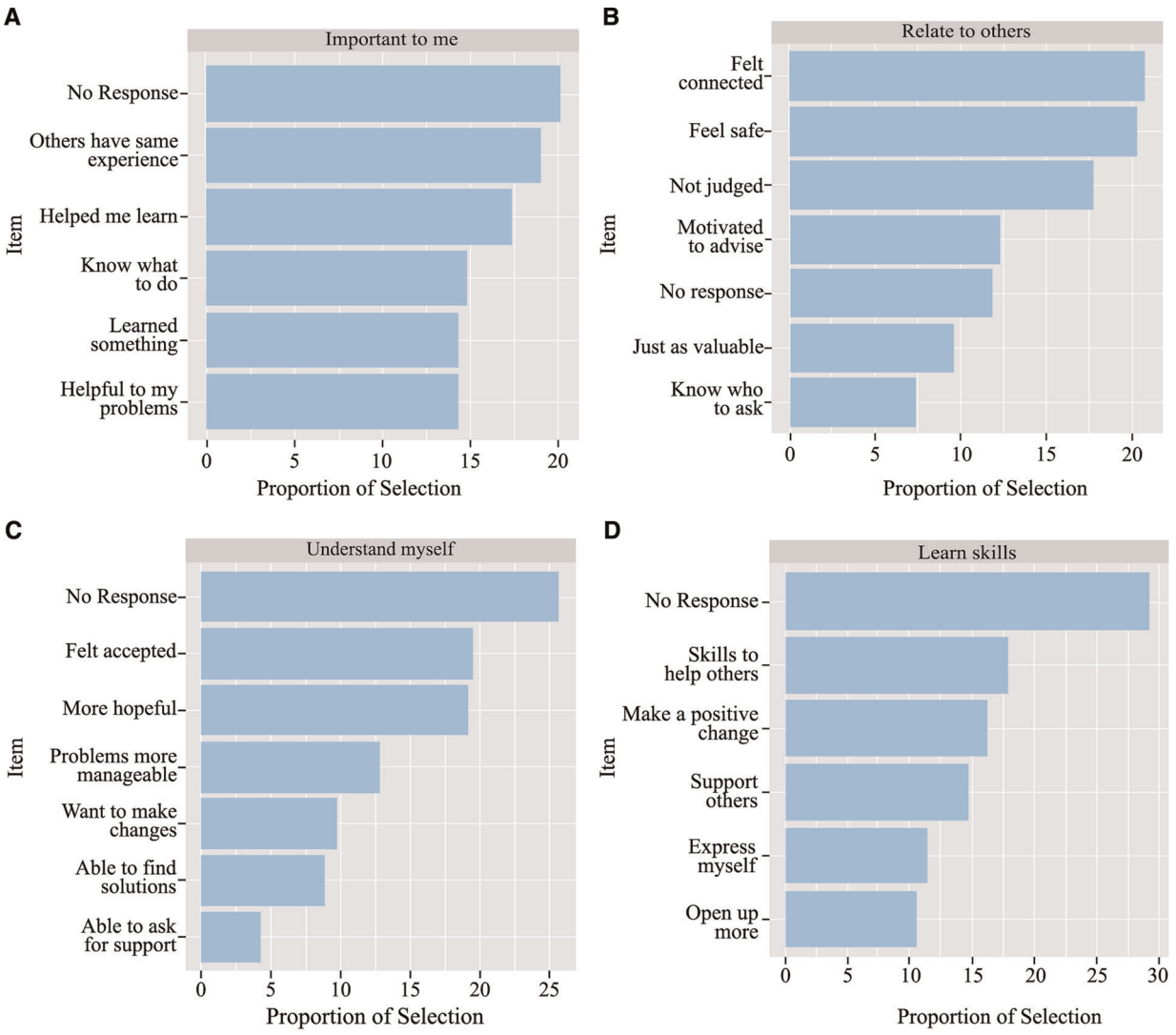


FIGURE 7
POCEM selection frequency of outcomes for each domain. Each panel shows the proportion of selection for outcomes selected by High-level domain of support selection: (A) "Important to me"; (B) "Relate to others"; (C) "Understand myself"; (D) "Learn skills".

young people within an online mental health peer community forum. We aim to provide a structure for design to improve experience of technology-enabled solutions in an online mental health service, whilst reflecting on lessons learnt, to support the future research of experience in other digital mental health contexts. We highlight the value of using mixed methods for an iterative design process with structured phases of data collection and synthesis.

Previous research has shown that role of online peer communities in supporting mental health is complex, and not always positive. There is a clear need for digital services to monitor the experience of service users engaging with online communities when the community is offered as part of digital mental health support, to understand whether the community is truly effective at providing mental health support as part of the service. An experience measure can provide an evaluation of the quality of care received by service users, uncovering what is and it is not working in a service (46). Using the Donabedian framework (30, 32) as the framework for measure design enabled an assessment of how the structure, processes, and outcomes within a specific mental health online community are experienced by young people and provided a theory-driven model for measure creation.

We conducted a phased approach for instrument development involving research with multiple stakeholders and mixed-method data collection activities, implemented in a real applied context (Kooth.com). A phased approach can help the

implementation of the instrument by two processes: (1) design and (2) evaluation, of each phase iteratively and gradually optimizing the solution for the technology-enabled service, to ultimately sustain it (47). The multi-phased design was structured in three study phases to answer specific research questions relevant to stages of measurement creation (48). First, experts to design the principles of the instrument, providing the foundation for a prototype. Second, young people as participants of user testing think-aloud protocols provided feedback on what they considered important to measure and their perceptions interacting with the measure as a prototype. Third, we evaluated in a pilot study the usage and completion of the measure within the platform with users as stakeholders.

The phased design involved multiple stakeholders, but each contributed to a singular phase. In the development of the measure itself, young people were consulted (phase 2), with much of their feedback influencing the final design. Involving young people across the development and design was essential for ensuring the instrument accurately reflected their needs within a peer online community, and therefore improved its acceptability (49). A phased approach for instrument development involving co-design participatory action research with multiple stakeholders can influence the structure and purpose of the measurement. Similar approaches are useful to influence the government policy on digital mental health in Australia (50). In the case of using community-led design for the development of an experience measure, findings and design decisions may be counterintuitive to the structure and administration of the instrument. For example, the design and administration may limit or breach assumptions to test measurement performance in psychological instrument research enhancing difficulties to understand the psychometric properties of the measure. On the other hand, ensuring the contribution of user stakeholders is embedded in the development increases the likelihood of acceptability and adoption within the given context. A consequence of placing service users, clinicians, and user experience experts at the centre of the design process may be an atypical structure to the measure or solution with competing the needs reflected in the process and the final solution and can add some complexity to the development process. This uncommon structure may hinder its generalizability and may not adhere to assumptions required to further investigate the quality of a measurement and its validity.

As a lesson learnt, it is important to carefully understand in which phases each stakeholder should influence and consider a wider involvement with participatory roles within each multi-phase design. This in turn may influence the time and complexity of this type of implementation in a technology-enabled service, as more complex data and synthesis will be involved.

This phased approach using the context-specific theory provided by the Theory of Change of the technology-enabled service (31), which should increase adoption and success on acceptability providing high response rates. Despite this, the pilot results found low completions rates, and a high drop-out rate as the stages of the measure progressed. Similar studies focusing on theory-driven instrument development found low response when surveying different services (51). This brings wider questions on how online communities and their mechanisms and outcomes should be investigated and subsequently measured. Future research aiming to refine the experience measure development process should consider how variations in the measure structure may impact acceptability by service users, alongside the transferability of the measure and the phased-design process to other communities and their own theory-driven frameworks.

Regarding instrument development, the first phase of the process focused in answering whether the theory-driven items generated by subject matter experts sufficiently represented the domains of support (also considered processes). The use of adapted Estimate-Talk Delphi rounds with experts allowed for a narrowed and improved content of the measurement. The approach was non-standard, with each Delphi round composed of two parts; an initial, independent assessment of the items followed by a group workshop. Unlike systematic approaches to Delphi rounds (52) the independent and anonymous polling incorporated both qualitative and quantitative feedback. Whilst most Delphi rounds included independent voting, only one round required panel members to rate the items. In the other rounds, panel members provided feedback only. Although rounds may have benefited from a psychometric systematic assessment like Content Validity Indexes (53), a dynamic approach to the Delphi provided rich qualitative feedback to influence the iterative design of the measure and enhance the design of the innovation (54). Feedback in the initial Delphi rounds suggested the three-part structure, with the helpfulness rating placed at the beginning, filtering the other parts of the instrument collecting information about mechanisms and outcomes of that community experience. The theory-driven items were carefully selected through this process and deemed relevant for the context, and a consensus was reached to directly map the items onto the processes of support identified in the Delphi rounds.

However, the three-part structure presented challenges on quality. The drop-out rates found in the pilot testing suggests that the measure structure was not adopted by service users. Whilst in many context-specific measure creations, experts will be at the core of the design and creation, researchers should sense check design by the target community, with key stakeholders consulted at multiple stages of the development process.

The User testing phase was considered by the research team a fundamental step for measure development in an online context, it aimed to acquire face validity by asking young

people about their understanding of what is being measured, and if people understand the measurement within the technology-enabled service. We recommend replicating the best fidelity prototype possible when conducting user testing research activities, despite low and high-fidelity prototypes have shown similar results when compared (55). We observed how young people benefited from structured activities and more realistic objects for the think-aloud exercises, which in turn can help to influence changes in the appearance and quality of the workshop outputs and its findings. User testing is a time-consuming process and the volume and complexity of data generated may contribute to longer periods of time and expertise needed for analysis. The KJ method can provide a good opportunity to analyse and synthesize findings but requires expertise and focus from researchers to facilitate the synthesis of the user testing activities. Affinity maps on the other hand can inform beyond the purpose of the research and provide ideas and improvements with general industrial value (e.g. user needs, product satisfaction). The findings from the user testing indicated that the participant understanding of what was being measured matched with the goal of the measure (to measure the experience interacting with content in the community).

In product development, usability research focuses on identifying areas where users struggle with a product or start to lose interest through observing people interacting with the product whilst trying to accomplish goals or tasks (56). For POCEM the participants were prompted to complete task but allowed exploration to gather their thoughts and cognitions about the wording and understanding of its purpose. This phase is likely to influence the item development process and provide further evidence on content validity of measurements in digital contexts.

User testing will often use a volunteer or purposive sample, added emphasis should be placed on finding participants normally underrepresented, as well as ensuring safety and ethical standards for research with vulnerable populations. User testing methods may present challenges integrating quantitative information like usability surveys, but this may help to improve researcher bias in the synthesis stage. User testing represents a new and additional phase for measure development that provides invaluable observations about the digital context and measure content (57). As wording of items were influenced by this phase findings, other studies should consider involving the target population in the initial item and theory development using community-based participatory research approaches (58).

Our pilot study set up to understand how acceptable the measure and how acceptability bias may influence scores. Results indicated that service users who completed the measure during the observational period had a positive experience when accessing the online peer community content, with the helpfulness ratings frequently positive. Those who contributed to the community as writers found it more helpful than those who consume the community as viewers, whilst younger service users rated the experience in general as more helpful. Given that most of the service users are aged between 10 and 14 years (72.6%), a lot of the content both written by service users and by the Kooth content team will be targeted towards younger service users. As such, it is not surprising the young service users may have a more positive experience within the community.

These initial results may indicate social desirability or acquiescence bias effects, previously found in digital contexts and scale creation (59, 60). The potential influence of an agreement bias effect was highlighted in the user testing phase, the majority of young people interviewed by think-aloud protocols reported that they were more likely to complete the measure if the specific content of the forum was considered helpful. Four out of the 11 participants indicated that they believed providing feedback would automatically notify the author contributor of the forum post. The user testing phase revealed worry from users about their responses being seen by other community members, changes in the instructions and text in the measure were applied in the pilot phase, including further instructions reinforcing anonymity of responses (**Figure 3**). The limited disclosure required to complete POCEM responses, along with the anonymity of the service, should help users to not anticipate a social consequence of their responses (61, 62) and promote completion and engagement with the measure.

The completion rates for POCEM during the study were low compared with the potential size of the community represented by the number of views during the pilot. Furthermore, from those who started the measure a respondent fatigue effect in each step of the measure was observed. The low competition rates and high drop-out rate of service users starting the measure presents a key challenge and threat to the acceptability of the measure. Users reading content in online communities are more likely to be "lurkers", individuals who will read community content but not actively participate (63). Therefore, digital environments might be more prone to missing or misleading data after administration, how missing data is treated can have consequences in psychometric testing and measure performance (64). Researchers in digital contexts should be aware of these issues, report missing data or exclusion rationale, and think in advance what psychometric properties or indicators of quality for the measure should be tested.

In regard to the frequency of selection of domains and outcomes, the pilot showed support domain "Relate to others" to be the most frequently selected process for those service users who perceived the resources as positive in terms of "helpfulness". The average helpfulness score for users selecting the domain was not significantly different to the other support domains, but service users were less likely to drop-

out at the item selection stage when selecting this domain. Previous research has illustrated key reasons for young people seeking out support in online communities are to feel less alone with their problems, find a space to talk with peers, and find a space where they feel less likely to be judged (26). Our pilot results show similar reasons for young people seeking support, with the most frequently selected outcomes in POCEM for positive experiences in the community being "Felt connection" and "Felt accepted". Similar outcomes for a supportive online community have been found previously (65) and demonstrate how online communities may help users to feel less isolated and more supported (66). The overall frequency of outcome selection at the outcome assessment part of POCEM will, at least partially, be a consequence of the differences in frequency selection of domains in the previous assessment. When respondents selected an outcome after selecting a domain in the process assessment, there was no significant difference between the average helpful scores for each outcome. The positive average of helpfulness during the pilot reflects a positive experience for service users in relationship to the outcomes selected in the instrument. On the other hand, it may also be a consequence of a ceiling effect in the measure (67).

There are several limitations to be considered for the development of the POCEM. This study offers insights into considerations that should be made in the early development of a measure for a digital context. By designing the instrument or measure with a specific service in mind, the ability to generalize the existing measure to other online peer communities is limited, and the use of experts from the same context may provide a limited view during Delphi rounds (68). Some of the lessons learnt in the development of the POCEM illustrate the benefits and challenges of designing and testing a measure in a digital environment within a multi-phased mixed methods approach. Further statistical and content validity testing is required, especially to understand how individual differences may affect the performance of the measure, and if biases of its design can be reduced by optimization cycles.

The POCEM can help us understand consumption and use of mental health supportive online communities beyond web-based analytics (e.g., how long people read, or contribute, and how frequently they engage). Further exploration on acceptability and completion rates in relationship with other digital phenotypes and instruments are required. Routinely collected information from this measure may help to understand the trends and commonalities deemed helpful in the community, it should also explore differences across population characteristics so the measure can be evaluated beyond the pilot, thus one should be mindful of the demographic differences observed and how they may affect future applications of the measurement and biases. If the measure is found in future research to have a sufficient level of acceptability by service users, then a further direction of research is to investigate the use of the measure by informing content recommendations. In the user testing phase, one participant believed that filling out the measure would result in personalised recommendations, suggesting that a personalisation experience may be expected or desired by service users.

Experience measures like POCEM can help services to understand the mechanisms and outcomes more frequently achieved by users of an online community. Online peer communities may use experience measures to understand what resources benefit or hinder the individual, so "active therapeutic engagement" can be monitored and better understood beyond digital analytics, so a positive and safe space and ecosystem can be maintained for peer support in a digital mental health community. Research aiming to replicate the development process for an experience measure in a different context should consider whether a greater level of involvement from service users could improve the acceptability and completion rates, as sustainability and adoption within the technology-enabled service will be more likely to be achieved.

4. Conclusions

Developing an experience measure for an online community requires a multi-phased systematic process, its development should be informed and structured involving stakeholders. Different stakeholders can contribute to pieces of information leading to key decisions on the design and development of the instrument. A phased approach with multiple methodologies and careful selection of stakeholders for the appropriate time and stage of measure development is recommended. Delphi expert rounds and think-aloud protocols provide rich data that can influence the structure and construct validity of the measure. Further studies are needed to understand psychosocial factors and causal explanations for supportive online communities' outcomes, particularly outcomes related to mental health and wellbeing. Measurement of self-reported helpfulness or experience of community content may serve as an indicator for "therapeutic active engagement" in a digital service and help to understand the main reasons users benefit from these communities and its content, but further psychometric testing and evaluation is required. The pilot findings collected on outcomes is supported by previous literature on online supportive communities highlighting its importance to reduce isolation and enhancing support. Further research is needed to improve the acceptability of an experience measure, including a focus on how service user stakeholders can be involved to a greater extent throughout the development process. Studies looking to

replicate the structure and development of the POCEM measure in other digital contexts should consider the extent to which the development process and measure is replicable. By understanding the main outcomes and mechanisms in online mental health communities, digital healthcare providers and funders will be better placed to enable online peer support communities for mental health.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent from the participants' legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements.

Author contributions

CM, TH and AS led on the conception and design of the study and measure. HB and KJ contributed to the design of the measure, and GS led on the user testing phase of the study. CM, LM, and SO organized the database, ran the analysis, and contributed to the manuscript. SO, LM and CM provided the overall contributions to the manuscript in respective order. LS and TH wrote sections of the manuscript. AS was responsible for supervision and project administration. All authors contributed to the article and approved the submitted version.

Funding

This work has been funded as part of service improvement and innovation by Kooth Plc. There has been no external funding sought to carry out this work. As the funder of the project Kooth Plc provided the resource to plan for, collect and analyse data, and write the report for this study. The funder was not involved in the study design, collection, analysis, interpretation of data, the writing of this article or the decision to submit it for publication. All authors and funders declare no other competing interests.

Acknowledgments

Tom Kayll provided data extraction and database support. Joseph Lloyd helped provide safeguarding for the service user activities and took notes in the workshops. The product team at Kooth Plc contributed to the development and implementation of the measure. The Kooth clinical team and content team, led by Lynne Green and Lex Young respectively for guiding the development and provide with clinical expertise and supervision, specially thanks to Laura Lynch Becherer to support the review process and. Danielle Anwan and Chris Smith for providing technical support with the figures of the manuscript.

Conflict of interest

CM, LM, AS, SG, GS, HB, LS and KJ are researchers employed and receive honorarium by Kooth plc. TH has no conflict of interest. The funder remained independent it did not influence the design or outcome of the study. The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fdgth.2022.872404/full#supplementary-material.

References

1. Lusk B. Digital natives and social media behavior: an overview. *Prev Res.* (2010) 17:3–7. doi: 10.1037/e510232011-001. link.gale.com/apps/doc/A246449154/AONE?u=ucl_ttda&sid=googleScholar&xid=6eac2c71 (Accessed December 9, 2021]).

2. Gray NJ, Klein JD, Noyce PR, Sesselberg TS, Cantrill JA. Health information-seeking behaviour in adolescence: the place of the internet. *Soc Sci Med*. (2005) 60:1467–78. doi: 10.1016/J.SOCSCIMED.2004.08.010

3. Dooley BA, Fitzgerald A. My World Survey : National Study of Youth Mental Health in Ireland. Dublin, Ireland (2012). https://researchrepository.ucd.ie/handle/10197/4286 (Accessed December 9, 2021).

4. Coleman JS. The structure of society and the nature of social research. *Knowledge*. (2016) 1:333–50. doi: 10.1177/107554708000100301

5. Gowen K, Deschaine M, Gruttadara D, Markey D. Young adults with mental health conditions and social networking websites: seeking tools to build community. *Psychiatr Rehabil J*. (2012) 35:245–50. doi: 10.2975/35.3.2012.245.250

6. Liu D, Baumeister RF, Yang C, Hu B. Digital communication media use and psychological well-being: a meta-analysis. *J Comput Commun*. (2019) 24:259–73. doi: 10.1093/JCMC/ZMZ013

7. O'Keeffe GS, Clarke-Pearson K, Mulligan DA, Altmann TR, Brown A, Christakis DA, et al. The impact of social media on children, adolescents, and families. *Pediatrics*. (2011) 127:800–4. doi: 10.1542/PEDS.2011-0054

8. Frison E, Eggermont S. Toward an integrated and differential approach to the relationships between loneliness, different types of facebook use, and adolescents' depressed mood. *Communic Res*. (2015) 47:701–28. doi: 10.1177/0093650215617506

9. Chou HTG, Edge N. "They are happier and having better lives than I am": the impact of using facebook on perceptions of others' lives. *Cyberpsychol Behav Soc Netw*. (2012) 15:117–21. doi: 10.1089/CYBER.2011.0324

10. 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) 24:90–3. http://www.ncbi.nlm.nih.gov/pubmed/22447092 (Accessed December 9, 2021).

11. de Vries DA, Peter J, de Graaf H, Nikken P. Adolescents' social network site use, peer appearance-related feedback, and body dissatisfaction: testing a mediation model. *J Youth Adolesc*. (2016) 45:211–24. doi: 10.1007/S10964-015-0266-4/TABLES/2

12. Custers K, Van den Bulck J. Viewership of pro-anorexia websites in seventh, ninth and eleventh graders. *Eur Eat Disord Rev*. (2009) 17:214–9. doi: 10.1002/ERV.910

13. Kvardova N, Machackova H, Smahel D. The direct and indirect effects of online social support, neuroticism, and web content internalization on the drive for thinness among women visiting health-oriented websites. *Int J Environ Res Public Heal*. (2020) 17:2416. doi: 10.3390/IJERPH17072416

14. Setoyama Y, Yamazaki Y, Nakayama K. Comparing support to breast cancer patients from online communities and face-to-face support groups. *Patient Educ Couns*. (2011) 85:e95–e100. doi: 10.1016/J.PEC.2010.11.008

15. Bender JL, Katz J, Ferris LE, Jadad AR. What is the role of online support from the perspective of facilitators of face-to-face support groups? A multi-method study of the use of breast cancer online communities. *Patient Educ Couns*. (2013) 93:472–9. doi: 10.1016/J.PEC.2013.07.009

16. Andalibi N, Ozturk P, Forte A. *Sensitive self-disclosures, responses, and social support on Instagram: the case of #depression*. *Proc 2017 ACM conf comput support coop work soc comput*. doi: 10.1145/2998181

17. Lawlor A, Kirakowski J. Online support groups for mental health: a space for challenging self-stigma or a means of social avoidance? *Comput Human Behav*. (2014) 32:152–61. doi: 10.1016/J.CHB.2013.11.015

18. Love B, Crook B, Thompson CM, Zaitchik S, Knapp J, Lefebvre L, et al. Exploring psychosocial support online: a content analysis of messages in an adolescent and young adult cancer community. *Cyberpsychol Behav Soc Netw*. (2012) 15:555–9. doi: 10.1089/CYBER.2012.0138

19. Saha K, Sharma A. Causal factors of effective psychosocial outcomes in online mental health communities. *Proc Int AAAI Conf Web Soc Media*. (2020) 14:590–601. doi: 10.1101/2020.08.15.20175836 https://ojs.aaai.org/index.php/ICWSM/article/view/7326 (Accessed December 10, 2021).

20. Whitlock JL, Powers JL, Eckenrode J. The virtual cutting edge: the internet and adolescent self-injury. *Dev Psychol*. (2006) 42:407–17. doi: 10.1037/0012-1649.42.3.407

21. Horne J, Wiggins S. Doing being "on the edge": managing the dilemma of being authentically suicidal in an online forum. *Sociol Health Illn*. (2009) 31:170–84. doi: 10.1111/J.1467-9566.2008.01130.X

22. Murray CD, Fox J. Do internet self-harm discussion groups alleviate or exacerbate self-harming behaviour? *Aust e-J Adv Ment Heal*. (2006) 5:225–33. doi: 10.5172/JAMH.5.3.225

23. Ellis LA, Campbell AJ, Sethi S, O'Dea BM. Comparative randomized trial of an online cognitive-behavioral therapy program and an online support group for depression and anxiety. *J Cyber Ther Rehabil*. (2011) 4:461–7. doi: 10.2/JQUERY.MIN.JS

24. Low KG, Charanasomboon S, Lesser J, Reinhalter K, Martin R, Jones H, et al. Effectiveness of a computer-based interactive eating disorders prevention program at long-term follow-up. *J Treat Prev*. (2006) 14:17–30. doi: 10.1080/10640260500403816

25. Horgan A, McCarthy G, Sweeney J. An evaluation of an online peer support forum for university students with depressive symptoms. *Arch Psychiatr Nurs*. (2013) 27:84–9. doi: 10.1016/J.APNU.2012.12.005

26. Jones R, Sharkey S, Ford T, Emmens T, Hewis E, Smithson J, et al. Online discussion forums for young people who self-harm: user views. *Psychiatrist*. (2011) 35:364–8. doi: 10.1192/pb.bp.110.033449

27. Prescott J, Rathbone AL, Brown G. Online peer to peer support: qualitative analysis of UK and US open mental health Facebook groups. *Digit Heal*. (2020) 6:2055207620979209. doi: 10.1177/2055207620979209

28. Andy A. Studying how individuals who express the feeling of loneliness in an online loneliness forum communicate in a nonloneliness forum: observational study. *JMIR Form Res*. (2021) 5:e28738. doi: 10.2196/28738

29. Ray S, Kim SS, Morris JG. The central role of engagement in online communities. *Inf Syst Res*. (2014) 25:528–46. doi: 10.1287/ISRE.2014.0525

30. Donabedian A. The quality of care: how can it be assessed? *JAMA*. (1988) 260:1743–8. doi: 10.1001/JAMA.1988.03410120089033

31. Hanley T, Sefi A, Grauberg J, Prescott J, Etchebarne A. A theory of change for web-based therapy and support services for children and young people: collaborative qualitative exploration. *JMIR Pediatr Parent*. (2021) 4(10):e23193. doi: 10.2196/23193

32. Donabedian A. Evaluating the quality of medical care. *Milbank Q*. (2005) 83:691. doi: 10.1111/J.1468-0009.2005.00397.X

33. Avison DE, Lau F, Myers MD, Nielsen PA. Action research. *Commun ACM*. (1999) 42:94–7. doi: 10.1145/291469.291479

34. MacDonald C. Understanding participatory action research: a qualitative research methodology option. *Can J Action Res*. (2012) 13:34–50. doi: 10.33524/CJAR.V13I2.37

35. Gustafson DH, Shukla RK, Delbecq A, Walster GW. A comparative study of differences in subjective likelihood estimates made by individuals, interacting groups, Delphi groups, and nominal groups. *Organ Behav Hum Perform*. (1973) 9:280–91. doi: 10.1016/0030-5073(73)90052-4

36. Wright B, Phillips H, Le Couteur A, Sweetman J, Hodkinson R, Ralph-Lewis A, et al. Modifying and validating the social responsiveness scale edition 2 for use with deaf children and young people. *PLoS One*. (2020) 15:e0243162. doi: 10.1371/JOURNAL.PONE.0243162

37. Zelmer J, van Hoof K, Notarianni M, van Mierlo T, Schellenberg M, Tannenbaum C. An assessment framework for e-mental health apps in Canada: results of a modified Delphi process. *JMIR mHealth UHealth*. (2018) 6(7):e10016. doi: 10.2196/10016

38. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol*. (2006) 3:77–101. doi: 10.1191/1478088706QP063OA

39. Sketch.com. https://www.sketch.com/about-us/ (Accessed December 10, 2021).

40. Lookback.io. https://www.lookback.com/ (Accessed December 10, 2021).

41. Larsen LB, Øvad T, Nielsen L, Larusdottir M. Remote user testing—experiences and trends. In: C Ardito, R Lanzilotti, A Malizia, H Petrie, A Piccinno, G Desolda, K Inkpen, editors. *INTERACT 2021: Human-computer interaction*. Cham: Springer (2021). p. 579–83. doi: 10.1007/978-3-030-85607-6_81

42. Joe J, Chaudhuri S, Le T, Thompson H, Demiris G. The use of think-aloud and instant data analysis in evaluation research: exemplar and lessons learned. *J Biomed Inform*. (2015) 56:284–91. doi: 10.1016/J.JBI.2015.06.001

43. Pernice K. *Affinity diagramming: collaboratively sort UX findings & design ideas*. California, USA: Nielsen Norman Group (2018) https://www.nngroup.com/articles/affinity-diagram/ (Accessed December 13, 2021).

44. Lucero A. Using affinity diagrams to evaluate interactive prototypes. In Abascal J, Barbosa S, Fetter M, Gross T, Palanque P, Winckler M, editors. *INTERACT 2015: Human-computer interaction*. Cham: Springer (2015). p. 231–248. doi: 10.1007/978-3-319-22668-2_19

45. Hardesty DM, Bearden WO. The use of expert judges in scale development. Implications for improving face validity of measures of unobservable constructs. *J Bus Res*. (2004) 57:98–107. doi: 10.1016/S0148-2963(01)00295-8

46. Lavela SL, Gallan AS. Evaluation and measurement of patient experience. *Patient Exp J*. (2014) 1:28–36. doi: 10.35680/2372-0247.1003

47. Mohr DC, Lyon AR, Lattie EG, Reddy M, Schueller SM. Accelerating digital mental health research from early design and creation to successful implementation and sustainment. *J Med Internet Res*. (2017) 19. doi: 10.2196/jmir.7725

48. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL. Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Front Public Heal*. (2018) 6:149. doi: 10.3389/FPUBH.2018.00149/XML/NLM

49. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Adm Policy Ment Health*. (2011) 38:65. doi: 10.1007/S10488-010-0319-7

50. Raphiphatthana B, Sweet M, Puszka S, Whitty M, Dingwall K, Nagel T. Evaluation of electronic mental health implementation in Northern territory services using the integrated "promoting action on research implementation in health services" framework: qualitative study. *JMIR Ment Heal*. (2020) 7(5): e14835. doi: 10.2196/14835

51. Finch TL, Mair FS, O'donnell C, Murray E, May CR. From theory to "measurement" in complex interventions: Methodological lessons from the development of an e-health normalisation instrument. (2012). http://www.biomedcentral.com/1471-2288/12/69.

52. Fink A, Kosecoff J, Brook RH. Consensus Methods: Characteristics and Guidelines for Use. (1984).

53. Fernández-Domínguez JC, Sesé-Abad A, Morales-Asencio JM, Sastre-Fullana P, Pol-Castañeda S, De Pedro-Gómez JE. Content validity of a health science evidence-based practice questionnaire (HS-EBP) with a web-based modified Delphi approach. *Int J Qual Heal Care*. (2016) 28:764–73. doi: 10.1093/INTQHC/MZW106

54. Zartha Sossa JW, Halal W, Hernandez Zarta R. Delphi method: analysis of rounds, stakeholder and statistical indicators. *Foresight*. (2019) 21:525–44. doi: 10.1108/FS-11-2018-0095/FULL/PDF

55. Walker M, Takayama L, Landay JA. *High-fidelity or low-fidelity, paper or computer? Choosing attributes when testing web prototypes*. *Proceedings of the human factors and ergonomics society annual meeting*. Los Angeles, CA: SAGE PublicationsSage CA (2016). p. 661–5. doi: 10.1177/154193120204600513

56. Ghasemifard N, Shamsi M, Reza A, Kenar R, Ahmadi V, Kenari R. A New View at Usability Test Methods of Interfaces for Human Computer Interaction. (2015).

57. de Ossorno Garcia S, Salhi L, Sefi A, Hanley T. The session wants and need outcome measure: the development of a brief outcome measure for single-sessions of web-based support. *Front Psychol*. (2021) 12:4900. doi: 10.3389/FPSYG.2021.748145/BIBTEX

58. Gonzalez J, Trickett EJ. Collaborative measurement development as a tool in CBPR: measurement development and adaptation within the cultures of communities. *Am J Community Psychol*. (2014) 54:112–24. doi: 10.1007/s10464-014-9655-1

59. Vesely S, Klöckner CA. Social desirability in environmental psychology research: three meta-analyses. *Front Psychol*. (2020) 11:1395. doi: 10.3389/FPSYG.2020.01395/BIBTEX

60. Kuru O, Pasek J. Improving social media measurement in surveys: avoiding acquiescence bias in Facebook research. *Comput Human Behav*. (2016) 57:82–92. doi: 10.1016/J.CHB.2015.12.008

61. Joinson A. Social desirability, anonymity, and intemet-based questionnaires. *Behav Res Methods Instrum Comput*. (1999) 31:433–8. doi: 10.3758/bf03200723

62. King MF, Bruner GC. Social desirability bias: a neglected aspect of validity testing. *Psychol Mark*. (2000) 17:79–103. doi: 10.1002/(SICI)1520-6793(200002)17:2<79::AID-MAR2>3.0.CO;2-0

63. Preece J, Nonnecke B, Andrews D. The top five reasons for lurking: improving community experiences for everyone. *Comput Human Behav*. (2004) 20:201–23. doi: 10.1016/j.chb.2003.10.015

64. Xu X, Xia L, Zhang Q, Wu S, Wu M, Liu H. The ability of different imputation methods for missing values in mental measurement questionnaires. *BMC Med Res Methodol*. (2020) 20(1):1–9. doi: 10.1186/S12874-020-00932-0

65. Coulson NS, Smedley R, Bostock S, Kyle SD, Gollancz R, Luik A, et al. The pros and cons of getting engaged in an online social community embedded within digital cognitive behavioral therapy for insomnia: survey among users. *J Med Internet Res*. (2016) 18(4):e5654. doi: 10.2196/JMIR.5654

66. White M, Dorman SM. Receiving social support online: implications for health education. *Health Educ Res*. (2001) 16:693–707. doi: 10.1093/HER/16.6.693

67. Gulledge CM, Grace Smith D, Ziedas A, Muh SJ, Moutzouros V, Makhni EC. Floor and ceiling effects, time to completion, and question burden of PROMIS CAT domains among shoulder and knee patients undergoing nonoperative and operative treatment. *JBJS Open Access*. (2019) 4(4):e0015.1–7. doi: 10.2106/JBJS.OA.19.00015

68. Hsu C-C, Sandford BA. The Delphi technique: making sense of consensus. *Pract Assessment Res Eval*. (2007) 12:10. doi: 10.7275/pdz9-th90

frontiers | Frontiers in Psychiatry

TYPE Original Research
PUBLISHED 20 October 2022
DOI 10.3389/fpsyt.2022.888524

OPEN ACCESS

EDITED BY
Brenna Renn,
University of Nevada, Las Vegas,
United States

REVIEWED BY
Arkers Kwan Ching Wong,
Hong Kong Polytechnic University,
Hong Kong SAR, China
Djawid Hashemi,
Charité – Universitätsmedizin Berlin,
Germany

*CORRESPONDENCE
Robert P. Nolan
rnolan@uhnres.utoronto.ca

SPECIALTY SECTION
This article was submitted to
Digital Mental Health,
a section of the journal
Frontiers in Psychiatry

RECEIVED 03 March 2022
ACCEPTED 29 September 2022
PUBLISHED 20 October 2022

CITATION
Fezza GC, Sansone S and Nolan RP
(2022) Therapeutic components
of digital counseling for chronic heart
failure.
Front. Psychiatry 13:888524.
doi: 10.3389/fpsyt.2022.888524

Therapeutic components of digital counseling for chronic heart failure

Gabriel C. Fezza[1,2], Stephanie Sansone[1,2] and Robert P. Nolan[1,3,4,5]*

[1]Behavioral Cardiology Research Unit, University Health Network (UHN), Toronto, ON, Canada, [2]Faculty of Health, York University, Toronto, ON, Canada, [3]Faculty of Medicine, University of Toronto, Toronto, ON, Canada, [4]Peter Munk Cardiac Centre, Toronto General Hospital Research Institute, University Health Network, Toronto, ON, Canada, [5]Ted Rogers Centre for Heart Research, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

Background: Task force statements support the use of cognitive behavioral therapy (CBT) and motivational interviewing (MI) to promote self-care in chronic heart failure (CHF) patients. Digital counseling interventions have the potential to complement conventional programs. However, therapeutic components of digital programs associated with improved outcomes are not clearly established.

Objective: Identify therapeutic components of the Canadian e-Platform to Promote Behavioral Self-Management in Chronic Heart Failure (CHF-CePPORT) protocol that were associated with improved health-related quality of life (HRQL).

Materials and methods: Ordinal logistic regression was used to identify therapeutic components of the CHF-CePPORT protocol. The primary outcome was the 12-month Kansas City Cardiomyopathy Questionnaire Overall Summary (KCCQ-OS) tertile. Logistic regressions determined the association between 12-month KCCQ-OS tertile, using logon hours for key segments of the protocol, modality of content delivery, and clinical themes.

Results: A total of 117 patients were enrolled in the e-Counseling arm of the CHF-CePPORT trial. Median age was 60 years (IQR 52–69). Total logon hours in the initial 4-month segment of CHF-CePPORT (Sessions 1–16) was associated with increased 12-month KCCQ-OS tertile (Odds Ratio, OR = 1.31, 95% CI, 1.1–1.5, P = 0.001). Within sessions 1–16, improved KCCQ-OS was associated with logon hours for self-assessment tools/trackers (OR = 1.49, 95% CI, 1.1–2.0, P = 0.007), and videos (OR = 1.57, 95% CI, 1.03–2.4, P = 0.04), but not for CHF information pages.

Conclusion: This study highlights the importance of using evidence-based guidelines from CBT and MI as core components of digital counseling, delivered through videos and interactive tools/trackers, to improve HRQL with CHF.

KEYWORDS

cognitive behavioral therapy, digital health, eHealth, heart failure, motivational interviewing, quality of life, self-care, telemedicine

Introduction

International task force statements from professional cardiovascular health societies emphasize the importance of patient self-care in the management of chronic heart failure (CHF) (1, 2). Engagement in self-care behaviors has been shown to have a positive effect on health-related quality of life (HRQL) and to reduce rates of CHF-related mortality and hospitalization (3). These task force statements endorse the use of behavioral counseling to improve CHF self-care and health status (1, 2). Protocols of behavioral counseling that promote CHF self-care have not been clearly established. However, key components of conventional face-to-face programs are consistent with well-established evidence-based models of counseling that include cognitive behavioral therapy (CBT) and motivational interviewing (MI) (1, 2, 4, 5).

In this period of the COVID-19 pandemic, the maintenance, monitoring, and management of self-care behaviors among patients with CHF has become increasingly important to reduce hospital readmissions and to maintain health status (2). Digital health interventions have become increasingly prominent (2, 6), and utilized to support patient CHF self-care (7). These digital interventions have the potential to complement conventional clinic-based treatment programs for CHF patients in a manner that is efficacious, accessible, and replicable (6, 8).

Digitally based counseling programs have been reported to improve HRQL in patients with CHF. Different modalities of digitally based counseling programs include telemonitoring, video monitoring, and home telehealth. To our knowledge, a detailed analysis has not been conducted of therapeutic benefit associated with individual components within these programs (9). Patient engagement in logging onto these programs is in the range of 55–62% (10, 11), which is similar to engagement rates for completing sessions in conventional behavioral programs (12). This moderate level of patient engagement highlights the need to specify therapeutic components of digital counseling for CHF patients. This in turn will increase the likelihood of improving the replicability and standardization of these programs. A recent policy paper for digital health highlights effective features of patient-centered models of care (13). Examples of effective features include goal setting, having a concrete behavioral goal for change, and an ability to monitor your progress. While those guidelines advocate the use of evidence-based models of counseling, they do not specify how these features are integrated into these models of digital counseling (14). The present study was undertaken to identify therapeutic components of an automated digital counseling program that were associated with improved HRQL in the Canadian e-Platform to Promote Behavioral Self-Management in Chronic Heart Failure trial (CHF-CePPORT: ClinicalTrials.gov NCT01864369) (15).

Materials and methods

Participants

This study focused on patients who were randomized to the e-Counseling + Usual Care intervention arm of the CHF-CePPORT trial. Patient recruitment began in January 2014 and the final 12-month assessment was completed in February 2018 (15). Informed written consent was given by participants during enrolment to participate in the trial. Inclusion criteria for CHF-CePPORT consisted of patients $\geq$ 18 years of age, with New York Heart Association (NYHA) class I–III and left ventricular ejection fraction (LVEF) $\leq$ 45. Patients were required to be stable for 12 months prior to enrolment with no worsening of CHF for 1 month prior to enrolment, as determined by the referring cardiologist.

Patients were included if they were not currently enrolled in a formal exercise program, had comprehension of English or French, and provided informed written consent. Patients were excluded if they had current symptomatic hypotension, persistent systolic or diastolic hypertension, or clinically significant comorbidities (e.g.,

CHF-CePPORT AUTOMATED DIGITAL COUNSELING PROTOCOL		
DIGITAL PROGRAM SEGMENT	**SESSION#**	**CLINICAL CONTENT THEMES**
1st Segment: Months 1-4 (Weekly)	1	**CHF-HRQL & CHF Self-Care**
	2	
	3	
	4	
	5	**Motivational Strategies**
	6	
	7	
	8	
	9	**Cognitive-Behavioural Self-Care Guide: Exercise, Diet, Fluid and Sodium Restrictions, Medication Adherence, and Smoke-Free Living**
	10	
	11	
	12	
	13	
	14	
	15	**Review of HRQL and CHF Self-Care Connection**
	16	
2nd Segment: Months 5-8 (Biweekly)	17	**Psychological Well-Being**
	18	
	19	
	20	**Social Factors and CHF Self-Care**
	21	
	22	
	23	
	24	
3rd Segment: Months 9-12 (Monthly)	25	**Maintenance of CHF Self-Care Lifestyle**
	26	
	27	
	28	

FIGURE 1
CHF-CePPORT digital counseling protocol. CHF, chronic heart failure; HRQL, health-related quality of life.

cancer, chronic kidney failure). Patients were also excluded if they were diagnosed with a major psychiatric disorder (e.g., psychosis).

Study design

This investigation was a sub-study of CHF-CePPORT, which has been reported previously for both the protocol and primary outcome (15, 16). Briefly, this trial was a phase two, multi-center randomized controlled trial with a two-parallel group, double blind design, and with repeated assessments at baseline, 4- and 12-months (15, 16). CHF-CePPORT was designed to evaluate the efficacy of an evidence-based and clinically organized e-counseling protocol that promoted adherence to recommended guidelines for exercise, diet, prescribed medications, and smoke-free living over a 12-month period. Eligible CHF patients were recruited across three Canadian sites: University Health Network (Toronto), Providence Health Care (Vancouver), and the Ottawa Heart Institute. Patient recruitment was voluntary and the content of CHF-CePPORT was complementary to usual care. It was introduced to participants as a as a research study. All participants were randomly assigned to either e-Counseling + Usual Care or e-Info Control + Usual Care. The primary endpoint of CHF-CePPORT was 12-month quality of life as assessed by the Kansas City Cardiomyopathy Questionnaire Overall Summary (KCCQ-OS) (17).

Study interventions and assessments

The automated counseling protocol for the e-Counseling arm of CHF-CePPORT has been previously described (15, 16). Briefly, the protocol was organized by 28 sessions that were sent to patients proactively *via* an email that contained a URL to the webpages for each randomized group. Emails were sent to patients weekly for months 1–4, bi-weekly for months 5–8, and monthly for months

TABLE 1 Background characteristics of e-Counseling + usual care arm of CHF-CePPORT trial (n = 117).

Characteristics	n, median	%, (IQR)
Declared gender, female	24	20.5
Age, years	60.0	(52–69)
Education:		
≤ Secondary	33	28.2
Post-secondary	84	71.8
LVEF%:		
< 35	57	48.7
35–40	31	26.5
41–45	29	24.8
NYHA functional class:		
1	45	41.3
2	48	44.0
3	16	14.7

LVEF, left ventricular ejection fraction; NYHA, New York Heart Association.

9–12. The Control intervention included session content that was based on an amalgamation of publicly available educational information on guidelines for self-managing CHF from the Canadian Heart Failure Association, American Heart Association, and the European Society of Cardiology. The Control and e-Counseling sessions included information aimed at improving self-help skills for adhering to recommended self-care behaviors for medications, exercise, fruit and vegetable intake, restriction of sodium and fluids, and smoke-free living. The e-Counseling intervention promoted adherence to self-care behaviors by utilizing core components of MI and CBT using different digital modes of presentation: information pages (comprised of narrative script with illustrations), interactive self-assessment tools/trackers, and videos (expert guidelines for self-care, dramatic vignettes, and peer discussion). For the e-Counseling sessions, key features from MI helped patients build their readiness for change through validating their stage of readiness and guiding them to identify goals for lifestyle change that were connected to their priorities for living well. In addition, core components of CBT provided a step-by-step guide to plan and initiate self-care behavior change, and patient efficacy was reinforced using performance-based feedback through interactive self-monitoring tools (e.g., self-assessment forms and interactive trackers) (18).

TABLE 3 Association between logon time for digital counseling modalities in sessions 1–16 and 12-month KCCQ-OS tertile.

Modality of content delivery	OR	95% CI	P-value
Information/education pages	1.01	(0.5–2.2)	0.80
Videos: Expert guideline, dramatic vignettes, and peer discussion	1.57	(1.03–2.4)	0.04
Self-assessment tools/trackers	1.49	(1.1–2.0)	0.007

Each OR, odds ratio corresponds to a separate ordinal logistic regression analysis.

Health-related quality of life outcomes

Health-related quality of life was assessed using the KCCQ-OS score at baseline, 4-, and 12-months. The KCCQ-OS incorporated patient reported symptoms of CHF, physical limitations, social function, and quality of life (17).

Statistical analysis

Mann–Whitney U tests were used to analyze continuous background variables, and Chi Square tests were used for categorical variables. The 12-month KCCQ-OS scores were transformed into tertiles (range, ≤ 74, 74.1–90.4, and 90.5–100) due to the severe skewness of scores and a clinically meaningful ceiling effect at baseline as detailed in the CHF-CePPORT primary outcome paper (15). The tertile ranges reflect fair, good, and excellent health status respectively. Scores in the higher range (two upper tertiles) of the KCCQ are well-established to predict decreased levels of morbidity and mortality in the CHF population (19).

Total patient logon time (hours) was used in ordinal logistic regression analyses to determine the components of the protocol that were associated with higher KCCQ-OS tertile at 12 months. These components included: program segment over 12 months, modality of content accessed by patients, and clinical content themes in logon sessions (**Figure 1**). All analyses were controlled for baseline KCCQ-OS tertiles, age, LVEF (< 35, 35–40, 41–45).

TABLE 2 Association between logon hours for CHF-CePPORT program segments and 12-month KCCQ-OS tertile.

Segment of program	Session	Schedule of contact	OR	95% CI	P-value
Month 1–4	Sessions 1–16	Weekly	1.31	(1.1–1.5)	0.001
Month 5–8	Sessions 17–24	Biweekly	1.26	(0.9–1.7)	0.13
Month 9–12	Sessions 25–28	Monthly	1.42	(0.8–2.7)	0.28

Each OR, odds ratio corresponds to a separate ordinal logistic regression analysis.

TABLE 4 Association between logon time for clinical content themes in sessions 1–16 and 12-month KCCQ-OS tertile.

Session	Clinical theme	Modality of content delivery	OR	95% CI	*P*-value
Session 5–8	Motivational interviewing	Self-assessment tools/trackers	39.9	(1.1–1413.0)	0.04
Session 9–14	CBT guide for CHF self-care	Videos: Expert guide and dramatic vignettes	464.7	(3.2–66778)	0.02
		Self-assessment tools/trackers	108.7	(2.1–5493.9)	0.02
Sessions 15–16	HRQL and self-care maintenance	Self-assessment tools/trackers	5.69	(1.5–22.2)	0.01

Each OR, odds ratio corresponds to a separate ordinal logistic regression analysis. CBT, cognitive behavioral therapy; CHF, chronic heart failure; HRQL, health-related quality of life.

The analyses were planned in a successive order according to the following objectives to identify the components of the program that were positively associated with higher KCCQ-OS tertile scores:

1. the association between 12-month KCCQ-OS tertile and logon time (hours) for the three successive periods of the CHF-CePPORT protocol: Baseline to 4 months, 4–8 months, and 8–12-months. These three time periods were characterized by proactive contact with patients on a weekly, biweekly, and monthly schedule, respectively.
2. the association between logon time for mode of content accessed by patients (information, video, or interactive tools/trackers) in automated digital counseling sessions and 12-month KCCQ-OS tertile.
3. the association between logon time for clinical themes such as quality of life, self-care behavior, and social functioning presented by digital counseling sessions and 12-month KCCQ-OS tertile.

Results

Patient characteristics

Chronic heart failure patients (n = 117) in this study received the e-Counseling intervention in the CHF-CePPORT trial. Median age was 60 years (IQR 52–69), who received the patient-centered e-counseling protocol (**Table 1**).

Analysis 1: Association between CHF-CePPORT program segment and 12-month KCCQ-OS

Total logon time (in hours) on the CHF-CePPORT platform during the initial 4 months of the program (sessions 1–16), when patient access to new digital sessions was scheduled weekly, was positively associated with the 12-month KCCQ-OS tertile (P = 0.001). Total logon time (in hours) was not associated with 12-month KCCQ-OS tertile for the second and third segments of the program, when access to new digital sessions was scheduled biweekly (sessions 17–24), and monthly (sessions 25–28), respectively (**Table 2** and **Supplementary Table 1** for details).

Analysis 2: Modality of digital counseling (information, videos, and tools/trackers) and 12-month KCCQ-OS

The association between digital counseling modality and 12-month KCCQ-OS was limited to the initial program segment (sessions 1–16), given the finding noted immediately above. The modality of content delivery that was associated with higher 12-month KCCQ-OS tertile were videos (P = 0.04) and interactive tools/trackers (P = 0.007). Patient logon time for digital pages that provided educational information on CHF was not associated with higher 12-month KCCQ-OS tertile (**Table 3** and **Supplementary Table 2** for details).

Analysis 3: Clinical content themes associated with improved 12-month KCCQ-OS

The association between digital counseling content themes and 12-month KCCQ-OS was limited to the digital counseling modalities of videos and interactive tools/trackers during sessions 1–16, due to the above findings. The clinical theme that focused on MI to improve patient readiness for change (sessions 5–8) was associated with higher 12-month KCCQ-OS tertile scores when patients utilized tools/trackers (P = 0.04). Within the program segment that focused on CBT guidelines for CHF self-care behaviors (sessions 9–14) patient logon time was associated with higher 12-month KCCQ tertile scores on digital pages that presented multimedia videos (P = 0.02), as well as tools/trackers (P = 0.02). In sessions that enabled patients to review the connection between their HRQL and CHF self-care maintenance (sessions 15–16), logon time was associated with higher 12-month KCCQ-OS tertile scores (P = 0.01) – **Table 4** and **Supplementary Table 3** for details.

Discussion

The objective of this study was to identify therapeutic components of the digital counseling arm of the CHF-CePPORT trial that were positively associated with the 12-month KCCQ-OS endpoint (17). CHF-CePPORT used core components of behavioral counseling from CBT and MI that were delivered through three digital modalities: interactive tools/trackers, videos, and information pages. Due to the clinical organization of CHF-CePPORT, we were able to readily identify therapeutic components over the 12-month period of the intervention. The initial 4-month segment of CHF-CePPORT (sessions 1–16) was associated with improved HRQL. During this period participants were sent a digital link to the intervention on a weekly basis, which may have reinforced a sustained pattern of engagement with digital counseling resources.

The median number of total sessions that patients accessed in CHF-CePPORT was 17 (61% of the full protocol of 28 sessions). Therefore, it appears that patient engagement in the initial 16 sessions accounted for most of the therapeutic effect in CHF-CePPORT. This level of patient engagement has been observed in other trials of digital health as a threshold that is associated with improved clinical outcomes (11). The present study provides a more granular analysis of this therapeutic effect. Patient logon time with core components of behavioral counseling delivered through interactive tools/trackers and videos was associated with higher HRQL, but this outcome was not observed for patient engagement with digital information pages (comprised of narrative script and illustrations). This finding raises a potential concern since conventional patient education websites for CHF are largely comprised of digital pages that are filled with narrative scripts and illustrations.

The present results are also consistent with findings from a systematic review, in which patient adherence to CHF self-care behavior was significantly enhanced with the use of video interventions (20). In CHF-CePPORT, the videos were designed to engage patients more holistically with dramatic vignettes, expert summaries of self-help tips, and peer discussion about self-care. It remains to be determined whether the therapeutic effect of these videos was attributable to features such as positive role modeling, comments from health professionals that validated patient efforts at lifestyle change, or the dynamic presentation of explicit guidelines for CHF self-care. With this, our study adds to the policy recommendations observed, by promoting evidence-based features within cognitive behavioral and motivational interviewing models of counseling (13).

A recent review of mobile health technologies for patients with CHF highlights the lack of sustained patient engagement with these interventions as a clinically challenging issue (21). This was supported by evidence from a separate systematic review of mobile health interventions for CHF, where program usage was observed to be consistently low, with some studies reporting attrition rates of 30–60% (22). Understanding the ways in which patients engage with digital health interventions over clinically meaningful time intervals is a priority for current research.

Some strategies have shown promise in improving both treatment efficacy and sustained patient engagement with digital interventions for health behavior change. An early meta-analytic study reported that outcomes were improved with dynamic tailoring that matched program goals with the participant's reported priorities for behavior change across repeated assessments. Dynamic tailoring with iterative feedback to patients, evoked greater treatment effects that remained significant in outcome assessments beyond 12-months (23). Current task force statements on digital counseling have not included explicit guidelines for dynamic tailoring (24), due in part to the limited availability of evidence. Nevertheless, the application of tailoring strategies in digital health has a clear potential to ensure that protocols for counseling and patient education are grounded within patient-centered goals for improved health status and quality of life.

The use of digital tailoring strategies to enhance the efficacy and usability of digital health interventions may become more prevalent with the emerging role of machine learning (ML) models in precision care for CHF. Predictive modeling based on ML is well-suited to identify how components of digital counseling programs can interface effectively with patient preferences. These preferences may be shaped by background attributes (e.g., socioeconomic status, education level, or severity of medical condition), health literacy level, motivation and skill for learning self-care behavior, and quality of social support for sustaining a lifestyle characterized by CHF self-care (25). The method of analysis used in this secondary study of CHF-CePPORT could be enhanced with the use of ML modeling.

Consistent with previous taskforce statements (1), it may be possible to better standardize and replicate positive outcomes from digital counseling programs when a theoretical framework is specified. In a recent meta-analysis on digital health interventions to manage hypertension, only 25% of trials identified a behavioral counseling model in their protocol (26). Moreover, heterogeneity in trial outcomes was significantly reduced and the treatment outcome was significantly improved among digital programs that specified a behavioral counseling model.

Study limitations

Findings of the present study were based on outcomes from the digital counseling arm of the CHF-CePPORT trial, which limits the generalizability of our results (15). The digital counseling protocol in CHF-CePPORT was organized according to a pre-set sequence for scheduling patient access to program components, and this feature may differentiate the present

digital counseling protocol from other digital programs of CHF self-care. Additionally, our analysis of patient usage of the various components of the CHF-CePPORT trial was limited to the segments of the trial (months 1–4, 5–8, and 9–12), the type of delivery (videos vs. tools/trackers vs. information pages), and the clinical themes. We were unable to provide a granular analysis of the specific types of videos (dramatic vignettes, expert summaries of self-help tips, peer discussion on self-care behavior and quality of life), interactive tools and trackers, or information pages that were utilized by patients in the digital counseling group. As reported in the primary outcome paper for CHF-CePPORT (15), enrolled patients presented with elevated baseline scores for the KCCQ-OS. Due to this ceiling effect, the primary analysis of CHF-CePPORT was not able to properly test whether the program was able to improve KCCQ outcomes over 12 months. Therefore, the primary outcome for CHF-CePPORT was a null finding. However, the follow-up analyses showed that the association between usage and 12-month improvement was significant for the treatment group but there was no association for the control group. This helps us to understand more clearly the potential therapeutic components for the intervention which are advisable to incorporate into subsequent trials. Furthermore, therapeutic components of digital counseling that were identified in this study may not be fully applicable to a sample of CHF patients that have greater impairment in health status. In addition, our sample had a positive balance between males and females; however, it was primarily Caucasian, and it did not include a large representation of individuals with low income. Further, education was elevated to a post-secondary level, which also affects the generalizability of our findings.

Conclusion

This sub-study of the CHF-CePPORT trial was conducted to specify therapeutic components of an automated digital counseling program for CHF self-care. Increased KCCQ-OS at 12 months was associated with logon time in the initial 4 months for videos and interactive tools/trackers that delivered key components of CBT and MI. These results confirm the importance of using evidence-based models of behavioral counseling to promote CHF self-care and HRQL in a digital counseling program. In sum, the present results highlight the need to develop a sophisticated analytic strategy (e.g., with ML modeling) to identify therapeutic components of digital counseling, and in turn improve the standardization and replicability of digital interventions for CHF.

Data availability statement

The datasets presented in this article are not readily available because the data are available from the corresponding author pending approval of Research Ethics Boards of participating institutions and on reasonable request received from qualified researchers trained in human subject confidentiality protocols. Requests to access the datasets should be directed to RN.

Ethics statement

The studies involving human participants were reviewed and approved by the Research Ethics Board, University Health Network. The patients/participants provided their written informed consent to participate in this study.

Author contributions

GF and SS: conceptualization, methodology, formal analysis, and writing – original draft, review, and editing. RN: conceptualization, methodology, formal analysis, writing – original draft, review, and editing, investigation (data contributed from the CHF-CePPORT trial), and supervision. All authors contributed to the article and approved the submitted version.

Acknowledgments

We would like to acknowledge the contribution of Luba Cermakova, Saba Vafaei-Nodeh, and Haya Abuzuluf from the CHF-CePPORT team of research associates in the collection of data that was used for this sub-study.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyt.2022.888524/full#supplementary-material

References

1. Riegel B, Moser DK, Anker SD, Appel LJ, Dunbar SB, Grady KL, et al. State of the science: promoting self-care in persons with heart failure: a scientific statement from the american heart association. *Circulation.* (2009) 120:1141–63. doi: 10.1161/CIRCULATIONAHA.109.192628

2. Jaarsma T, Hill L, Bayes-Genis A, La Rocca HPB, Castiello T, Èelutkienë J, et al. Self-care of heart failure patients: practical management recommendations from the heart failure association of the European society of cardiology. *Eur J Heart Fail.* (2021) 23:157–74. doi: 10.1002/ejhf.2008

3. Lee CS, Bidwell JT, Paturzo M, Alvaro R, Cocchieri A, Jaarsma T, et al. Patterns of self-care and clinical events in a cohort of adults with heart failure: 1 year follow-up. *Hear Lung.* (2018) 47:40–6. doi: 10.1016/j.hrtlng.2017.09.004

4. Peng Y, Fang J, Huang W, Qin S. Efficacy of cognitive behavioral therapy for heart failure: a meta-analysis of randomized controlled studies. *Int Heart J.* (2019) 60:665–70. doi: 10.1536/ihj.18-408

5. Masterson Creber R, Patey M, Lee CS, Kuan A, Jurgens C, Riegel B. Motivational interviewing to improve self-care for patients with chronic heart failure: MITI-HF randomized controlled trial. *Patient Educ Couns.* (2016) 99:256–64. doi: 10.1016/j.pec.2015.08.031

6. Farwati M, Riaz H, Tang WHW. Digital health applications in heart failure: a critical appraisal of literature. *Curr Treat Options Cardiovasc Med.* (2021) 23:885. doi: 10.1007/s11936-020-00885-z

7. Kitsiou S, Vatani H, Paré G, Gerber BS, Buchholz SW, Kansal MM, et al. Effectiveness of mobile health technology interventions for patients with heart failure: systematic review and meta-analysis. *Can J Cardiol.* (2021) 2021:15. doi: 10.1016/j.cjca.2021.02.015

8. Manteghinejad A, Javanmard S. Challenges and opportunities of digital health in a post COVID19 world. *J Res Med Sci.* (2021) 26:2021–6. doi: 10.4103/jrms.JRMS_1255_20

9. Bashi N, Karunanithi M, Fatehi F, Ding H, Walters D. Remote monitoring of patients with heart failure: an overview of systematic reviews. *J Med Internet Res.* (2017) 19:1–14. doi: 10.2196/jmir.6571

10. Beishuizen CRL, Stephan BCM, Van Gool WA, Brayne C, Peters RJG, Andrieu S, et al. Web-based interventions targeting cardiovascular risk factors in middle-aged and older people: a systematic review and meta-analysis. *J Med Internet Res.* (2016) 18:5218. doi: 10.2196/jmir.5218

11. Inglis SC, Clark RA, Dierckx R, Prieto-Merino D, Cleland JGF. Structured telephone support or non-invasive telemonitoring for patients with heart failure. *Cochrane Database Syst Rev.* (2017) 103:255–7. doi: 10.1136/heartjnl-2015-309191

12. Cooper LB, Mentz RJ, Sun JL, Schulte PJ, Fleg JL, Cooper LS, et al. Psychosocial factors, exercise adherence, and outcomes in heart failure patients: insights from heart failure: a controlled trial investigating outcomes of exercise training (HF-ACTION). *Circ Hear Fail.* (2015) 8:1044–51. doi: 10.1161/CIRCHEARTFAILURE.115.002327

13. Bhavnani SP, Parakh K, Atreja A, Druz R, Graham GN, Hayek SS, et al. 2017 roadmap for innovation—ACC health policy statement on healthcare transformation in the era of digital health, big data, and precision health: a report of the american college of cardiology task force on health policy statements and systems of care. *J Am Coll Cardiol.* (2017) 70:2696–718.

14. Krist AH, Davidson KW, Mangione CM, Barry MJ, Cabana M, Caughey AB, et al. Behavioral counseling interventions to promote a healthy diet and physical activity for cardiovascular disease prevention in adults with cardiovascular risk factors: US preventive services task force recommendation statement. *JAMA J Am Med Assoc.* (2020) 324:2069–75. doi: 10.1001/jama.2020.21749

15. Nolan RP, Ross HJ, Farkouh ME, Huszti E, Chan S, Toma M, et al. Automated E-counseling for chronic heart failure: CHF-CePPORT trial. *Circ Hear Fail.* (2021) 2021:1–12. doi: 10.1161/CIRCHEARTFAILURE.120.007073

16. Nolan RP, Payne AY, Ross H, White M, D'Antono B, Chan S, et al. An internet-based counseling intervention with email reminders that promotes self-care in adults with chronic heart failure: randomized controlled trial protocol. *JMIR Res Protoc.* (2014) 3:e5. doi: 10.2196/resprot.2957

17. Spertus JA, Jones PG. Development and validation of a short version of the kansas city cardiomyopathy questionnaire. *Circ Cardiovasc Qual Outcomes.* (2015) 8:469–76. doi: 10.1161/CIRCOUTCOMES.115.001958

18. Nolan RP, Feldman R, Dawes M, Kaczorowski J, Lynn H, Barr SI, et al. Randomized controlled trial of e-counseling for hypertension: REACH. *Circ Cardiovasc Qual Outcomes.* (2018) 11:1–10. doi: 10.1161/CIRCOUTCOMES.117.004420

19. Kosiborod M, Soto GE, Jones PG, Krumholz HM, Weintraub WS, Deedwania P, et al. Identifying heart failure patients at high risk for near-term cardiovascular events with serial health status assessments. *Circulation.* (2007) 115:1975–81. doi: 10.1161/CIRCULATIONAHA.106.670901

20. Tuong W, Larsen ER, Armstrong AW. Videos to influence: a systematic review of effectiveness of video-based education in modifying health behaviors. *J Behav Med.* (2014) 37:218–33. doi: 10.1007/s10865-012-9480-7

21. Mortara A, Vaira L, Palmieri V, Iacoviello M, Battistoni I, Iacovoni A, et al. Would you prescribe mobile health apps for heart failure self-care? An integrated review of commercially available mobile technology for heart failure patients. *Card Fail Rev.* (2020) 6:11. doi: 10.15420/cfr.2019.11

22. Cajita MI, Gleason KT, Han H-RA. Systematic review of mhealth-based heart failure interventions. *J Cardiovasc Nurs.* (2016) 31:E10–22. doi: 10.1097/JCN.0000000000000305

23. Krebs P, Prochaska JO, Rossi JS. Computertailored interventions for health. *Prev Med.* (2010) 51:214–21. doi: 10.1016/j.ypmed.2010.06.004.DEFINING

24. Norine Walsh M, Bove AA, Cross RR, Ferdinand KC, Forman DE, Freeman AM, et al. ACCF 2012 health policy statement on patient-centered care in cardiovascular medicine: a report of the American college of cardiology foundation clinical quality committee. *J Am Coll Cardiol.* (2012) 59:2125–43. doi: 10.1016/j.jacc.2012.03.016

25. Johnson KW, Torres Soto J, Glicksberg BS, Shameer K, Miotto R, Ali M, et al. Artificial intelligence in cardiology. *J Am Coll Cardiol.* (2018) 71:2668–79. doi: 10.1016/j.jacc.2018.03.521

26. Stogios N, Kaur B, Huszti E, Vasanthan J, Nolan RP. Advancing digital health interventions as a clinically applied science for blood pressure reduction: a systematic review and meta-analysis. *Can J Cardiol.* (2020) 36:764–74. doi: 10.1016/j.cjca.2019.11.010

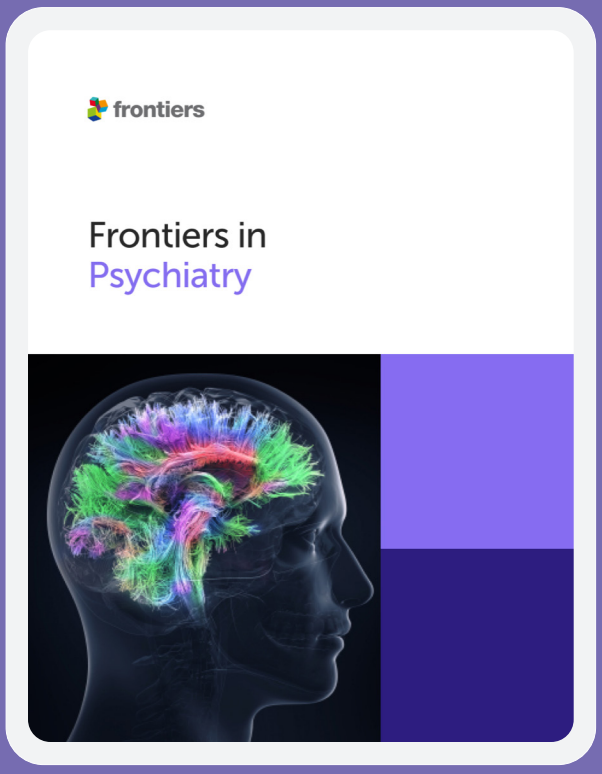
Frontiers in Psychiatry

Explores and communicates innovation in the field of psychiatry to improve patient outcomes

The third most-cited journal in its field, using translational approaches to improve therapeutic options for mental illness, communicate progress to clinicians and researchers, and consequently to improve patient treatment outcomes.

Discover the latest Research Topics

See more →



Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

Contact us

+41 (0)21 510 17 00
frontiersin.org/about/contact