JMIR Formative Research

Impact Factor (2023): 2.0
Volume 5 (2021), Issue 10 ISSN 2561-326X Editor in Chief: Amaryllis Mavragani, PhDc

Contents

Original Papers

Corrigenda and Addendas

Original Paper

Assessing the Acceptability and Feasibility of a Web-Based Screening for Psychoactive Substance Users Among a French Sample of University Students and Workers: Mixed Methods Prospective Study

Emmanuelle Anthoine[1], MSc; Julie Caillon[2], PhD; Xavier Deparis[3], MD; Michel Blanche[4], MD; Maxime Lebeaupin[1], BSc; Marc-Antoine Brochard[5], MSc; Jean-Luc Venisse[5], MD, PhD; Leïla Moret[1,6], MD, PhD

[1]PHU11 Public Health Department, Centre Hospitalier Universitaire (CHU) de Nantes, Nantes, France

[2]Addictions Department, Centre Hospitalier Universitaire (CHU) de Nantes, Nantes, France

[3]Service de Santé au Travail de la Région Nantaise, Nantes, France

[4]Service de Santé des Étudiants (SUMPPS), University of Nantes, Nantes, France

[5]Pulsio Santé, Stimulab Company, Paris, France

[6]UMR 1246 SPHERE (Methods in Patient-Centered Outcomes and Health Research), University of Nantes and Tours, Nantes, France

Corresponding Author:
Leïla Moret, MD, PhD
PHU11 Public Health Department
Centre Hospitalier Universitaire (CHU) de Nantes
Hopital Saint-Jacques
85 Rue Saint-Jacques
Nantes, 44093
France
Phone: 33 2 40 84 69 24
Email: lmoret@chu-nantes.fr

Abstract

Background: Early detection in the prevention of addictive behaviors remains a complex question in practice for most first-line health care workers (HCWs). Several prevention measures have successfully included a screening stage followed by a brief intervention in case of risk-related use or referral to an addiction center for problematic use. Whereas early detection is highly recommended by the World Health Organization, it is not usually performed in practice.

Objective: The aim of this study was to assess the acceptability and feasibility of a web-based app, called Pulsio Santé, for health service users and first-line prevention HCW and to carry out an exhaustive process of early detection of psychoactive substance use behaviors.

Methods: A mixed methods prospective study was conducted in 2 departments: HCWs from the regional occupational health department and from the university department of preventive medicine dedicated to students were invited to participate. Participants 18 years or older who had been seen in 2017 by a HCW from one of the departments were eligible. The study procedure comprised 5 phases: (1) inclusion of the participants after a face-to-face consultation with an HCW; (2) reception of a text message by participants on their smartphone or by email; (3) self-assessment by participants regarding their substance use with the Pulsio Santé app; (4) if participants agreed, transfer of the results to the HCW; and (5) if participants declined, a message to invite them to get in touch with their general practitioner should the assessment detect a risk. Several feasibility and acceptability criteria were assessed by an analysis of a focus group with the HCW that explored 4 themes (usefulness and advantages, problems and limitations, possible improvements, and finally, integration into routine practice).

Results: A total of 1474 people were asked to participate, with 42 HCWs being involved. The percentage of people who agreed to receive a text message or an email, which was considered as the first level of acceptability, was 79.17% (1167/1474). The percentage of participants who clicked on the self-assessment link, considered as the second level of acceptability, was 60.24% (703/1167). The percentage of participants who completed their self-evaluation entirely, which was considered as the first level of feasibility, was 76.24% (536/703). The percentage of participants who shared the results of their evaluation with the HCWs, considered as the second level of feasibility, was 79.48% (426/536). The qualitative study showed that there were obstacles on

the side of HCWs in carrying out the recommended interventions for people at risk based on their online screening, such as previous training or adaptations in accordance with specific populations.

Conclusions: Quantitative results showed good acceptability and feasibility of the Pulsio Santé app by users and HCWs. There is a need for further studies more directly focused on the limitations highlighted by the qualitative results.

(JMIR Form Res 2021;5(10):e15519) doi:10.2196/15519

KEYWORDS

feasibility study; preventive medicine; addiction; screening and brief intervention; web-based app

Introduction

Today, use of psychoactive substances is a major problem everywhere in the world and is insufficiently dealt with and taken into account too late from the perspective of damage prevention and risk reduction [1-3]. Besides tobacco, other examples of psychoactive substances include alcohol, cannabis, cocaine, amphetamines, and hallucinogenic drugs. In France, for example, it is estimated that 14 million people are daily smokers, 3.4 million engage in risky alcohol consumption, and 1.5 million use cannabis regularly (according to data produced in 2019 by the French Monitoring Center for Drugs and Drug Addiction) [4]. Moreover, 75,000 tobacco-related deaths and 41,000 alcohol-related deaths occur each year. Psychoactive substance use opens the door to addictive behaviors, which can be divided into 2 types: risk-related use and problematic use [5].

Risk-related use is defined as use involving a particular risk according to the characteristics of its usage (quantity, frequency, and nature), the context in which it is used (eg, when driving a car or doing precision work on hazardous machines), and the presence of particular vulnerability factors (eg, young age, pregnancy or psychiatric pathology, treatment). Problematic use corresponds to harmful usage or the causing of detectable damage, whether somatic, psychological, socio-professional, familial, or that characterized by a syndrome of pathological dependence [6].

Among existing prevention strategies, certain prevention programs have successfully included a detection stage followed by a brief intervention stage in case of risk-related use. Miller's team [7] defined a brief intervention in 6 points grouped under the acronym FRAMES *(*Feedback, Responsibility, Advise, Menu for Change, Empathy, and Enhancing Self-Efficacy). The brief intervention includes detailed feedback of the results of the evaluation and the risks incurred. It further includes the friendly assertion that the responsibility for change lies with the service user concerned, that he or she is in the best position to design and program this change, and that the reduction in consumption is an accessible objective within the means of the person's own resources [8]. Brief interventions are usually conducted in a one-on-one situation in primary care. They can be implemented anywhere on the intervention continuum and are based on a short, evidence-based, structured conversation consisting of personalized feedback and counseling adjusted to the patient's risk.

Although early detection has been highly recommended for 20 years by the World Health Organization (WHO) [9], primary health care professionals, specifically general practitioners, face difficulties in identifying the use of psychoactive substances, particularly alcohol and cannabis. Despite tobacco use often appearing in patients' records, there is little documentation of alcohol or cannabis use. Existing tracking questionnaires are little used in practice. Moreover, general practitioners currently face difficulties in the feasibility of additional systematic screening, and the implementation of a brief intervention would double the length of a consultation. In fact, general practitioners are not yet sufficiently assuming their role of prevention, which is often attributable to a lack of training [10,11].

It is thus reasonable to believe a web-based app could facilitate the work of primary health care professionals by helping them with early detection and brief interventions. Indeed, a web-based app could simplify the work of professionals by offering patients anonymous but standardized assessments of behaviors that are potentially risky for substance users and by providing tailored backup for health professionals in their interventions.

There is a growing body of literature exploring the use of apps to enhance the feasibility, reliability, and cultural acceptance of screening tests for possible overuse of psychoactive substances. Focusing mostly on a single substance (first and foremost alcohol), they have extended their scope of study to all other substances—medical drugs included—thanks to the WHO-developed Alcohol, Smoking and Substance Involvement Screening Test (ASSIST) [12,13]

Researchers also investigated how much these web-based apps can facilitate the realization of recommended brief interventions once a risk-prone usage has been identified—face-to-face with the patient or online [14-18]

The role of such apps in mediating the relationship between a service user and a health professional still needs to be evaluated, most notably when it comes to the level of autonomy and responsibility it provides to the service user (eg, giving service users the choice to share his or her results). The conditions enabling the realistic implementation of these apps in the real-life practice of professionals in the environments where they consult has also yet to be scrutinized to complement the latest publications [19].

The following study is therefore the first French study aimed at tackling these issues.

Pulsio Santé and Centre Hospitalier Universitaire de Nantes committed to an agreement to perform the study through a collaborative research project.

The main objective of this study was to assess the acceptability and feasibility of the Pulsio Santé web-based app by service users and first-line health professionals involved in prevention for exhaustive early detection of psychoactive substance use behaviors. The secondary objective was to assess whether the Pulsio Santé app could contribute to facilitating the recourse to brief interventions or referral to addiction centers if the results of the assessment pointed to the app's usefulness.

Methods

A mixed methods prospective study was conducted. It was carried out by a project team composed of senior and junior psychiatrists (JLV and JC, respectively), a senior public health physician (LM), a computer scientist (MAB), a project manager (ML), and a statistician (EA). The head of the Occupational Health Department (XD) and the head of the University Department of Preventive Medicine (MB) also contributed to the study.

Characteristics of the Sample

Departments and Health Professionals

Two departments of health professionals volunteered to participate in the study: (1) The Occupational Health Department for the Nantes Region (Service de Santé au Travail de la Région Nantaise [SSTRN]), which is the biggest occupational health department in the county district and is responsible for managing the health of salaried workers in the region—indeed, the issue of risk-related and problematic psychoactive substance use has become a major concern in the business world [20]; (2) The Nantes City University Department of Preventive Medicine and Health Promotion (Service de Santé des Étudiants [SUMPPS]), one of the health departments within the university dedicated to students who are particularly concerned by this type of consumption due to the students' age range [21,22]. In each department, all health professionals (physicians and nurses) were invited to participate as investigators in the study without altering their work schedules. Participating in the study as investigators was voluntary and unpaid.

Participants

Service users 18 years or older who were seen by a health professional either in the SSTRN or in the SUMPPS in the fourth quarter of 2017 were eligible. Participation in the study was voluntary and unpaid to prevent any monetary bias.

The following exclusion criteria were used to screen out unsuitable study participants: insufficient knowledge of the French language, no access or inability to use a web-based app, under 18 years old or under guardianship, ongoing treatment in a specialized addiction care unit, or disorder of the higher level cognitive functions or cognitive dysfunction (particularly in the area of psychoses) that would make the use of a web-based app difficult.

Description of the Pulsio Santé Web-Based App

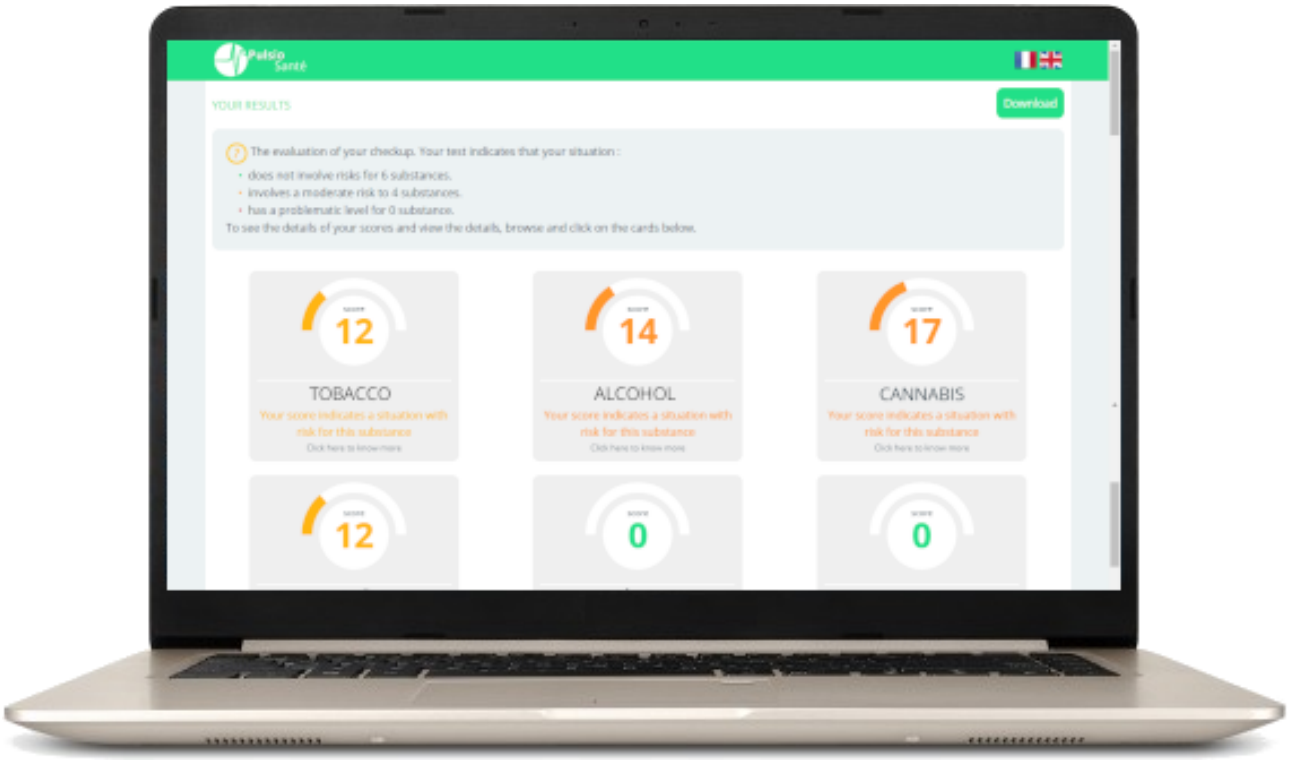
The Pulsio Santé web-based app can be accessed from either a laptop, tablet, or smartphone. Apart from ownership of one of these devices, it only requires internet access. The app does not have to be downloaded, and no account creation by the participant is necessary. For this ease of access, it offers completion of a whole screening test, provides one's results, and enables understanding of the underlying risks without having to enter any identifying information. Only at the end of this process does it offer those participants who would like to discuss their results with a health professional the possibility of creating an account and requesting an appointment. All data are stored on health data–compliant servers located in France.

The Pulsio Santé study comprised 4 implementation stages: inclusion by health professionals, self-assessment by the users of their substance use, analysis of the results, and implementation of a tailored intervention if required. The app is based on the ASSIST developed by the WHO [23] and also validated in a digital version (Audio Computer-Assisted Self Interview [ACASI]) [24] which enables a thorough assessment for the service user of all psychoactive substance use. The system provides offline recruitment and online inclusion by health professionals, one-text messaging via a smartphone or by email, the completion of the self-assessment by clicking on the proposed link, sharing of the results with the health professional if the patient agrees, and the offer of a brief intervention or referral to an addiction center in case of identification of risk-prone usage.

The ASSIST enables 9 scores to be calculated corresponding to the different types of psychoactive substances (tobacco, alcohol, cannabis, cocaine, amphetamines, solvents, tranquilizers, hallucinogenic drugs, and opiates). A specific score for each substance was calculated by summing the scores for 6 questions. For example, scores of >4 (alcohol >10) reflect a moderate risk, and a high risk is indicated by scores >27 [23]. Depending on the risk profile, the participant is directed to a specific page. In case of low (or no) risk, a reassuring message inviting people to take the test again in a year's time is displayed. If the participant presents a moderate or high risk, they are encouraged to see the health professional again to discuss the results. To facilitate this, participants are given the opportunity to share their level of risk with the aforementioned health professional.

After receiving results from a participant, the health professional should invite him or her to make an appointment (brief intervention or addiction center) within 2 weeks, provided that the assessment was determined to be some form of risk-related usage. To do this, the professional can use a digital guide on his or her Pulsio Santé app account and consult the participant's responses to the self-assessment in detail, provided the person has agreed to share these responses (Figure 1).

Figure 1. Screenshot of the Pulsio Santé application app.



The Pulsio Santé app has also been made commercially available to hospitals, occupational health institutions, and medical institutions to deploy prevention services for people at risk of addiction.

This study is the result of a tender response to a regional public health institution looking for subventions for operational and research projects dedicated to the prevention of addiction in several populations (employees, youth, low-income individuals) in real-life settings and through consultations. The study was therefore also aimed at demonstrating the soundness of the research to the public health institutions regarding the feasibility and efficacy of these types of novel digital health interventions (including but not limited to Pulsio Santé) deployed by private companies in real-life settings.

Quantitative Study

Quantitative data were collected electronically. All actions, from the inclusion of participants, to the access to the screening test, to its completion, took place within the Pulsio Santé app. Therefore, use of the app, acceptance of the test (through connections to the app), and test results were all collected on the Pulsio Santé app.

Qualitative data from focus groups were collected and analyzed with the Sphinx iQ2 software.

Study Procedure

In September 2017, health professionals who volunteered to participate as investigators were invited to a meeting organized by the project team to receive information about the Pulsio Santé app and to standardize the manner in which the app was presented to participants.

An information and the consent form was added to the Pulsio Santé app. Service users could only be enrolled after having read the information and agreed to participate.

A case report form was created for each service user. All the information required by the protocol was gathered as encoded data. By signing this protocol, the principal investigator (JLV) and all the health professionals undertook to keep participant identities confidential.

As this self-assessment was initiated online by a text message or email request and outside the health professional's consultation (but nevertheless linked to the consultation), the participants' commitment to the procedure was ensured. It was an essential criterion for the participants' access to the app. Similarly, any transmission of the results of the self-assessment to the health professional required the participants' consent.

Users who opened the link were connected to a platform that allowed them to answer the ASSIST in its digital version (ACASI) in less than 5 minutes. If participants then agreed, they could transfer their results to their health professional. If the service users had not initially agreed to this, messages appeared allowing them to do so after completion of their self-evaluation or would invite them to get in touch with their general practitioner should the assessment detect a risk.

Evaluation Criteria

Acceptability of the Pulsio Santé Web-Based App

Two levels of appreciation of the acceptability were assessed. They concerned the health professionals' ability to propose the app to service users they encounter in their daily practice and the users' ability to use it. The percentage of salaried workers and students who agreed to receive a text message (level 1) and

the percentage of those who clicked on the proposed link (level 2) were thus collected.

Feasibility of the Pulsio Santé Web-Based App

Two levels of appreciation of the feasibility of the Pulsio Santé app were explored. They concerned the ability of the app to enable those to whom the app was offered to carry out a complete assessment of their psychoactive substance use and to share their results with the health professional who had offered it to them. The percentage of salaried workers and students who filled in the ASSIST completely (level 1) and the percentage of users who shared their results with their health professional (level 2) were thus collected.

Furthermore, to meet the secondary objective, the percentage of service users who agreed to a brief intervention was calculated in cases of risk-related status, and the percentage of those presenting problematic substance use who agreed to be referred to an addiction center was also calculated.

Data Collection

The data were collected by health professionals from the SSTRN and SUMPPS involved as investigators during their day-to-day work with workers and students as well as by service users themselves through completing their screening autonomously.

All data were collected already encoded and stored in data servers (European HIPAA [Health Insurance Portability and Accountability Act]-like regulations) located in France.

Statistical Analysis

Statistical analyses used the usual techniques of descriptive statistics (frequency, mean, and SD). The ASSIST-specific substance scores were computed according to the scoring method [23]. A *t* test was performed to anonymously compare substance scores, and a chi-square test was used to compare percentages across the groups of sharers and nonsharers. The statistical significance was set at α =.05. Statistical analyses were performed with R version 3.4.3 (The R Foundation for Statistical Computing).

Qualitative Study

Data Collection

A panel of health professionals participated in semistructured focus groups (or 1 interview) on the acceptability and feasibility of the web-based app between February 2018 and March 2018. A first focus group with the SUMPPS professionals was organized with 5 nurses. The participation rate was 83.3% (5/6). In the SSTRN department, a focus group was organized with 7 SSTRN professionals (4 nurses and 3 physicians), and 1 face-to-face interview with a nurse. The participation rate of the SSTRN professionals was 22.0% (8/36). Focus group topics covered the following: usefulness and advantages of the app; problems, limitations, and frustrations; necessary improvements to the app; and integration into routine practice. The content was recorded and transcribed for analysis. After each focus group, the interviewers checked the transcripts for quality and clarity.

Qualitative Analysis

A multicategory content analysis of the focus group corpus using a frequency distribution table method was conducted to form an exhaustive corpus of what had been said and to identify the main themes and subthemes around which the representations of professionals and participants were organized. A theme frequency analysis was conducted, and possible links between the categories were sought. Results were compared by 2 researchers (EA and ML) to ensure intercoder validity.

Qualitative analyses were performed on Sphinx iQ2 survey and data analysis software.

Ethics Review

All study procedures fully complied with the requirements of French law. The relevant authority for this research was the National Data Protection Committee (Commission Nationale de l'Informatique et des Libertés [CNIL]). The CNIL declaration of conformity of the Study was registered under reference number 2094055 dated August, 28, 2017. No ethics Approval or registration of clinical trials was necessary for this research, as it is not within the scope of application of the French Public Health Code.

Results

Characteristics of the Sample

A total of 1474 service users who came to see either a health professional in the SSTRN department (n=1067) or a nurse in the SUMPPS department (n=407) were asked to participate.

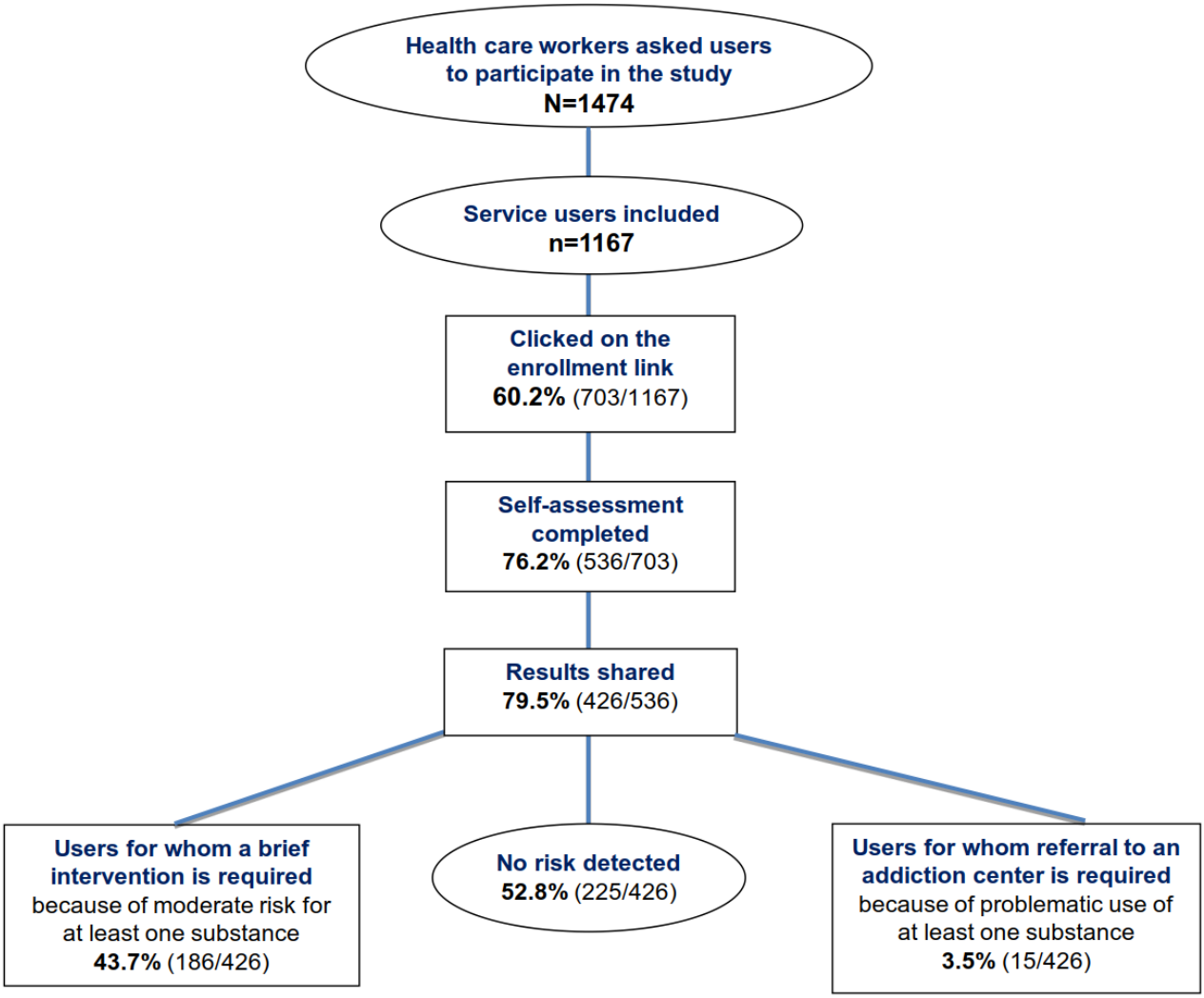
Forty-two health professionals participated as investigators in the study. Among them, 14% (6/42) were working in the SUMPPS and 86% (36/42) in the SSTRN.

Acceptability of the Pulsio Santé Web-Based App

The percentage of participants who agreed to receive a text message or an email, considered as the first level of acceptability, was 79.17% (1167/1474; Figure 2). The participation rates for salaried workers and students were 76.85% (820/1067) and 85.2% (347/407), respectively. The sample comprised 70.26% (820/1167) salaried workers and 29.73% (347/1167) students.

The percentage of participants who clicked on the subscriber link, considered as the second level of acceptability, was 60.24% (703/1167; Figure 2). The participation rates for salaried workers and students were 54.0% (443/820) and 74.9% (260/347), respectively. The sample comprised 63.0% (443/703) salaried workers and 37.0% (260/703) students.

Figure 2. Flowchart of the acceptability and feasibility study - 20210928.



Feasibility of the Pulsio Santé Web-Based App

The percentage of participants who completed their self-evaluation entirely, considered as the first level of feasibility, was 76.2% (536/703; Figure 2). The participation rates for salaried workers and students were 73.8% (327/443) and 80.4% (209/260), respectively. The sample comprised 61.0% (327/536) salaried workers and 39.0% (209/536) students.

The percentage of participants who shared the results of their evaluation with their health professional, considered as the second level of feasibility, was 79.5% (426/536; Figure 2). The participation rates for salaried workers and students were 80.1% (262/327) and 78.5% (164/209), respectively. The sample comprised 61.5% (262/426) salaried workers and 38.5% (164/426) students.

Students clicked on the registration link significantly more often (260/347, 74.9%) than did salaried workers (443/820, 54%; P<.001), but there was no statistical difference between students and salaried workers for the number of completed self-evaluations (P=.06) or the number of shared self-evaluation results (P=.72). For all substances except tobacco (P=.02), the students' scores were not significantly different from those of the salaried workers' scores.

Health Professionals' Perceptions of the Feasibility of the Web-Based App

Focus groups with the health professionals explored 4 themes to evaluate their points of view on the acceptability and feasibility of the Pulsio Santé app.

First, the usefulness and the advantages of the app were studied. It appeared that the app was easy and quick to propose and that participants were easy to enroll in the study (see quotation 1, Table 1).

Table 1. Quotations illustrating the professionals' perceptions about the acceptability and the feasibility of the app.

Quotation number	English translation with original French quotation below
1	
English	Including patients was kind of easy; it didn't really take long.
French	Les inclusions, c'était assez facile, ça ne prenait vraiment pas longtemps.
2	
English	It added a little extra to the individual consultation with the patient by opening up the exchange a little.
French	Ça a apporté un plus au niveau de la consultation du rendez-vous individuel, en ouvrant un petit peu plus l'échange.
3	
English	I'd say it's rather beneficial to patients themselves; self-assessment is interesting on a personal level; it helps to know where we are.
French	J'aurais plutôt dit l'avantage pour la personne elle-même, l'autotest aussi qui est intéressant, pour soi, de savoir où on en est.
4	
English	It also encourages people to think.
French	Ça amène les gens à réfléchir aussi.
5	
English	I wonder whether we shouldn't offer the survey before consultations… well in advance.
French	Moi je me pose la question de savoir si on ne devrait pas proposer l'enquête avant les rendez-vous en fait.
6	
English	It was the same for me. I didn't get any response to the appointments I suggested.
French	Moi c'est pareil, je n'ai pas eu de réponses au rendez-vous proposé.
7	
English	We can do a bit of digging around the addiction problems, but then, what do we do next? If the person does not come back for a brief intervention… question mark. Screening, why not? But then, what means do we have as health professionals?
French	On va aller creuser un petit peu sur les addictions et puis après, qu'est-ce qu'on en fait ? Si la personne ne revient pas en intervention brève, point d'interrogation. Faire du dépistage, ok, mais après, qu'est-ce qu'on a, nous, comme moyens ?
8	
English	Even if people don't share their result, it can still be interpreted. I mean, used it even so. But to do this, we need to know if the person enrolled and gave his answers or enrolled and did nothing. That's important. Honestly, I was surprised; we couldn't see if he had answered.
French	Même sans partager, ça peut être interprété, enfin utilisé quand même. Mais pour ça, il faut qu'on voie s'il s'est inscrit et s'il a répondu, ou s'il s'est inscrit et qu'il n'a rien fait. Ça c'est important. Franchement, moi ça m'a étonnée quand même, on ne voit pas s'il a répondu.
9	
English	He scored as "at risk for tobacco use" despite the fact that he also uses speed, cannabis, and cocaine, which worries me far more, but these did not score as risky. There is definitely something wrong that needs to be looked at.
French	Il m'est apparu comme « situation à risque pour le tabac », alors qu'il consomme des amphétamines, du cannabis, de la cocaïne, qui m'interpellent largement plus et qui, là, indique qu'il n'y a pas de risque. Il y a quelque chose à revoir vraiment là-dessus.
10	
English	I think it would be good for these people to be given links to prevention websites at the end of the questionnaire or a list of institutions that might help. Given that we are including people via a web-based app, it would be good to have links that enable them to know which institutions specialize in treating addictions.
French	Je pense qu'à la fin du questionnaire, ce serait bien qu'il y ait des liens qui leur permettent d'accéder à des sites de sensibilisation, à des organismes qui peuvent venir en aide. Dans la mesure où l'on est dans une transmission par voie numérique, c'est effectivement de mettre des liens qui permettraient d'avoir connaissance des organismes d'addictologie existant.
11	

Quotation number	English translation with original French quotation below
English	I would be willing to use this tool in my daily practice, providing adaptations are made to our area of work, adapted to occupational medicine.
French	Je serais prête à utiliser cet outil au quotidien, avec une adaptation par rapport à notre métier, adapté santé au travail.
12	
English	I experienced some difficulties introducing the tool, when to mention it… So I tried different approaches.
French	J'avais des difficultés pour présenter l'outil, à quel moment dire... Donc j'ai un peu tâtonné.

The app facilitated discussion on the use of substances, complemented the medical consultation (see quotation 2, Table 1), and made participants think about their substance use (see quotation 3 and 4, Table 1).

Second, problems, limitations, and frustrations were discussed. Self-evaluation after the medical consultation was not considered as an optimal strategy (see quotation 5, Table 1). The messages (mail, email, or phone calls) sent by the health professionals to users with moderate or high risk were not followed up (see quotation 6, Table 1). Health professionals also frequently wondered what to do with a positive result from a participant (see quotation 7, Table 1).

Health professionals were then interviewed on the necessary improvements of the app. They suggested making a distinction between nonresponders and nonsharers (see quotation 8, Table 1). This is impossible, ethically speaking, if the patient refuses to share his or her results.

A discrepancy between the results on the WHO measure and actual use by participants was highlighted (see quotation 9, Table 1). At the end of the self-evaluation, several professionals suggested that links to government preventive medicine and public health websites should be created (see quotation 10, Table 1).

Finally, the integration into routine practice was debated. Focus group participants pointed out that the app should be customized according to the medical department concerned (see quotation 11, Table 1) and that professionals should be trained before offering the app to their service users (see quotation 12, Table 1).

Health Professionals' Practice

The other secondary objective focused on the way the Pulsio Santé app could facilitate access to a brief intervention or referral to an addiction center when these are recommended on the basis of the evaluation results. Health professionals offered 51 brief intervention appointments to 186 substance users identified as being at moderate risk for at least 1 substance (51/186, 27.4%). A total of 5.4% (10/186) of the brief interventions considered necessary were carried out by the health professionals, half of which were mediated via the Pulsio Santé app. Health professionals offered 2 addiction center appointments to 15 substance users identified as having problematic use of at least 1 substance (2/15,13.3%). The 2 proposed appointments were accepted by the users, and all were mediated by the Pulsio Santé app. More descriptive results are available in Multimedia Appendix 1.

Discussion

Principal Results

This acceptability and feasibility study, carried out on a large size sample across 2 institutions with different service user profiles, demonstrated high acceptability of the Pulsio Santé web-based app by health professionals, as well as by the salaried workers and students to whom it was offered. In terms of feasibility, the data were similar across workers and students: once the service users had clicked on the enrolment link, a high participation rate in the self-assessment of (536/703, 76%) and a high rate of results shared with the health professional (426/536. 79.5%) were observed. The fact that this applied to both students and occupational medicine participants, where reluctance to confide for fear of their employer being informed is often observed, is also worth noting. It is also important to note that the service users who filled in their self-assessment completely without sharing their results with their health professionals were not very different from those who did share, except for those who indicated tobacco use. It is finally worth noting that the percentage of service users in a problematic situation who did not wish to share their results is not significantly different from those who did, which reinforces the relevance of the app.

The fact that this assessment can be performed in a few minutes as an online self-assessment, with only 1 text message needed to access the screening test and outside consultation time with the health professional, guarantees that it will be acceptable to users. The presence of the health professional offers the possibility for help or care if necessary.

The ASSIST and its digital version (ACASI) enable a quick and precise assessment of the use of all the main psychoactive substances (including medications that are misused).

Limitations

There were no missing values in the screening questionnaire, as service users had to answer all the questions for the psychoactive substances they had checked at the first question. They could not go to the following steps and, therefore, to the final results if they did not answer all their questions. This can be interpreted as a limitation given that even partial results and scores on certain substances might be sufficient to trigger the willingness to get some help. We could imagine that some people bored with the whole questionnaire or misunderstanding parts of it would like to submit their partial answers to get partial results (and act on them). In our study, their only choice in this

situation was to drop out completely or persist to obtain the full results.

Furthermore, while early detection is clearly facilitated by this app, there are nevertheless many unanswered questions. First, inclusion refusals, although not particularly numerous, were to be documented by health professionals, which was not in fact done. Second, the risk thresholds that were defined a decade ago for the ASSIST raise issues of relevance. Indeed, nearly half of the service users were at moderate risk with at least 1 substance, and some substance users who had a score above 20 for at least 3 psychoactive substances did not appear as "problematic". Finally, the proposals (brief intervention or addiction center) justified by the results of the assessment were very often not implemented. Although this was not the main objective of the study, it is nevertheless a crucial issue for the impact of early detection in public health prevention.

Among the improvements suggested during the focus groups, the possibility of offering the app before rather than during the meeting with a health professional was mentioned by several professionals. This would allow for the implications of the assessment to be quickly taken into account. Besides the purely practical and ethical questions raised by this possibility (such as collective answers by groups of students in the waiting room), studies on other populations should be carried out. To this end, the Pulsio Santé app has been deployed in larger populations (such as rural areas) and will be evaluated in those new settings.

Moreover, defining the stage when the screening test should be introduced during follow-up and adapting the web-based app to the aims of each department and to their service users were among the strong expectations that were expressed during the focus groups. These two points indeed seem essential to guaranteeing efficient coordination between detection and care when the latter is indicated.

Conclusions

This study on a large-size sample showed good acceptability and feasibility of the Pulsio Santé web-based app by service users and health professionals.

The qualitative study, however, showed that there were obstacles on the professionals' side, which could justify the need for further studies more directly focused on these limitations and the web-based app as a mediator-facilitator; studies on other populations should be carried out and designed to extend to purely behavioral addictions (ie, with no psychoactive substance use).

Acknowledgments

Financial support was provided by a grant from the Mission Interministérielle de Lutte contre la Drogue Et les Conduites Addictives, an institution attached to the French Prime Minister's office.

The funding source had no role in the design or conduct of the study; in the collection, management, analysis, or interpretation of the data; in the preparation, review, or approval of the manuscript; or in the decision to submit the manuscript for publication.

Conflicts of Interest

Two authors declare having an interest in the Pulsio Santé company. JLV and MAB, who participated in the drafting, are both cofounders of the Pulsio Santé web-based app, the web-based app used for this study.

Multimedia Appendix 1
Quantitative results from the screening test.
[DOCX File , 31 KB - formative_v5i10e15519_app1.docx]

References

1. UNODC World Drug Report 2020: Global drug use rising; while COVID-19 has far reaching impact on global drug markets. United Nations Office on Drugs and Crime. 2020 Jun 25. URL: https://www.unodc.org/unodc/press/releases/2020/June/media-advisory---global-launch-of-the-2020-world-drug-report.html [accessed 2021-07-12]
2. Peacock A, Leung J, Larney S, Colledge S, Hickman M, Rehm J, et al. Global statistics on alcohol, tobacco and illicit drug use: 2017 status report. Addiction 2018 Oct;113(10):1905-1926. [doi: 10.1111/add.14234] [Medline: 29749059]
3. Rehm J, Manthey J, Shield KD, Ferreira-Borges C. Trends in substance use and in the attributable burden of disease and mortality in the WHO European Region, 2010-16. Eur J Public Health 2019 Aug 01;29(4):723-728. [doi: 10.1093/eurpub/ckz064] [Medline: 31008515]
4. Rapport national à l'EMCDDA – 2019, OFDT. Observatoire français des drogues et des toxicomanies. URL: https://www.ofdt.fr/publications/collections/rapports/rapports-nationaux/rapport-national-ofdt-2019/ [accessed 2021-07-21]
5. Babor T, Campbell R, Room R, Saunders J. Lexicon of Alcohol and Drug Terms. Geneva: World Health Organization; 1994.
6. Diagnostic and Statistical Manual of Mental Disorders. 5th Edition. Washington DC: American Psychiatric Association; 2013.
7. Bien T, Miller W, Tonigan JS. Brief interventions for alcohol problems: a review. Addiction 1993 Mar;88(3):315-335. [doi: 10.1111/j.1360-0443.1993.tb00820.x] [Medline: 8461850]

8. O'Donnell A, Anderson P, Newbury-Birch D, Schulte B, Schmidt C, Reimer J, et al. The impact of brief alcohol interventions in primary healthcare: a systematic review of reviews. Alcohol Alcohol 2014;49(1):66-78 [FREE Full text] [doi: 10.1093/alcalc/agt170] [Medline: 24232177]
9. Heather N, editor. WHO Phase IV Collaborative Project on Implementing Country-Wide Early Identification and Brief Intervention Strategies in Primary Health Care. Geneva: World Health Organization; 2006.
10. Falcón M, Navarro-Zaragoza J, García-Rodríguez RM, Nova-López D, González-Navarro MD, Mercadal M, et al. Perceived barriers to implementing screening and brief intervention for alcohol consumption by adolescents in hospital emergency department in Spain. Adicciones 2018 Jul 12;30(3):189-196 [FREE Full text] [doi: 10.20882/adicciones.896] [Medline: 28749525]
11. Seigneuret L, Bouzillé G, Moirand R. Obstacles au repérage précoce et intervention brève en alcoologie chez les aînés. Alcoologie et Addictologie 2018 Mar;40(1):6 [FREE Full text]
12. Saitz R, Helmuth ED, Aromaa SE, Guard A, Belanger M, Rosenbloom DL. Web-based screening and brief intervention for the spectrum of alcohol problems. Prev Med 2004 Nov;39(5):969-975. [doi: 10.1016/j.ypmed.2004.04.011] [Medline: 15475031]
13. Donoghue K, Patton R, Phillips T, Deluca P, Drummond C. The effectiveness of electronic screening and brief intervention for reducing levels of alcohol consumption: a systematic review and meta-analysis. J Med Internet Res 2014;16(6):e142 [FREE Full text] [doi: 10.2196/jmir.3193] [Medline: 24892426]
14. Schulenberg J, Johnston L, O'Malley P, Bachman J, Miech R, Patrick M. Monitoring the Future, National Survey Results on Drug Use, 1975-2017. Ann Arbor: The University of Michigan Institute for Social Research; Jul 31, 2018. [doi: 10.3998/2027.42/146531]
15. Carey KB, Carey MP, Henson JM, Maisto SA, DeMartini KS. Brief alcohol interventions for mandated college students: comparison of face-to-face counseling and computer-delivered interventions. Addiction 2011 Mar;106(3):528-537 [FREE Full text] [doi: 10.1111/j.1360-0443.2010.03193.x] [Medline: 21059184]
16. Schwartz RP, Gryczynski J, Mitchell SG, Gonzales A, Moseley A, Peterson TR, et al. Computerized versus in-person brief intervention for drug misuse: a randomized clinical trial. Addiction 2014 Jul;109(7):1091-1098 [FREE Full text] [doi: 10.1111/add.12502] [Medline: 24520906]
17. Garnett C, Crane D, West R, Brown J, Michie S. Identification of behavior change techniques and engagement strategies to design a smartphone app to reduce alcohol consumption using a formal consensus method. JMIR Mhealth Uhealth 2015;3(2):e73 [FREE Full text] [doi: 10.2196/mhealth.3895] [Medline: 26123578]
18. Arnaud N, Baldus C, Elgán TH, De PN, Tønnesen H, Csémy L, et al. Effectiveness of a web-based screening and fully automated brief motivational intervention for adolescent substance use: a randomized controlled trial. J Med Internet Res 2016 May 24;18(5):e103 [FREE Full text] [doi: 10.2196/jmir.4643] [Medline: 27220276]
19. Lord S, Moore SK, Ramsey A, Dinauer S, Johnson K. Implementation of a substance use recovery support mobile phone app in community settings: qualitative study of clinician and staff perspectives of facilitators and barriers. JMIR Ment Health 2016 Jun 28;3(2):e24 [FREE Full text] [doi: 10.2196/mental.4927] [Medline: 27352884]
20. Bush DM, Lipari RM. The CBHSQ Report: Substance use and substance use disorder, by industry. Substance Abuse and Mental Health Services Administration, Center for Behavioral Health Statistics and Quality. 2015 Apr 16. URL: https://www.samhsa.gov/data/sites/default/files/report_1959/ShortReport-1959.pdf [accessed 2021-06-01]
21. Palfai TP, Saitz R, Winter M, Brown TA, Kypri K, Goodness TM, et al. Web-based screening and brief intervention for student marijuana use in a university health center: pilot study to examine the implementation of eCHECKUP TO GO in different contexts. Addict Behav 2014 Sep;39(9):1346-1352 [FREE Full text] [doi: 10.1016/j.addbeh.2014.04.025] [Medline: 24845164]
22. Gryczynski J, Kelly SM, Mitchell SG, Kirk A, O'Grady KE, Schwartz RP. Validation and performance of the Alcohol, Smoking and Substance Involvement Screening Test (ASSIST) among adolescent primary care patients. Addiction 2015 Feb;110(2):240-247 [FREE Full text] [doi: 10.1111/add.12767] [Medline: 25311148]
23. Humeniuk R, Ali R, Babor T, Farrell M, Formigoni M, Jittiwutikarn J, et al. Validation of the Alcohol, Smoking And Substance Involvement Screening Test (ASSIST). Addiction 2008 Jun;103(6):1039-1047. [doi: 10.1111/j.1360-0443.2007.02114.x] [Medline: 18373724]
24. McNeely J, Strauss SM, Rotrosen J, Ramautar A, Gourevitch MN. Validation of an audio computer-assisted self-interview (ACASI) version of the alcohol, smoking and substance involvement screening test (ASSIST) in primary care patients. Addiction 2016 Feb;111(2):233-244. [doi: 10.1111/add.13165] [Medline: 26360315]

Abbreviations

ACASI: Audio Computer-Assisted Self Interview
ASSIST: Alcohol, Smoking and Substance Involvement Screening Test
CNIL: Commission Nationale de l'Informatique et des Libertés
HIPAA: Health Insurance Portability and Accountability Act
SSRTRN: Service de Santé au Travail de la Région Nantaise

SUMPPS: Service de Santé des Étudiants
WHO: World Health Organization

Edited by G Eysenbach; submitted 16.07.19; peer-reviewed by J Lopez-Rodriguez, L O'Brien, M Winter; comments to author 07.09.20; revised version received 18.12.20; accepted 11.03.21; published 01.10.21.

Please cite as:
Anthoine E, Caillon J, Deparis X, Blanche M, Lebeaupin M, Brochard MA, Venisse JL, Moret L
Assessing the Acceptability and Feasibility of a Web-Based Screening for Psychoactive Substance Users Among a French Sample of University Students and Workers: Mixed Methods Prospective Study
JMIR Form Res 2021;5(10):e15519
URL: https://formative.jmir.org/2021/10/e15519
doi:10.2196/15519
PMID:34596571

Original Paper

A Novel Mobile App ("CareFit") to Support Informal Caregivers to Undertake Regular Physical Activity From Home During and Beyond COVID-19 Restrictions: Co-design and Prototype Development Study

Kieren J Egan[1], BSc, PhD; William Hodgson[2], BSc; Mark D Dunlop[1], BSc, PhD; Gennaro Imperatore[1], BSc, PhD; Alison Kirk[2], BSc, PhD; Roma Maguire[1], BN, MSc, PhD

[1]Department of Computer and Information Science, University of Strathclyde, Glasgow, United Kingdom
[2]School of Psychological Sciences and Health, University of Strathclyde, Glasgow, United Kingdom

Corresponding Author:
Kieren J Egan, BSc, PhD
Department of Computer and Information Science
University of Strathclyde
Livingstone Tower
26 Richmond Street
Glasgow, G1 1XH
United Kingdom
Phone: 44 0141 548 3138
Email: kieren.egan@strath.ac.uk

Abstract

Background: Informal caregivers, or carers (unpaid family members and friends), are instrumental to millions worldwide for the ongoing delivery of health and well-being needs. The risk of crisis points (eg, hospitalizations) for caregivers increases with the absence of physical activity. The COVID-19 pandemic is highly likely to have increased the risk of crisis points for caregivers by increasing the amount of time spent indoors due to shielding and lockdown restrictions. Thus, accessible evidence-based tools to facilitate physical activity for caregivers indoors are urgently needed.

Objective: The aim of this study was to co-design and develop a novel mobile app to educate and support carers in the undertaking of regular physical activity at home during and beyond COVID-19 restrictions via integration of the transtheoretical model of behavior change and UK physical activity guidelines.

Methods: We co-designed a mobile app, "CareFit," by directly involving caregivers, health care professionals, and social care professionals in the requirements, capturing, and evaluation phases of three Agile Scrum design and development sprints. Seven participants representing multistakeholder views took part in three co-design sessions, each of which was followed by a development sprint. Requirements for CareFit were grounded in a combination of behavioral change science and UK government guidelines for physical activity.

Results: Participants identified different barriers and enablers to physical activity, such as a lack of time, recognition of existing activities, and concerns regarding safely undertaking physical activity. Requirements analysis highlighted the importance of simplicity in design and a need to anchor development around the everyday needs of caregivers (eg, easy-to-use video instructions). Our final prototype app integrated guidance for undertaking physical activity at home through educational, physical activity, and communication components.

Conclusions: Integrating government guidelines with models of behavioral change into a mobile app to support the physical activity of carers is novel. We found that integrating core physical activity guidelines into a co-designed smartphone app with functionality such as a weekly planner and educational material for users is feasible. This work holds promise to fill the gap of effective physical activity solutions for caregivers both during and beyond the COVID-19 pandemic. Further work is now needed to explore the feasibility, acceptability, and usability of the approach in real-world settings.

(JMIR Form Res 2021;5(10):e27358) doi:10.2196/27358

KEYWORDS

physical activity; Android; COVID-19; intervention; co-design; exercise; app; development; support; caregiver

Introduction

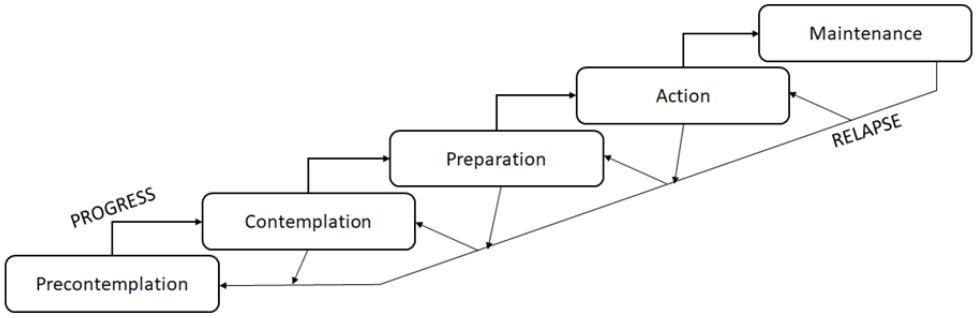
Informal caregivers or carers—those providing unpaid care for friends or family—constitute a vital lifeline to millions of people worldwide. In the United Kingdom alone, there are an estimated 6.5 million carers, and across Europe, up to 80% of all long-term care is understood to be delivered by carers [1,2]. Although some carers benefit and achieve a sense of fulfillment from caring roles [3], there is now strong evidence that caregiving may have an adverse impact on health and wellness both in the short and long term [4,5]. Preventable crisis points (eg, hospitalizations, significant worsening of mental or physical health, irreversible changes to caring circumstances) are commonplace (even in the absence of COVID-19) and frequently cause irreversible deterioration in health for the carer and those cared for [6,7]. As the global population ages, and the health and social care workforce shrinks [8], it appears inevitable that the reliance placed on caregivers will only increase. A public health priority is to raise quality of life and prevent crisis points. Furthermore, the COVID-19 pandemic substantially increased pressures and time spent at home, and reduced opportunities and motivations for physical activity [6,9,10].

The unmet needs of caregivers are considerable and diverse. There have been many innovations in recent years to aid caregivers in areas such as support, care coordination, telehealth/diagnostics, and digital care delivery [11]. Solutions aimed at caregiver support have mainly focused on targeting mental health (eg, burden, anxiety, depression) through face-to-face, telephone, and digital interventions [12,13]. Less established solutions are those aimed at improving physical health. Systematic review work in this area identified only 14 studies to date [14], with interventions mainly delivered face to face and/or by telephone-based approaches. Across these studies, improvements were observed in physical activity levels, distress, well-being, quality of life, and sleep quality. Such targeted solutions are yet to make the "leap" into the digital spectrum and mass impact potential of smartphone apps.

The potential of the digital spectrum is now emerging for all populations (eg, automated data collection, machine learning, augmented reality), and there are key questions as to whether vulnerable groups such as informal caregivers will also be able to enjoy the benefits. Advantages could include simply raising awareness of physical activity guidelines through mobile apps (such as those from the United Kingdom, which suggest a variety of different types of activities per week according to age group). More sustainable and greater impacts may also be realized through using evidence-based models of behavioral change. The well-established transtheoretical model (TTM) of behavioral change [15,16] postulates that the more sustainable changes in behaviors are those that are altered habitually and through a cyclical process of specific stages (see Figure 1). However, it remains to be explored precisely how to design and integrate a solution capable of translating such key messages in a feasible, acceptable, and usable manner for more vulnerable groups such as caregivers.

Recent survey data suggest that (even in the absence of COVID-19) 81% of carers are not able to perform as much physical activity as they would like [17]. There is therefore an imperative need to continue researching innovations for caregivers, and to explore what empowering evidence-based tools could be delivered at home both during the COVID-19 pandemic and beyond. We here present a rapid response project to produce a novel evidence-based mobile app designed to empower caregivers to undertake regular physical activity at home during (and beyond) the COVID-19 outbreak. We designed our app, "CareFit," with a co-design team of user experts, and using robust and well-established scientific knowledge (eg, the TTM [15,16], government guidelines [18], and sports and exercise specialist knowledge).

Figure 1. Overview of the transtheoretical model across the different stages of precontemplation, contemplation, preparation, action, and maintenance alongside relapse.



Methods

Recruitment, Consent, and Ethical Approval

Participants were recruited using convenience sampling (connections across both Carers Scotland and the University of Strathclyde). We aimed to identify 6-8 participants to a co-design group to maximize the depth of conversation achievable with our discussions [19]. We contacted three known professionals who had an interest in caregivers. Caregivers were recruited specifically though links with Carers UK (Scotland), whereby (a few/targeted) local carer centers across Scotland were asked to approach carers who would be suitable for the study and interested in being part of a working group. All invited

participants accepted the offer to take part in this work. Inclusion criteria were that participants were aged 18 years and over and interested in contributing to current knowledge of digital innovations for caregivers. Individuals were asked to commit no more than 7 hours in total to the co-design process. The co-design sessions took place between July and August 2020 (ie, during the COVID-19 pandemic). The University of Strathclyde Ethics Committee approved the study protocol. As per standard ethical procedures, each member of the group signed an individual consent form after being given an information sheet and opportunity to ask any questions about their overall involvement in the study.

Study Design

We developed CareFit using an Agile Scrum co-design methodology [20,21] (see the App Development and Testing section below), and had to work within national and local restrictions imposed by COVID-19 pandemic measures, which represented a considerable challenge given the face-to-face nature of traditional co-design sessions. We viewed our co-design participants as architects and partners of this work (see Figure 2), as outlined by Sanders and Stappers [22]: "creativity of designers and people not trained in design working together in the design development process." Our stakeholders consisted of carers, employers, physical health experts, and health care professionals. This multidisciplinary team was involved throughout the design and evaluation stages of the co-design process [20,22]. In total, three versions of the app were developed iteratively, and the final version of the app was released and evaluated by our participants (caregivers and caregiver-related professionals). Our focus was on people in the contemplation or preparation stages of the TTM, which includes those only thinking about being more active, and those who have thought about and taken some steps to becoming more physically active. Our goal was to help participants form and regularly "action" intentions to be more active. In terms of the software/tools used, we conducted three co-design video call sessions (using Zoom and simulations of notice boards/post-its [MURAL]), and complemented collective discussions with three individual questionnaires (incorporating around 20 to 30 questions on Qualtrics online software) as a basis for design sprints. If a participant could not attend group meetings, one-to-one calls were offered as an alternative. See Table 1 for an overview of the meetings. The time immediately after our co-design meetings was dedicated to our development "sprints," each lasting 2-3 weeks.

Figure 2. Overview of the co-design process across the three co-design meetings.

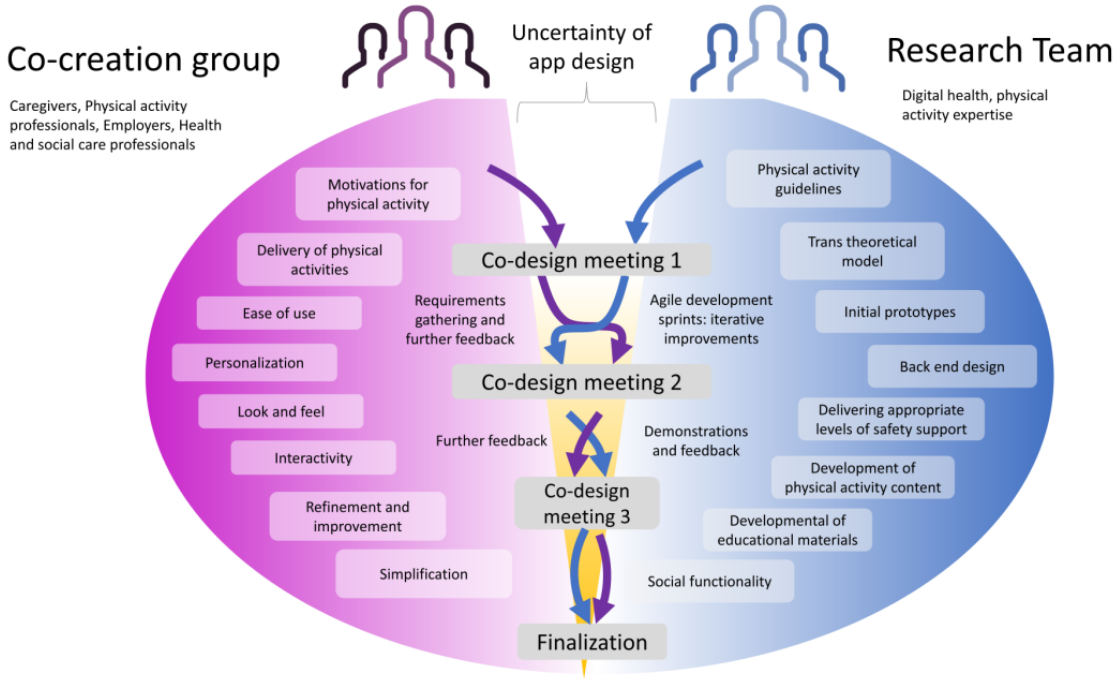


Table 1. Overview of the co-design meetings.

Sprint (number of questionnaires completed)	Focus/aim	Detail of meeting used to guide sprint
1 (N=5)	To critique and present a simple initial app prototype; to collectively present the principles of the transtheoretical model (TTM) and the UK government national physical activity guidelines.	We explored the topics of motivation, goals, physical activity guidelines, delivery options, health, and safety. We also explored "keep, lose, change," and asked our participants to prioritize needs according to the MoSCoW[a] methodology.
2 (N=6)	To review the feedback from meeting 1 and progress during design sprint 1.	We explored how to deliver details within the educational, physical activity, and communication components, including the "keep, lose, change" format. We presented future options of the physical activities using videos and subsequent feedback.
3 (N=6)	To review and finalize the app design in preparation for a 3-week real-world study.	Final, detailed discussion on the presentation of the revised app developed, and further discussion of the education, physical activity, and communication sections.

[a]MoSCoW: "Must Have, Should Have, Could Have, Won't Have this time" prioritization.

App Development and Testing

The CareFit app was developed for Android (versions 7 to 10). The app was mainly developed in Java, with the exception of the education section, which was developed in HTML/CSS/JavaScript and integrated into the main app. Extensive unit and user testing was undertaken using Android phone simulators and a range of different Android physical devices of different ages, specifications, and display sizes. Test versions of CareFit were distributed to users through online emulators in advance of the co-design meeting sessions to improve the requirements-capturing discussions and involve target users in the development process. CareFit was developed using the Agile Scrum methodology. Agile Scrum is an iterative software development process in which software development takes place in short and fast periods of development formally defined as sprints. Before each development sprint, requirements from the previous sprint are improved or new requirements were captured using feedback given to developers by users. We categorized functional and nonfunctional requirements using the FURPS+ (Functional, Usability, Reliability, Performance, where the "+" is used to indicate additional requirements such as programming language and other constraints) approach [23]. Such requirements were guided throughout by our co-design team using MoSCoW ("Must Have, Should Have, Could Have, Won't Have this time") [21] and "keep, lose, change" [22] methodologies, and informed by the TTM [15]. Our design supported users by enabling them to report errors and crashes easily through a dedicated email address. The email address was displayed on the main screen of the app at all times and, where possible, we provided an immediate response to users with technical issues. CareFit had the following "+" requirements: being developed for Android OS (support ranged from Android 7.0 known as "Nougat" to Android 10 known as "Pie").

Data Handling and Prioritization

The structure of co-design sessions consisted of an online white board (MURAL), online conference calls (Zoom), and online questionnaires (Qualtrics). All of our online meetings involved the presentation of slides and/or prototype mockups/video "walkthroughs." The first development sprint involved requirements-capturing using the MoSCoW methodology prioritization method, which ranks requirements as "must have," "should have," "could have," or "won't have" [24]. During the co-design and sprints 1 and 2, we also used the "keep, lose, change" approach to offer our participants freedom to decide on fundamental aspects of the app where required [25]. Whenever the majority of the group expressed clear and strong preferences, these were integrated in the app design. There was a small number of occasions where user suggestions conflicted with physical activity guidelines. Any discrepancies to MoSCoW preferences are explained within the text. When the co-design group did not reach consensus, the academic team reviewed and reached a final decision. For qualitative data, quotations were examined by two researchers (BH, KE) who analyzed data and identified core themes. Disagreements were discussed with a third researcher (RM).

Results

Co-design Group

Our co-design group consisted of four different stakeholder groups: caregivers (n=4), a health care professional (n=1), an expert in physical activity (n=1), and an employer representative who supported caregivers on a regular basis (n=1). Our sample included six women and one man. All our participants resided in Scotland. Our three group meetings involved discussions about all aspects of the app design and were hosted by three researchers including our lead developer. Follow-up to group sessions involved the cumulative delivery of more than 100 questions delivered in the format of online questionnaires. This work was undertaken as a rapid response to the COVID-19 pandemic, and was carried out over 6 months (July to November 2020).

Co-design Meeting 1 and Design Sprint 1

As part of co-design meeting 1, we presented a simple exercise app based on National Health Service guidance on exercises without integration of any behavioral change models or national guidelines (see Multimedia Appendix 1 for an example questionnaire). Participants highlighted that barriers and enablers to physical activity included lack of time, motivation, safety,

recognition of achievements, and a need for personalization (see Textbox 1).

We asked participants what they would like to "keep, lose, (or) change" after reviewing a basic prototype that contained some physical activity exercises of different intensities, a very basic reminder system, and text instruction of exercises with an accompanying timer. Participants wished to "lose" the timer for strength exercises and "keep" aspects regarding icons. "Change" included the addition of videos to demonstrate safe ways to complete exercises, and to find ways to capture progress (eg, "include planner and progress chart" and "add a video to demonstrate a safe way of completing exercises"). We further explored how best to deliver safety information to participants. In total, 60% of users preferred a disclaimer about a risk of injury and/or a summary about safety only on the "first login." The addition of instructions for safe exercise on "every login" was supported by 60% of users.

Textbox 1. Barriers and enablers discussed during our first co-design session with representative quotes for each theme.

Enablers

- Incorporate daily living activities as physical activity opportunities
- Explore user support
- Motivational strategies
- Provide users with physical activity advice and safe practice

Example quote, "Time: making it short and simple and able to do in their own time; reminders to motivate; peer support." [Participant 1.5]

Barriers

- (Lack of) peer support
- Poor mental health
- Lack of education
- Changing definition of wellness
- Lack of recognition
- Lack of individualized approaches
- Support missing to receive coping strategies

Example quote, "Time constraints, financial pressures, physical impact of caring (eg, back injury), emotional barriers (eg, guilt over leaving loved one), lack of respite opportunities, ineffective coping strategies, lack of motivation (exacerbated by depression)." [Participant 1.3]

Participants were primarily interested in the app supporting delivery of physical activity elements (47.8%) and secondarily interested in the education and social/community components (31.8% and 20.4%, respectively). Participants stated a desire for education and physical activity information to be displayed graphically (ie, less text-based). Use of icons, graphs, and videos was a particularly popular approach (40% of participants stated visual elements were a "must have" feature, with 20% stating that audio and video elements were also "must haves"). In terms of personalization, 20% of the participants stated that use of the first name of the carer was a "must have," whereas none of the participants classed displaying the name of the person cared for as a "must have" feature.

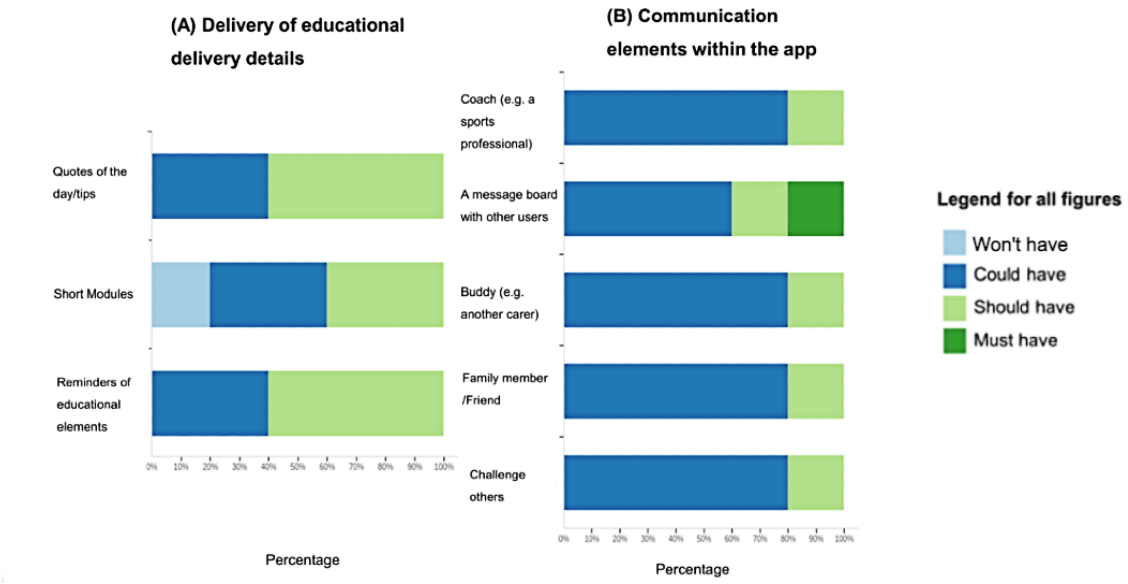
For the physical activity/motivation elements, most of the participants (80%) stated that routine builders were a "must have" component of the app (see Figure 3A). Key requirements for users (ie, 40% of users stating that these were "must have") included goal/target setting and identifying improvements (Figure 3A). Participants were presented with three potential designs (Figure 3B) for functionality to measure physical activity progress. Despite reacting positively to the idea, there was no consensus on precisely how this could be implemented.

Participants recognized the need for caregivers to undertake different types of physical activities, with cardiovascular physical activity identified as the greatest need (60% of participants stating this as a "must have" feature; Figure 3C). Other prominent features included muscle (endurance and strength), flexibility, and breaking up sedentary behaviors (20% of participants respectively stated these were "must have" elements).

Respondents were divided as to what the educational elements should look like. There were no clear interactive features that were recognized as a requirement for all users (Figure 3D); however, there was some preference to add functions such as "meditation exercises," "triggers and relapse prevention exercises," and elements regarding "time planning." Some participants (40%) thought that tips or quotes of the day were requirements that the app "could have," as opposed to 60% who thought the app "should have" these elements (Figure 4A). There was no consensus on how communication elements of the app could be delivered. A variety of formats were suggested, such as the presence of a coach, a message board, and challenging other users (Figure 4B).

Figure 3. Feedback received using MoSCoW methodology across motivational elements (A, top left), measuring physical activity (B, top right), specific types of activities needed (C, bottom left), and focus within education sections (D, bottom right). Abbreviations used: 6 Minute walk test (6MW), Government Guidelines (Gov. guidelines).

Figure 4. Feedback from participants using the MoSCoW method for (A) delivery method for educational details and (B) communication elements within the app.



Implementation of Design Sprint 1

To implement the requirements gathered in co-design meeting 1, we reviewed the data gathered as a research group and improved our app accordingly (see Table 2). The majority of user requirements (eg, routine builder, time planner) were implemented through the development of a weekly planner, physical activity content, and education and communication elements. Physical activity plans were simplified from the UK guidelines as much as possible so that “muscle and balance” could encompass aspects of muscle strength and endurance alongside flexibility. Educational materials were influenced by both carer needs and previous paper-based resources developed for the diabetes field (adapted for use with caregivers). We took an academic decision not to include GPS functionality step counting, as carers may not always carry their phones and therefore could potentially lose recognition for the physical activity undertaken. There are also many existing apps that focus on this type of physical activity (eg, running, walking). We still planned to incorporate outdoor exercise aspects into the app. We also took the decision to support carers using the app with a user guide in addition to the guidance delivered in educational sections. This represents our aim to develop an app without requiring significant external training to use it.

Table 2. Requirements identified and developed within co-design stage 1.

Requirements	Development/implementation details
Physical activity needs	
Need to develop a simple evidence-based physical activity plan[a,b]	An easy-to-use planner where an entire week would be visible, ideally reflecting (1) cardiovascular, (2) muscle and balance, and (3) sedentary breakers according to the UK physical activity guidelines. This "weekly planner" was the cornerstone of the app's physical activity functionality where users could make, revise, view, and review their plan for the week(s) ahead.
Users would like to record any cardiovascular activity (ie, at home and outside)[a,b]	For cardiovascular activities, we built a simple dialogue system that could record time and intensity. We also incorporated "active living" activities through a drop-down menu for adding further detail.
Muscle and balance simplicity[a,b]	We devised a system that incorporated 3 to 5 different muscle and balance activities (with the precise content yet to be determined), allowing personalization.
Underline importance of health and safety[a,b]	Users are supported with information about how to undertake safe exercises both through an initial information and disclaimer screen, alongside some brief information within each physical activity video.
Capturing sedentary activity[a,b]	Users can optionally record sedentary activity.
Educational needs: Increase awareness of the activity guidelines and behavioral change[a,b,c]	Initial educational content was developed on PowerPoint for subsequent transfer to the app. The format follows the activity consultation (built in part from existing resources within the group for diabetes, and includes interactive elements based on the TTM[d]).
Communication needs: Flexibility on how social media/messaging could be implemented[a]	As "communication" was a lower priority feature, we remained open to comments and considerations from the group. Our plan was to be agile in our development. We concentrated our efforts primarily around exploring links to social media and message boards.
Look and feel	
The app should be simple to navigate and personalized.[a,e]	Controls were clearly marked with labels. For the educational section users could choose the font size to facilitate reading.
App colors that are familiar/associated with trust to users should be used.[a,e]	User interface was designed to keep the different sections of the app compartmentalized both visually and functionally, while the look and feel of the app was kept consistent; by using different colors and clear labels, users were always kept aware of which section of the app they were in.
The components around education and physical activities should be clearly distinct.[a]	To improve user experience, the educational section was implemented in HTML/JS/CSS as this section was primarily text-based.

[a]Based on co-design discussions.

[b]Based on UK activity guidelines.

[c]Based on models of behavioral change.

[d]TTM: transtheoretical model.

[e]Based on user design principles.

Co-design Meeting 2 and Design Sprint 2

Overview

We presented a revised prototype to participants based on feedback from design sprint 1. Feedback was generally positive; in particular, strengths of the work mentioned included the simplicity of design and user-friendliness of the app. Elements suggested to "keep" included the overall app look and feel such as "the simplicity of selecting exercises." There were no "lose" elements suggested. "Change" elements included themes regarding flexibility/personalization such as "the ability to move exercises as things come up on certain days." Other feedback from participants concerned the colors of the app. Many participants reported that they liked the simplicity of the app (eg, "I think it looks good and it's concise and to the point.") To improve usability, participants suggested that many different types of elements could be added, including short and focused educational materials. Some users suggested further improvements to the user friendliness of the app and that a user guide/video introduction could be a useful introduction to the app for carers (eg, "I think make it as user friendly as possible; less is more").

Physical Activity Elements

We asked participants about how they would like reminders to function. There was no consensus about when the best time of the day or week to deliver these would be. Further comments came back from several respondents that more personalization holds value to carers, including "the user could choose this to suit their individual needs such as evenings/weekends." During co-design meeting 2, we presented to the group an existing short "sedentary breaker" video produced from the University of Strathclyde aimed at staff members. Feedback on the video included that the informality of the activities is a strength and that we should consider increasing the clarity of instructions.

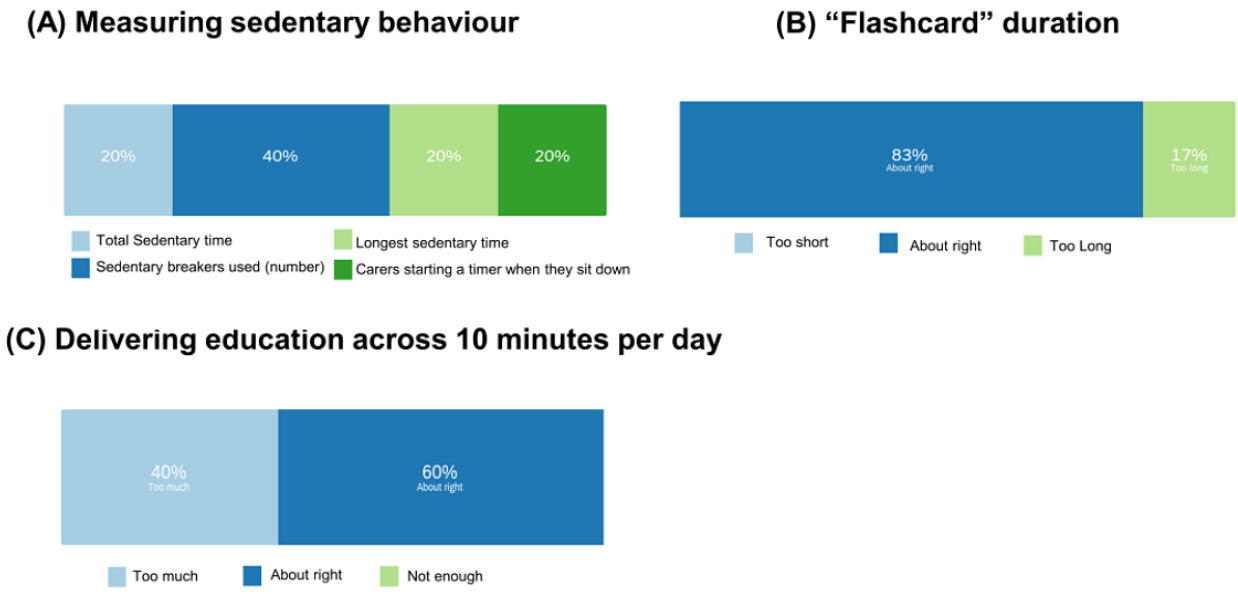
I feel that the style of the video is sufficient but maybe a subtitle on the video of what the carer should also be doing. [Co-design participant 2.4]

Good to show in a home setting and using equipment from around the home. [Co-design participant 2.1]

The videos are great and also good to have a written description at the side. Would suggest a commentary with each exercise to give advice on exercise and, for example, what muscles you should feel stretching to minimize issues. [Co-design participant 2.5]

For selecting a unit of measurement for sedentary achievements, the "number" of sedentary breakers was the top choice from four options (Figure 5A). The majority of our participants (83% of respondents) considered flashcards of around 5 minutes duration to be the most suitable (Figure 5B). Participants requested a wide range of different cardiovascular activities possible (eg, walking the dog, running), some of which could take place outside the home. Participants were also interested to see a broad mixture of different muscle exercises delivered (eg, upper body, lower body), and all of our respondents wanted to see physical activity specific to caregivers incorporated into the overall app design.

Figure 5. User preferences for components of app design in co-design meeting 2. Participants provided feedback on how best to (A) measure sedentary behavior, (B) deliver flashcard duration, and (C) deliver education across 10 minutes per day.



Educational Elements

We asked our participants if up to 10 minutes a day of interacting with educational materials was feasible for caregivers to carry out; 60% of our participants stated that this was "about right," compared to 40% of participants who said it was "too much" (Figure 5C). All components of the education fit well with participants' expectations across the seven original elements proposed: (1) "Introduction," (2) "Relationships and Physical Activity," (3) "Managing Time," (4) "Goals and Rewards," (5) "Physical Activity and Consequences," (6) "The Mind and Body," and (7) "Knowledge Quiz."

Communication Elements

Participants provided feedback that components of the app could be useful to share with friends and family, including goals, activities, barriers, support, and others such as sharing achievements. However, there was no clear consensus on which single aspect would be most useful to share. Participants were interested to integrate their activities with many different platforms, including Facebook (100% of respondents), WhatsApp (83% of respondents), and Twitter/Instagram (50% respondents each).

Implementation of Design Sprint 2

To implement requirements gathered in co-design meeting 2 (Table 3), the physical activity functionality was refined much further, allowing users to now add up to three activities on any given day of the week in the planner. Cardiovascular activities would be delivered in the app or as "ad hoc," where details could be recorded in a drop-down menu (eg, walking the dog). Users could also select an intensity of activity ranging from low to moderate to high. We further refined educational sections to incorporate a target of 10 minutes per section, and explored the use of visuals and breaking up text. To support the widest social media integration, we created image-based certificates for user achievements and integrated them through Android's "Share with" functionality. This meant that our user could share their progress on any social media platform as well as through email or MMS.

Table 3. Details of design sprint 2 following requirements identification within co-design session 2.

Requirements	Development/implementation details
Physical activity needs/themes	
Improve the clarity of the videos, including the use of text on the screen[a]	Develop bespoke videos for each of the physical activities supported by the app. This would include text on the screen and audio guides of how to undertake each activity. Videos will cover a wide range of different activities across cardiovascular, sedentary breakers, and muscle and balance work. Three bespoke videos for each activity group will be developed, guided by a physical activity specialist.
Participants would like to measure progress in sedentary behavior using number of days [a]	Implement simple drop-down menu options to record the number of sedentary breakers used per day. This would allow users to set a target for sedentary breakers each day and record progress accordingly.
Participants would like to be able to move activities onto the next day[a,b]	A feature will be added to the weekly planner so that users can move an activity forward if not completed at the intended time.
Participants would like to set their own reminders as required[a,b,c]	Support users to add reminders for activities as required within the planner. There will also be additional support within the app to allow users to review all reminders set at the same time.
Muscle and balance activities need to exercise many different muscle groups within the same activity[a,b]	We would explore the feasibility of developing "flashcards" that would present a sequence of random activities. This could include building more holistic exercise sets within an individual 5-minute video.
Educational needs: "Lessons" need to last up to 10 minutes per day and deliver the 7 lessons as intended, but the terminology could be off-putting[a,c]	Materials developed for up to 10 minutes a day, and all proposed elements on the app. All educational elements are to be optional and termed "stages" to avoid overly formal language. Development of rules of the education sections, including how to provide consistency of content and delivery.
Communication needs: Allow participants flexibility on the modality of sharing information[a]	Our app must support many different modalities of sharing user progress, and may be more functionally suited to Android system sharing.
Other	
User guide required for participants[a]	User guide will be accessible through the app.
Look and feel of the app, including color scheme, need to be revised[a,b,d]	Implement consistent use of logos and color scheme across the different app components based on the activity guidelines and UK National Health Service colors.

[a]Based on co-design discussions.

[b]Based on UK activity guidelines.

[c]Based on models of behavioral change.

[d]Based on user design principles.

Co-design Meeting 3 and Design Sprint 3

Overview

During this last co-design meeting and resulting sprint, we finalized the app design. We used information already presented to the group and built the final design on key examples (Table 4). Overall, participants responded positively to the design of CareFit's home screen, most participants (67%) describing it as "very user friendly" and the remaining (33%) describing it as a "little user friendly." Free-text feedback from participants suggested positive reception of the activity planner. For example, comments described the planner as "easy to understand" or having a "simple layout which is simple to follow"; however, there were some concerns raised by some describing the planner as "busy and hard to follow." Feedback also highlighted the importance of personalization; for example, participants suggested allowing users to select/design elements of the user interface: "everyone is different and should choose their own color scheme if they can" (co-design participant 3.6).

After showing participants our proposed design for the user interface, all participants found the icons suitable, including 17% who found it "very suitable" (Figure 6A). Designs presented in discussion included the icons proposed for specific activities. In response, 33% of the participants indicated that use of an "arm flexing" icon was not appropriate for caregivers to signify strength and balance activities. Other feedback indicated that the icons were "…simple and easy to recognize and follow." For the planner (Figure 6B), there was a general preference for rounded circular icons (as opposed to squares or rounded squares). Most participants (67%) thought that the overall app logo design was very suitable (Figure 6C and Figure 7A).

Table 4. Details of design sprint 3 following requirements identification within co-design session 3.

Requirements	Development/implementation details
Physical activity needs/requirements	
Participants requested that we alter the icons used (bicep) for muscle and balance	An alternative graphic was selected, more suitable for the carer demographic.
Participants would like to access physical activities (eg, sedentary behaviors) from within the education sections[a]	Implement a link between the educational and physical activity components to link the two.
Videos delivered with clarity, supported by text. There was no consensus on university branding; the academic group decided to proceed with videos using the university logo[a]	Videos are supported with slow, clear narration; safety messages; and on-screen text. A link to each video must be accessible within the app delivered when both planning and undertaking activities.
Participants with physical activity expertise recommended that delivering "muscle and balance" activities with significant variation of targeted areas within each video.	Deliver, record, and integrate videos that support all physical activity types: sedentary activity, cardiovascular activity, and muscle and balance. We will develop 3 short videos (2 to 5 minutes).
Appropriate measurement of physical activities and progress[a,b,c]	For cardiovascular activities, users measure time and intensity; for sedentary breakers, users measure the number per day; and for muscle and balance, users can measure the number of events. Timing of cardiovascular activities will be measured using a start/stop timer dialogue.
Educational needs/requirements	
As per sprint 2, ensure that educational sections last around 10 minutes or less[a,c]	Split initial educational sections so that there are 8 sections overall: "Introduction" now becomes "Welcome and Introduction" and "Physical activity: Beginners Guide."
Increase accessibility of the educational materials [a,c]	Use more visuals and break up education text
Communication needs/requirements: As per sprint 2	Deliver the ability to share progress across different social media/communication tools.
Other/look and feel of the app	
Participants liked the overall color scheme and logo formats suggested [a,b,d]	Look and feel includes colors from activity guidelines and those familiar within the UK National Health Service.
Personalization of app [a,b,d]	User guide will be developed. Users can increase/decrease the font size of the educational sections as required. Content delivered included "personas" relevant to a Scottish context.
Integration of reminders [a,b,d]	Users can set reminders any time through clicking on planned activity. A prompt will be given to users when originally setting an event.

[a]Based on co-design discussions.

[b]Based on UK activity guidelines.

[c]Based on models of behavioral change.

[d]Based on user design principles.

Figure 6. Feedback from participants using the MoSCoW method. (A) Preference for the icon type within the weekly planner; (B) feedback on planner design buttons for use; (C) suitability of our proposed app logo; (D) response on the number of interactive tasks within the educational sections.

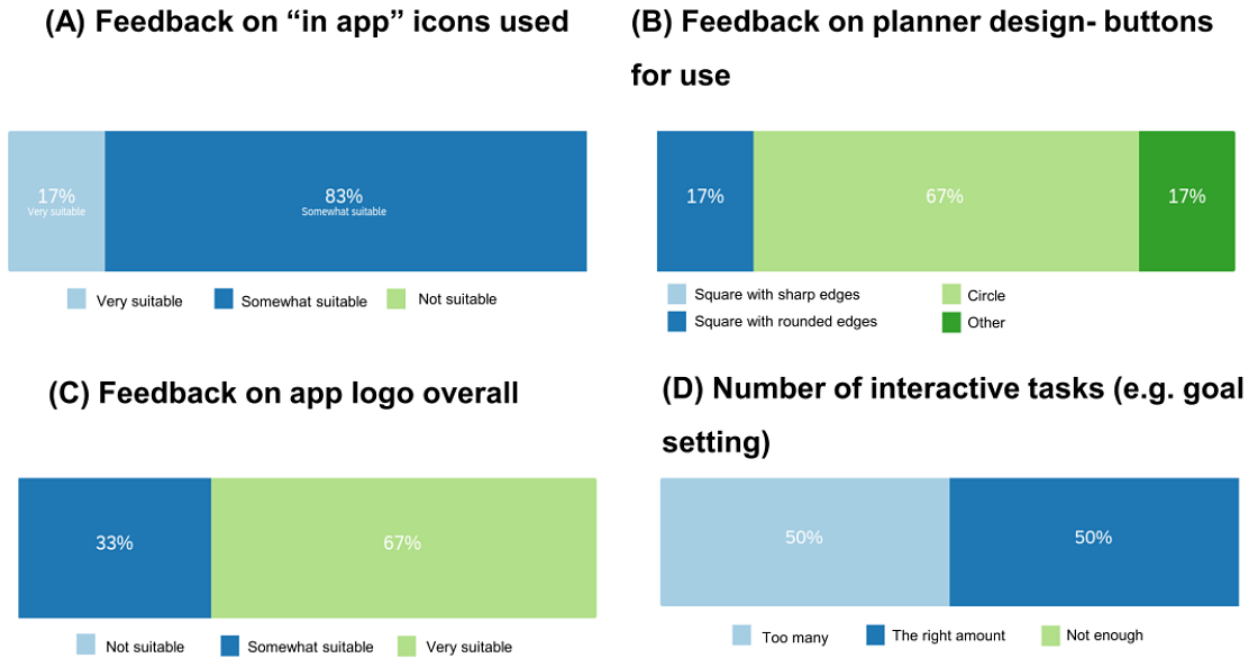
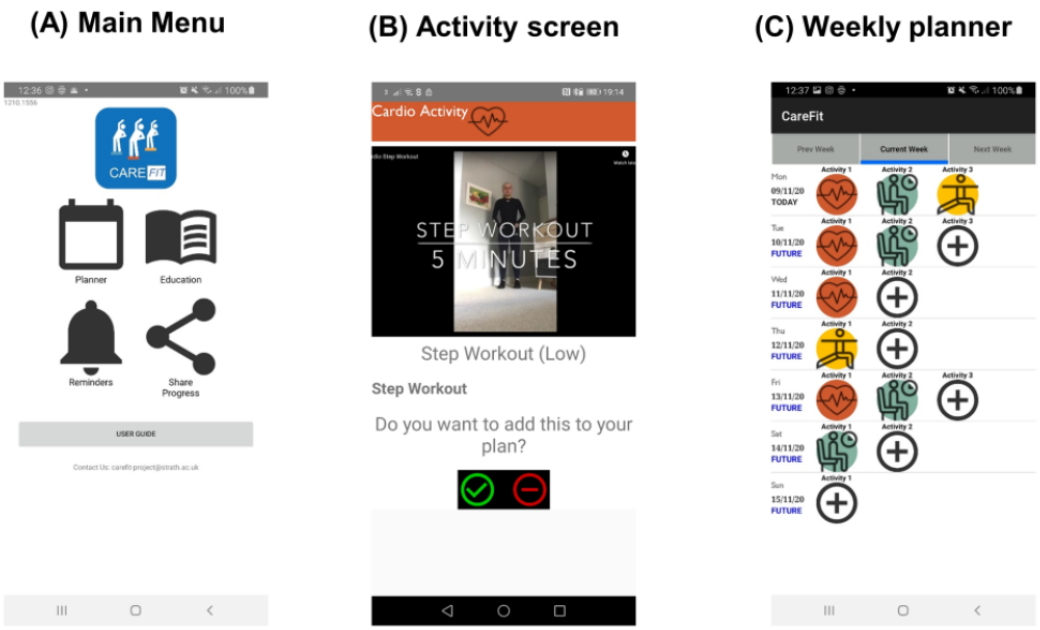


Figure 7. Screenshots from the final app design, including (A) main menu page, (B) example of an exercise page, (C) weekly planner.



Physical Activity Elements

Our proposed methods for recording muscle and balance activity by number of days completed were well received by participants: 67% found the approach "very suitable" and 33% found the approach "somewhat suitable." We had similar responses for our approach to measure cardiovascular activities using "time" and "intensity," where 60% of respondents found the approach to be "very suitable." For sedentary breaker activities, all respondents wanted these to be accessible from both the educational and physical activity sections of the app. In terms of the "look and feel" of the instructional videos, there was no consensus on whether to use university branded clothing or casual (without branding) clothing while giving instructions. Unfortunately, the link to a prototype video had stopped working for 3 users by the time several of the questionnaire respondents completed their feedback, and therefore we could not explore the responses for this question.

Educational Elements

During co-design online meeting 3, we presented full draft sections of the educational elements, overall structure, and proposed rules for development (eg, developing some standards/formatting requirements). In our follow-up questionnaire, we asked participants how relevant the educational elements were for our target group, where 83% of respondents stated that the content was "very relevant." In terms of usability, 67% of the respondents deemed that the educational materials as presented were "very user friendly," and 33% thought they were "a little user friendly." Three participants gave further feedback that the app should incorporate carer "experience and voices," and participants could see value in the developments that had taken place since the previous development sprints: "You can see it developing and coming together from previous stages. This is much improved." Participants encouraged the use of audio delivery of materials alongside visual presentation of materials such as images and videos. Other feedback included that the number of (optional) interactive tasks was "the right amount" for 50% of respondents (Figure 6D).

Communication Elements

After analyzing feedback from participants, we decided to give priority to the implementation of the educational and physical activity elements over communication elements. Participants were asked what sort of information they would like to share on social media in future versions of the app. A variety of communication mechanisms were suggested, such as forums (eg, "In future, I think a forum where they can support each other or a buddy system would be beneficial"); use of emojis ("How you feel after exercise such as smiley faces/thumbs up, small bite size text such as a Twitter-style link, or share with friends family or other social media, challenge others"); and finally progress-sharing mechanisms (eg, "I think the progress page is suitable to be able to share").

Implementation of Design Sprint 3

Our last development sprint finalized the app design as requested by the group. We made every effort to address any aspects for which clear and feasible changes had been requested by our co-design group. A major focus of this sprint was the creation of the physical activity videos, taking place from the home setting by an "Active Lifestyle" officer based at the university. Videos developed were no more than 5 minutes long, considering both the stage of change and lack of free time of caregivers. These videos (accessible via the planner) were integrated into the final app version. All team members were involved in testing the final app functionality. Using a task checklist, we evaluated elements of consistency, error prevention, and clarity. Although many of these passed user testing, we did notice that the code added to allow font resizing as an accessibility feature failed on some phones, and the videos displayed were too small on others. This reinforced the need for extensive testing on a wide range of devices and, in the event that something is missed, we put in place ongoing procedures to update the software.

Final Prototype Developed

The CareFit final prototype (see Figure 7 for screenshots) was designed to be used for the duration of a 3-week study. Users could navigate to the different parts of CareFit via the following main menu options: Weekly Planner, Education, Reminders, Share Progress.

The Weekly Planner allowed planning of physical activities for up to 2 weeks ahead. Users could also view activities planned and completed during the previous week. The planner allowed users to plan up to three types of physical activities (with a bespoke icon and individual screen for each) on any day of the week. When users were unable to complete an activity as intended, they had the ability to move the activity to another date of their choice. CareFit users could choose from the following types of physical activities based on current guidelines: (1) cardiovascular activities plus a daily activities option where the activity took place outside of the app-delivered elements (where the user could set the intensity and duration level and/or use custom activities); (2) muscle and balance activities (where the user sets the intensity level); and (3) three sedentary breaker activities that users were free to choose from.

Instructions on how to perform exercises were delivered via videos hosted on YouTube. The videos were focused on developing functional fitness while acknowledging daily life constraints imposed by being a caregiver. The education section was structured as follows: (1) Welcome and Introduction, (2) Physical Activity: Beginners Guide, (3) Relationships and Physical Activity, (4) Managing Time, (5) Goals and Rewards, (6) Physical Activity and Consequences, (7) The Mind and Body, and (8) Knowledge Quiz. The reminders section of the app let users manage reminders for activities they had planned. Once a reminder was set in the planner, users could use the reminders section to view their reminders or delete unwanted reminders. The "Share progress" functionality let users share a summary image of physical activities/achievements completed to be shared across a variety of social media/phone platforms.

Discussion

Principal Findings

Regular physical activity is important for everyone; however, many groups are underserved by existing guidance and targets [26]. Globally, we lack sustainable formats for the delivery of physical activity instructions for those on the lower end of the physical activity level spectrum [27]. Caring responsibilities can push individuals needlessly toward becoming a "syndemic" statistic (ie, being vulnerable due to the effects of widespread noncommunicable disease) [6], including cases where individuals lack the time, tools, or motivation to undertake regular physical activity. Cumulative data from more than 80,000 people and 64 studies suggest that the COVID-19 pandemic has been associated with an increase in sedentary behavior and a decrease in physical activity [28], where lack of physical activity (and its associated effects) will remain a critical concern for chronic disease [29-31]. It is not simply the risk of mortality or poor health from future pandemics that is of concern, but it is also the seemingly inevitable poor quality of life, deterioration of health, hospitalizations, and other crisis

points that can affect both the carer and those cared for [6,7]. Perhaps one of the most striking lessons of the COVID-19 pandemic is that caregivers are irreplaceable. Here, we have presented a rapid response project that is a first in digital health: a prototype app co-designed by carers that delivers a personalized approach to behavioral change science aimed at improving physical activity in the home.

The development of this app offers several opportunities for further learning. The use of co-design in caregiver research is growing and aligns well with other emerging work. Our strategy was to equip our participants with a variety of different stakeholder viewpoints through discussion before completing questionnaires [32]. Such co-design was successfully used previously by Xu et al [32] when designing an app for caregivers of children with atopic dermatitis to develop functionalities such as login, disease diary, journal, chatbot, forum, and disease monitor. As part of this work, participants helped us to identify several different barriers and enablers to physical activity from the home, including lack of time, finding a way to recognize efforts, and being able to conduct activities safely. Similar findings have been replicated elsewhere both in physically active and inactive populations [33,34]. For example, Hoare et al [33] surveyed a total of 894 Australian adults aged 25 to 54 years, who were both active and inactive, and found that lack of time, lack of enjoyment, and a preference to do other things were key barriers toward physical activity. Mulligan et al [34] systematically searched for personal barriers of physical activity participation for people with neurological conditions, and found that safety, confidence, and lack of support were key contributors to lack of physical activity.

Our results demonstrate the utility of online co-design: carers and care professionals have made measurable contributions to the project at every stage of the design process, taking the app from a "fuzzy" concept to the implementation and evaluation stages [20]. A key theme (and enabler) within the app design is to value the role of the carer within the framework of activity guidelines (eg, a few minutes of activity is better than none), and to recognize that common caring activities such as cleaning, lifting, and moving have inherent value for physical health [18,35]. We have designed the app wherever possible to be supportive. There is no pressure put on the carer to undertake physical activity, and personalization is possible through making individual plans, exercises, and engaging with the education sections as and when required. We also supported caregiving tasks wherever possible (eg, lifting, carrying). Simplicity of the design (both in terms of content and technology use/delivery) is a core element of the solution. Physical activity guidelines and behavioral change models are distilled into manageable, daily tasks.

The theoretical underpinnings of this app are of considerable interest to future work and practice. Our use of the TTM allowed for several personal reflective exercises to be developed that were suited to the stage of change our participants were at (eg, goal setting and list of pros and cons). We are not the first to develop elements of the TTM into a digital app. There is evidence to suggest that this model of behavioral change can allow up to 6 months of positive behaviors within a "GreyMatters" app study [36]. The context of the study was to support individuals with healthy lifestyle factors that reduce the chances of developing dementia (eg, targeting holistic health needs across cognition, diet, physical health, sleep, social, and stress). App use was supported by a coach that incorporates both personal and simplified generic goals. Although there are similarities with CareFit (including scope to expand CareFit to support more holistic health care needs), the populations served by these apps remain largely distinct. While the design of the app aligns well with the TTM overall, the precise modality of interaction that works best now needs to be researched further. For example, previous literature has shown that goal setting is not straightforward, and certain app features such as "trophies" and "ribbons" in themselves are insufficient to motivate participants to undertake physical activity on a regular basis [37]. Further complicating matters is that components of the UK national physical activity guidelines can be difficult to put into action. There is no specific "dose" of muscle and balance activity work, only a recommendation that the activity should take place 2 days a week. Future related work could explore other stages of the TTM (eg, action stage) in greater depth, including over a longer duration (more than 3 weeks). There are also future options to expand CareFit by integrating wearable technology, supporting further outdoor activities, and increasing educational information available. Other interesting areas for future exploration include understanding how individuals can be supported in undertaking exercises correctly and how the app could identify those who are most at risk of complications from being overweight/obese [38]. Finally, the digital divide remains a significant risk to reaching the caregiver population, which must be accounted for [39].

There are some limitations of note relating to this work. CareFit was developed as a rapid response to COVID-19 (6-month project duration) in the middle of a global pandemic where convenience sampling may have skewed our feedback. Participant engagement was structured to genuinely collate the opinions of our co-design participants; however, prioritization through online MoSCoW methodology with supporting online meetings is not infallible [40]. Further research is required to test the external validity of our approach. Despite the short timeframe of this project, we managed to integrate many requirements stemming from participants' feedback. However, combining different sources of information still requires researcher-based decision-making. The evidence-based materials used (eg, behavioral change, government guidelines, educational activities) have not been synthesized and delivered in this manner before, and the extent to which individual caregivers can guide themselves through the materials needs further appraisal. Not least is the barrier of caregivers being left with "another" task in their busy schedules: physical activities may work best where unmet needs are addressed holistically [41]. Our users did not extensively test the final prototype built, as our focus for such questions is reserved for a real-world trial.

Key Lessons and Future Recommendations

Key lessons from this work are as follows. Primarily, this work emphasizes the value of the co-design process and the importance of involving carers and care professionals in research and practice. In addition, the feasibility of co-designing evidence-based physical activity apps for caregivers with a small

development team is demonstrated, even with the limitations imposed by COVID-19 restrictions. Our results also highlight the importance of synergy among theory, expert knowledge, and target users' personal experience in developing bespoke solutions for special populations such as caregivers. The need for assistive technologies to move from computer solutions to portable device–based solutions is further emphasized. We have also shown that developing a user-centered digital health app to improve the quality of life of caregivers is feasible. Nevertheless, the digital literacy of caregivers will vary significantly, and further exploration will be needed to understand what works in practice in terms of confidence and support. There are also gaps in current knowledge regarding physical activity guidelines to be addressed, such as whether caregivers are receiving information and how to measure components objectively. The constraints of the Android environment can be a limitation to user experience, especially with respect to difficulties in updating app versions. Overall, feedback from our participants demonstrates the strength of the co-design process as opposed to universal design apps.

Conclusion

We have demonstrated the utility of the co-design process to develop a novel approach to combine national physical guidelines and behavioral change models into a personalized app for carers. Further work is now required to explore the acceptability, usability, and feasibility of this app within a real-world setting.

Acknowledgments

The study was funded by a Chief Scientist Office, Scottish Government (COV/SCL/20/06) research grant awarded to KE, MD, AK, and RM. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. We would like to acknowledge all of our co-design group experts who helped shape this project with the time and valuable experiences that they shared. We are grateful to Catherine Mann and the Centre for Sports and Exercise at the University of Strathclyde who supported this work through the development of nine high-quality physical activity videos. We would also like to acknowledge the strong role that Carers Scotland has played throughout this work, and in particular, Patricia Clark who helped review iterations of the app and provided feedback throughout.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Example of the first co-design postdiscussion survey.
[DOCX File , 36 KB - formative_v5i10e27358_app1.docx]

References

1. Hoffmann F, Rodrigues R. Informal carers: who takes care of them? European Centre for Social Welfare and Policy Research. URL: https://www.euro.centre.org/publications/detail/387 [accessed 2021-06-02]
2. Missing out: the identification challenge. Carers UK. URL: https://www.carersuk.org/for-professionals/policy/policy-library/missing-out-the-identification-challenge [accessed 2021-06-02]
3. Pristavec T. The burden and benefits of caregiving: a latent class analysis. Gerontologist 2019 Nov 16;59(6):1078-1091 [FREE Full text] [doi: 10.1093/geront/gny022] [Medline: 29659788]
4. Stansfeld S, Smuk M, Onwumere J, Clark C, Pike C, McManus S, et al. Stressors and common mental disorder in informal carers--an analysis of the English Adult Psychiatric Morbidity Survey 2007. Soc Sci Med 2014 Nov;120:190-198 [FREE Full text] [doi: 10.1016/j.socscimed.2014.09.025] [Medline: 25259657]
5. Metzelthin SF, Verbakel E, Veenstra MY, van Exel J, Ambergen AW, Kempen GIJM. Positive and negative outcomes of informal caregiving at home and in institutionalised long-term care: a cross-sectional study. BMC Geriatr 2017 Oct 10;17(1):232 [FREE Full text] [doi: 10.1186/s12877-017-0620-3] [Medline: 29017453]
6. Egan K. Digital technology, health and well-being and the Covid-19 pandemic: it's time to call forward informal carers from the back of the queue. Semin Oncol Nurs 2020 Dec 15;36(6):151088 [FREE Full text] [doi: 10.1016/j.soncn.2020.151088] [Medline: 33229183]
7. Ewing G, Austin L, Jones D, Grande G. Who cares for the carers at hospital discharge at the end of life? A qualitative study of current practice in discharge planning and the potential value of using The Carer Support Needs Assessment Tool (CSNAT) Approach. Palliat Med 2018 Feb 28;32(5):939-949. [doi: 10.1177/0269216318756259]
8. Maynard A. Shrinking the state: the fate of the NHS and social care. J R Soc Med 2017 Feb 01;110(2):49-51 [FREE Full text] [doi: 10.1177/0141076816686923] [Medline: 28169585]
9. Kent EE, Ornstein KA, Dionne-Odom JN. The family caregiving crisis meets an actual pandemic. J Pain Symptom Manage 2020 Jul;60(1):e66-e69 [FREE Full text] [doi: 10.1016/j.jpainsymman.2020.04.006] [Medline: 32283220]

10. Thomas GP, Saunders CL, Roland MO, Paddison CA. Informal carers' health-related quality of life and patient experience in primary care: evidence from 195,364 carers in England responding to a national survey. BMC Fam Pract 2015 May 15;16(1):62 [FREE Full text] [doi: 10.1186/s12875-015-0277-y] [Medline: 25975608]
11. Lindeman D, Kim K, Gladstone C, Apesoa-Varano E. Technology and caregiving: emerging interventions and directions for research. Gerontologist 2020 Feb 14;60(Suppl 1):S41-S49 [FREE Full text] [doi: 10.1093/geront/gnz178] [Medline: 32057082]
12. Blom MM, Zarit SH, Groot Zwaaftink RBM, Cuijpers P, Pot AM. Effectiveness of an internet intervention for family caregivers of people with dementia: results of a randomized controlled trial. PLoS One 2015;10(2):e0116622 [FREE Full text] [doi: 10.1371/journal.pone.0116622] [Medline: 25679228]
13. Egan KJ, Pinto-Bruno Á, Bighelli I, Berg-Weger M, van Straten A, Albanese E, et al. Online training and support programs designed to improve mental health and reduce burden among caregivers of people with dementia: a systematic review. J Am Med Dir Assoc 2018 Mar;19(3):200-206. [doi: 10.1016/j.jamda.2017.10.023] [Medline: 29306605]
14. Lambert SD, Duncan LR, Kapellas S, Bruson A, Myrand M, Santa Mina D, et al. A descriptive systematic review of physical activity interventions for caregivers: effects on caregivers' and care recipients' psychosocial outcomes, physical activity levels, and physical health. Ann Behav Med 2016 Dec;50(6):907-919. [doi: 10.1007/s12160-016-9819-3] [Medline: 27439530]
15. Prochaska JO, DiClemente CC. Stages and processes of self-change of smoking: toward an integrative model of change. J Consult Clin Psychol 1983;51(3):390-395. [doi: 10.1037/0022-006X.51.3.390]
16. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. Am J Health Promot 1997 Sep;12(1):38-48. [doi: 10.4278/0890-1171-12.1.38] [Medline: 10170434]
17. State of caring: A snapshot of unpaid care in the UK. Carers UK. URL: http://www.carersuk.org/images/News__campaigns/CUK_State_of_Caring_2019_Report.pdf [accessed 2021-06-02]
18. UK Chief Medical Officers' Physical Activity Guidelines. Department of Health and Social Care. 2019 Sep 07. URL: https://tinyurl.com/3tp894ne [accessed 2021-09-14]
19. Doria N, Condran B, Boulos L, Curtis Maillet DG, Dowling L, Levy A. Sharpening the focus: differentiating between focus groups for patient engagement vs. qualitative research. Res Involv Engagem 2018;4:19 [FREE Full text] [doi: 10.1186/s40900-018-0102-6] [Medline: 29983994]
20. Sanders EB, Stappers PJ. Probes, toolkits and prototypes: three approaches to making in codesigning. CoDesign 2014 Mar 06;10(1):5-14. [doi: 10.1080/15710882.2014.888183]
21. Agile Development. Interaction Design Foundation. URL: https://www.interaction-design.org/literature/topics/agile-development [accessed 2021-06-02]
22. Sanders EB, Stappers PJ. Co-creation and the new landscapes of design. CoDesign 2008 Mar;4(1):5-18. [doi: 10.1080/15710880701875068]
23. Nakajo T, Sasabuchi K, Akiyama T. A structured approach to software defect analysis. Hewlett-Packard J 1989;40(2):50 [FREE Full text]
24. Clegg D, Barker R. Case Method Fast-Track: A RAD Approach. Boston, MA: Addison-Wesley Longman Publishing Co, Inc; 1994.
25. Frohlich DM, Lim CSC, Ahmed A. Keep, lose, change: Prompts for the re-design of product concepts in a focus group setting. CoDesign 2014 Jan 24;10(2):80-95. [doi: 10.1080/15710882.2013.862280]
26. Guthold R, Stevens GA, Riley LM, Bull FC. Worldwide trends in insufficient physical activity from 2001 to 2016: a pooled analysis of 358 population-based surveys with 1·9 million participants. Lancet Glob Health 2018 Oct;6(10):e1077-e1086 [FREE Full text] [doi: 10.1016/S2214-109X(18)30357-7] [Medline: 30193830]
27. Craike M, Wiesner G, Hilland TA, Bengoechea EG. Interventions to improve physical activity among socioeconomically disadvantaged groups: an umbrella review. Int J Behav Nutr Phys Act 2018 May 15;15(1):43 [FREE Full text] [doi: 10.1186/s12966-018-0676-2] [Medline: 29764488]
28. Stockwell S, Trott M, Tully M, Shin J, Barnett Y, Butler L, et al. Changes in physical activity and sedentary behaviours from before to during the COVID-19 pandemic lockdown: a systematic review. BMJ Open Sport Exerc Med 2021 Feb 01;7(1):e000960 [FREE Full text] [doi: 10.1136/bmjsem-2020-000960] [Medline: 34192010]
29. Booth FW, Roberts CK, Laye MJ. Lack of exercise is a major cause of chronic diseases. Compr Physiol 2012 Apr;2(2):1143-1211 [FREE Full text] [doi: 10.1002/cphy.c110025] [Medline: 23798298]
30. Albanese E, Launer LJ, Egger M, Prince MJ, Giannakopoulos P, Wolters FJ, et al. Body mass index in midlife and dementia: Systematic review and meta-regression analysis of 589,649 men and women followed in longitudinal studies. Alzheimers Dement (Amst) 2017;8:165-178 [FREE Full text] [doi: 10.1016/j.dadm.2017.05.007] [Medline: 28761927]
31. LaMonte MJ, Durstine JL, Addy CL, Irwin ML, Ainsworth BE. Physical activity, physical fitness, and Framingham 10-year risk score: the cross-cultural activity participation study. J Cardiopulm Rehabil 2001;21(2):63-70. [doi: 10.1097/00008483-200103000-00001] [Medline: 11314285]
32. Xu X, Griva K, Koh M, Lum E, Tan WS, Thng S, et al. Creating a smartphone app for caregivers of children with atopic dermatitis with caregivers, health care professionals, and digital health experts: participatory co-design. JMIR Mhealth Uhealth 2020 Oct 29;8(10):e16898 [FREE Full text] [doi: 10.2196/16898] [Medline: 33118949]

33. Hoare E, Stavreski B, Jennings GL, Kingwell BA. Exploring motivation and barriers to physical activity among active and inactive Australian adults. Sports (Basel) 2017 Jun 28;5(3):47 [FREE Full text] [doi: 10.3390/sports5030047] [Medline: 29910407]
34. Mulligan HF, Hale LA, Whitehead L, Baxter GD. Barriers to physical activity for people with long-term neurological conditions: a review study. Adapt Phys Activ Q 2012 Jul;29(3):243-265. [doi: 10.1123/apaq.29.3.243] [Medline: 22811565]
35. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med 2020 Dec 25;54(24):1451-1462 [FREE Full text] [doi: 10.1136/bjsports-2020-102955] [Medline: 33239350]
36. Hartin PJ, Nugent CD, McClean SI, Cleland I, Tschanz JT, Clark CJ, et al. The empowering role of mobile apps in behavior change interventions: The Gray Matters randomized controlled trial. JMIR Mhealth Uhealth 2016 Aug 02;4(3):e93 [FREE Full text] [doi: 10.2196/mhealth.4878] [Medline: 27485822]
37. Munson S, Consolvo S. Exploring goal-setting, rewards, self-monitoring, and sharing to motivate physical activity. 2012 Presented at: 6th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops; May 2012; San Diego, CA p. 21-24. [doi: 10.4108/icst.pervasivehealth.2012.248691]
38. Chatterjee A, Gerdes MW, Martinez SG. Identification of risk factors associated with obesity and overweight-A machine learning overview. Sensors (Basel) 2020 May 11;20(9):2734 [FREE Full text] [doi: 10.3390/s20092734] [Medline: 32403349]
39. McAuley A. Digital health interventions: widening access or widening inequalities? Public Health 2014 Dec;128(12):1118-1120. [doi: 10.1016/j.puhe.2014.10.008] [Medline: 25458115]
40. Ma Q. The effectiveness of requirements prioritization techniques for a medium to large number of requirements: a systematic literature review PhD Thesis. URL: https://tinyurl.com/erwmd9jk [accessed 2021-09-14]
41. Chen Y, Ngo V, Park S. Caring for caregivers: designing for integrality. New York, NY: Association for Computing Machinery; 2013 Presented at: 2013 Conference on Computer Supported Cooperative Work; February 23-27, 2013; San Antonio, TX. [doi: 10.1145/2441776.2441789]

Abbreviations

FURPS+: Functional, Usability, Reliability, Performance and additional requirements
MoSCoW: "Must Have, Should Have, Could Have, Won't Have this time" prioritization technique
TTM: transtheoretical model of behavioral change

Edited by G Eysenbach; submitted 23.02.21; peer-reviewed by S Rottenberg, A Chatterjee; comments to author 22.03.21; revised version received 30.04.21; accepted 04.08.21; published 01.10.21.

Please cite as:
Egan KJ, Hodgson W, Dunlop MD, Imperatore G, Kirk A, Maguire R
A Novel Mobile App ("CareFit") to Support Informal Caregivers to Undertake Regular Physical Activity From Home During and Beyond COVID-19 Restrictions: Co-design and Prototype Development Study
JMIR Form Res 2021;5(10):e27358
URL: https://formative.jmir.org/2021/10/e27358
doi:10.2196/27358
PMID:34406969

Original Paper

Use of a Guided Imagery Mobile App (See Me Serene) to Reduce COVID-19–Related Stress: Pilot Feasibility Study

Judith S Gordon[1], PhD; David Sbarra[2], PhD; Julie Armin[3], PhD; Thaddeus W W Pace[1], PhD; Chris Gniady[4], PhD; Yessenya Barraza[1], BS, BA

[1]College of Nursing, University of Arizona, Tucson, AZ, United States

[2]Department of Psychology, University of Arizona, Tucson, AZ, United States

[3]Department of Family and Community of Medicine, College of Medicine, University of Arizona, Tucson, AZ, United States

[4]Department of Computer Science, University of Arizona, Tucson, AZ, United States

Corresponding Author:
Judith S Gordon, PhD
College of Nursing
University of Arizona
1305 North Martin Avenue
Tucson, AZ, 85721
United States
Phone: 1 5206264970
Email: judithg@email.arizona.edu

Abstract

Background: The SARS-CoV-2 pandemic has led to concerns about mental health resulting from regional and national lockdowns, social isolation, job loss, and concern about disease exposure.

Objective: We describe results of the pilot feasibility study of the See Me Serene mHealth app. The app provides users with immersive, vivid, nature experiences to reduce stress and anxiety related to COVID-19 and other isolation. The goals of the study were to develop the See Me Serene app and test the feasibility and acceptability of study procedures, and explore the potential impact of the app on stress and anxiety.

Methods: We developed and tested the See Me Serene app and our study procedures for feasibility, and gathered preliminary data with a goal of 100 participants. The research was conducted in 2 phases: (1) development and internal testing of the app; and (2) feasibility and pilot testing with participants recruited online through earned media (eg, news stories), presentations at a university campus, and social media (eg, online sharing of earned media and presentations). The feasibility study employed a mixed methods, within-subjects, pre-/posttest design. At baseline and 30-day follow-up, we assessed stress-related variables via validated self-report measures and saliva samples for determination of cortisol concentrations.

Results: We met or surpassed all our feasibility benchmarks for recruitment (101 participants recruited), retention (91% [90/99] of 30-day assessment completed), and data collection (99 participants completed all baseline data; 85% [84/99] of salivary cortisol samples returned). Participants adhered to the intervention. On average, participants listened to 48.2 audio files over 30 days or approximately 1.6 audio files per day. Participants were satisfied with the app, with 87% (78/90) rating the app as helpful in dealing with stress and anxiety. The app showed the potential to reduce stress, anxiety, loneliness, and worry. We did not find significant differences (P=.41) in cortisol levels over time. Our findings suggest that future research is warranted to test the efficacy of the See Me Serene app with a representative, diverse sample.

Conclusions: There is a need for evidence-based and easily disseminable stress-reduction interventions. See Me Serene is a feasible intervention and has the potential to reduce stress related to COVID-19 and other forms of social isolation. More research on See Me Serene is warranted.

(JMIR Form Res 2021;5(10):e32353) doi:10.2196/32353

KEYWORDS

COVID-19; stress; anxiety; isolation; intervention; guided imagery; mobile app

Introduction

The SARS-CoV-2 pandemic is a major social upheaval associated with considerable lifestyle changes and economic burdens around the globe—including lockdowns, social isolation, and job loss. The pandemic has hastened concerns about mental illness and emotional distress worldwide, especially among frontline and essential workers who are directly impacted by disease exposures and the psychological sequelae of these potential exposures [1]. A recent meta-analytic work suggested that the prevalence of major depressive disorder has increased substantially since the inception of the pandemic [2]. New evidence suggests that rates of loneliness, the subjective experience of social isolation, have increased dramatically during the COVID-19 pandemic [2-4]. The psychological consequences of long-term home confinement and quarantine are extensive, including anxiety, stress, and insomnia, among other aversive states [5]. It is well known that these psychological states have biological correlates [6] and stress-related psychological changes (eg, disruption of diurnal cortisol rhythm) are associated with risk for a range of poor health outcomes, including cancer, heart disease, and metabolic disease [7]. The need to address the behavioral health consequences of COVID-19 is urgent.

Cortisol is a hormone commonly measured as a biomarker of stress [8]. Diurnal cortisol rhythm is the fluctuation of circulating cortisol over the course of the day, which can be assessed by measuring cortisol concentrations in saliva [9]. Saliva concentrations of cortisol are highest in the morning just before waking, and lowest in the evening before bedtime. Greater cortisol concentrations are available in the morning when it is needed to initiate key physiological processes (eg, metabolism, central nervous system arousal). Diurnal cortisol rhythm is distinct from other patterns of cortisol secretion (eg, phasic responses to acute stress challenge, or a meal). Relative disruptions in diurnal cortisol rhythm (ie, less change over the course of the day) have been associated with chronic conditions (eg, asthma, obesity, and major depression) [7], while a more dynamic change in saliva concentrations of cortisol from morning to evening has been associated with improved physical well-being [10].

Guided imagery can reduce anxiety, chronic pain, and change lifestyle behaviors [11-23]. Guided imagery is unique from mindfulness meditation in that it is an immersive experience in which participants use all their senses and emotions to imagine a scene or situation. In our previous studies, we delivered guided imagery interventions using digital health technologies, including an mHealth app (*See Me Smoke-Free;* NCI, grant number R21CA174639) for promoting smoking cessation, healthy eating, and physical activity, and via telephone (*Be Smoke Free*; NCCIH, grant number R34AT008947) [24,25]. Guided imagery for stress reduction may appeal to both those who embrace other forms of meditation for relaxation and those who use guided imagery for improving athletic performance [26,27].

Stay-at-home orders, lockdowns, and other forms of social isolation may remove people from nature and the outdoors [28]. Several recent reviews indicate that exposure to nature has a positive effect on stress, depression, and anxiety [29-31]. In a systematic review of 12 randomized controlled trials and cross-sectional studies, Shuda and colleagues [31] found that the exposure to nature reduced both self-reported stress (as measured by the Perceived Stress Scale [PSS] and other validated instruments) and physiologic stress (as measured by salivary cortisol, blood pressure, functional magnetic resonance imaging, skin conductance, heart rate variability, and other standardized measures). In a literature review of 113 articles, Berto [29] summarized that exposure to natural scenes mediates the negative effects of stress and the cognitive deficits associated with stress, and improves mood. Corazon and colleagues [30] conducted a systematic review of 36 studies conducted between 2012 and 2018 on the effects of outdoor, nature-based exposure on stress. They found that interventions focused on outdoor exposures showed strong evidence for positive outcomes on perceived stress, while the results for physiologic measures were more varied [30]. Overall, the studies contained in these reviews indicate that there is evidence to support the use of outdoor imagery to reduce perceived and physiologic stress [29-31]. The definition of outdoor, nature-based experiences is broad. Therefore, they could include more traditional nature scenes (eg, mountains, beach) as well as more "adventure-oriented" scenes (eg, playing golf, riding a bicycle) which may be relaxing to specific individuals.

Although technologies exist to provide immersive experiences (eg, virtual reality), they require specialized equipment and can be cost-prohibitive. Combining guided imagery with smartphone technology is a low-cost, scalable, and disseminable method for allowing people to engage in evocative, immersive experiences in nature—from sitting in a forest to cycling on a bike path. The vast majority of Americans across age, race/ethnicity, and socioeconomic status own smartphones; and of those, most use mobile apps [32,33]. Several studies have shown that people experiencing stress are more likely to use meditation apps (eg, Calm) [34-37]. Using a simple mHealth app to deliver guided imagery audio files has the potential for simulating outdoor experiences for people experiencing lockdowns and social isolation, as well as for individuals who are homebound due to other reasons.

Therefore, we developed and tested the feasibility and potential impact of the *See Me Serene* mHealth app. See Me Serene provides users with immersive, vivid, nature experiences designed to reduce stress and anxiety related to social isolation. Herein, we describe the results of this pilot feasibility study.

Methods

Study Overview

We conducted the pilot study between May 1, 2020, and December 31, 2020. The study was approved by the University of Arizona Institutional Review Board (Protocol #2005625231). The primary goals of the study were to develop the See Me Serene app and test our study procedures for feasibility. Our secondary goal was to gather preliminary data regarding the potential impact of the app on reducing perceived and physiologic stress; the study was not designed nor powered to

test efficacy. The research was conducted in 2 phases: (1) development and internal testing of the app; and (2) feasibility and pilot testing with participants recruited online through earned media (eg, new stories), presentations on a university campus, and social media (eg, online sharing of earned media and presentations). The feasibility study employed a mixed methods, within-subjects, pre-/posttest design. Feasibility outcomes included metrics for participant recruitment (goal of 100 within 2 months), participant retention (at least 75% at 1 month), and data collection (at least 90% at baseline and 75% at the 1-month follow-up). Exploratory outcomes included changes in self-reported measures of stress, anxiety, loneliness, and worry, and diurnal cortisol rhythm (collected via salivary cortisol assays) from baseline to 1-month follow-up. We also collected consumer satisfaction at the 1-month assessment.

Intervention Development and Description

Program Development

In Phase 1, we developed and conducted internal testing on the See Me Serene app. The app was developed based on feedback regarding guided imagery for relaxation obtained in a previous study [25]. Initial guided imagery scripts were reviewed by the investigators and project staff and vetted by an advisory board that consisted of faculty, staff, and students at the University of Arizona. During this process, the advisory board suggested the addition of outdoor activities to relaxing nature scenes to enhance the appeal of the app to a more diverse population of users and differentiate See Me Serene from other relaxation apps in the marketplace. The app was programmed by undergraduate students in the Department of Computer Science with oversight by the investigators (CG and JG). We developed Android and iOS versions of the app using the Thunkable platform, which allows simple design and app development. This process streamlined development and accelerated deployment by automatically generating native apps for both Android and iOS platforms. We used Google's Firebase which provides database and authentication services targeted for mobile apps. The app audio and image files were stored in the Firebase database along with user preferences. This allowed for easier updates of guided imagery files without requiring app updates. Beta testing of the See Me Serene app was conducted by the project team and members of the Advisory Board. Once complete, the app was deployed to the App Store and the Google Play Store. See Me Serene was updated twice during the study to fix bugs and make improvements to usability.

See Me Serene Description

The See Me Serene app allows users to choose from a variety of guided imagery audio files and associated photos of immersive, evocative nature scenes. Users can select audio files from guided imagery categories such as "outdoor adventures" (eg, motorcycling along a country road, hiking in the desert, fishing) and "serene scenes" (eg, in the mountains, at the lake, at the beach). Each file contains detailed, vivid descriptions of the scenarios, including sights, sounds, smells, tastes, tactile sensations, and emotions. Each audio file starts with a brief relaxing breathing exercise and instructions to "release any tension in your body and mind." The audio files in the app, which were tested during this study, included soft background music and were approximately 5 minutes in length. Users receive notifications to listen to the files once each day. See Me Serene tracks the user's mood each time the user logs in with 4 questions assessing how often/much the user has felt stressed, anxious, lonely, or worried "today" (0=not at all to 4=very often/extremely). The app tracks and displays the user's profile information and the number of guided imagery audio files listened to. See Me Serene provides a list of clickable links to free or low-cost, evidence-based, mental health resources, some of which are available 24/7. The app also contains a list of frequently asked questions and answers addressing common technical or content-related questions, and an About Us page describing the study team.

Feasibility Pilot Study

In Phase 2, we conducted a within-subjects, pre-/posttest, pilot feasibility study to recruit a convenience sample of 100 participants recruited nationally through earned media (eg, news stories), presentations on a university campus, and social media (eg, sharing of news stories and presentations online). Recruitment took place between August 3, 2020, and October 5, 2020. Potential participants were eligible for the study if they were over aged 18; had a valid phone number and email address; owned a smartphone with internet access; and spoke English. Potential participants were excluded if they had ever been told by a health care provider that they had an adrenal gland disorder such as Cushing disease or Addison disease, or if they regularly used any corticosteroid medications such as hydrocortisone, prednisone, or fludrocortisone acetate. Recruitment of the convenience sample took place via earned media (ie, stories carried by local, regional, and national news outlets on television, radio, and the internet), presentations at a university campus, and social media (ie, reposting of earned media stories and presentations). Recruitment materials described guided imagery and the See Me Serene app in general terms and invited potential participants to visit the project website for more information or download the app from the App Store or the Google Play Store. After signing into the app, users were given the option to participate in the study or just use the app.

Participants completed a baseline assessment online or by phone. We also collected salivary cortisol (a biomarker of stress) samples pre- and posttest in the home setting. Saliva sample kits were sent to participants by overnight mail. Participants were instructed to collect 2 samples per day (1 in the morning immediately after waking, and 1 at night right before bed) for 2 consecutive days, and record the time of day when they collected their samples on an instructions form that was included in the saliva sample kits. Participants were instructed to not brush their teeth, eat, exercise, smoke, take a cold shower, or consume alcohol or caffeine until after they had collected a saliva sample. We also instructed participants to not collect a sample until at least 60 minutes after eating dinner in the evening. After samples were collected, participants stored them in their home freezer until the posttest, at which time they repeated the collection process and returned all 8 samples and instructions form with collection times to the research team by prepaid, overnight mail. Study staff provided reminders to participants to collect samples at both baseline and follow-up.

Participants were instructed to use the app at least once each day for 4 weeks. Follow-up assessments (online or telephone-delivered self-report surveys and salivary cortisol assays) were conducted at 30 days after access to the app. Use of the app was collected automatically within the Firebase database (in which specific app use data were stored), and through Google Analytics and Apple Developer Tools (eg, app downloads and current users). Upon completion of all study activities, participants received a US $50 electronic gift card.

Assessments

Data Collected

At baseline, we collected demographics including age, gender, race, ethnicity, education, marital status, number of people and pets in the household, mental health diagnoses, and amount of moderate to vigorous physical activity per week. We also collected data on COVID-19–related variables including stay-at-home orders, essential personnel status, length of restrictions, leave the house every day status, amount of time spent outdoors, and amount of time spent doing fun outdoor activities.

At baseline and 30-day follow-up, we assessed stress-related variables via validated self-report measures and saliva for determination of cortisol concentrations.

Perceived Stress Scale (PSS)

Self-reported stress was measured via the PSS [38]. The PSS is a 10-item scale that assesses the perception of psychological stress and the degree to which people feel events in their life are unpredictable and uncontrollable, as well as the extent to which the events tax their coping resources over the past month. Each item is scored (or reverse scored from 0 to 4), and the total score is summed across all items. Population norms for the PSS are mean (SD) 12.1 (5.9) for males and 13.7 (6.6) for females. The PSS has excellent psychometric properties and is widely used.

Overall Anxiety Severity and Impairment Scale (OASIS)

The Overall Anxiety Severity and Impairment Scale (OASIS) is a 5-item scale for assessing anxiety symptoms associated with multiple anxiety disorders [39]. Each item is measured on a 5-point Likert scale (0=none/no anxiety to 4=constant/extreme/all the time anxiety) with respondents reflecting on their symptoms in the past week. Participants rate how often they have felt anxious in the past week, the intensity of their anxiety, their degree of avoidance of situations and activities, and the degree of interference in daily activities and social relationships. The items are summed for a total score. In the validation research, a score of 8 or greater correctly classified 87% of the sample as having an anxiety disorder or not [40].

Impact of Events Scale—Revised (IES-R)

The Impact of Events Scale—Revised (IES-R) [41] assesses subjective emotional distress, typically over the prior 7 days, following a traumatic or stressful event. Our focus was on the COVID-19 pandemic. Example items include statements such as, “Any reminder brought back feelings about it,” and “I feel irritable and angry.” The scale has 22 questions using a 5-point Likert scale rating system ranging from 0 (Not at all) to 4 (Extremely). The total IES-R has high internal consistency, and the measure is widely used in the literature. On the IES-R, a cutoff score of 1.5 (or a total score of 33) provides the highest diagnostic power with a sensitivity of 0.91 against the Posttraumatic Stress Disorder (PTSD) checklist [41].

University of California Los Angeles Loneliness Scale

The 3-item University of California Los Angeles (UCLA) Loneliness Scale [42] includes the following items, “How often do you feel you lack companionship?” “How often do you feel left out?,” and “How often do you feel isolated from others?.” The items are scored on a 3-point Likert scale from “hardly ever” to “often.” This scale has excellent psychometric properties and is widely used in the literature.

The Brief Penn State Worry Questionnaire (PSWQ)

The ultra-brief Penn State Worry Questionnaire (PSWQ) [43] was used to measure worry and consists of 3 validated items on a 5-point scale, with individual items scored between 0 and 5 and a composite score of 0-15 (0=no worry and 15=high worry). The items focus on the situations that make people work, the extent of the worry, and the controllability of worry. In a treatment-seeking sample, the mean of the brief PSWQ was 10.42 (SD 3.44).

Cortisol

Finally, we assessed diurnal rhythm changes over time in cortisol, a biomarker of stress, via saliva concentrations of cortisol determined with enzyme immunoassay kits (obtained from Salimetrics), according to manufacturer instructions. Samples were assayed in duplicate per sample, and the lower limit of sensitivity of the assay was 0.006 µg/dL. Intra- and interassay coefficients of variation of the cortisol enzyme immunoassay were 9.6% and 4.7%, respectively.

Guided Imagery and Other Meditation

We also measured the use of guided imagery (number of minutes per day in the last week); use of breathing, meditation, or progressive relaxation (number of minutes per day in the last week); and 2 items adapted from the Borkovec and Nau Credibility/Expectancy Questionnaire [44] regarding perceived confidence in guided imagery to reduce stress, and perceived logic for the use of guided imagery to reduce stress (rated on a 5-point Likert scale, where 1=very unconfident/very illogical to 5=very confident/very logical) [24,25].

Physical Activity

Exercise (past week) was measured using the 3-item Godin Leisure-Time Exercise Questionnaire (LTEQ) [45,46]. The LTEQ is sensitive enough to differentiate mild, moderate, and strenuous exercise, can be scored in total number of minutes of activity or as a metabolic equivalents, and differentiates between household, occupational, and leisure-time physical activity and seasonal variations [47].

User Satisfaction

At the 30-day follow-up, we also assessed usability and participant satisfaction with the app. We assessed 6 items of usability, including: (1) “How organized was the app?”; (2) “How much did you like the format/design of the app?”; (3)

"How easy was the app to use?"; (4) "How well did the app address issues you have with anxiety/stress?"; (5) "How much new information did this app provide?"; and (6) "How helpful did you find the features of this app?"

We asked 3 satisfaction items: (1) "Overall, how would you rate this app in helping you deal with stress/anxiety?"; (2) "How likely would you be to recommend this app to other people?"; and (3) "How likely are you to continue using this app?" We also elicited open-ended feedback about the app with the question, "What suggestions do you have for improving the app?"

App Use

We automatically collected use of the app, including number of guided imagery audio files listened to and number of times answering the daily questions. We also tracked the number of downloads and current users from Google Analytics and Apple Developer Tools.

Feasibility

Finally, we assessed feasibility outcomes by achievement of our feasibility benchmarks, including meeting the recruitment/accrual goal of 100 participants within 2 months; collection of at least 90% of baseline self-report data; collection of at least 75% of follow-up self-report surveys and cortisol samples; and at least 75% participant retention at 30 days.

Analyses

Study data were collected and managed using REDCap electronic data capture tools hosted at The University of Arizona [48,49]. REDCap (Research Electronic Data Capture) is a secure, web-based software platform designed to support data capture for research studies, providing (1) an intuitive interface for validated data capture; (2) audit trails for tracking data manipulation and export procedures; (3) automated export procedures for seamless data downloads to common statistical packages; and (4) procedures for data integration and interoperability with external sources. Descriptive statistics were compiled using the REDCap reports and stats function. Our feasibility analyses focused on change in the main psychological outcome variables over time using paired-samples *t* tests. Open-ended responses to the question, "What suggestions do you have to improve the app?" were summarized using a matrix analysis approach, which is an efficient method of descriptively organizing and interpreting qualitative findings for program integration [50]. For diurnal cortisol rhythm, we computed the slope of the change in cortisol from morning to afternoon to evening for each collection day, taking into consideration the time of sample collection [51]. In cases where sample collection time was missing, average sample collection time was used [51]. We then averaged cortisol slopes across both collection days to develop a metric of diurnal cortisol rhythm.

Results

Participant Characteristics

We recruited 101 participants from August 3, 2020, to October 5, 2020. Following enrollment, 2 participants withdrew from the study (1 due to technical issues and 1 due to a family emergency), leaving 99 participants with complete data at baseline. As displayed in Table 1, the majority of participants were non-Hispanic White (83/99, 84%) females (85/99, 86%) who had at least some college education (93/99, 94%), and were married or living with a partner (66/99, 67%). The average age of participants was 51.4 (SD 14.7) years. COVID-19 restrictions affected 68% (67/99) of participants, and of those under stay-at-home orders, 92% (61/66) reported that they had been staying at home for more than 3 months. Participants reported diagnoses of generalized anxiety disorder (24/99, 24%), major depression (18/99, 18%), and PTSD (12/99, 12%). Participants had an average score of 3.03 (SD 0.54) on the PSS; 8.05 (SD 4.03) on the OASIS; 2.29 (SD 0.58) on the IES-R; 2.73 (SD 0.77) on the UCLA Loneliness Scale; and 8.94 (SD 3.15) on the PSWQ. Most participants reported using less than 10 minutes of guided imagery (90/99, 91%) or other relaxation techniques (73/99, 74%) per day in the previous week. Still, they expressed confidence in the use of guided imagery for reducing stress (86/99, 87%) and found it logical to use guided imagery for stress reduction (93/99, 94%).

Table 1. Baseline participant characteristics[a].

Characteristics	Values[b]
Age (years), mean (SD)	51.4 (14.7)
Female sex	85 (86)
Non-Hispanic White	83 (84)
Hispanic ethnicity	17 (17)
At least some college	93 (94)
Married or living with a partner	66 (67)
Generalized anxiety disorder	24 (24)
Major depression	18 (18)
Posttraumatic stress disorder	12 (12)
At least two people in household	79 (80)
At least one pet in household	68 (69)
Currently under stay-at-home order or restricting where you go	67 (68)
Staying at home for more than 3 months (N=66)	61 (92)
Not exempt from stay-at-home orders	73 (74)
Leave house each day (sometimes to often)	61 (62)
Go outdoors (sometimes to often)	86 (87)
Do fun outdoor activities (sometimes to often)	53 (54)
Get <3 hours of moderate/vigorous physical activity per week	65 (66)
Perceived Stress Scale score, mean (SD)	19.16 (4.58)
Overall Anxiety Severity and Impairment Scale score, mean (SD)	8.05 (4.03)
Impact of Events Scale—Revised score, mean (SD)	2.29 (0.58)
University of California Los Angeles Loneliness Scale score, mean (SD)	2.73 (0.77)
Brief Penn State Worry Questionnaire score, mean (SD)	8.94 (3.15)
Morning cortisol (µg/dL), mean (SD)	0.33 (0.36)
Evening cortisol (µg/dL), mean (SD)	0.05 (0.03)
Cortisol slope, mean (SD)	–0.43 (0.47)
<10 minutes per day in the past week practiced guided imagery	90 (91)
<10 minutes per day in the past week practiced other relaxation techniques	73 (74)
Confidence in guided imagery to reduce stress (somewhat to very)	86 (87)
Logical to use guided imagery to reduce stress (somewhat to very)	93 (94)

[a]N=99 unless otherwise noted.

[b]Values shown are n (%) unless otherwise noted.

Feasibility

As displayed in Table 2, we met or surpassed all our feasibility benchmarks. We were able to recruit 100 participants within 2 months, collected 98% (99/101) of baseline survey data, 91% (90/99) of 30-day surveys, and 85% (84/99) of cortisol samples were returned. Five samples contained some missing data and were excluded from analyses for a total of 80 (79/99) usable samples. Using a conservative definition of retention (complete survey and cortisol data at 30 days), we surpassed our goal with an 85% (84/99) retention rate.

Table 2. Feasibility outcomes.

Outcome	Value, n/N (%)	Feasibility criteria	Feasible?
Participant recruitment, n	101	Enroll 100 participants in 2 months	Yes
Baseline surveys	99/101 (98)	Collect at least 90% of survey data	Yes
30-Day surveys	90/99 (91)	Collect at least 75% of survey data	Yes
Cortisol samples returned	84/99 (85)	Collect at least 75% of cortisol data	Yes
Participant retention (conservatively defined as those with complete data at 30 days)	84/99 (85)	At least 75% participant retention at 30 days	Yes

Adherence

Adherence was measured by the number of times participants listened to guided imagery audio files during the 4-week intervention period. On average, participants listened to 48.2 audio files (range 0-280) over 30 days, or approximately 1.6 audio files per day. This finding establishes the feasibility of listening to at least one audio file per day as instructed. We also found a significant increase in time spent each day using guided imagery (t_{89}=–9.94, P<.001) and other forms of meditation (t_{89}=–7.34, P<.001).

Usability and Satisfaction

Overall, the 90 participants retained at the 30-day follow-up found the app usable and were satisfied with it. See Table 3 for individual ratings of usability and satisfaction. Suggestions for improving the app included increasing audio file length, providing more variety in the guided imagery scenarios (eg, more nature-focused guided imagery), customizing the narrator's voice and background music or sounds, allowing users to save their favorite audio files to a library within the app, and providing users the ability to track their use of the app and easily return to where they were.

Table 3. Participant ratings (somewhat to very) of usability and satisfaction with the app (N=90).

Question	Values, n (%)
How organized did you find this app?	88 (98)
How much did you like the format/design of the app?	84 (93)
How easy was the app to use?	88 (98)
How well did the app address issues you have with anxiety/stress?	81 (90)
How much new information did the app provide?	66 (73)
How helpful did you find the features of this app?	80 (89)
Overall, how would you rate this mobile app in helping you deal with stress/anxiety?	78 (87)
How likely would you be to recommend this app to others?	71 (79)
How likely are you to continue using this app?	66 (73)

Potential Impact on Outcome Measures

As shown in Table 4, we found preliminary evidence that the See Me Serene app may have the potential to reduce anxiety. Although this study was not designed to test efficacy, we found significant reductions in all outcome variables with small to medium effect sizes. Participants reported reductions in self-reported stress (t_{89}=3.38, P=.001), symptoms of PTSD (t_{89}=4.02, P<.001), anxiety (t_{89}=3.80, P<.001), loneliness (t_{89}=3.74, P<.001), and worry (t_{89}=2.77, P=.007) from pre- to posttest. We did not find significant differences (P=.41) in cortisol levels over time.

Table 4. Changes in outcome variables pre-/post test.

Measure	Preintervention, mean (SD)	Postintervention, mean (SD)	*t*	*df*	*P*	Cohen *d*
Perceived Stress Scale score	19.16 (4.58)	17.5 (5.11)	3.38	89	.001	0.37
Overall Anxiety Severity and Impairment Scale score	8.05 (4.02)	6.63 (3.44)	3.80	89	<.001	0.40
Impact of Events Scale—Revised score	2.29 (0.58)	2.01 (0.51)	4.02	89	<.001	0.42
University of California Los Angeles Loneliness Scale score	2.73 (0.77)	2.51 (0.85)	3.74	89	<.001	0.40
Brief Penn State Worry Questionnaire score	8.95 (3.15)	8.82 (2.79)	2.77	89	.007	0.29
Cortisol (morning)	0.33 (0.36)	0.48 (1.35)	–0.99	79	.32	–0.11
Cortisol (evening) average	0.06 (0.08)	0.14 (0.62)	–1.25	80	.21	–0.14
Average cortisol slope	–0.44 (0.47)	–0.54 (1.22)	0.81	79	.41	0.09

Discussion

Principal Results

Feasibility

The results of this study suggest that the See Me Serene app is feasible to deliver and test. Our results appear to be similar to other studies that have found mHealth apps feasible to deliver and easy to disseminate [52-55]. We met or surpassed all of our feasibility benchmarks, including the recruitment and retention of participants and the collection of both self-report data and biochemical samples. Our ability to easily recruit participants solely through earned media indicates the great need for novel, easily disseminable, stress-reduction interventions. Our high retention and data collection rates suggest that participants are eager to help find solutions for managing stress. Finally, we found that it was feasible to conduct the study entirely remotely, and to collect biological samples even during a pandemic when the logistics of mailing and receiving shipments were challenging.

Participants were adherent to the study protocol and listened to an average of 2 files each day, surpassing the minimum requirement of listening to 1 audio file per day. Our results compare favorably with other studies that have shown modest adherence or engagement with mHealth stress management apps [52,55,56]. Based on participant feedback, our adherence results may have been due in part to the length of the guided imagery audio files because each file was approximately 5 minutes long. The length of the guided imagery was based on those used in our previous studies [24,25]. However, participants expressed a desire for longer audio files, and they may have listened to multiple files or the same file multiple times each day to achieve this result. Participants significantly increased the number of minutes per day in the last week that they practiced guided imagery as well as other forms of relaxation, such as meditation, deep breathing, or progressive muscle relaxation. The guided imagery did include a brief breathing exercise at the beginning of each audio file but did not focus solely on breathing, nor did the audio files teach meditation or progressive muscle relaxation. Therefore, it appears that the use of this app may have encouraged users to seek out and engage in other forms of relaxation techniques as well.

Usability and Satisfaction

Overall, participants found See Me Serene easy to use and useful. Our results were similar to previous studies of mHealth apps aimed at reducing stress, anxiety, and depression among a variety of different populations [55-59]. There were few reported technical issues, which were resolved with updates to the app. Although 87% (78/90) of participants found the app helpful in reducing stress and would recommend the app to others, only 73% (66/90) reported that they would continue to use the app. This may have been due to several factors: (1) stay-at-home orders were lifted during the study, allowing for less isolation; (2) easing of restrictions may have reduced the stress faced by the participants, thus they had less need for the app; (3) the app was effective at reducing participants' stress, thus reducing their need to use the app; (4) participants may have wanted changes to the app to enhance their experience, as suggested by the qualitative data.

The open-ended question on the follow-up survey asking for suggestions provided the team with input regarding app revisions and offered insight into how the app was being used by participants, which can help us tailor revisions to their needs. For instance, several participants noted that they used the audio files before bed to relax for sleep. These individuals generally wanted longer audio files to include music that would continue after the guided imagery stopped. In general, the comments supported longer guided imagery files that allowed listeners to practice breathing and develop the mental imagery described in the narration. This recommendation may result from participants' lack of familiarity with guided imagery (eg, only 9% [9/99] reported using guided imagery more than 10 minutes a day at baseline), and the revision could make the app content more accessible to users who are not familiar with guided imagery. Participants also wanted more variety of guided imagery, and the ability to personalize the app to their needs. Providing more and different types of guided imagery, and features such as the ability to choose the narrator's voice, save favorites, and track mood and audios listened to over time may improve engagement with the app and encourage continued use.

Potential Impact

Although this study was not designed to test efficacy, we found significant reductions in all outcome variables with small to

medium effect sizes. The evidence regarding effectiveness of mHealth apps is limited and mixed [53,54,56-64]. While several randomized controlled trials have shown these apps to be effective at reducing stress, anxiety, depression, or PTSD symptoms [53,54,57,59], some have not reported positive results [60,61,64]. In addition, many studies report only significant pre–post results [56,58,62,63], thus limiting generalizability of the findings. Finally, several reviews have found that the majority of currently available mHealth apps are not evidence based, and lack a strong theoretical basis and research to support efficacy [65-68].

Although our study was not designed to test efficacy, we found significant reductions in all outcome variables with small to medium effect sizes. Participants in our study reported reductions in self-reported stress, symptoms associated with PTSD and anxiety, loneliness, and worry. However, this was an open within-subjects design which did not have a control group, and we recruited a convenience sample. Therefore, participants knew the purpose of the app and that it used guided imagery. For example, participants may have been more inclined to use guided imagery versus other types of stress reduction methods. At baseline, almost all participants reported confidence in and logic of guided imagery to manage stress.

We did not find significant differences in cortisol levels over time. This could be because of the relatively heterogeneous sample with respect to age and baseline psychological well-being characteristics, as well as the possibility that life stressors associated with the pandemic (ie, social isolation, limited time outdoors, myriad sources of psychological distress) may have prevented the intervention from exerting an effect on diurnal cortisol rhythm that could be detected across the whole sample, especially in the short term. Controlling or stratifying for psychological and physical well-being and relevant sources of stress, including later assessment time points and increasing the sample size, would all likely help reveal intervention efficacy at the level of diurnal cortisol rhythm [9,10]. However, it is possible that a “light touch” intervention such as ours may not produce immediate physiologic changes in cortisol. This variability in physiologic outcomes is consistent with the results of the systematic review by Corazon and colleagues [30], suggesting the need for more research. Although our findings are promising, more research is needed to determine whether See Me Serene is effective at reducing stress-related symptoms and objective measures of stress-related biology.

Limitations

This study had several limitations. First, we used a within-subjects design without a control group, which limited our ability to test the feasibility of an active control condition as well as to determine efficacy as a result of specific (as opposed to nonspecific) intervention components. The research team has already developed a control app that will serve as an active attention comparator, and a pilot study is planned. In addition, we were not powered to detect differences in mood or other outcomes based on usage of different types of audio files (eg, “adventure scenes” versus “relaxing scenes”). In the future, we plan to conduct a fully powered efficacy trial which will allow for subanalyses. Second, we recruited a convenience sample of participants, which limits the generalizability of the findings. Participants were primarily college educated, White, and women. The lack of diversity may have resulted from reliance on earned media (eg, news stories) and presentations on a university campus due to budget constraints. Future studies will employ methods aimed at recruiting a more diverse sample, including more males and racial/ethnic groups. We will actively recruit low-income and racial/ethnic minority participants to demonstrate feasibility and usability of the See Me Serene app with these groups. We have used these methods successfully in previous trials to yield a more representative sample. Third, we followed participants for only 1 month. Subsequent studies will follow participants for longer periods to assess the duration of the intervention effect. We have successfully followed participants for 6 months in our previous trials [24,25]. Fourth, this study was conducted during the COVID-19 pandemic, which caused some delays and logistic challenges with the collection and return of the saliva samples to determine cortisol concentrations. However, overall, these challenges (eg, managing overnight shipping, providing replacement saliva collection supplies) were overcome, with the result being a large percentage of saliva samples collected and returned in a manner acceptable for laboratory analysis. Finally, we could not collect data from the 9 participants who dropped out of the study regarding why they did not complete the study or what they thought about the app. However, a 9% (9/99) dropout rate is very low compared with that of many mHealth studies which often have higher dropout rates than in-person studies.

Conclusions

The study results suggest that delivering and testing the See Me Serene mHealth app is feasible and that the app has the potential to reduce stress related to COVID-19 and other forms of social isolation. We learned valuable lessons during the pilot that can be applied to future studies. There is a great need for evidence-based and easily disseminable stress-reduction interventions, as the majority of available stress management apps are not evidence based. More research is warranted to test the efficacy of See Me Serene.

Acknowledgments

This study was funded by the University of Arizona, TRIF BIO5 Seed Grant number 5833210 (to JG and CG, multiple principal investigators). We gratefully acknowledge the contributions of Yessenya Barraza, Chelsey Bergmann, Chloe Dillingham, Jacob Rafko, Lauren Taylor, and Jialiang Zhang, without whom this project would not have been successful.

Authors' Contributions

All authors contributed to the writing of this manuscript.

Conflicts of Interest

None declared.

References

1. Greenberg N, Docherty M, Gnanapragasam S, Wessely S. Managing mental health challenges faced by healthcare workers during covid-19 pandemic. BMJ 2020 Mar 26;368:m1211. [doi: 10.1136/bmj.m1211] [Medline: 32217624]
2. Bueno-Notivol J, Gracia-García P, Olaya B, Lasheras I, López-Antón R, Santabárbara J. Prevalence of depression during the COVID-19 outbreak: A meta-analysis of community-based studies. Int J Clin Health Psychol 2021 Jan;21(1):100196 [FREE Full text] [doi: 10.1016/j.ijchp.2020.07.007] [Medline: 32904715]
3. Mencacci C, Salvi V. Expected effects of COVID-19 outbreak on depression incidence in Italy. J Affect Disord 2021 Jan 01;278:66-67 [FREE Full text] [doi: 10.1016/j.jad.2020.09.043] [Medline: 32950845]
4. Killgore WD, Cloonan SA, Taylor EC, Dailey NS. Loneliness: A signature mental health concern in the era of COVID-19. Psychiatry Res 2020 Aug;290:113117 [FREE Full text] [doi: 10.1016/j.psychres.2020.113117] [Medline: 32480121]
5. Pfefferbaum B, North CS. Mental Health and the Covid-19 Pandemic. N Engl J Med 2020 Aug 06;383(6):510-512. [doi: 10.1056/nejmp2008017]
6. Cohen S, Janicki-Deverts D, Miller GE. Psychological stress and disease. JAMA 2007 Oct 10;298(14):1685-1687. [doi: 10.1001/jama.298.14.1685] [Medline: 17925521]
7. Adam EK, Quinn ME, Tavernier R, McQuillan MT, Dahlke KA, Gilbert KE. Diurnal cortisol slopes and mental and physical health outcomes: A systematic review and meta-analysis. Psychoneuroendocrinology 2017 Sep;83:25-41 [FREE Full text] [doi: 10.1016/j.psyneuen.2017.05.018] [Medline: 28578301]
8. Crosswell AD, Lockwood KG. Best practices for stress measurement: How to measure psychological stress in health research. Health Psychol Open 2020;7(2):2055102920933072 [FREE Full text] [doi: 10.1177/2055102920933072] [Medline: 32704379]
9. Kirschbaum C, Hellhammer DH. Salivary cortisol in psychoneuroendocrine research: Recent developments and applications. Psychoneuroendocrinology 1994 Jan;19(4):313-333. [doi: 10.1016/0306-4530(94)90013-2]
10. Stafford M, Ben-Shlomo Y, Cooper C, Gale C, Gardner MP, Geoffroy M, et al. Diurnal cortisol and mental well-being in middle and older age: evidence from four cohort studies. BMJ Open 2017 Oct 12;7(10):e016085 [FREE Full text] [doi: 10.1136/bmjopen-2017-016085] [Medline: 29025828]
11. Giacobbi P, Long D, Nolan R, Shawley S, Johnson K, Misra R. Guided imagery targeting exercise, food cravings, and stress: a multi-modal randomized feasibility trial. J Behav Med 2018 Feb 1;41(1):87-98 [FREE Full text] [doi: 10.1007/s10865-017-9876-5] [Medline: 28766183]
12. Giacobbi P, Dreisbach KA, Thurlow NM, Anand P, Garcia F. Mental imagery increases self-determined motivation to exercise with university enrolled women: A randomized controlled trial using a peer-based intervention. Psychology of Sport and Exercise 2014 Jul;15(4):374-381. [doi: 10.1016/j.psychsport.2014.03.004]
13. Martin KA, Hall CR. Using mental imagery to enhance intrinsic motivation [Utilisation de l ' imagerie mentale pour ameliorer la motivation intrinseque]. Journal of Sport & Exercise Psychology 1995;17(1):54-69. [doi: 10.1123/jsep.17.1.54]
14. Morris T, Spittle M, Watt AP. Imagery in Sport. Champaign-Urbana, IL: Human Kinetics; 2005.
15. Wynd CA. Guided health imagery for smoking cessation and long-term abstinence. J Nurs Scholarsh 2005 Sep;37(3):245-250. [doi: 10.1111/j.1547-5069.2005.00042.x] [Medline: 16235865]
16. Zernig G, Wallner R, Grohs U, Kriechbaum N, Kemmler G, Saria A. A randomized trial of short psychotherapy versus sustained-release bupropion for smoking cessation. Addiction 2008 Dec;103(12):2024-2031. [doi: 10.1111/j.1360-0443.2008.02348.x] [Medline: 19469746]
17. Posadzki P, Ernst E. Guided imagery for musculoskeletal pain: a systematic review. The Clinical Journal of Pain 2011;27(7):648-653. [doi: 10.1097/ajp.0b013e31821124a5]
18. Forward JB. Effect of structured touch and guided imagery for pain and anxiety in alective joint replacement patients—a randomized controlled trial: M-TIJRP. permj 2015 Nov 2:19. [doi: 10.7812/tpp/14-236]
19. Foji S, Tadayonfar MA, Mohsenpour M, Rakhshani MH. The study of the effect of guided imagery on pain, anxiety and some other hemodynamic factors in patients undergoing coronary angiography. Complement Ther Clin Pract 2015 May;21(2):119-123. [doi: 10.1016/j.ctcp.2015.02.001] [Medline: 25733083]
20. Apóstolo JLA, Kolcaba K. The effects of guided imagery on comfort, depression, anxiety, and stress of psychiatric inpatients with depressive disorders. Arch Psychiatr Nurs 2009 Dec;23(6):403-411. [doi: 10.1016/j.apnu.2008.12.003] [Medline: 19926022]
21. Holden-Lund C. Effects of relaxation with guided imagery on surgical stress and wound healing. Res Nurs Health 1988 Aug;11(4):235-244. [doi: 10.1002/nur.4770110405] [Medline: 3043570]
22. Watanabe E, Fukuda S, Hara H, Maeda Y, Ohira H, Shirakawa T. Differences in relaxation by means of guided imagery in a healthy community sample. Altern Ther Health Med 2006;12(2):60-66. [Medline: 16541998]
23. Hart J. Guided Imagery. Alternative and Complementary Therapies 2008 Dec;14(6):295-299. [doi: 10.1089/act.2008.14604]

24. Gordon JS, Armin J, Hingle MD, Giacobbi P, Cunningham JK, Johnson T, et al. Development and evaluation of the See Me Smoke-Free multi-behavioral mHealth app for women smokers. Transl Behav Med 2017 Jun 2;7(2):172-184 [FREE Full text] [doi: 10.1007/s13142-017-0463-7] [Medline: 28155107]
25. Gordon J, Bell ML, Armin JS, Giacobbi PR, Nair US. A telephone-based guided imagery tobacco cessation intervention: results of a randomized feasibility trial. Transl Behav Med 2021 Mar 16;11(2):516-529 [FREE Full text] [doi: 10.1093/tbm/ibaa052] [Medline: 32542352]
26. Jose J, Joseph MM. Imagery: Its effects and benefits on sports performance and psychological variables: A review study. International Journal of Physiology, Nutrition and Physical Education 2018;3(2):A-193 [FREE Full text]
27. Utay J, Miller M. Guided imagery as an effective therapeutic technique: a brief review of its history and efficacy research. Journal of Instructional Psychology. Journal of Instructional Psychology 2006;33(1):e1 [FREE Full text]
28. Rice W, Mateer TJ, Reigner N, Newman P, Lawhon B, Derrick Taff B. Changes in recreational behaviors of outdoor enthusiasts during the COVID-19 pandemic: analysis across urban and rural communities. Journal of Urban Ecology 2020;6(1):juaa020 [FREE Full text] [doi: 10.1093/jue/juaa020]
29. Berto R. The role of nature in coping with psycho-physiological stress: a literature review on restorativeness. Behav Sci (Basel) 2014 Oct 21;4(4):394-409 [FREE Full text] [doi: 10.3390/bs4040394] [Medline: 25431444]
30. Corazon SS, Sidenius U, Poulsen DV, Gramkow MC, Stigsdotter UK. Psycho-physiological stress recovery in outdoor nature-based interventions: a systematic review of the past eight years of research. Int J Environ Res Public Health 2019 May 16;16(10):1711 [FREE Full text] [doi: 10.3390/ijerph16101711] [Medline: 31100773]
31. Shuda Q, Bougoulias ME, Kass R. Effect of nature exposure on perceived and physiologic stress: A systematic review. Complement Ther Med 2020 Sep;53:102514. [doi: 10.1016/j.ctim.2020.102514] [Medline: 33066853]
32. Pew Research. Mobile Fact Sheet. 2021. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-05-07]
33. Statista. Reach of Leading US Smartphone Apps. 2021. URL: https://www.statista.com/statistics/281605/reach-of-leading-us-smartphone-apps/ [accessed 2021-05-07]
34. Bhuiyan N, Puzia M, Stecher C, Huberty J. Associations Between Rural or Urban Status, Health Outcomes and Behaviors, and COVID-19 Perceptions Among Meditation App Users: Longitudinal Survey Study. JMIR Mhealth Uhealth 2021 May 26;9(5):e26037 [FREE Full text] [doi: 10.2196/26037] [Medline: 33900930]
35. Green J, Huberty J, Puzia M, Stecher C. The effect of meditation and physical activity on the mental health impact of COVID-19-related stress and attention to news among mobile app users in the United States: cross-sectional survey. JMIR Ment Health 2021 Apr 13;8(4):e28479 [FREE Full text] [doi: 10.2196/28479] [Medline: 33788698]
36. Rauschenberg C, Schick A, Goetzl C, Roehr S, Riedel-Heller SG, Koppe G, et al. Social isolation, mental health, and use of digital interventions in youth during the COVID-19 pandemic: A nationally representative survey. Eur Psychiatry 2021 Mar 09;64(1):e20 [FREE Full text] [doi: 10.1192/j.eurpsy.2021.17] [Medline: 33686930]
37. Huberty J, Vranceanu A, Carney C, Breus M, Gordon M, Puzia ME. Characteristics and Usage Patterns Among 12,151 Paid Subscribers of the Calm Meditation App: Cross-Sectional Survey. JMIR Mhealth Uhealth 2019 Nov 03;7(11):e15648 [FREE Full text] [doi: 10.2196/15648] [Medline: 31682582]
38. Cohen S, Kamarck T, Mermelstein R. A Global Measure of Perceived Stress. Journal of Health and Social Behavior 1983 Dec;24(4):385. [doi: 10.2307/2136404]
39. Norman SB, Cissell SH, Means-Christensen AJ, Stein MB. Development and validation of an Overall Anxiety Severity And Impairment Scale (OASIS). Depress Anxiety 2006;23(4):245-249. [doi: 10.1002/da.20182] [Medline: 16688739]
40. Campbell-Sills L, Norman SB, Craske MG, Sullivan G, Lang AJ, Chavira DA, et al. Validation of a brief measure of anxiety-related severity and impairment: the Overall Anxiety Severity and Impairment Scale (OASIS). J Affect Disord 2009 Jan;112(1-3):92-101 [FREE Full text] [doi: 10.1016/j.jad.2008.03.014] [Medline: 18486238]
41. Creamer M, Bell R, Failla S. Psychometric properties of the Impact of Event Scale - Revised. Behav Res Ther 2003 Dec;41(12):1489-1496. [doi: 10.1016/j.brat.2003.07.010] [Medline: 14705607]
42. Hughes ME, Waite LJ, Hawkley LC, Cacioppo JT. A Short Scale for Measuring Loneliness in Large Surveys: Results From Two Population-Based Studies. Res Aging 2004 Aug 19;26(6):655-672 [FREE Full text] [doi: 10.1177/0164027504268574] [Medline: 18504506]
43. Berle D, Starcevic V, Moses K, Hannan A, Milicevic D, Sammut P. Preliminary validation of an ultra-brief version of the Penn State Worry Questionnaire. Clin Psychol Psychother 2011 Aug 30;18(4):339-346. [doi: 10.1002/cpp.724] [Medline: 20806422]
44. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. Journal of Behavior Therapy and Experimental Psychiatry 2000 Jun;31(2):73-86. [doi: 10.1016/s0005-7916(00)00012-4]
45. Godin G, Jobin J, Bouillon J. Assessment of leisure time exercise behavior by self-report: a concurrent validity study. Can J Public Health 1986;77(5):359-362. [Medline: 3791117]
46. Godin G, Shephard RJ. A simple method to assess exercise behavior in the community. Can J Appl Sport Sci 1985 Sep;10(3):141-146. [Medline: 4053261]
47. Paffenbarger RS, Blair SN, Lee I, Hyde RT. Measurement of physical activity to assess health effects in free-living populations. Med Sci Sports Exerc 1993 Jan;25(1):60-70. [doi: 10.1249/00005768-199301000-00010] [Medline: 8423758]

48. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
49. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: Building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
50. Averill JB. Matrix Analysis as a Complementary Analytic Strategy in Qualitative Inquiry. qual health res 2002 Jul 01;12(6):855-866. [doi: 10.1177/10432302012006011]
51. Saxbe DE, Repetti RL, Nishina A. Marital satisfaction, recovery from work, and diurnal cortisol among men and women. Health Psychol 2008 Jan;27(1):15-25. [doi: 10.1037/0278-6133.27.1.15] [Medline: 18230009]
52. Jaworski BK, Taylor K, Ramsey KM, Heinz A, Steinmetz S, Pagano I, et al. Exploring Usage of COVID Coach, a Public Mental Health App Designed for the COVID-19 Pandemic: Evaluation of Analytics Data. J Med Internet Res 2021 Mar 01;23(3):e26559 [FREE Full text] [doi: 10.2196/26559] [Medline: 33606656]
53. Khademian F, Aslani A, Bastani P. The effects of mobile apps on stress, anxiety, and depression: overview of systematic reviews. Int J Technol Assess Health Care 2020 Dec 14;37(1):e4. [doi: 10.1017/s0266462320002093]
54. Matis J, Svetlak M, Slezackova A, Svoboda M, Šumec R. Mindfulness-based programs for patients with cancer via eHealth and mobile health: systematic review and synthesis of quantitative research. J Med Internet Res 2020 Nov 16;22(11):e20709 [FREE Full text] [doi: 10.2196/20709] [Medline: 33196452]
55. Golden EA, Zweig M, Danieletto M, Landell K, Nadkarni G, Bottinger E, et al. A resilience-building app to support the mental health of health care workers in the COVID-19 era: design process, distribution, and evaluation. JMIR Form Res 2021 May 05;5(5):e26590 [FREE Full text] [doi: 10.2196/26590] [Medline: 33872189]
56. Fuller-Tyszkiewicz M, Richardson B, Little K, Teague S, Hartley-Clark L, Capic T, et al. Efficacy of a smartphone app intervention for reducing caregiver stress: randomized controlled trial. JMIR Ment Health 2020 Jul 24;7(7):e17541 [FREE Full text] [doi: 10.2196/17541] [Medline: 32706716]
57. Moberg C, Niles A, Beermann D. Guided self-help works: randomized waitlist controlled trial of Pacifica, a mobile app integrating cognitive behavioral therapy and mindfulness for stress, anxiety, and depression. J Med Internet Res 2019 Jun 08;21(6):e12556 [FREE Full text] [doi: 10.2196/12556] [Medline: 31199319]
58. Roy A, Druker S, Hoge EA, Brewer JA. Physician Anxiety and Burnout: Symptom Correlates and a Prospective Pilot Study of App-Delivered Mindfulness Training. JMIR Mhealth Uhealth 2020 Apr 01;8(4):e15608 [FREE Full text] [doi: 10.2196/15608] [Medline: 32234708]
59. van der Meer CAI, Bakker A, van Zuiden M, Lok A, Olff M. Help in hand after traumatic events: a randomized controlled trial in health care professionals on the efficacy, usability, and user satisfaction of a self-help app to reduce trauma-related symptoms. Eur J Psychotraumatol 2020 Mar 10;11(1):1717155 [FREE Full text] [doi: 10.1080/20008198.2020.1717155] [Medline: 32284818]
60. Faieta J, Sheehan J, DiGiovine C. Mhealth interventions to improve health and quality of life related outcomes for informal dementia caregivers: A scoping review. Assist Technol 2021 Mar 05:1-13. [doi: 10.1080/10400435.2020.1829174] [Medline: 33006526]
61. Fiol-DeRoque MA, Serrano-Ripoll MJ, Jiménez R, Zamanillo-Campos R, Yáñez-Juan AM, Bennasar-Veny M, et al. A Mobile Phone-Based Intervention to Reduce Mental Health Problems in Health Care Workers During the COVID-19 Pandemic (PsyCovidApp): Randomized Controlled Trial. JMIR Mhealth Uhealth 2021 May 18;9(5):e27039 [FREE Full text] [doi: 10.2196/27039] [Medline: 33909587]
62. Gál É, tefan S, Cristea IA. The efficacy of mindfulness meditation apps in enhancing users' well-being and mental health related outcomes: a meta-analysis of randomized controlled trials. J Affect Disord 2021 Jan 15;279:131-142. [doi: 10.1016/j.jad.2020.09.134] [Medline: 33049431]
63. Goreis A, Felnhofer A, Kafka JX, Probst T, Kothgassner OD. Efficacy of Self-Management Smartphone-Based Apps for Post-traumatic Stress Disorder Symptoms: A Systematic Review and Meta-Analysis. Front Neurosci 2020 Jan 24;14:3 [FREE Full text] [doi: 10.3389/fnins.2020.00003] [Medline: 32038153]
64. Lee RA, Jung ME. Evaluation of an mHealth App (DeStressify) on University Students' Mental Health: Pilot Trial. JMIR Ment Health 2018 Jan 23;5(1):e2 [FREE Full text] [doi: 10.2196/mental.8324] [Medline: 29362209]
65. Blázquez Martín D, De La Torre I, Garcia-Zapirain B, Lopez-Coronado M, Rodrigues J. Managing and Controlling Stress Using mHealth: Systematic Search in App Stores. JMIR Mhealth Uhealth 2018 May 09;6(5):e111 [FREE Full text] [doi: 10.2196/mhealth.8866] [Medline: 29743152]
66. Hwang WJ, Ha JS, Kim MJ. Research Trends on Mobile Mental Health Application for General Population: A Scoping Review. Int J Environ Res Public Health 2021 Mar 02;18(5):2459 [FREE Full text] [doi: 10.3390/ijerph18052459] [Medline: 33801537]
67. 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 May 22;8(5):e17798 [FREE Full text] [doi: 10.2196/17798] [Medline: 32357125]

68. Wang K, Varma DS, Prosperi M. A systematic review of the effectiveness of mobile apps for monitoring and management of mental health symptoms or disorders. J Psychiatr Res 2018 Dec;107:73-78. [doi: 10.1016/j.jpsychires.2018.10.006] [Medline: 30347316]

Abbreviations

IES-R: Impact of Events Scale—Revised
LTEQ: Leisure-Time Exercise Questionnaire
OASIS: Overall Anxiety Severity and Impairment Scale
PSS: Perceived Stress Scale
PSWQ: Penn State Worry Questionnaire
UCLA: University of California Los Angeles

Edited by G Eysenbach; submitted 23.07.21; peer-reviewed by M Weigensberg, P Giacobbi, Jr.; comments to author 01.09.21; revised version received 08.09.21; accepted 18.09.21; published 04.10.21.

Please cite as:
Gordon JS, Sbarra D, Armin J, Pace TWW, Gniady C, Barraza Y
Use of a Guided Imagery Mobile App (See Me Serene) to Reduce COVID-19–Related Stress: Pilot Feasibility Study
JMIR Form Res 2021;5(10):e32353
URL: https://formative.jmir.org/2021/10/e32353
doi:10.2196/32353
PMID:34546941

Original Paper

Perceptions and Attitudes Toward the Use of a Mobile Health App for Remote Monitoring of Gingivitis and Willingness to Pay for Mobile Health Apps (Part 3): Mixed Methods Study

Guy Tobias[1], DMD, MPH, PhD; Harold Sgan-Cohen[1], DMD, MPH; Assaf B Spanier[2], PhD; Jonathan Mann[1], MSc, DMD

[1]Department of Community Dentistry, Faculty of Dental Medicine, The Hebrew University-Hadassah School of Dental Medicine, Jerusalem, Israel
[2]Department of Software Engineering, Azrieli College of Engineering, Jerusalem, Israel

Corresponding Author:
Guy Tobias, DMD, MPH, PhD
Department of Community Dentistry
Faculty of Dental Medicine
The Hebrew University-Hadassah School of Dental Medicine
Ein Kerem
Jerusalem, 91120
Israel
Phone: 972 52 705 2333
Email: guy.tobias@mail.huji.ac.il

Related Articles:

Companion article: https://mhealth.jmir.org/2020/8/e19433/

Companion article: https://mhealth.jmir.org/2021/9/e24955/

Abstract

Background: Gum infection, known as gingivitis, is a global issue. Gingivitis does not cause pain; however, if left untreated, it can worsen, leading to bad breath, bleeding gums, and even tooth loss, as the problem spreads to the underlying structures anchoring the teeth in the jaws. The asymptomatic nature of gingivitis leads people to postpone dental appointments until clinical signs are obvious or pain is evident. The COVID-19 pandemic has necessitated social distancing, which has caused many people to postpone dental visits and neglect gingival health. iGAM is a dental mobile health (mHealth) app that remotely monitors gum health, and an observational study demonstrated the ability of iGAM to reduce gingivitis. We found that a weekly dental selfie using the iGAM app reduced the signs of gingivitis and promoted oral health in a home-based setting.

Objective: The aim of this mixed methods study is to assess perceptions, attitudes, willingness to pay, and willingness to use an mHealth app.

Methods: The first qualitative phase of the study included eight semistructured interviews, and the second quantitative phase included data collected from responses to 121 questionnaires.

Results: There was a consensus among all interviewees that apps dealing with health-related issues (mHealth apps) can improve health. Three themes emerged from the interviews: the iGAM app is capable of improving health, the lack of use of medical apps, and a contradiction between the objective state of health and the self-definition of being healthy. Participants were grouped according to how they responded to the question about whether they believed that mHealth apps could improve their health. Participants who believed that mHealth apps can enhance health (mean 1.96, SD 1.01) had a higher willingness to pay for the service (depending on price) than those who did not believe in app efficacy (mean 1.31, SD 0.87; t_{119}=−2417; P=.02). A significant positive correlation was found between the amount a participant was willing to pay and the benefits offered by the app (rs=0.185; P=.04).

Conclusions: Potential mHealth users will be willing to pay for app use depending on their perception of the app's ability to help them personally, provided they define themselves as currently unhealthy.

(JMIR Form Res 2021;5(10):e26125) doi:10.2196/26125

KEYWORDS

mHealth; public health; oral health promotion; gum health; willingness to pay; willingness to use; willingness; perception; attitude; mouth; oral health; dentist; app; monitoring; mixed method

Introduction

COVID-19 and Gingivitis

Gingivitis [1] is a reversible gum inflammation, characterized by red, swollen, and bleeding gums. More than 80% of the global adult population is affected by gingivitis periodically [2]. Usually, it heals within 10 days of at-home oral hygiene practices that include twice-daily brushing and interproximal cleaning using dental floss, toothpicks, and mouthwash [3]. In dental clinics, gingivitis is usually treated in a single cleaning session where plaque and calculus are removed. Untreated gingivitis develops into periodontitis [4,5], the irreversible stage of periodontal disease, in which bacterial toxins and the immune response to them destroy the tissues that support the teeth. Treatment for periodontitis is complex and requires multiple appointments and sometimes surgical intervention. As the disease progresses, teeth may become mobile or lost [6].

Most people recognize the early signs of gingival inflammation, namely bleeding while brushing or eating something hard, an unpleasant odor, or swollen gums [1]. However, because gingivitis is not painful, appointments tend to be postponed until the disease is more advanced [3]. Currently, we are in the midst of a major public health crisis. The COVID-19 pandemic [7] has necessitated new behaviors and rules to limit the spread of the infection, such as wearing masks, hand hygiene, and maintaining 2 m of space between people (social distancing).

Dentistry is a field with close contact between patients and the clinical team, with a high risk of infection transmission [8]. Keeping in mind that most people only seek dental treatment when experiencing pain, we assume that because of the current restrictions, people will come even less, resulting in oral health deterioration, especially in those with gingivitis.

Remote Patient Monitoring

Remote patient monitoring (RPM) [9] is a term describing technologies that enable patient-therapist interaction without a physical meeting, for example, when the patient is at home and the care provider is elsewhere. RPM eliminates problems of geographical distance, improves access to treatment, and reduces the indirect costs of traditional clinical treatment, such as travel time, fuel, and parking. RPM improves treatment access to individuals living far from medical centers and to populations with limited mobility [10]. Furthermore, RPM encourages people to seek treatment for nonurgent issues that tend to be postponed because of travel and time issues [11].

Incorporating RPM in the management of chronic diseases significantly improves patient quality of life, as less time is spent in the doctor's office while maintaining contact and follow-up [12]. RPM is delivered using advanced technologies that quickly and effectively detect deterioration of the patient's condition and update the treating staff and the patient. Early warning reduces the number of hospitalizations and hospitalization days and improves quality of life [13]. The components [9] of RPM are (1) disease-specific sensors that monitor physiological and pathological changes; (2) storage of relevant information accessible to attending physicians; (3) an encrypted server where personal information is stored securely; and (4) diagnostic tools, software that facilitates information retrieval and helps develop therapeutic recommendations based on the patient's data.

In dentistry, as far as we know, no work has been done regarding RPM, and there are currently no technologies for remote oral health monitoring.

iGAM App

iGAM [14] is a dental mobile health (mHealth) app that remotely monitors gingival health, and our previous article described the development process followed by a pilot study to evaluate the acceptance of remotely monitoring gum health.

The iGAM mHealth app has three main features:

- A self-completion questionnaire that deals with knowledge and attitudes toward oral hygiene habits.
- Text accompanied by illustrations describing brushing techniques as well as short articles about the importance of maintaining oral health in general and during pregnancy, the implications of smoking on gum health, and the connection between gum diseases and general health.
- Feature for self-photography of the gums using the rear camera of the smartphone, that is, the *dental selfie*.

We performed an 8-week observational study, with 126 participants divided into 2 groups. The first group photographed their gums weekly, and the second group took only 2 pictures, one at the start of the study and the second 8 weeks later. We found that a weekly dental selfie using the iGAM app reduced the signs of gingivitis and promoted oral health. These findings were published in part 2 of this series of articles [15].

Health Belief Model and Willingness to Pay for mHealth Apps

The fact that an mHealth app is capable of improving health does not guarantee that the public will use it. Some degree of health literacy and a mobile device with advanced technological capabilities, availability of storage space, and money are needed. The Health Belief Model (HBM) [16] is a psychological model developed in 1950 by social psychologists Hochbaum, Rosenstock, and Kegels, which tries to explain and predict the adoption of health behaviors by focusing on the attitudes and beliefs of the individual.

The HBM has five content categories [17]:

- Perceived susceptibility—understanding the possibility of developing a specific disease. Those who estimate that the

likelihood of getting sick is low deny the possibility that they will become unwell, whereas those with a high degree of sensitivity estimate that they are in real danger of getting sick.
- Perceived severity—understanding the severity and seriousness of a specific disease and its consequences. This category reflects an individual's beliefs regarding the difficulties that illness may cause, such as pain, discomfort, and financial burden.
- Perceived benefits—believing that recommended health behavior will be beneficial by preventing the disease or reducing its effects.
- Perceived barriers—the costs of or obstacles to performing the recommended behavior, including tangible costs (eg, time, money, availability, and skill acquisition) and the psychological costs associated with performing the behavior (eg, pain, feeling anxious, pessimistic, and embarrassed). A low perception of barriers increases the likelihood of adopting such behavior.
- Cues to action—the circumstances that inspire the readiness to act.

The model assumes that people take action when they perceive a threat to their health and consider whether the benefits of a new behavior are greater than the obstacles to performing it. The more benefits the individual believes there are from the behavior, the greater the chance of adopting it [18].

The principles of the HBM have been used as a conceptual basis for many studies. It should be noted that many studies address only some of their components.

Willingness to pay (WTP) [19] describes the maximum amount a consumer is prepared to pay for a particular service or product. It was found that the more the consumer perceives that the technology promotes health, the more they will be inclined to pay for it. A study [20] examining the willingness of young Iranians to pay higher prices for organic food products found that perceived benefits and barriers were significant predictors of the WTP. Those believing that consumption of organic food was beneficial were more willing to purchase organic food at higher prices. WTP research has two main branches [21]: (1) RP-revealed preference, where the WTP is estimated by real market analysis, and (2) SP-stated preference, where the WTP is estimated by consumer responses to hypothetical scenarios in which consumers are asked to indicate their preferences. The second method is used when it is impossible to obtain real data about consumer preferences, for example, when determining the demand for a new product. Wong et al [22] used a cross-sectional study with 1159 participants and asked about their WTP based on the HBM model. They found that most participants described the observed benefit from a corona vaccine as the main reason they would be willing to pay more for it.

The aim of this mixed methods study is to use qualitative and quantitative methods to examine attitudes and acceptance of mHealth apps in general and an app for remotely monitoring gingivitis (iGAM) in particular. Another goal is to examine the WTP for mHealth apps as a way to characterize responsiveness to mHealth app use.

Methods

Overview

This mixed methods study was conducted between March 2020 and July 2020 at the Department of Community Dentistry Faculty of Dental Medicine, the Hebrew University, Hadassah School of Dental Medicine. The protocol was approved by the Hadassah Research Ethics Committee (institutional review board, 0212-18-HMO), and written informed consent was obtained from all participants. There was no payment for participation. The study was conducted in two phases: a qualitative study based on semistructured interviews that reached saturation and highlighted the need for further data that was gathered in a quantitative study based on answers to questionnaires.

Qualitative Study Design and Data Collection

A total of 10 individuals from 126 participants in the 8-week observational study were randomly selected. They were asked via an SMS text message sent through the app whether they would participate in a 50- to 60-minute phone interview (the interview could not be done in person because of COVID-19 social distancing restrictions) to discuss their experience with the iGAM app and their thoughts regarding mHealth apps. Approximately 3 days later, a follow-up SMS text message was sent through the app. Of the 10 people, 8 agreed to the interview (one did not reply and one had schedule limitations). The semistructured interview was developed with guidance from two qualitative research methodology experts (Multimedia Appendix 1). The questions were divided into four sets: (1) opening questions—Tell me about yourself, What do you expect from your cellular phone? Describe your cellular phone use, etc; (2) questions regarding cellular apps in general and medical apps in particular: Describe the apps currently on your phone, How often do you use cellular apps? What sort of mHealth apps are you familiar with? In your opinion, what is the purpose of using mHealth apps? (3) Questions related to iGAM mHealth: describe your expectations of the app before the study, describe your experience using the app, do you think that using the app improved your oral health? In your opinion, who would this app help most? (4) Personal information, age, address, and occupation.

The interviews were audio recorded and transcribed using Microsoft Word 2016 (Microsoft Corporation). The primary author performed a qualitative data analysis. Each interview was read several times to identify the codes. The list of codes was grouped into categories based on content similarity, determined by counting the frequency that the interviewees talked about them. Examination of the categories enabled the identification of themes. At the end of the process, all authors reviewed the themes that emerged from the interviews.

Quantitative Study Design and Data Collection

On completion of the observational study, questionnaires were sent to all 126 participants via the iGAM app, and an SMS text message was sent. The participants were told that they were not obligated to answer the questions and that their answers needed to be submitted within 3 weeks. A second SMS text message

reminder was sent a week before the deadline. The questionnaire included 17 statements (Multimedia Appendix 2) on mHealth apps in general. The participant had to circle the response by matching their points of view. The statements were divided into two sets. The first set dealt with (1) beliefs about the ability of mHealth to improve health; (2) perceptions regarding the ability of mHealth to effectively monitor blood pressure, obesity, physical fitness, and oral health; (3) attitudes regarding searching for medical information on the internet; (4) personal use of medical apps; (5) statements about personal perceptions of health and state of health; and (6) reasons for not using mHealth apps. The second set dealt with a WTP for medical apps with costs between US $15 and US $150.

Statistical Methods

Data were analyzed using SPSS Statistics software version 27.0. The significance level was set at a *P* value of .05. *t* tests for independent samples were used to examine WTP. A Pearson correlation was conducted to examine the relationship between the tendency to use apps and the WTP depending on price. Spearman correlations were conducted to examine the relationship between the amount the participant is willing to pay for an app with optimal functions and the tendency to use the apps and WTP according to the app features. All independent variables were included in the enter regression test.

Results

Qualitative Results

Eight interviews were conducted after 10 participants in an 8-week study were asked if they were willing to be interviewed. One interviewee did not respond to the two requests, and the second interviewee canceled at the last minute and was not interested in rescheduling. Interviews ranged from 32.25 to 55.75 minutes, with an average of 46.25 (SD 2.5) minutes. Analysis of the interviewees revealed that one is a physician aged 33 years with 3 years of experience. Two work as nurses, one aged 41 years with 10 years of experience and the other aged 30 years with 4 years of experience. Three were students, one was a fourth-year dental student aged 26 years and two were first-year biology students aged 25 and 23 years. One interviewee aged 32 years worked as a computer engineer with 5 years of experience, and one interviewee was an electronic technician aged 36 years with 9 years of experience.

A total of three main themes emerged from the interviews: (1) the iGAM app is capable of improving health, (2) lack of use of medical apps, and (3) a contradiction between the objective state of health and self-definition of being healthy.

The iGAM App Is Capable of Improving Health

There was a consensus among all interviewees that apps dealing with health-related issues (mHealth apps) can improve health; 2 participants added that mHealth apps may be able to improve access for those in remote areas and be beneficial. An analysis of interview content revealed that most interviewees considered health apps to be desirable and be able to improve the provision of health care by better allocation of resources:

> *In my opinion, this device can alert and improve, this device can say here it's time to go to the dentist. I think it's something that is very very good, it's first of all will improve the patient's health.*
>
> *I would like to say that this app is good, it will serve both doctors and patients, both in Israel and in third world countries, as I told you, with low medicine, even in countries with advanced but expensive medicine, in the end it is in everyone's interest.*
>
> *I believe this app will have an impact on better health products.*

Lack of mHealth App Use

When questioned about apps installed in addition to those that came with the device, apps from four content worlds arose: (1) social apps (WhatsApp, Facebook, and Instagram); (2) apps related to household management, bank account management, payments, deliveries, and orders from supermarkets; (3) apps related to leisure: Netflix, books, holiday booking, etc; (4) and apps for editing videos and photos. None of the participants installed additional medical apps on their device, and they did not use the apps that were already on the phone (eg, the pedometer).

When asked why there were no medical apps for them, 2 participants answered that they had technical limitations of phone memory space, and the remaining 6 explained that they had no need for a medical app as they were healthy:

> *I have WhatsApp, I have a camera, Facebook, a bank account, Maccabi, Wizz, booking, cheap vacations, photos.*
>
> *I have no medical apps, other than what I got with the phone, pedometer because I do not need and I have no place on the phone.*
>
> *No I do not need; I am healthy.*
>
> *I do not use for example, speaking of blood pressure, once I measure myself sporadically and not consistently and once I understand that I do not need it because I am a healthy person, it is quite difficult to say that using the phone would improve my blood pressure, because I am healthy And feel fine.*

The Contradiction Between the Objective State of Health and the Self-definition of Being Healthy

When the interviewees were asked about their state of health, they all said that they were healthy. When asked to expand on the subject, their perception of being healthy may not be accurate, 3 reported uncontrolled hypertension, 2 reported episodes of high blood glucose, and 1 reported frequent headache:

> *I'm healthy. I'm healthy, uh, my blood pressure occasionally jumps but it's normal.*
>
> *I sometimes feel headaches, occasionally I also get sugar, it does not say that I am sick.*
>
> *No, I would not say I have a medical problem, I measure my blood pressure, and sometimes there are jumps, it does not mean it is unbalanced.*

Quantitative Results

Descriptive Statistics

Of the 126 participants asked to fill in the questionnaire, 121 responded positively. The 5 remaining individuals were contacted again (through the app) and did not reply. Of the 121 respondents, 91 (75.2%) were male; the age range was 19-66 years, with an average of 27.65 (SD 8.93) years; and the most common occupation was *student* (53/121, 44.2%). The results were analyzed by dividing the respondents into 2 groups based on their response to the yes/no question about the belief that mobile apps can improve health. A total of 13.2% (16/121) individuals were in group 1 and "lacked faith that apps can improve health," and the remaining 86.8% (105/121) were in group 2 and "believed that mobile apps can improve health." Most of the members of both groups responded that they first seek medical information on the web (12/16, 75% vs 90/105, 85.7%) and then go to a physician (11/16, 69% vs 63/105, 60%; group 1 versus 2, respectively). In addition, most members of both groups replied that they do not use mHealth apps to improve their health (14/16, 87% vs 70/105, 66%); however, if they had a long-term illness and the app associated with their illness was free, they would download it to their mobile device (12/16, 75% vs 97/105, 92.4%, group 1 versus 2, respectively). If they had to pay 50 NIS (currency of Israel–the New Israeli Shekel [NIS] $1=approximately 3.5 NIS) on a one-time basis, most of the group that did not believe that apps can improve their health would not buy it (9/16, 56%) and if it cost 500 NIS (US $150), none of this group would purchase it. Overall, 55.2% (58/105) of the group that believed that mobile apps can improve health stated that they would pay 50 NIS (US $15) for an app but not 500 NIS (89/105, 84.8%). When given the option to purchase an app that is able to (1) detect a health condition, (2) offer a personalized treatment plan, and (3) inform your personal physician when problems occur, 44% (7/16) versus 46% (48/105) of participants from group 1 and 2, respectively, stated that they would be willing to pay the maximum amount asked: *100 NIS*. Most participants believed that apps can be effective in obesity monitoring (13/16, 81% vs 100/105, 95%) and fitness monitoring (11/16, 69% vs 96/105, 91%) but not for monitoring hypertension (10/16, 62% vs 85/105, 80.9%) in group 1 versus 2, respectively. In group 2, 88.6% (93/105; those who believed that mobile apps can improve health) believed that apps can effectively monitor oral health versus 44% (7/16) of those in group 1 (those who lacked faith that apps can improve health; Multimedia Appendix 2).

To understand the significance of our findings and based on consultations with two experts in qualitative research methods, we used the responses (on a 4-point scale ranging from do not agree to completely agree) to sets of statements to determine the variables "Willingness to pay depending on price" and responses of agree or disagree to "Tendency to use applications" (on an 8-point scale). (1) "Willingness to pay depending on price," based on responses to statements: If I were sick with a chronic illness, I would download a free medical app related to the disease; If I were sick with a chronic illness, I would download a 50 NIS (one-time fee) medical app related to the disease; If I were sick with a chronic illness, I would download a 500 NIS (one-time fee) medical app related to the disease. (2) "Tendency to use applications," based on responses to statements: I believe that cell phone apps can improve health; When I need medical information about myself, I search the internet first; I use mobile applications to improve my health; If I were sick with a chronic illness, I would download a free medical app related to the disease; I have an interest in medical apps: Hypertension, Obesity, Fitness, Oral health.

The responses to the research variable "Willingness to pay depending on price" were given a score from 0 (do not agree) to 4 (completely agree), and the average score was 1.87 (SD 1.02). One point was given for each response of *agree* to the statements relating to the research variable "Tendency to use applications," creating a score between 0 and 9. The average score was 6.02 (SD 1.44). The answers to the question, "If I were sick with a chronic illness I would pay a maximum amount of NIS (one-time fee) for an optimal medical app," describing an ideal medical app was worded as follows: if you were diagnosed with a specific chronic disease and there was an app could monitor your illness and provide information about the stage of the disease you were in, recommend a reliable treatment plan, and contact your treating physician when changes in your health occur, ie, an optimal app—the maximum I would pay (one-time fee) is (a)10 NIS; (b)50 NIS; (c)100 NIS; and (d) I have no interest in this type of app—ranged from 0 (no interest in this type of app) to 4 (willing to pay 100 NIS), with an average of 2.82 (SD 1.30; Table 1).

Table 1. Means and standard deviations of the study variables.

Research variable	Value, mean (SD)
Willingness to pay depending on price	1.02 (1.87)
Willingness to pay for an optimal app	2.82 (1.30)
Tendency to use app	6.02 (1.44)

Analytical Statistics

Participants responding positively to the question, "when I need medical information about myself, I use the internet first" (mean 1.97, SD 1.06) showed a significantly greater tendency to use mHealth apps than those responding differently (mean 1.37, SD 0.60; t_{119}=−2.407; *P*=.02; Figure 1). Furthermore, participants who first searched for medical information on the internet (mean 6.16, SD 1.48) showed a significantly higher WTP for an app than those who did not search for information on the internet (mean 5.31, SD 0.95; t_{119}=−2.372; *P*=.02; Figure 1).

Participants who believed that mHealth apps can improve health (mean 1.96, SD 1.01) were significantly more willing to pay for an app (depending on price) than those who did not believe

in the efficacy of apps (mean 1.31, SD 0.87; t_{119}=−2.417; *P*=.017; Figure 2).

A significant positive relationship was found between the tendency to use apps and WTP depending on price (r=0.442; *P*<.001). In addition, individuals interested in paying for apps (mean 6.21, SD 1.32) had a significantly greater tendency to use apps than those who have no interest in paying (mean 5.58, SD 1.64; $t_{55,416}$=−2.035; *P*=.047; Figure 3).

Figure 1. Average positive responses based on seeking or not seeking medical information on the internet.

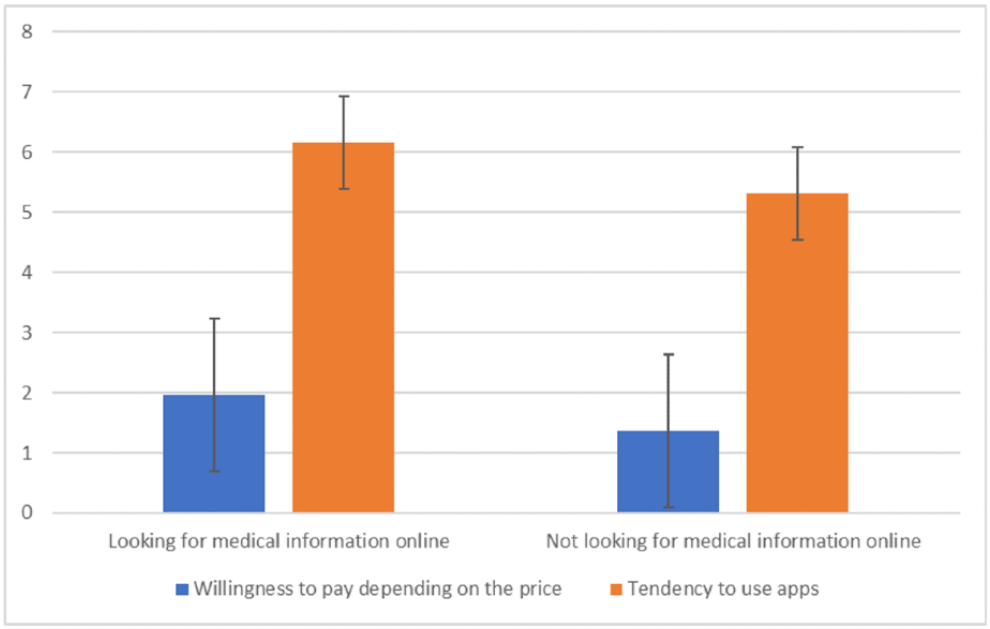


Figure 2. Willingness to pay depending on price according to the belief or lack of belief that mobile apps can improve health based on average positive responses.

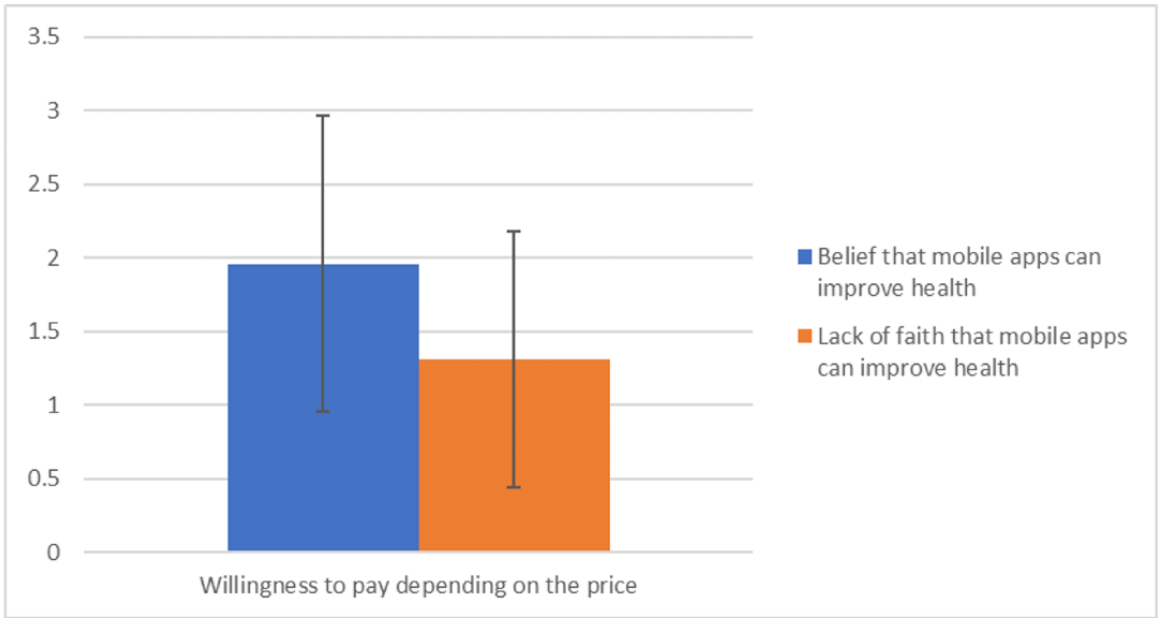
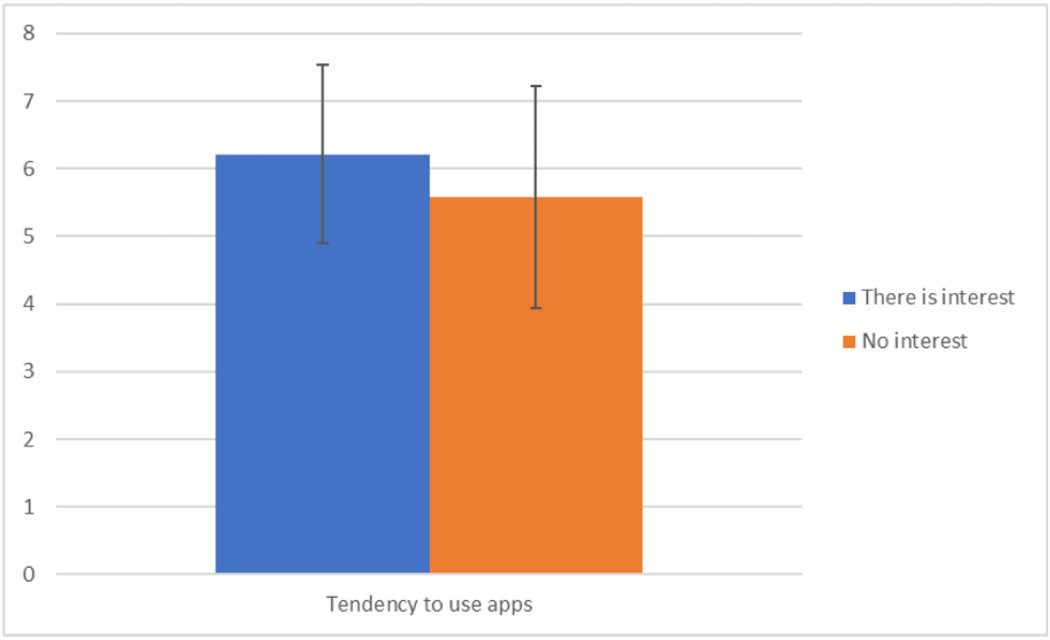


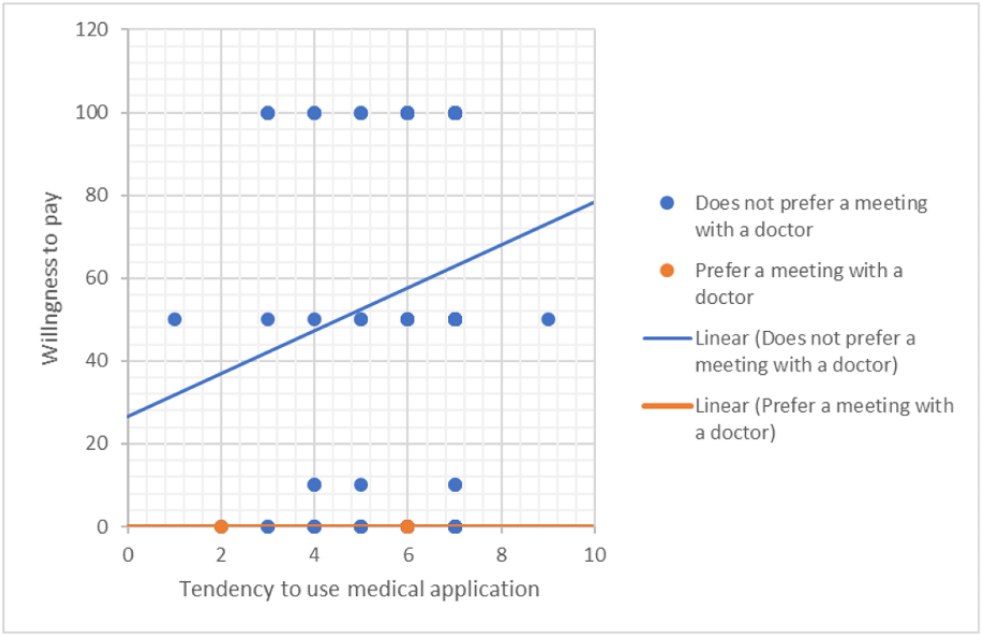
Figure 3. Tendency to use apps according to interest or no interest in paying based on average positive responses.



A significant positive relationship between the amount the participant was willing to pay for an optimal app and (1) the propensity to use apps (rs=0.185; *P*=.04) and (2) the WTP depending on price (rs=0.478; *P*<.001) was found. The regression test for predicting the WTP for an optimal app showed statistical significance ($F_{1,119}$=207.272; *P*<.001) with an explained variance of 63.5%. Predicting a tendency to use apps was significant (*P*<.001); that is, the higher the tendency to use apps and the WTP depending on price, the higher the amount the participant is willing to pay for an optimal app.

The question, "If you answered that you have no interest in medical applications, what is the reason for this?" examined the reasons for the lack of interest in using mHealth apps. The regression test showed statistical significance in predicting the use of available apps. The predictive variable with significance was "preference for an appointment with a doctor versus a lack of preference for an appointment with a doctor," and participants who preferred to see a physician were less likely to use medical apps compared with those with the opposite preference (Figure 4).

Figure 4. Tendency to use a medical app depending on willingness to pay. X axis: average positive responses to calculating the variable, "Tendency to use medical application." Y axis: price (in New Israeli Shekel).

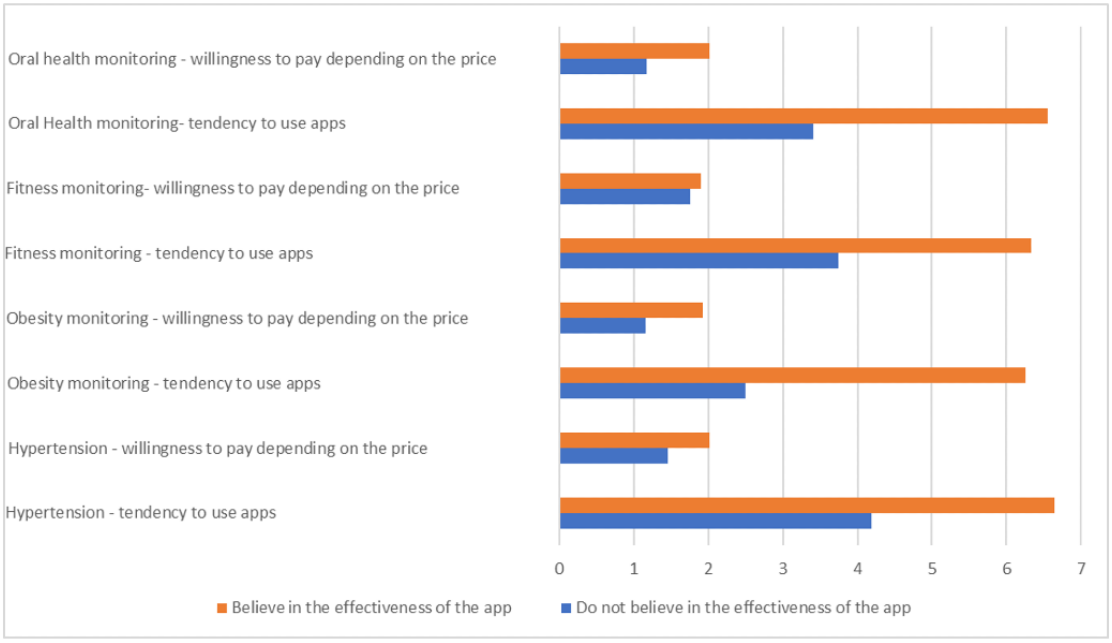


The predictive variable, "I'm healthy so I do not need medical apps / other, and I have / I have no interest in paying," was found to be statistically significant (*P*<.001). Participants who stated that they were healthy and did not need medical apps had no interest in payment.

The final four questions examined which medical field participants would be willing to pay for an app (depending on price) to monitor them:

1. Hypertension: those who believed in the effectiveness of apps to monitor hypertension showed a higher tendency to use an app in this field (mean 6.65, SD 1.44) than those who did not believe in their efficacy (mean 4.19, SD 1.44; t_{115}=−2.488; *P*=.01; Figure 5). Those who believed (mean 2.01, SD 1.01) in the effectiveness of apps for hypertension showed a higher WTP depending on price than those who did not believe in their efficacy (mean 1.46, SD 0.95; Figure 5; $t_{28.971}$=−8.363; *P*=<.001; Figure 5).
2. Obesity: those who believed (mean 6.26, SD 1.13) in the effectiveness of obesity tracking apps showed a higher tendency to use an app in this field than those who did not believe in their efficacy (mean 2.50, SD 1.04; t_{117}=−1.838; *P*=.04; Figure 5). Those who believed (mean 1.93, SD 0.99) in the effectiveness of obesity tracking apps showed a higher WTP depending on price than those who did not believe in their efficacy (mean 1.16, SD 1.17; t_{119}=−7.950; *P*<.001; Figure 5).
3. Physical fitness: those who believed (mean 6.33, SD 1.12) in the efficiency of fitness tracking apps showed a higher tendency to use an app in this field than those who did not believe in their efficacy (mean 3.75, SD 1.48; t_{117}=−7.308; *P*<.001; Figure 5). Interestingly, no statistically significant difference was found between those who believed (mean 1.90, SD 1.02) in the effectiveness of fitness tracking apps and those who did not believe in their efficacy (mean 1.75, SD 0.96) regarding WTP depending on price (t_{117}=−.505; *P*=.62; Figure 5).
4. Oral health: those who believed (mean 6.56, SD 0.78) in the effectiveness of oral health monitoring apps showed a higher tendency to use an app in this field than those who did not believe in their efficacy (mean 3.41, SD 1.12; $t_{18.739}$=−11.125; *P*<.001; Figure 5). Those who believed (mean 2.01, SD 0.97) in the effectiveness of oral health monitoring apps showed a higher WTP depending on price than those who did not believe in their efficacy (mean 1.17, SD 1.01; t_{115}=−3.257; *P*=.001; Figure 5).

Figure 5. Tendency to use and willingness to pay for medical apps depending on belief in their efficacy.



Discussion

Principal Findings

The purpose of this mixed methods study was to identify the attitudes, needs, WTP, and perceptions of use in mHealth apps. This study has two sections. The qualitative part included 8 participants of both sexes with different ages and professions. The second, quantitative part, was based on the responses of 121 individuals to a questionnaire sent to them after participation in an 8-week observational study.

The qualitative theme analysis revealed that all the interviewees believed in the ability of a medical app to improve health outcomes in general, and the iGAM app for remote monitoring of gingivitis in particular. The interviewees shared that they do not use medical apps because they believe that they are healthy and have no use for this type of app. Surprisingly, when questioned about their health, most of the interviewees had objective problems, such as hypertension and unbalanced blood glucose. The category analysis showed a contradiction between self-definition and reality.

A systematic review [23] regarding patient perceptions about medical app use found that most participants expressed a positive attitude toward mHealth app use and believed that these apps can promote health. In addition, a survey [24] of 500 physicians and 1000 app users found that 46% of physicians welcomed the medical apps into their lives and believed that they would improve communication between doctors and patients, and 72% of the physicians felt that medical apps would improve patient control over their health. Overall, 96% of the

app users believed that medical apps can improve health. Despite the statistically significant belief in the health-promoting abilities of apps, there are barriers preventing app use. Studies [25-29] show that the major issues discouraging medical apps use are related to privacy and security. Other obstacles include app costs and low technological and health literacy [30].

An Australian qualitative study [31] included interviews with 20 general practitioners and 15 older adult patients. They found that among the physicians, the barriers to app use were related to technology, such as fear of prescribing medications through the app, the amount of time required to learn how to use the app, secure storage of private patient information, etc. For patients, the main issues were about ease of use. Interestingly, we discovered a seemingly novel barrier, namely the individual's definition of themselves as healthy, and therefore not needing a medical app.

The data for the quantitative part of the study came from questionnaires that all participants in our observational study were asked to answer. There were 2 groups in the 8-week observational using the iGAM app. One group took weekly photographs, and the other only took photographs at the beginning and end of the study. The aim of this quantitative study was to understand the perceptions of mHealth app users. 86.8% (105/121) of the survey respondents believed that mobile apps could improve health. Most of the participants were in the third decade of their lives, men, and students. Studies [26,32-34] on the characteristics of medical app users have reported similar findings. The possible selection bias in our study is discussed below (see the *Limitations* section).

A significantly higher tendency to use mHealth apps was found among participants who responded that they first turn to the internet for medical information. These individuals were also willing to pay higher prices for apps than those not using the internet as their first source for clarifying health issues and believed that medical apps could improve health. Consistent with the findings of the qualitative part of the study, most of the respondents noted that they do not currently use mHealth apps but would if they were sick. When asked the general question of whether the participant would be willing to pay for an app related to a disease that they hypothetically have, with no additional information given about the app's features, a statistically significant majority of both groups, 92.4% (97/105) of the believing group and 75% (12/16) of the nonbelieving group, would use a free app, but as the price rose, the differences between the groups became apparent. The believing group was willing to pay between US $15 (58/105, 55.2%) and US $150% (16/105, 15.2%), whereas in the nonbelieving group, most (9/16, 56%) were not willing to pay US $15, and none agreed to pay US $150.

As we changed the questions slightly, by adding app features making an optimal app, that is, the ability to monitor, offer treatment suggestions, contact the treating physician, and most participants in both groups were willing to pay the maximum amount. This finding contributes to understanding what potential users are looking for in medical apps that they are willing to purchase. WTP increased not only with the tendency to adopt technologies but also with understanding the ability of the app to promote the health of the user. Three components from the HBM [16] can be applied: the perceived benefits, perceived barriers, and cues to action are significant predictors of WTP for mHealth apps.

We found two main reasons for the lack of interest in using medical apps:

- Individuals who prefer an in-person meeting with their physician are less inclined to use applications, and their tendency to pay decreases as the price increases.
- Self-defined as healthy: the participants in this study were relatively young and stated that they were healthy and that this was the reason for the lack of interest in using medical apps. Older populations have more diseases, and younger people tend to be healthier and define themselves as such [35]. Other factors include the findings that different age groups perceive health differently and that older people are more concerned about their health and have more contact with sick friends and family than young people [36]. This may be because younger people are preoccupied with matters, such as starting a family, establishing a career, etc.

When we examined the willingness to use and pay for apps that monitor hypertension, obesity, and physical fitness, those who believe that apps can improve health had a significantly higher propensity to use an app for all these issues. However, when examining WTP, the results were different. A statistically significant difference in WTP was only found for monitoring hypertension, obesity, and oral health monitoring, and no significant difference was found for fitness tracking. This may be because of the fact that most smartphones come with free fitness–related apps [37,38]; therefore, people may be reluctant to spend money on another similar app.

Limitations

The main limitation of this study is the difficulty of generalizing the results because of a possible selection bias among the volunteer participants. First, the participants were young, and perceptions of wellness and health may be different in older age groups. Second, most of the participants were students who may be more comfortable with technology and have a certain level of health and technological literacy. These issues can be resolved by including individuals of different ages in the study groups.

It is possible that there was a selection bias regarding the characteristics of the interviewees (4 out of 8 were involved in the health sector: 1 physician, 2 nurses, and 1 dental student). Therefore, we decided that all study participants should complete the quantitative questionnaire. Nevertheless, there is a need for further research to examine the acceptance of iGAM mHealth apps in groups with different levels of literacy. Another potential limitation is that the qualitative data were coded by one person; therefore, to increase credibility, an independent expert examined the process of the entire data analysis and themes.

The fact that all participants defined themselves as healthy is a substantial limitation of this study. The questionnaire may not have been sensitive enough, and we may need to include individuals that classify themselves as *sick* or *unwell* or those taking medication regularly, for example, those treated for high blood pressure or diabetes. However, we follow the broad

definition of health of the World Health Organization, namely "Health is a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity," and assumed that our qualitative and quantitative methods gave the participants the leeway needed to define their own health status and yield results that accurately represent their attitudes toward the use of medical apps.

Conclusions

This study found that people believe an mHealth app can be used to monitor gingivitis to improve gum health. Furthermore, people are willing to use and pay for an mHealth app depending on their perceptions of their health requirements, state of health, and the level of active involvement of the app promoting health for the particular disease they have.

Conflicts of Interest

None declared.

Multimedia Appendix 1
iGAM interview plan.
[DOCX File , 15 KB - formative_v5i10e26125_app1.docx]

Multimedia Appendix 2
Characteristics of the study population.
[DOCX File , 19 KB - formative_v5i10e26125_app2.docx]

References

1. Page RC. Gingivitis. J Clin Periodontol 1986 May;13(5):345-359. [doi: 10.1111/j.1600-051x.1986.tb01471.x] [Medline: 3522644]
2. Marcenes W, Kassebaum NJ, Bernabé E, Flaxman A, Naghavi M, Lopez A, et al. Global burden of oral conditions in 1990-2010: a systematic analysis. J Dent Res 2013 Jul;92(7):592-597 [FREE Full text] [doi: 10.1177/0022034513490168] [Medline: 23720570]
3. Chapple IL, Van der Weijden F, Doerfer C, Herrera D, Shapira L, Polak D, et al. Primary prevention of periodontitis: managing gingivitis. J Clin Periodontol 2015 Apr;42 Suppl 16:S71-S76. [doi: 10.1111/jcpe.12366] [Medline: 25639826]
4. Robinson PJ. Gingivitis: a prelude to periodontitis? J Clin Dent 1995;6 Spec No:41-45. [Medline: 8593192]
5. Slots J. Periodontitis: facts, fallacies and the future. Periodontol 2000 2017 Oct;75(1):7-23. [doi: 10.1111/prd.12221] [Medline: 28758294]
6. Graziani F, Karapetsa D, Alonso B, Herrera D. Nonsurgical and surgical treatment of periodontitis: how many options for one disease? Periodontol 2000 2017 Oct;75(1):152-188. [doi: 10.1111/prd.12201] [Medline: 28758300]
7. Ahn D, Shin H, Kim M, Lee S, Kim H, Myoung J, et al. Current status of epidemiology, diagnosis, therapeutics, and vaccines for novel coronavirus disease 2019 (COVID-19). J Microbiol Biotechnol 2020 Mar 28;30(3):313-324 [FREE Full text] [doi: 10.4014/jmb.2003.03011] [Medline: 32238757]
8. Coulthard P. Dentistry and coronavirus (COVID-19) - moral decision-making. Br Dent J 2020 Apr;228(7):503-505. [doi: 10.1038/s41415-020-1482-1] [Medline: 32277203]
9. Stewart H. Remote Patient Monitoring. JEMS 2015 Nov;Suppl:5-6. [Medline: 26710386]
10. Walker RC, Tong A, Howard K, Palmer SC. Patient expectations and experiences of remote monitoring for chronic diseases: systematic review and thematic synthesis of qualitative studies. Int J Med Inform 2019 Apr;124:78-85 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.01.013] [Medline: 30784430]
11. Flodgren G, Rachas A, Farmer AJ, Inzitari M, Shepperd S. Interactive telemedicine: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2015 Sep 7;2015(9):CD002098 [FREE Full text] [doi: 10.1002/14651858.CD002098.pub2] [Medline: 26343551]
12. Basch E, Barbera L, Kerrigan CL, Velikova G. Implementation of patient-reported outcomes in routine medical care. Am Soc Clin Oncol Educ Book 2018 May 23;38:122-134 [FREE Full text] [doi: 10.1200/EDBK_200383] [Medline: 30231381]
13. Weenk M, Koeneman M, van de Belt TH, Engelen LJ, van Goor H, Bredie SJ. Wireless and continuous monitoring of vital signs in patients at the general ward. Resuscitation 2019 Mar;136:47-53. [doi: 10.1016/j.resuscitation.2019.01.017] [Medline: 30685546]
14. Tobias G, Spanier AB. Developing a mobile app (iGAM) to promote gingival health by professional monitoring of dental selfies: user-centered design approach. JMIR Mhealth Uhealth 2020 Aug 14;8(8):e19433 [FREE Full text] [doi: 10.2196/19433] [Medline: 32795985]
15. Tobias G, Spanier A. Using an mHealth App (iGAM) to Reduce Gingivitis Remotely (Part 2): Prospective Observational Study. JMIR Mhealth Uhealth 2021 Sep;9(9):e24955 [FREE Full text] [doi: 10.2196/24955] [Medline: 34528897]
16. Janz NK, Becker MH. The health belief model: a decade later. Health Educ Q 1984;11(1):1-47. [doi: 10.1177/109019818401100101] [Medline: 6392204]

17. Rosenstock IM, Strecher VJ, Becker MH. Social learning theory and the health belief model. Health Educ Q 1988;15(2):175-183. [doi: 10.1177/109019818801500203] [Medline: 3378902]
18. Champion VL. Instrument refinement for breast cancer screening behaviors. Nurs Res 1993;42(3):139-143. [Medline: 8506161]
19. Frost N, Freeman J, Brixner D, Mort J, Clem J, Ngorsuraches S. Patients' preferences and willingness-to-pay for disease-modifying therapies. Mult Scler Relat Disord 2019 Oct;35:55-60. [doi: 10.1016/j.msard.2019.07.005] [Medline: 31323483]
20. Yazdanpanah M, Forouzani M, Hojjati M. Willingness of Iranian young adults to eat organic foods: application of the Health Belief Model. Food Qual Prefer 2015 Apr;41:75-83. [doi: 10.1016/j.foodqual.2014.11.012]
21. Barber S, Pavitt S, Khambay B, Bekker H, Meads D. Eliciting preferences in dentistry with multiattribute stated preference methods: a systematic review. JDR Clin Trans Res 2018 Oct;3(4):326-335. [doi: 10.1177/2380084418780324] [Medline: 30931788]
22. Wong LP, Alias H, Wong P, Lee HY, AbuBakar S. The use of the health belief model to assess predictors of intent to receive the COVID-19 vaccine and willingness to pay. Hum Vaccin Immunother 2020 Sep 1;16(9):2204-2214 [FREE Full text] [doi: 10.1080/21645515.2020.1790279] [Medline: 32730103]
23. Vo V, Auroy L, Sarradon-Eck A. Patients' perceptions of mhealth apps: meta-ethnographic review of qualitative studies. JMIR Mhealth Uhealth 2019 Jul 10;7(7):e13817 [FREE Full text] [doi: 10.2196/13817] [Medline: 31293246]
24. Mcaskill R. Study finds patient engagement opportunities high with mHealth. Xtelligent Healthcare Media. URL: https://mhealthintelligence.com/news/study-finds-patient-engagement-opportunities-high-with-mhealth [accessed 2021-09-03]
25. Zhou L, Bao J, Watzlaf V, Parmanto B. Barriers to and facilitators of the use of mobile health apps from a security perspective: mixed-methods study. JMIR Mhealth Uhealth 2019 Apr 16;7(4):e11223 [FREE Full text] [doi: 10.2196/11223] [Medline: 30990458]
26. Krebs P, Duncan DT. Health app use among us mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 4;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
27. Atienza AA, Zarcadoolas C, Vaughon W, Hughes P, Patel V, Chou WS, et al. Consumer attitudes and perceptions on mhealth privacy and security: findings from a mixed-methods study. J Health Commun 2015;20(6):673-679. [doi: 10.1080/10810730.2015.1018560] [Medline: 25868685]
28. Peng W, Kanthawala S, Yuan S, Hussain SA. A qualitative study of user perceptions of mobile health apps. BMC Public Health 2016 Nov 14;16(1):1158 [FREE Full text] [doi: 10.1186/s12889-016-3808-0] [Medline: 27842533]
29. Prasad A, Sorber J, Stablein T, Anthony D, Kotz D. Exposing privacy concerns in mHealth. In: USENIX Workshop on Health Security (HealthSec). 2011 Presented at: 2nd USENIX Workshop on Health Security and Privacy; Aug 9, 2011; San Francisco, CA.
30. Holtz B, Vasold K, Cotten S, Mackert M, Zhang M. Health care provider perceptions of consumer-grade devices and apps for tracking health: a pilot study. JMIR Mhealth Uhealth 2019 Jan 22;7(1):e9929 [FREE Full text] [doi: 10.2196/mhealth.9929] [Medline: 30668515]
31. Byambasuren O, Beller E, Hoffmann T, Glasziou P. Barriers to and facilitators of the prescription of mHealth apps in australian general practice: qualitative study. JMIR Mhealth Uhealth 2020 Jul 30;8(7):e17447 [FREE Full text] [doi: 10.2196/17447] [Medline: 32729839]
32. Hung MC, Jen WY. The adoption of mobile health management services: an empirical study. J Med Syst 2012 Jun;36(3):1381-1388. [doi: 10.1007/s10916-010-9600-2] [Medline: 20878452]
33. Zhang X, Guo X, Lai K, Guo F, Li C. Understanding gender differences in m-health adoption: a modified theory of reasoned action model. Telemed J E Health 2014 Jan;20(1):39-46. [doi: 10.1089/tmj.2013.0092] [Medline: 24161004]
34. Bhuyan SS, Lu N, Chandak A, Kim H, Wyant D, Bhatt J, et al. Use of mobile health applications for health-seeking behavior among US adults. J Med Syst 2016 Jun;40(6):153. [doi: 10.1007/s10916-016-0492-7] [Medline: 27147516]
35. Kim Y, Park I, Kang S. Age and gender differences in health risk perception. Cent Eur J Public Health 2018 Mar;26(1):54-59 [FREE Full text] [doi: 10.21101/cejph.a4920] [Medline: 29684299]
36. Bowling A. Do older and younger people differ in their reported well-being? A national survey of adults in Britain. Fam Pract 2011 Apr;28(2):145-155 [FREE Full text] [doi: 10.1093/fampra/cmq082] [Medline: 21084566]
37. Pre-installed apps: Samsung Galaxy S20 Ultra 5G. T-Mobile. URL: https://www.t-mobile.com/support/devices/android/samsung-galaxy-s20-ultra-5g/pre-installed-apps-samsung-galaxy-s20-ultra-5g [accessed 2021-09-03]
38. Pre-installed apps: Samsung Galaxy S8. T-Mobile. URL: https://www.t-mobile.com/support/devices/android/samsung-galaxy-s8/pre-installed-apps-samsung-galaxy-s8 [accessed 2021-09-03]

Abbreviations

HBM: Health Belief Model
NIS: New Israeli Shekel
RPM: remote patient monitoring
WTP: willingness to pay

Edited by G Eysenbach, L Buis; submitted 29.11.20; peer-reviewed by J Gaspar, M Touré; comments to author 10.12.20; revised version received 13.12.20; accepted 07.08.21; published 05.10.21.

Please cite as:
Tobias G, Sgan-Cohen H, Spanier AB, Mann J
Perceptions and Attitudes Toward the Use of a Mobile Health App for Remote Monitoring of Gingivitis and Willingness to Pay for Mobile Health Apps (Part 3): Mixed Methods Study
JMIR Form Res 2021;5(10):e26125
URL: https://formative.jmir.org/2021/10/e26125
doi:10.2196/26125
PMID:34609320

Original Paper

Supporting Mental Health During the COVID-19 Pandemic Using a Digital Behavior Change Intervention: An Open-Label, Single-Arm, Pre-Post Intervention Study

Charlotte Summers[1], BSc; Philip Wu[2], PhD; Alisdair J G Taylor, PhD

[1]DDM Health, Coventry, United Kingdom

[2]School of Management, Royal Holloway, University of London, Egham, United Kingdom

Corresponding Author:
Charlotte Summers, BSc
DDM Health
Technology House
Science Park, University of Warwick
Coventry, CV4 7EZ
United Kingdom
Phone: 44 7969091134
Email: charlotte@ddm.health

Abstract

Background: The COVID-19 pandemic is taking a toll on people's mental health, particularly as people are advised to adhere to social distancing, self-isolation measures, and government-imposed national lockdowns. Digital health technologies have an important role to play in keeping people connected and in supporting their mental health and well-being. Even before the COVID-19 pandemic, mental health and social services were already strained.

Objective: Our objective was to evaluate the 12-week outcomes of the digitally delivered Gro Health intervention, a holistic digital behavior change app designed for self-management of mental well-being, sleep, activity, and nutrition.

Methods: The study used a quasi-experimental research design consisting of an open-label, single-arm, pre-post intervention engagement using a convenience sample. Adults who had joined the Gro Health app (intervention) and had a complete baseline dataset (ie, 7-item Generalized Anxiety Disorder scale, Perceived Stress Scale, and 9-item Patient Health Questionnaire) were followed up at 12 weeks (n=273), including 33 (12.1%) app users who reported a positive COVID-19 diagnosis during the study period. User engagement with the Gro Health platform was tracked by measuring total minutes of app engagement. Paired *t* tests were used to compare pre-post intervention scores. Linear regression analysis was performed to assess the relationship between minutes of active engagement with the Gro Health app and changes in scores across the different mental health measures.

Results: Of the 347 study participants, 273 (78.67%) completed both the baseline and follow-up surveys. Changes in scores for anxiety, perceived stress, and depression were predicted by app engagement, with the strongest effect observed for changes in perceived stress score ($F_{1,271}$=251.397; R^2=0.479; P<.001).

Conclusions: A digital behavior change platform that provides remote mental well-being support can be effective in managing depression, anxiety, and perceived stress during times of crisis such as the current COVID-19 pandemic. The outcomes of this study may also support the implementation of remote digital health apps supporting behavior change and providing support for low levels of mental health within the community.

(JMIR Form Res 2021;5(10):e31273) doi:10.2196/31273

KEYWORDS

stress; mental health; COVID-19; digital therapy; mHealth; support; behavior; intervention; online intervention; outcome; wellbeing; sleep; activity; nutrition

Introduction

The COVID-19 pandemic has created an unprecedented global health challenge and has strained health care systems worldwide [1]. To minimize the risk of infection and spread, and to protect the most vulnerable groups of people and the population at large, governments across the world have advised people at high risk to leave their homes for very limited purposes [2].

There is risk that the COVID-19 outbreak will create a *second pandemic* of mental health crises across health systems and communities [3]. Prior to the COVID-19 pandemic, mental health and social services were already stretched. Depression is the second leading cause of disability worldwide, and by 2030, it is expected to be the leading contributor to the global burden of disease [4]. Efforts to contain the spread of COVID-19, including prolonged social distancing and self-isolation, may trigger or exacerbate social, mental, and physical health problems, such as anxiety, relationship breakdowns, domestic violence, substance abuse or withdrawal, and obesity [5-7]. This could be especially serious for those with preexisting medical and psychological conditions [8].

Times are unprecedented, and the COVID-19 pandemic has created global uncertainty. During times of uncertainty, people are more likely to be stressed, depressed, and anxious [9-12]. In 2014, when an outbreak of the Middle East respiratory syndrome coronavirus (MERS-CoV) was reported, anxiety levels were associated with an increased perception of susceptibility to infection and social avoidance behaviors related to travel and being in public places [13]. Data originating from Wuhan city in the Hubei province of China, collected through the National Health Commission of China, showed that there was a correlation between the rapidly increasing numbers of confirmed cases and deaths and psychological problems, including anxiety, depression, and stress experienced by medical staff and the public [14].

The unpredictability and uncertainty of the COVID-19 pandemic and the resulting economic breakdown could increase the risk of mental health problems and exacerbate health inequalities [15]. Preliminary findings suggest adverse mental health effects in previously healthy people and in people with pre-existing mental health disorders [16]. Health care disparities will disproportionately affect socially disadvantaged patients, including those from ethnic minorities who have a relatively worse access to health care and receive poorer quality care than their Caucasian counterparts [17].

Approximately 5% of the general population experiences generalized anxiety disorder (GAD) at least once in their lifetime, and the estimated lag time to treatment for GAD can range between 9 and 23 years [18,19]. A key challenge to delivering interventions is low level of engagement [20,21].

Similarly, depression is the single largest contributor to global disability, with an estimated 300 million people affected worldwide [22]. Major depression has been found to impair the quality of life of people [23], as well as their psychosocial functioning [24]. However, resources to address these challenges are limited—for instance, the World Health Organization Mental Health Atlas 2017 reported a global median of 9 mental health workers per 100,000 people [25].

Face-to-face therapy and guided self-help techniques such as cognitive behavioral therapy and mindfulness have been shown to be effective in treating depression and anxiety [26-28]. However, face-to-face therapy traditionally provides point-in-time support, and there is struggle to scale fast enough to address growing mental health challenges [29]. With restrictions in interactions and activities during the COVID-19 pandemic, innovative delivery methods are required to augment care.

Our study aims to add to the research and evidence base on the effectiveness and engagement levels of a digital behavior change app (Gro Health) in the context of the COVID-19 pandemic. Previous research has shown that when mental health apps are properly designed, they can be cost-effective and scalable solutions for the treatment of anxiety, stress, and depression [30]. Moreover, meta-analyses of randomized controlled trials have shown that mental health apps can help alleviate symptoms of anxiety and depression, as well as assist patients to self-manage their conditions [31-34]. However, there is still limited research on the effectiveness of mental health apps in promoting behavior changes and improving health outcomes [35]. Health apps aiming at *lifestyle interventions* hold great promise, but evidence of their use and efficacy amid the COVID-19 pandemic is sparse [36].

To address this gap, we designed a quasi-experimental study to evaluate the 12-week outcomes of the Gro Health platform, a behavior change intervention for self-management of mental well-being, sleep, activity, and nutrition. The intervention provides education with modules such as stress management; building mental well-being and resilience; benefits of meditation; and guided activities to maintain positive mental well-being, including guided mindfulness-based meditations, classical music, 360-degree immersive guided relaxation videos, and facilitated yoga classes.

The primary study objective was to determine the effectiveness of delivering mental well-being activities using the Gro Health app on self-reported symptoms of anxiety (7-item Generalized Anxiety Disorder scale [GAD-7]), depression (9-item Patient Health Questionnaire [PHQ-9]), and perceived stress (Perceived Stress Scale [PSS]) [37-39]. We posited that greater engagement in the guided activities to maintain positive mental well-being (eg, yoga Tai Chi, Qi Gong, and guided mindfulness) would lead to improvements in anxiety, depression, and perceived stress compared to the baseline.

Methods

Intervention

Gro Health is a digital health intervention that provides behavioral change support through structured education and guided activities in the areas of mental well-being, nutrition, sleep, and activity. The app utilizes a similar behavior change architecture as the Low Carb Program, which has been demonstrated to achieve long-term engagement and sustainable behavior change [21].

Gro Health provides educational and therapeutic behavioral change support in the following four therapeutic areas: (1) mental well-being, (2) sleep, (3) activity, and (4) nutrition. Of particular interest for this study, engagement with the well-being/function of the app was examined. The well-being function of the app provides structured educational modules on topics such as stress management, building mental well-being and resilience, and benefits of meditation, delivered via video and text. These modules are supported with behavior change tools and resources, such as guided mindfulness-based meditations, Qi Gong, Tai Chi, classical music, 360-degree immersive guided relaxation videos, and facilitated yoga classes. Guided activities are of varying lengths (approximately 7-24 minutes long) and are presented in video and podcast format. Please see Multimedia Appendix 1 for example screenshots.

On March 1, 2020, the Gro Health app was updated to include education around minimizing the risk of infection and spread of COVID-19, in response to the feedback received from app users. The education syllabus is detailed in Table 1. The digital platform also provides digital tools for self-monitoring data activity (eg, steps and distance), body weight, blood pressure, heart rate, mood, food intake, body weight, and blood glucose levels. Participants can converse with coaches should they have questions and speak to peers in a moderated peer-to-peer community. Weekly automated feedback is provided to users based on their use of the program through email notifications, and participants are notified weekly to engage within the app. Prior studies demonstrate that peer support may help prevent stress and burnout, anxiety, and symptoms of depression [40-43]. The platform uses artificial intelligence to facilitate conversation and social support within the peer-to-peer support community by presenting community discussions to users that match their interests and demographics, including age, gender, and their self-selected goal.

Table 1. Syllabus of the COVID-19 educational program within the Gro Health app.

Module number	Title	Learning objectives
1	Safety notes	To ensure appropriate clinical safety context is provided and understood
2	Introduction to COVID-19	To understand what COVID-19 is, its origins, and current understandings
3	Symptoms of COVID-19	To define the symptoms of COVID-19 with the latest available evidence.
4	How to stay safe and prevent contracting and spreading the virus	To understand the protocol around social distancing
5	Washing your hands	To ensure washing hands is efficient and being completed often with the most effective technique
6	What to do if you feel unwell	To understand the latest protocol for illness during the COVID-19 pandemic
7	If I get infected with coronavirus, will I get better?	To share the latest available information about the COVID-19 pandemic, in particular the recovery rates
8	What to do if you need to go to hospital or see a doctor	To understand the latest protocol for hospital or doctor appointments during the COVID-19 pandemic
9	Living in self-isolation	Creating a routine to support social distancing while self-isolating (eg, food delivery services)
10	Managing stress	Understanding steps that can be taken to minimize stress levels during the COVID-19 pandemic
11	Mindfulness	Utilizing mindfulness-based practices to support well-being
12	Mental well-being	Understanding steps that can be taken to support mental and emotional well-being during the COVID-19 pandemic
13	What is meditation?	Understanding what meditation is and how it could be incorporated into a daily routine

The content and strategies used in the program are reviewed by primary care physicians and built off prior research and theory [21]. The program encourages participants to select a goal on registration (eg, lose weight, improve fitness, healthier life for family, reduce stress, improve dietary choices, be happier, and improve a health condition). Participants are periodically prompted to consider how close they are to attaining their goal.

Study Design

The study used a quasi-experimental research design consisting of an open-label, single-arm, pre-post intervention. Ethics approval was obtained from the Royal Holloway, University of London ethics review board. Participants were not paid for their participation and accessed the Gro Health app for free. Participants downloaded the app and agreed to terms of service and privacy policy of the Gro Health app, which included consent to use anonymized data for research purposes. Minimal de-identified user data required for the analyses were collected.

Study Participants

We collected a convenience sample of 347 participants aged 22-70 years (mean 49.6, SD 9.24 years) who signed up on the Gro Health app between February 26, 2020, and March 27, 2020. Just over half of participants were female (162/273, 59.3%). All participants were based in the United Kingdom. A total of 40.3% (110/273) of them had full-time employment,

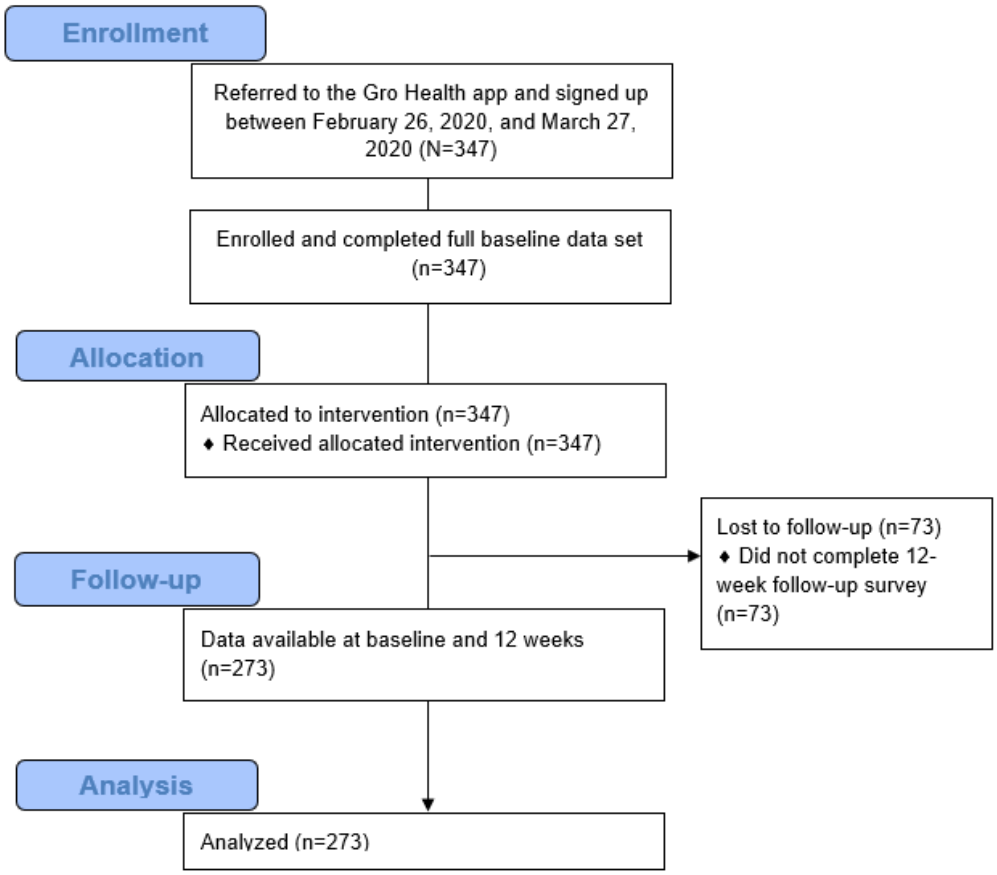
68.1% (186/273) were White, and 80.2% (219/273) reported being obese. An a priori power analysis using G*Power (version 3.1; Heinrich-Heine-Universität Düsseldorf) indicated that a sample size of 270 people would be sufficient to detect a medium effect size (r=0.3) with 80% power, using a linear bivariate regression with α=.05. Thus, our proposed sample size of N=347 was more than adequate for detecting an effect of the linear predictor (well-being engagement) separately based on the outcomes of PHQ, GAD-7, and PSS score changes.

See Table 2 for baseline characteristics and Figure 1 for the participant flowchart of the study. At the baseline, participants' mean age was 49.6 (SD 9.2) years.

Table 2. Baseline characteristics of study participants (N=347).

Characteristic	Value
Age (years), mean (SD)	49.6 (9.2)
Gender, n (%)	
Male	111 (40.7)
Female	162 (59.3)
Employment, n (%)	
Full-time employment	110 (40.3)
Part-time employment	65 (23.8)
Retired	82 (30)
Student	1 (0.4)
Unemployed	15 (5.5)

Figure 1. Participant flow chart used for this study.



Study Measures

Upon sign-up (baseline), participants (ie, app users) were asked to report their age, gender, health goal, and diagnosis of any pre-existing health conditions. They were also asked to complete the following scales: GAD-7, PHQ-9, and PSS. At 12 weeks, participants were asked to complete the same scales again, in the same format. User engagement with the Gro Health app was monitored and recorded as the total number of minutes of active engagement with the app across the 12-week study period. Of

the 347 participants that provided full baseline data, 273 (78.67%) completed the follow-up surveys at 12 weeks.

Statistical Analyses

Analyses were performed using the SPSS software (version 22.0; SPSS Inc). First, paired-sample *t*-tests were performed to compare mean changes in the three outcome measures (ie, anxiety, depression, and perceived stress), as measured by GAD-7, PHQ-9, and PSS, respectively, between the baseline and the 12-week follow-up. Second, a linear regression analysis was used to calculate how in-app engagement in the well-being function of the app (recorded in minutes) predicts participants' change in mental health status, as well as changes in anxiety, depression, and perceived stress scores. Change scores for anxiety were calculated by subtracting follow-up anxiety scores from baseline anxiety scores, with a positive calculated score indicating a reduction in anxiety. Change scores for depression and stress were calculated in the same way. To control for potential effects of demographics and other health-related variables, age, gender, and COVID-19 self-diagnosis were included in the regression as control variables. Occupation status and ethnicity, included as multicategorical variables, were used as factors in one-way analysis of variance (ANOVA) with the three outcome measures. Bonferroni posthoc tests followed up any significant effects. Relevant statistical assumptions were assessed prior to the analysis. The normal distribution of the outcome measures was met, indicating the data was suitable for parametric analyses. Additionally, the assumptions of independence and normal distribution of residuals, linearity, and homoscedasticity were tenable, meaning the data were appropriate for a linear regression analysis.

Results

Overview

Of the 273 study participants who completed both the baseline and follow up surveys, 12.1% (n=33) reported that they had received a positive diagnosis of COVID-19. App engagement was measured through total minutes of use, an analytic indicator used in prior studies to evaluate the effective engagement of digital health apps [44,45]. The mean number of engaged minutes with the well-being function of the Gro Health app was 36.74 (SD 25.9) minutes, as recorded during the 12-week study period.

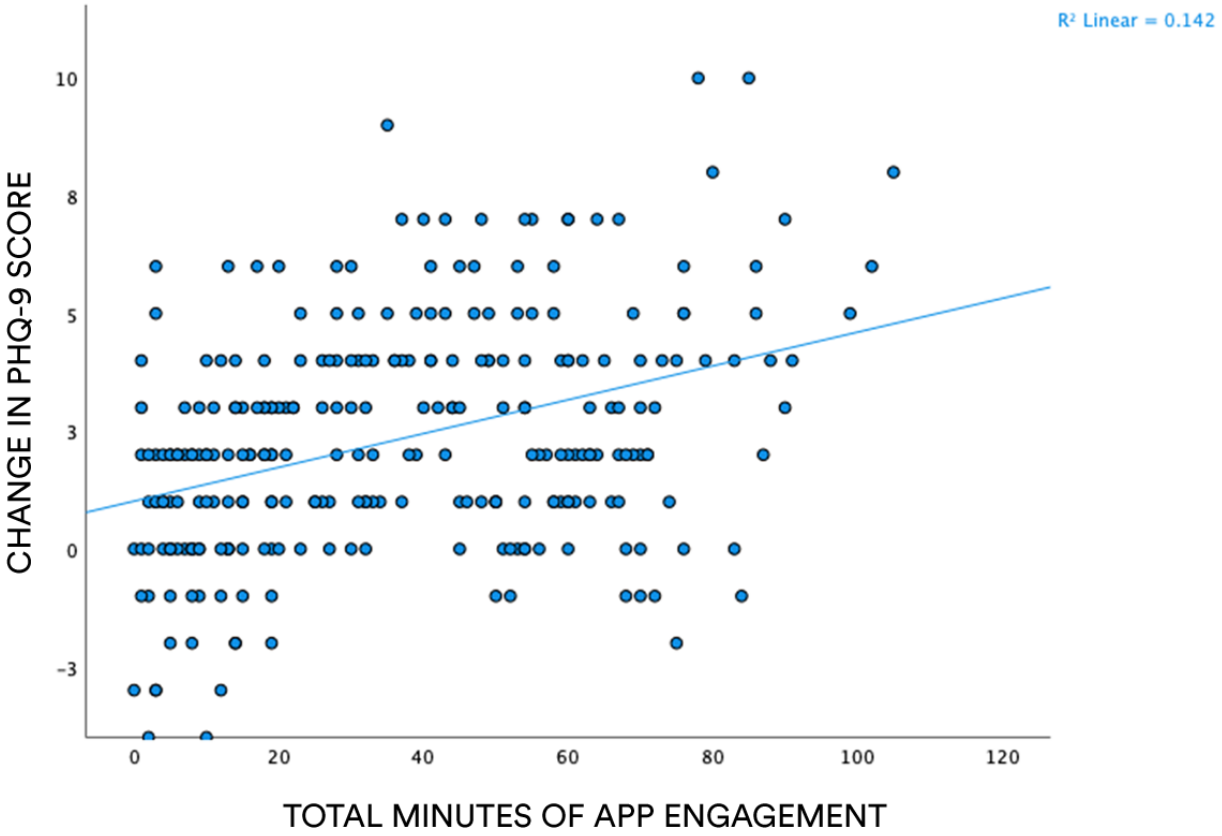
Changes in Depression

Across the 12-week study period, there was a statistically significant change in PHQ-9 scores (reduction in score: mean 2.33, SD 2.97), which is a 32.95% reduction from the baseline mean score of 7.07 (SD 4.62) to the follow-up mean score of 4.74 (SD 3.82) (t_{272}=15.6; 95% CI 2.04, 2.63; P<.001).

As shown in Figure 2, a positive relationship exists between well-being engagement and change in PHQ-9 scores. This observation suggests that individuals who engaged for more time with the app also experienced the greatest reduction in markers of depression. A simple linear regression was calculated to predict participants change in PHQ-9 scores based on well-being engagement, while controlling for demographic and health-related variables. Demographic variables (eg, age and gender) and health-related variables (eg, COVID-19 diagnosis) were first entered into the model as controls, with Gro Health app engagement entered into a separate block as the predictor variable. The final model significantly accounted for 19.5% of the variance in PHQ-9 change scores (R=0.442; R^2=0.195; $F_{1,271}$=10.74; P<.001). Well-being engagement significantly positively predicted PHQ-9 change scores (β=.378; t_{271}=6.8; 95% CI 0.026-0.046; P<.001), and it accounted for 14% of variance in PHQ-9 change scores when controlling for the demographic and prior COVID-19 diagnosis. For every additional minute of engagement with the app, PHQ-9 change scores increased by B=0.036 (unstandardized beta coefficient), implying that participants' depression levels decreased when using the app.

A second linear regression was conducted to evaluate whether well-being engagement could predict PHQ-9 change scores in participants with higher levels of depression (n=50), that is, among app users with scores ≥12 [38]. The regression indicated that similar to the previous analysis, well-being engagement significantly positively predicted PHQ-9 change scores (β=.488; t_{48}=3.654; 95% CI 0.021-0.073; P<.001) and that it accounted for 21% of variance in PHQ-9 change scores, when controlling for the demographic and prior COVID-19 diagnosis. For every additional minute of well-being engagement with the app, PHQ-9 change scores changed by B=0.047, meaning participants with higher engagement levels within the app saw a greater improvement in their PHQ-9 scores.

Figure 2. Scatter plot showing correlation between total time of engagement with the well-being function of the Gro Health app and pre-post test difference in PHQ-9 scores. PHQ-9: 9-item Patient Health Questionnaire.



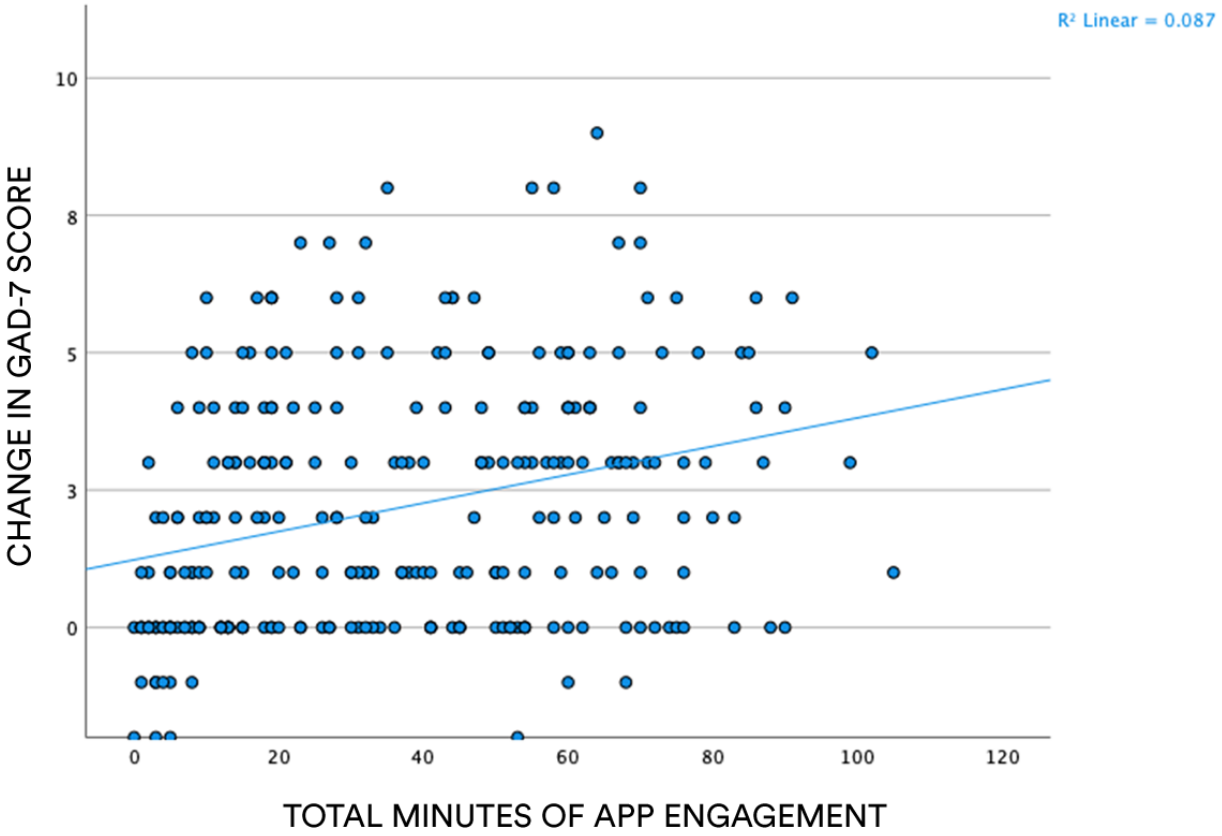
Changes in Anxiety

Across the 12-week study period, there was a statistically significant change in GAD-7 scores (reduction: mean 2.18, SD 2.26), which is a 31.82% reduction from the mean baseline score of 6.85 (SD 3.25) to the mean follow-up score of 4.67 (SD 3.08) (t_{272}=15.9; 95% CI 1.91-2.45; P<.001).

As can be seen in Figure 3, a positive relationship exists between total minutes of app engagement and change in GAD-7 scores, suggesting that, as expected, those individuals who engaged more with the app were also those who experienced a greater reduction in anxiety levels.

A simple linear regression was performed to predict participants change in GAD-7 scores based on Gro Health app engagement, while controlling for demographic and health-related variables. As before, demographic and health-related variables were entered into the model first, followed by the Gro Health app engagement data. The final model significantly accounted for 11.9% of the variance in GAD-7 change scores (R=0.345; R^2=0.11.9; $F_{6,272}$=5.977; P<.001). Gro Health app engagement significantly positively predicted GAD-7 change scores (β=.287; t_{271}=4.938; 95% CI 0.015-0.035; P<.001), and it accounted for 8% of variance in GAD-7 change scores, while controlling for the demographic and prior COVID-19 diagnosis. For every additional minute of engagement with the app, GAD-7 change scores increased by 0.025, implying that participants' anxiety levels decreased when using the app. COVID-19 self-diagnosis (β=.137, t_{271}=–2.3; 95% CI 0.138-1.772; P<.001) predicted change in GAD-7 scores. Individuals self-diagnosed with COVID-19 had greater reduction in anxiety scores (measured by GAD-7: B=0.955) after app engagement than those who were not self-diagnosed with COVID-19.

Figure 3. Scatter plot showing correlation between total time of engagement with the well-being function of the Gro Health app and pre-post test difference in GAD-7 scores. GAD-7: 7-item Generalized Anxiety Disorder scale.



Changes in Perceived Stress

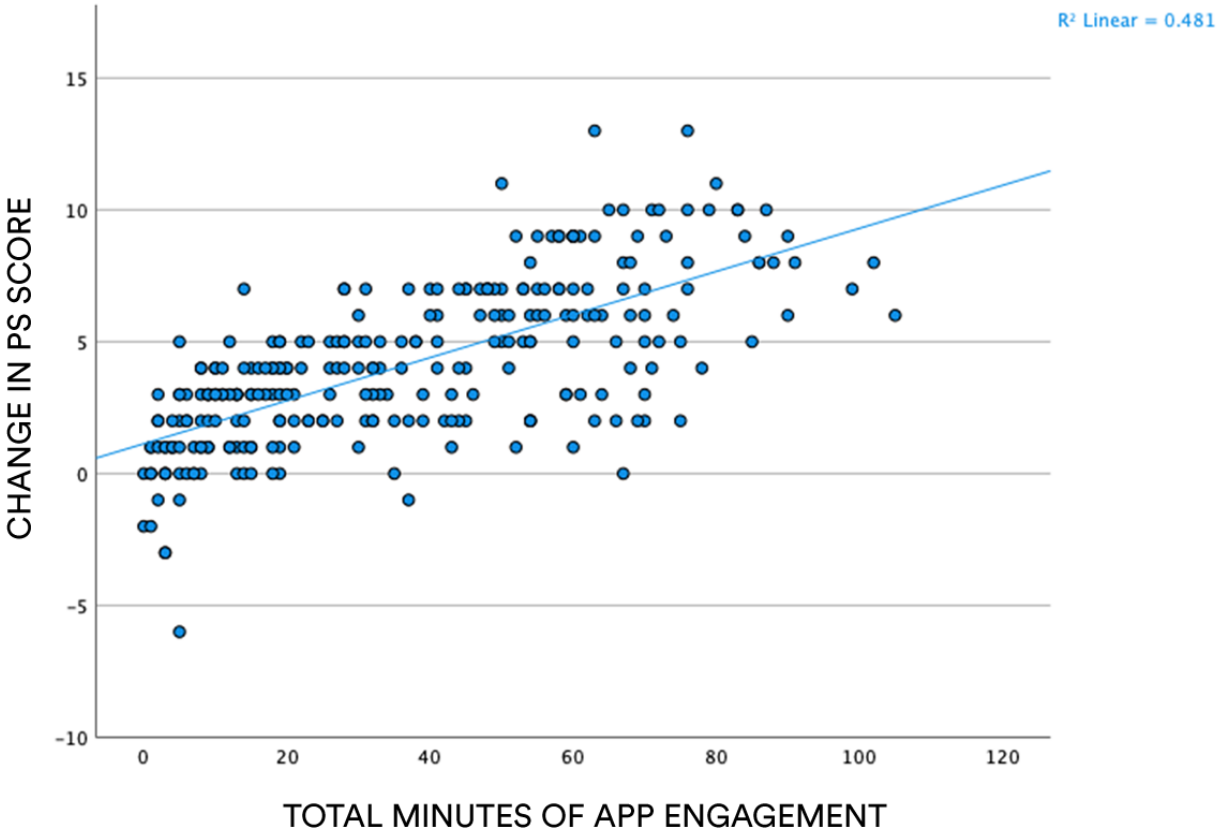
Across the 12-week study period, there was a statistically significant change in perceived stress scores (mean 4.13, SD 3.03), which was a 23.95% reduction from the mean baseline score of 17.24 (SD 3.43) to the mean follow-up score of 13.11 (SD 2.87) (t_{272}=22.4; 95% CI 3.77-4.5; *P*<.001).

As described in Figure 4, a positive relationship exists between the total minutes of app engagement and change in PS scores, suggesting that, as predicted, those individuals who engaged more with the Gro Health app were also those who experienced a greater reduction in their stress levels.

A simple linear regression was performed to predict participants' change in perceived stress scores based on Gro Health app engagement, while controlling for demographic and health-related variables. Similar to the previous analyses, demographics and prior COVID-19 diagnoses were entered into the model first, followed by the Gro Health app engagement predictor. The final model significantly accounted for 49% of the variance in perceived stress change scores (*R*=.70; R^2=.49; $F_{6,272}$=42.61; *P*<.001). Gro Health app engagement significantly positively predicted perceived stress change scores (β=.684; t_{271}=15.461; 95% CI 0.07-0.091; *P*<.001), and it accounted for 45.8% of variance in PS change scores, while controlling for the demographic and prior COVID-19 diagnosis. For every additional minute of engagement with the app, perceived stress change scores increased by 0.081, implying that participants' stress levels decreased when using the app.

A series of one-way ANOVAs were conducted to explore the potential effect of employment status on changes in PHQ-9, GAD-7, and PSS scores. The only significant effect found was for changes in perceived stress scores ($F_{2,270}$=4.969; *P*<.01). A Bonferroni posthoc comparison revealed that change in perceived stress scores were significantly greater for retired participants (mean 5, SD 3.07) than for part-time and full-time individuals (mean 3.78, SD 3.05; *P*<.01). Therefore, those who were retired had greater improvements in perceived stress than presently employed. A series of one-way ANOVAs were conducted to explore potential differences in PHQ-9, GAD-7, and PSS changes scores as a function of ethnicity. The only significant effect found was for changes in PHQ-9 scores ($F_{3,269}$=12.5; *P*<.001). A Bonferroni posthoc test found that White participants reported a significantly smaller change in PHQ-9 scores (mean 1.76, SD 2.25) than did participants of mixed ethnicity (mean 4, SD 1.52; *P*<.001) and those of Indian and Asian ethnicities (mean 3.45, SD 2.88; *P*<.001). Moreover, participants of mixed ethnicity had the greatest improvement in depression levels overall. Overall, Gro Health app engagement had the greatest impact on reducing stress scores (ie, change in perceived stress scores).

Figure 4. Scatter plot showing correlation between total time of engagement with the well-being function of the Gro Health app and pre-post test difference in PS scores. PS: perceived stress.



Discussion

Principal Findings

The results of this study are consistent with prior research on the use of digital health interventions for mental health, showing improvements in the symptoms of perceived stress, depression, and anxiety [46-48]. Over three-quarters (273/347, 78.67%) of the participants who signed up for the Gro Health app completed both the baseline and follow-up surveys. There were statistically significant interactions between app engagement and change in mental health scores, with the greatest effect observed in perceived stress scores.

The findings indicate that digital health solutions such as the Gro Health app that provide mental well-being resources could be of significant benefit when provided at scale to help address the growing mental health crises faced by global health services in the wake of the COVID-19 pandemic [49,50].

Greater changes in perceived stress scores were reported among participants who were retired compared to the rest of the population. The results from this study contradict prior evidence on the differences in employment status during the COVID-19 pandemic on mental health and well-being. Prior research evaluating employment and mental health during the initial stages of the pandemic found that compared to their unemployed counterparts, individuals who were employed at the start of the pandemic reported lower levels of mental health distress, higher levels of psychosocial well-being, better overall quality of life, and lower levels of overall loneliness, social loneliness, and emotional loneliness [51]. Further research should explore the variances in impact of this intervention in different employment groups.

There were differences in the impact of the app on depression scores between ethnic groups. For instance, individuals of Indian or Asian ethnicity reported greater improvements in their depression scores. This may be due to several factors, including the fact that mindfulness-based approaches may be more acceptable due to their grounding benefits in Eastern traditions [52]. Additionally, these results support prior research demonstrating that to engage people of various cultures and ethnicities within digital health solutions, they must be adapted to satisfy individual needs [53].

Individuals self-diagnosed with COVID-19 had greater reduction in anxiety scores after engaging with the app compared to those who were not self-diagnosed. Further research should explore whether the behavior change techniques provided by the app may help the longer-term mental health crisis that may occur in the aftermath of the pandemic both among those with a prior diagnosis of COVID-19 and those living with “long COVID.”

Limitations

Our study has several limitations. We encouraged participants to engage in guided activities; however, we did not control for participants using other functionality of the app alongside the mental well-being tools. We also measured health outcomes (ie, PHQ-9, GAD-7, and PSS) by using patient self-report, rather than measuring those through medical records. However, previous research has found that these self-reported health

outcomes can be quite close to actual values [54]. Similarly, patterns of engaged minutes with the app were not analyzed. Although outside the scope of this study, further analysis should examine whether patterns of app engagement impact the levels of change in mental health scores.

This was not a randomized controlled trial, so we cannot compare the 12-week results to a control or standard-of-care group. Therefore, the results of our trial should be interpreted cautiously because the study used convenience sampling, a single-arm design, and pre-post self-reported outcomes. Criticism for convenience sample use includes the lack of control over potential intervening variables (such as other stress-mitigating or mental well-being–boosting activities) that active participants may have been engaging in. Since the sample is not representative of the population, these results are not representative of the entire population. The study results have low external validity. However, a significant proportion of participants were from hard-to-reach populations (eg, retired, unemployed, and ethnic groups) or those diagnosed with type 2 diabetes, hypertension, or high cholesterol. Individuals with these health conditions are more at-risk of COVID-19 and its complications, which makes this sample of particular interest [55].

The sample contained 33 participants who self-reported a COVID-19 diagnosis. As the research context is set in the pandemic, analysis of covariance (ANCOVA) was conducted to evaluate whether app use had a greater effect on those who had COVID-19. Mixed results were obtained from these analyses. When app engagement time was controlled, a statistically significant difference was observed with regard to GAD-7 scores (P=.004) between the two groups, but not with regard to PHQ-9 and PSS scores.

Despite several limitations, the study suggests a digital platform that provides remote mental well-being support can be extremely effective in managing depression, anxiety, and perceived stress levels during times of crisis such as the current COVID-19 pandemic. Future studies with 6- and 12-month post-tests will provide a stronger assessment of the impact of the intervention on mental health outcomes.

Further research should explore the most appropriate mechanisms by which such digital health interventions can be scaled to help manage and mitigate the mental health demands that will inevitably follow a global pandemic of the magnitude of COVID-19 [49,50]. Additional research should also evaluate the use of the intervention in patients diagnosed with long-COVID. Emerging evidence suggests 20% of patients diagnosed with COVID experience symptoms of long COVID, including anxiety, breathlessness, and fatigue [56]. Furthermore, studies should explore whether the behavior change techniques recommended by the Gro Health app could also help alleviate the symptoms or burden of long COVID.

Although our design does not support causal conclusions, further research should investigate the particular in-app guided activities that might have an impact on users' perceived stress, depression, and anxiety scores. In addition, it is important to conduct further research into identifying for which participants self-guided digital interventions may be sufficient in supporting mental well-being, and those who may need triaging to receive additional face-to-face or intensive support.

Conclusions

A digital platform providing remote mental well-being support (Gro Health) can be effective in managing depression, anxiety, and perceived stress during times of crisis, such as the current COVID-19 pandemic. Further research should investigate how best to implement such digital health solutions at scale to mitigate the burden on national health services during pandemics.

Acknowledgments

We thank Dr Peter Foley for his medical governance and support.

Conflicts of Interest

CS is employed by DDM, which operates the Gro Health app. The rest of the authors declare no conflicts of interest.

Multimedia Appendix 1
Gro Health app screenshots showing well-being function, educational modules, and mindfulness-based behavior change activity.
[PNG File , 522 KB - formative_v5i10e31273_app1.png]

References

1. Emanuel EJ, Persad G, Upshur R, Thome B, Parker M, Glickman A, et al. Fair allocation of scarce medical resources in the time of COVID-19. N Engl J Med 2020 May 21;382(21):2049-2055. [doi: 10.1056/nejmsb2005114]
2. People at higher risk from coronavirus (COVID-19). NHS (National Health Service) UK. URL: https://www.nhs.uk/conditions/coronavirus-covid-19/advice-for-people-at-high-risk/ [accessed 2020-05-01]
3. Choi KR, Heilemann MV, Fauer A, Mead M. A second pandemic: mental health spillover from the novel coronavirus (covid-19). J Am Psychiatr Nurses Assoc 2020 Apr 27;26(4):340-343. [doi: 10.1177/1078390320919803]
4. Depression: A Global Crisis World Mental Health Day, October 10 2012. World Federation for Mental Health. 2012. URL: https://www.who.int/mental_health/management/depression/wfmh_paper_depression_wmhd_2012.pdf [accessed 2020-03-31]

5. Yao H, Chen J, Xu Y. Patients with mental health disorders in the COVID-19 epidemic. Lancet Psychiatry 2020 Apr;7(4):e21. [doi: 10.1016/s2215-0366(20)30090-0]
6. Duan L, Zhu G. Psychological interventions for people affected by the COVID-19 epidemic. Lancet Psychiatry 2020 Apr;7(4):300-302. [doi: 10.1016/s2215-0366(20)30073-0]
7. Gabbat A. 'Social recession': how isolation can affect physical and mental health. The Guardian. 2020 Mar 18. URL: https://www.theguardian.com/world/2020/mar/18/coronavirus-isolation-social-recession-physical-mental-health [accessed 2020-03-31]
8. Sartorius N. CS01-01 - Comorbidity of mental and physical disorders: a central challenge for medicine of the 21st century. European Psychiatry 2012 Jan;27:1. [doi: 10.1016/s0924-9338(12)74058-0]
9. Miceli M, Castelfranchi C. Anxiety as an "epistemic" emotion: an uncertainty theory of anxiety. Anxiety, Stress & Coping 2005 Dec;18(4):291-319. [doi: 10.1080/10615800500209324]
10. Carleton RN, Mulvogue MK, Thibodeau MA, McCabe RE, Antony MM, Asmundson GJ. Increasingly certain about uncertainty: intolerance of uncertainty across anxiety and depression. J Anxiety Disord 2012 Apr;26(3):468-479. [doi: 10.1016/j.janxdis.2012.01.011] [Medline: 22366534]
11. Greco V, Roger D. Uncertainty, stress, and health. Personality and Individual Differences 2003 Apr;34(6):1057-1068. [doi: 10.1016/s0191-8869(02)00091-0]
12. Mishel MH. Perceived uncertainty and stress in illness. Res Nurs Health 1984 Sep;7(3):163-171. [doi: 10.1002/nur.4770070304] [Medline: 6567948]
13. AlNajjar NS, Attar LM, Farahat FM, AlThaqafi A. Psychobehavioural responses to the 2014 Middle East respiratory syndrome-novel corona virus (MERS CoV) among adults in two shopping malls in Jeddah, western Saudi Arabia. East Mediterr Health J 2017 Feb 01;22(11):817-823 [FREE Full text] [doi: 10.26719/2016.22.11.817] [Medline: 28177112]
14. Xiang Y, Yang Y, Li W, Zhang L, Zhang Q, Cheung T, et al. Timely mental health care for the 2019 novel coronavirus outbreak is urgently needed. Lancet Psychiatry 2020 Mar;7(3):228-229 [FREE Full text] [doi: 10.1016/s2215-0366(20)30046-8]
15. Roca M, Gili M, Garcia-Campayo J, García-Toro M. Economic crisis and mental health in Spain. The Lancet 2013 Dec;382(9909):1977-1978. [doi: 10.1016/s0140-6736(13)62650-1]
16. 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 Sep;7(9):813-824. [doi: 10.1016/s2215-0366(20)30307-2]
17. Cook BL, Trinh N, Li Z, Hou SS, Progovac AM. Trends in racial-ethnic disparities in access to mental health care, 2004–2012. PS 2017 Jan;68(1):9-16. [doi: 10.1176/appi.ps.201500453]
18. Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, et al. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States. Results from the National Comorbidity Survey. Arch Gen Psychiatry 1994 Jan 01;51(1):8-19. [doi: 10.1001/archpsyc.1994.03950010008002] [Medline: 8279933]
19. Feldner MT, Zvolensky MJ, Schmidt NB. Prevention of anxiety psychopathology: a critical review of the empirical literature. Clinical Psychology: Science and Practice 2004;11(4):405-424. [doi: 10.1093/clipsy.bph098]
20. Panesar A, Panesar H. Artificial intelligence and machine learning in global healthcare. In: Haring R, editor. Handbook of Global Health. New York: Springer; May 12, 2021:1775-1813.
21. Summers C, Curtis K. Novel digital architecture of a "Low Carb Program" for initiating and maintaining long-term sustainable health-promoting behavior change in patients with type 2 diabetes. JMIR Diabetes 2020 Mar 04;5(1):e15030 [FREE Full text] [doi: 10.2196/15030] [Medline: 32130113]
22. Depression and Other Common Mental Disorders Global Health Estimates. Geneva, Switzerland: World Health Organization; 2017. URL: https://apps.who.int/iris/bitstream/handle/10665/254610/WHO-MSD-MER-2017.2-eng.pdf [accessed 2021-09-13]
23. Saarijärvi S, Salminen JK, Toikka T, Raitasalo R. Health-related quality of life among patients with major depression. Nord J Psychiatry 2002 Jul 12;56(4):261-264. [doi: 10.1080/08039480260242741] [Medline: 12470316]
24. Pulcu E, Elliott R. Neural origins of psychosocial functioning impairments in major depression. Lancet Psychiatry 2015 Sep;2(9):835-843. [doi: 10.1016/s2215-0366(15)00237-0]
25. Mental Health Atlas 2017. Geneva, Switzerland: World Health Organization; 2018:68.
26. Sethi S, Campbell AJ, Ellis LA. The use of computerized self-help packages to treat adolescent depression and anxiety. J Technol Hum Serv 2010 Aug 31;28(3):144-160. [doi: 10.1080/15228835.2010.508317]
27. Van't Hof E, Cuijpers P, Stein DJ. Self-help and Internet-guided interventions in depression and anxiety disorders: a systematic review of meta-analyses. CNS Spectr 2009 Feb 07;14(2 Suppl 3):34-40. [doi: 10.1017/s1092852900027279] [Medline: 19238128]
28. Cuijpers P, Donker T, van Straten A, Li J, Andersson G. Is guided self-help as effective as face-to-face psychotherapy for depression and anxiety disorders? A systematic review and meta-analysis of comparative outcome studies. Psychol Med 2010 Apr 21;40(12):1943-1957. [doi: 10.1017/s0033291710000772]
29. Johnson S, Dalton-Locke C, Vera San Juan N, Foye U, Oram S, Papamichail A, COVID-19 Mental Health Policy Research Unit Group. Impact on mental health care and on mental health service users of the COVID-19 pandemic: a mixed methods survey of UK mental health care staff. Soc Psychiatry Psychiatr Epidemiol 2021 Jan;56(1):25-37 [FREE Full text] [doi: 10.1007/s00127-020-01927-4] [Medline: 32857218]

30. Chandrashekar P. Do mental health mobile apps work: evidence and recommendations for designing high-efficacy mental health mobile apps. Mhealth 2018;4:6 [FREE Full text] [doi: 10.21037/mhealth.2018.03.02] [Medline: 29682510]
31. Rathbone AL, Prescott J. The use of mobile apps and SMS messaging as physical and mental health interventions: systematic review. J Med Internet Res 2017 Aug 24;19(8):e295 [FREE Full text] [doi: 10.2196/jmir.7740] [Medline: 28838887]
32. Lecomte T, Potvin S, Corbière M, Guay S, Samson C, Cloutier B, et al. Mobile apps for mental health issues: meta-review of meta-analyses. JMIR Mhealth Uhealth 2020 May 29;8(5):e17458 [FREE Full text] [doi: 10.2196/17458] [Medline: 32348289]
33. Firth J, Torous J, Nicholas J, Carney R, Rosenbaum S, Sarris J. Can smartphone mental health interventions reduce symptoms of anxiety? a meta-analysis of randomized controlled trials. J Affect Disord 2017 Aug 15;218:15-22 [FREE Full text] [doi: 10.1016/j.jad.2017.04.046] [Medline: 28456072]
34. Gleitman H, Gross J, Reisberg D. Psychology, 8th edition. New York, NY: W.W. Norton & Company; Apr 16, 2010.
35. Milne-Ives M, Lam C, De Cock C, Van Velthoven MH, Meinert E. Mobile apps for health behavior change in physical activity, diet, drug and alcohol use, and mental health: systematic review. JMIR Mhealth Uhealth 2020 Mar 18;8(3):e17046 [FREE Full text] [doi: 10.2196/17046] [Medline: 32186518]
36. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
37. Williams N. The GAD-7 questionnaire. Occupational Medicine 2014 Apr 03;64(3):224-224. [doi: 10.1093/occmed/kqt161]
38. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatric Annals 2002 Sep;32(9):509-515. [doi: 10.3928/0048-5713-20020901-06]
39. Cohen S, Kessler R, Gordon L. Measuring stress: a guide for health and social scientists. Oxford: Oxford University Press; Dec 12, 1997.
40. Peterson U, Bergström G, Samuelsson M, Asberg M, Nygren A. Reflecting peer-support groups in the prevention of stress and burnout: randomized controlled trial. J Adv Nurs 2008 Sep;63(5):506-516. [doi: 10.1111/j.1365-2648.2008.04743.x] [Medline: 18727753]
41. Parent N, Fortin F. A randomized, controlled trial of vicarious experience through peer support for male first-time cardiac surgery patients: impact on anxiety, self-efficacy expectation, and self-reported activity. Heart Lung 2000 Nov;29(6):389-400. [doi: 10.1067/mhl.2000.110626] [Medline: 11080319]
42. Heisler M. Overview of peer support models to improve diabetes self-management and clinical outcomes. Diabetes Spectrum 2007 Oct 01;20(4):214-221. [doi: 10.2337/diaspect.20.4.214]
43. Prevatt B, Lowder EM, Desmarais SL. Peer-support intervention for postpartum depression: participant satisfaction and program effectiveness. Midwifery 2018 Sep;64:38-47. [doi: 10.1016/j.midw.2018.05.009] [Medline: 29908406]
44. Pham Q, Graham G, Carrion C, Morita PP, Seto E, Stinson JN, et al. A library of analytic indicators to evaluate effective engagement with consumer mHealth apps for chronic conditions: scoping review. JMIR Mhealth Uhealth 2019 Jan 18;7(1):e11941 [FREE Full text] [doi: 10.2196/11941] [Medline: 30664463]
45. Lin P, Grambow S, Intille S, Gallis JA, Lazenka T, Bosworth H, et al. The association between engagement and weight loss through personal coaching and cell phone interventions in young adults: randomized controlled trial. JMIR Mhealth Uhealth 2018 Oct 18;6(10):e10471 [FREE Full text] [doi: 10.2196/10471] [Medline: 30341051]
46. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2006 Nov 20;37(03):319. [doi: 10.1017/s0033291706008944]
47. Kählke F, Berger T, Schulz A, Baumeister H, Berking M, Cuijpers P, et al. Efficacy and cost-effectiveness of an unguided, internet-based self-help intervention for social anxiety disorder in university students: protocol of a randomized controlled trial. BMC Psychiatry 2019 Jun 25;19(1):197 [FREE Full text] [doi: 10.1186/s12888-019-2125-4] [Medline: 31238907]
48. Domhardt M, Geßlein H, von Rezori RE, Baumeister H. Internet- and mobile-based interventions for anxiety disorders: A meta-analytic review of intervention components. Depress Anxiety 2019 Mar 19;36(3):213-224. [doi: 10.1002/da.22860] [Medline: 30450811]
49. Galea S, Merchant RM, Lurie N. The mental health consequences of COVID-19 and physical distancing: the need for prevention and early intervention. JAMA Intern Med 2020 Jun 01;180(6):817-818. [doi: 10.1001/jamainternmed.2020.1562] [Medline: 32275292]
50. Estes KD, Thompson RR. Preparing for the aftermath of COVID-19: Shifting risk and downstream health consequences. Psychol Trauma 2020 Aug;12(S1):S31-S32. [doi: 10.1037/tra0000853] [Medline: 32478540]
51. Ruffolo M, Price D, Schoultz M, Leung J, Bonsaksen T, Thygesen H, Geirdal. Employment uncertainty and mental health during the COVID-19 pandemic initial social distancing implementation: a cross-national study. Glob Soc Welf 2021 Jan 07;8(2):1-10 [FREE Full text] [doi: 10.1007/s40609-020-00201-4] [Medline: 33432284]
52. Amaro A. A holistic mindfulness. Mindfulness 2015 Jan 4;6(1):63-73. [doi: 10.1007/s12671-014-0382-3]
53. Ouhbi S, Fernández-Alemán JL, Carrillo-de-Gea JM, Toval A, Idri A. E-health internationalization requirements for audit purposes. Comput Methods Programs Biomed 2017 Jun 01;144:49-60. [doi: 10.1016/j.cmpb.2017.03.014] [Medline: 28495006]

54. Løvaas KF, Cooper JG, Sandberg S, Røraas T, Thue G. Feasibility of using self-reported patient data in a national diabetes register. BMC Health Serv Res 2015 Dec 15;15(1):553 [FREE Full text] [doi: 10.1186/s12913-015-1226-0] [Medline: 26666413]
55. Zaki N, Alashwal H, Ibrahim S. Association of hypertension, diabetes, stroke, cancer, kidney disease, and high-cholesterol with COVID-19 disease severity and fatality: A systematic review. Diabetes & Metabolic Syndrome: Clinical Research & Reviews 2020 Sep 01;14(5):1133-1142. [doi: 10.1016/j.dsx.2020.07.005]
56. The prevalence of long COVID symptoms and COVID-19 complications.: Office for National Statistics; 2020 Dec 16. URL: https://www.ons.gov.uk/news/statementsandletters/theprevalenceoflongcovidsymptomsandcovid19complications [accessed 2021-06-04]

Abbreviations

ANCOVA: analysis of covariance
ANOVA: analysis of variance
GAD: generalized anxiety disorder
GAD-7: 7-item Generalized Anxiety Disorder scale
MERS-CoV: Middle East respiratory syndrome coronavirus
PHQ-9: 9-item Patient Health Questionnaire
PSS: Perceived Stress Scale

Edited by G Eysenbach; submitted 16.06.21; peer-reviewed by PH Lin; comments to author 06.08.21; revised version received 12.08.21; accepted 24.08.21; published 06.10.21.

Please cite as:
Summers C, Wu P, Taylor AJG
Supporting Mental Health During the COVID-19 Pandemic Using a Digital Behavior Change Intervention: An Open-Label, Single-Arm, Pre-Post Intervention Study
JMIR Form Res 2021;5(10):e31273
URL: https://formative.jmir.org/2021/10/e31273
doi:10.2196/31273
PMID:34459740

Original Paper

Development of a Mobile Health Application for HIV Prevention Among At-Risk Populations in Urban Settings in East Africa: A Participatory Design Approach

Wilhellmuss Mauka[1,2], MSc, MD; Christopher Mbotwa[3,4], BSc, MSc; Kåre Moen[5], MD, PhD; Hanne Ochieng Lichtwarck[5], MD, MPhil; Inga Haaland[5], BA, MPhil; Method Kazaura[3], PhD; Germana H Leyna[6], MD, PhD; Melkizedeck T Leshabari[1], PhD; Elia J Mmbaga[3,5], MD, PhD

[1]Department of Behavioural Science, Muhimbili University of Health and Allied Sciences, Dar es Salaam, United Republic of Tanzania

[2]Ministry of Health, Community Development, Gender, Elderly and Children, Dodoma, United Republic of Tanzania

[3]Department of Epidemiology and Biostatistics, Muhimbili University of Health and Allied Sciences, Dar es Salaam, United Republic of Tanzania

[4]Mbeya College of Health and Allied Sciences, University of Dar es Salaam, Mbeya, United Republic of Tanzania

[5]Department of Community Medicine and Global Health, Faculty of Medicine, Institute of Health and Society, University of Oslo, Oslo, Norway

[6]Tanzania Food and Nutrition Centre, Dar es Salaam, United Republic of Tanzania

Corresponding Author:
Wilhellmuss Mauka, MSc, MD
Department of Behavioural Science
Muhimbili University of Health and Allied Sciences
PO Box 65015
Dar es Salaam
United Republic of Tanzania
Phone: 255 763225717
Email: wilhemauka@yahoo.com

Abstract

Background: There is limited evidence in Africa on the design and development of mobile health (mHealth) applications to guide best practices and ensure effectiveness. A pragmatic trial for HIV pre-exposure prophylaxis roll-out among key populations in Tanzania is needed.

Objective: We present the results of the development of a mobile app (Jichunge) intended to promote adherence to pre-exposure prophylaxis (PrEP) among men who have sex with men (MSM) and female sex workers (FSW) in Tanzania.

Methods: A participatory design approach was employed and guided by the information system research framework. MSM and FSW were the target populations. A total of 15 MSM and 15 FSW were engaged in the relevance and design cycles, while the piloting phase included 10 MSM and 20 FSW.

Results: The relevance cycle enabled the description of the existing problem, provided the compatible app features for the target population, and identified the need to develop an mHealth app that provides health services in a stigmatizing and discriminating environment. User involvement in the app's design and evaluation provided an opportunity to incorporate social, cultural, and community-specific features that ensured usability. In addition, the participants suggested valuable information to inform the app, text message services, medication registration, and chat platform designs.

Conclusions: The participatory design approach in the development of mHealth apps is useful in identifying and validating population-specific functional features, improve usability, and ensuring future health impacts. Through this participatory process, the Jichunge app took end-user needs, perspectives, and experiences into account, eliciting enthusiasm regarding its potential role in supporting pre-exposure prophylaxis adherence for HIV and related behavioral change promotion.

Trial Registration: International Clinical Trials Registry Platform PACTR202003823226570; https://trialsearch.who.int/Trial2.aspx?TrialID=PACTR202003823226570

(JMIR Form Res 2021;5(10):e23204) doi:10.2196/23204

KEYWORDS

mHealth application; participatory design; HIV; pre-exposure prophylaxis; Africa; female sex workers; sex and gender minorities

Introduction

As part of a pragmatic trial in Tanzania, this paper describes the process of developing a mobile health (mHealth) application to improve adherence to pre-exposure prophylaxis (PrEP) against HIV among men who have sex with men (MSM) and female sex workers (FSW).

Over the last decade, there has been a steady increase in global mobile phone ownership, including sub-Saharan Africa, which has led to remarkable growth in mobile device use in curative and preventive health care service delivery [1,2]. Short text–based messaging was the mainstay of mobile health communication in the past. However, smartphone applications with increased functionality have become more commonplace in recent years. They have been used to prevent disease and promote health, including sexual and reproductive health [3-7]. mHealth apps have been reported to be useful in disease self-management, customized to clients' and local needs, and providing on-demand interventions as well as communitywide health promotion [8,9].

Among other things, mHealth apps have been used to improve drug adherence in HIV care by reminding clients to take a particular medication, providing health education, and counseling and services [10-12].

A systematic review on mHealth apps in low-income and middle-income countries has reported increased use of various apps focusing on patient care, supporting the health system, and general population health [13,14].

In 2015, the World Health Organization recommended PrEP for HIV prevention among at-risk populations [15]. Like other parts of the world, poor adherence to medication among the high-risk population has been a significant challenge to implementing PrEP intervention in Africa [16-18]. In 2018, Tanzania initiated plans for a PrEP roll-out, but suboptimal adherence due to stigma, stock shortages, forgetfulness in taking daily pills, or attending clinics for medication refills pose a challenge to its effectiveness [19,20].

Since more than half of the mobile subscribers in Tanzania have access to the internet [21], and social network usage is common [22], mHealth apps offer an opportunity to promote adherence to PrEP in the country.

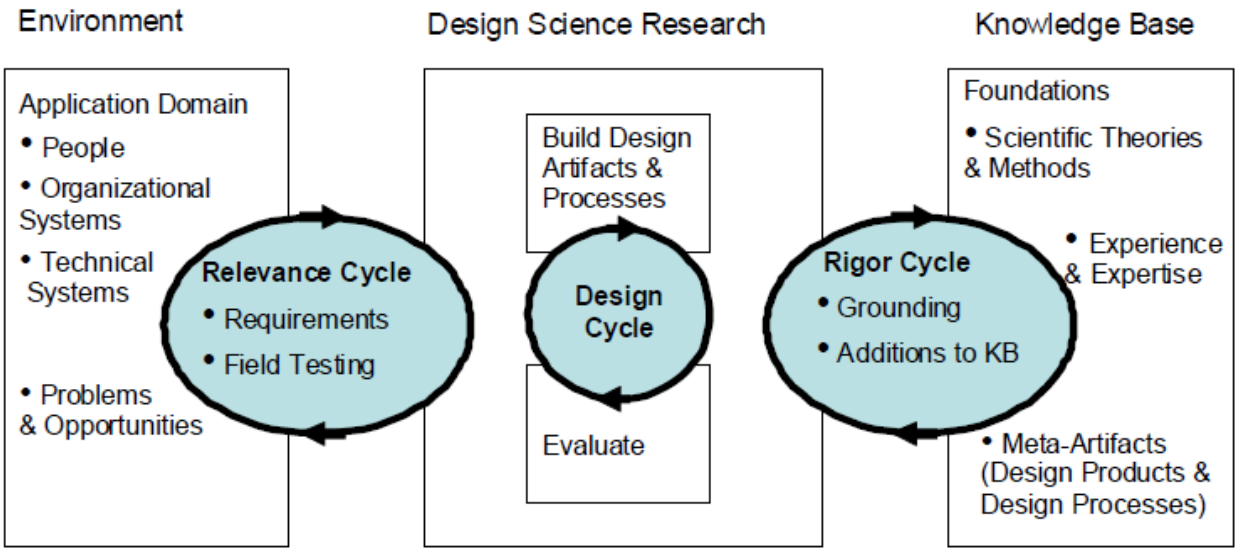
The reported increase in mobile apps usage for health calls for the documentation of best practices in the development process of these interventions. Such evidence can guide future development and increase the effectiveness of mHealth apps in promoting health. Various development approaches have been evaluated in Asia and the United States, several of which have underlined the critical importance of end-user involvement in the design and development process [23-25]. End-user involvement is paramount when developing mHealth apps targeting populations that face a high level of stigma and discrimination, such as MSM and FSW. While mobile apps have increasingly come into use among stigmatized and criminalized people in sub-Saharan Africa, there is limited data on the development process to inform best practices in the region and among these populations [26]. This paper aims to present the development process of an mHealth app (Jichunge app) for PrEP adherence among MSM and FSW in Tanzania.

Methods

Design

A participatory design approach guided by a framework reminiscent of the information system research framework (ISR) was employed in the Jichunge app development [27]. ISR emanates from design science and system engineering commonly used to develop complex information systems [28]. The framework urges end-users' greater inclusion and active participation in designing and evaluating application systems [23,29]. The ISR framework consists of three overarching user participatory design cycles. The relevance cycle determines end-user requirements; the design cycle involves prototype development and evaluation; the rigor cycle focuses on assessing underpinning theories (Figure 1). This paper focuses on the relevance and design cycles that are critical in end-product acceptability and usability. Therefore, user involvement was paramount in these undertakings. We supported end-user involvement in the conceptualization, design, and testing by considering their views, perspectives, comments, and suggestions at different levels of the participatory continuum [30].

Figure 1. The information system research framework. KB: Knowledge Base.



Population

MSM (men who regularly or occasionally have sex with men) and FSW (women who exchange sex for money or goods) living in Dar es Salaam, Tanzania, participated in this study as members of the target population or end-users. Eligibility involved being at least 18 years and a resident of Dar es Salaam, the largest metropolitan city in Tanzania. The resident was defined as having an address and having lived in the city for the past 6 months preceding the study.

Additionally, at each stage of development, the app was reviewed by a team of experts in HIV behavior and research in key populations. The prototype evaluators involved a multidisciplinary team consisting of PhD students, medical researchers, epidemiologists, biostatisticians, public health specialists, and clinical research coordinators.

Participant Sampling and Recruitment

Purposive sampling was used to enroll members of the target population (participants) into the study. Based on the investigators' previous knowledge of working with the target populations, participants were recruited based on their potential to provide rich information on the issues related to PrEP use, challenges, and opportunities. Moreover, different strata of the population by age and geographical location in the city, mobile phone ownership, leadership position within target population grassroots organizations, and economic structures were critical in providing diverse views on PrEP and mobile phone use to address health challenges among the target population.

Development Process

Relevance Cycle

The study gathered and synthesized information related to the Jichunge app's relevance to the target populations through the relevance cycle step. Firstly, the investigators consulted members of the target populations for problem definition and justification of the app's relevance, including need and potential challenges. End-users' consultation aimed to explore their views on adherence to PrEP and whether a mobile app would be relevant in addressing those challenges. Secondly, using investigators' experience working with these population subgroups [31-34], a detailed and appropriate literature review was conducted to validate the challenges in PrEP adherence raised and explore the proposed Jichunge app's needs to address PrEP adherence among the members of the target populations. Furthermore, researchers reviewed literature on mobile phone access and usage involving the target population, using available social media platforms to inform the design and assess future application challenges and opportunities.

Design Cycle

Creating an App Prototype

In creating the app prototype, consultation with target populations led to the consideration of developing an app that accommodates both iOS and Android smartphone operating systems available in Tanzania. Due to the potential for limited mobile data among the key population, users recommended developing an application that allows access to app features such as notification messages (reminding them to take their medication) even in the absence of mobile data. Similarly, the app should enable the client to register daily pills taken for drug adherence, monitoring, and virtualizing the pill calendar even when the user lacks mobile data. An app's development with these innovations was led by a company based in Norway with vast software development experiences [35].

Stage I of Testing the Prototype and Modifications: Prototype Preliminary Reviews

The initial prototype was installed into investigators' mobile phones as a demonstration version with the essential functions. Next, the investigators reviewed the various app functionalities and evaluated their relevance per planned use among the intended study populations.

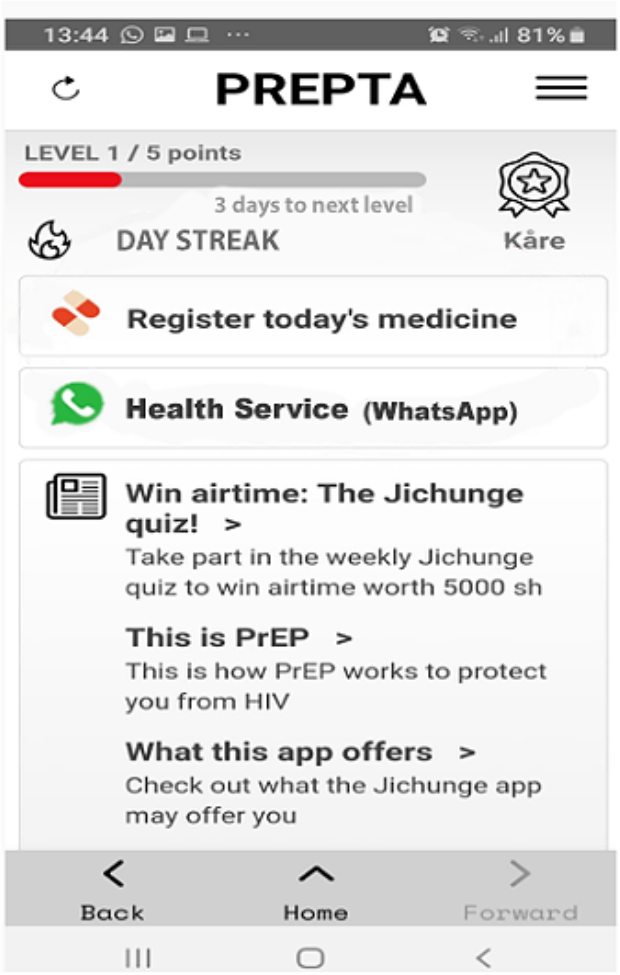
Stage II of Testing the Prototype and Modifications: Target Populations Design Inputs

Various external outlooks and logos were designed (Figure 2) and shared with 15 members from each end-user group (MSM and FSW) in informal, exploratory individual and small group discussions. The participants were FSW and MSM with a mean age of 26 years (IQR 20-48), the majority (18/30; 60%) had completed primary school, and 6.6% (2/30) had a college diploma. In addition, most of the participants (21/30; 70%) were self-employed.

Figure 2. Proposed and designed logos.

User involvement at this design stage aimed to receive user perspectives, ensure their needs are met, and explore anticipated challenges to inform the app design. End-users' discussion topics in small groups included the app name, layout, and functions. Figure 3 illustrates the sample outlooks and app name as proposed by end-users. All suggestions provided by the end-users were submitted to the software developer for app modification. The intended outcome was to develop an outlook and application name, which is original, user-friendly, and stigma and discrimination-free.

Figure 3. Previous outlook and name of the app. PrEP: pre-exposure prophylaxis.



Rigor Cycle

Piloting of the Jichunge App

The pilot was carried out in Dar es Salaam, where the app will be rolled out and evaluated. FSW and MSM end-user populations were recruited through their peers to participate in the app's piloting phase.

We conducted 2 focus group discussions (FGD) before and after using the app to explore participants' perceptions, experiences, and challenges associated with using the Jichunge app. The FGDs involved 20 members of the HIV at-risk populations (10 FSW and 10 MSM). The participants' median age was 27 years (IQR 21-44), the majority (12/20; 70%) had completed primary school, and 65% (13/20) were self-employed. Two PhD students led the FGDs, one as a facilitator and the other as a notetaker.

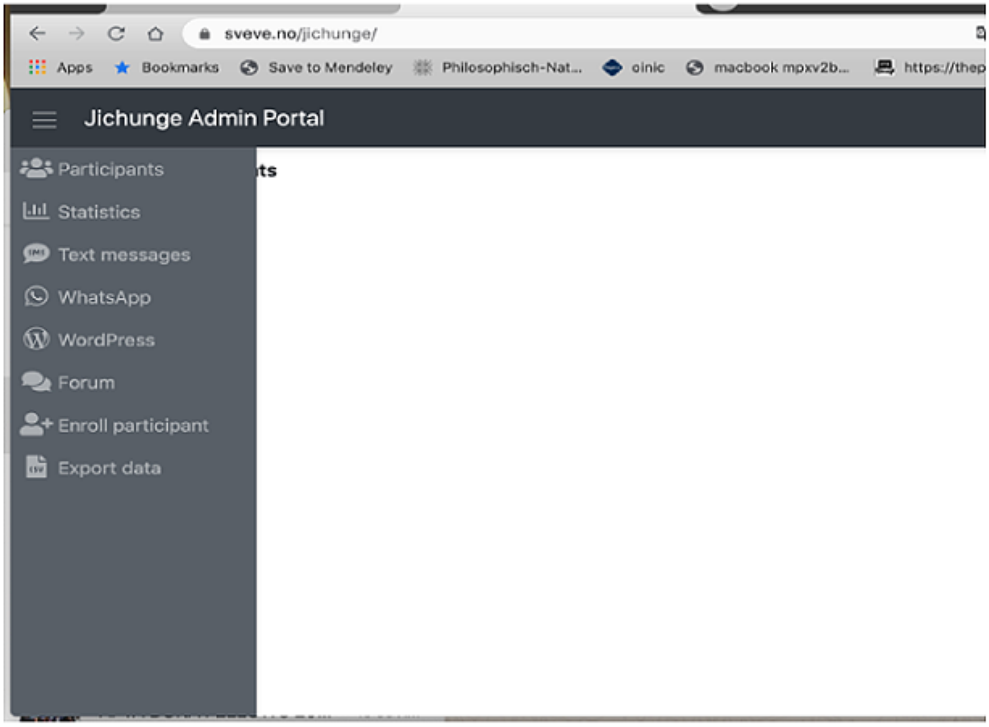
A demonstration of the app's functionalities and operations was performed for the users, and they were later invited to discuss those functionalities, including expectations regarding the app use. Each of the FGDs took an average of one hour.

Following the FGDs, the Jichunge app was provided to the participants. The app is only available by invitation to ensure that only the target populations can access it and maintain confidentiality. Participants were registered in the web-based portal during the installation, using their mobile phone numbers as illustrated in Figure 4. The installation process took 10 to 20 minutes, depending on the phone type and internet connectivity.

After installation, participants were given 2 weeks to use the app, and they had the opportunity to communicate with a doctor, one of the researchers, and a designated peer educator. During this period, they were also able to consult the research team and the software company.

At the end of the second week, participants were invited for the second round of FGDs. Two FGDs, one for MSM and one for FSW, were conducted. Discussions focused on their experiences, perspectives, and challenges regarding using each of the app's functionalities and features. Suggestions for improvement were also presented to the software developer for improvement.

Figure 4. Jichunge app web-based admin portal.



Usability Measures

The usability was quantitatively measured through user statistics of various app features recorded in the app web-based server. Data were extracted from the web-based server into Microsoft Excel, then entered into Stata (version 13.1; StataCorp) for analysis. The mean number of visits to each of the 9 app features per day was estimated and used to measure and monitor the app's usability.

Data Analysis

All qualitative data from FGDs and informal exploratory discussions were done in Swahili, the language spoken by all Tanzanians. Data from FGDs were recorded and transcribed verbatim. Field notes were taken during the exploratory individual or small group discussions. Data were later translated to the English language before conducting a content analysis. We applied an open coding approach during the initial data collection stage, where emerging subthemes were identified. Various emerging and related subthemes were subsumed into broader themes relevant to the app's design, functions, and usability. These themes were later summarized and discussed by at least 2 researchers. Relevant quotes for each of the themes were summarized and used in the manuscript. Descriptive analysis was done for quantitative data. Proportions were calculated for categorical variables, and mean or median were estimated for continuous variables as deemed appropriate. The study was approved by the National Health Research Ethics committee, Tanzania (NIMR/HQ/R.8a/Vol.IX/3454).

Results

Relevance Cycle

The results are presented based on the ISR framework cycles adopted in the Jichunge app development process. End-users' consultation revealed that the high demand for developing and implementing innovative approaches that can assist key populations in health-related matters, including improving uptake and adherence to PrEP. Members of the target population enthusiastically received the idea and expressed their willingness to support such development. Users indicated the need to have a culturally and legally sensitive application that would provide anonymity and safety and yet offer a high level of desired functions to accommodate such anonymity. Following users' suggestions, the investigators proposed indistinct terms to label the target population groups in the app. The literature review indicated limited access to health services among HIV at-risk populations, including FSW and MSM in the African setting [33]. Access to health information and services among these populations in mainstream public health facilities is associated with many difficulties due to stigma and discrimination. Poor adherence to PrEP among at-risk populations was also reported and threatened the effectiveness of this preventive intervention [26,34,36].

Jichunge App Main Functionalities

The final prototype of the Jichunge app has 10 various functionalities or features, as presented in Textbox 1. The outlook of the app features to the users is displayed in Figure 5. The Jichunge app can provide information on the participant's level of adherence to daily PrEP pills, allow gamification and winning of prizes, register pills taken, allow setting up of timed

reminder messages, send a customized reminder message to the user, provide a platform for real-time communication between the user with a peer educator or a health care worker, provide health education, and allow participation in various quizzes.

Textbox 1. A summary of the Jichunge app's main functionalities. PrEP: pre-exposure prophylaxis.

Levels and points: Shows a person's level of adherence per month.

Gamification: Provides a virtual view of the medication registration progress and displays a winning trophy that automatically enters the user in gamification, allowing them to win various prizes, including mobile airtime, data, and others.

Drug registration: The user registers each time they take the medication for records.

Medication time reminder: The user can set up a personal time to be reminded to take the medication daily.

Notification: It notifies the user to take and register the medication per the medication reminder.

Communication with peers (WhatsApp feature): It offers the users an opportunity to communicate in real-time with a trained peer educator about PrEP and other health service-related opportunities and challenges. They are also able to discuss different social issues of relevance to the key population for HIV.

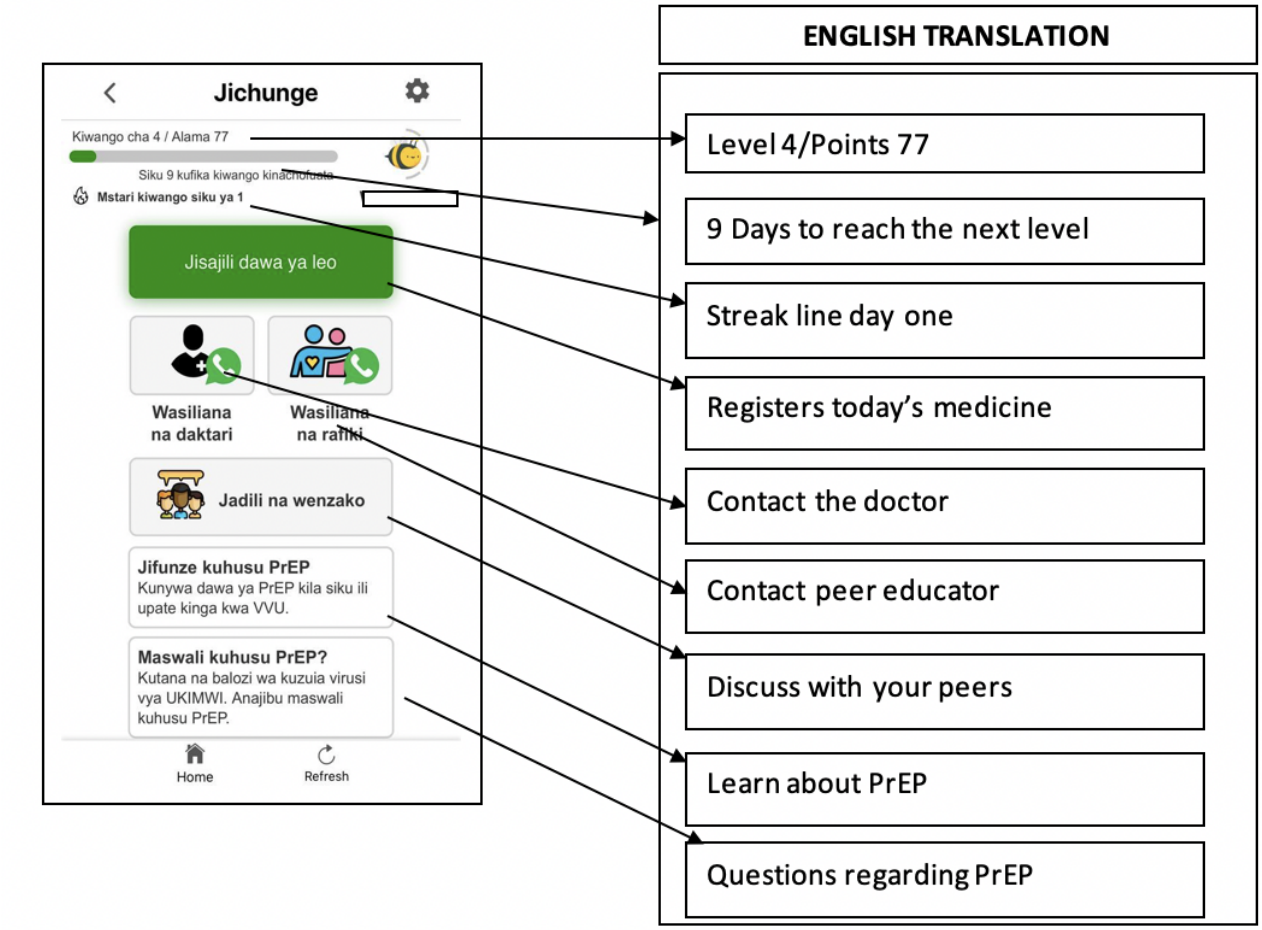
Communication with a health care provider (WhatsApp): It offers users an opportunity to communicate in real-time with a health care provider (doctor) on PrEP and other health-related issues.

Discussion forum (chat room feature): A platform where users can chat with peers anonymously using autogenerated aliases.

Educational materials (WordPress feature): The Jichunge app will contain educational documents and pictorial presentations providing information on learning about PrEP and common questions regarding PrEP. This platform provides users with information on PrEP-related issues.

Jichunge quiz: Users can take a quiz and have a chance to win airtime.

Figure 5. Final outlook and functionalities of Jichunge app. PrEP: pre-exposure prophylaxis.



Rigor Cycle

Participants' Reported Experiences Using the Jichunge App

Overall, participants expressed their satisfaction with the Jichunge app and expressed their readiness to continue app use. Some reported having proposed the app's usefulness to their peers who were not part of the pilot study. Participants praised the reminder messages as being extremely useful. Even if they did not have internet connectivity or airtime to receive the message, one participant stated:

> *I have never received the message but, since I like to use the phone, once I see the app's logo, I remember to take medicine.* [FSW FGD, participant 1]

Access to educational content about PrEP and other HIV-related information was of most interest to the participants. In addition, the app served as a reference and source of information whenever misunderstandings among members of the at-risk population arose.

> *…I learn and know more about PrEP, you find many things, considering there are things we are asked in the group. If there is a question that has been asked and you don't know the answer, you go there (into the app), check, and come back to tell your colleagues.* [MSM FGD, participant 1]

Both MSM and FSW were excited about the app. They reported that the outlook (logo) of the app, its features, and built-in anonymity (members only) make it secure and confidential, protecting their discussions and identities. An FSW from the FGD narrated.

> *…This is going to keep our secrets because they will not be able to just get it like eh! this WhatsApp. For us, it looks like a baby app, the logo leaf, we are mothers, so this is good and secretive.* [FSW FGD; participant 6]

Users who provided feedback generally expressed positive attitudes and acceptability of the app design and functions. Suggestions for app improvement included redesigning the discussion forum to match the WhatsApp platform, with which many study participants were acquainted. Users also had some suggestions on the features that should be improved, including the appearance of chat forums on its appearance and showing current online users.

> *…This discussion forum is not somehow appealing…Once you enter, you should at least know how many are online…Sometimes you don't see the message of other people.* [MSM FGD, participant 3]

Moreover, users proposed changes in the gamification score indicators to resemble everyday objects in their environments, such as a ladder, speedometer, or a trophy used in football games. End-users endorsed the app's name, Jichunge, a Swahili word meaning "protect yourself." Population-specific and sensitive language translation from English to Swahili (Tanzania's national language) was also incorporated in the app.

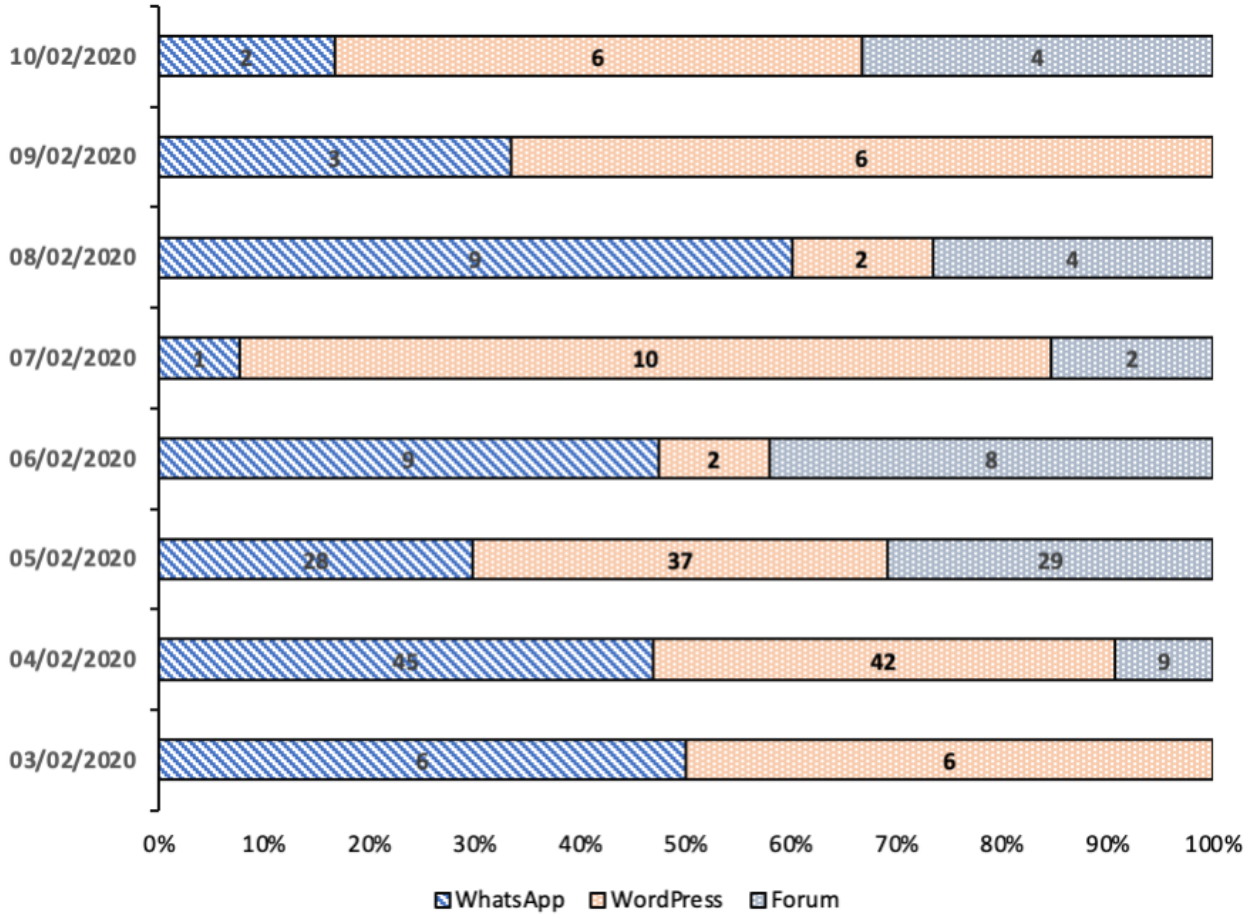
Usability Measures

The Jichunge app pilot study recorded 68 pill registrations during the first 7 days, and on average, half of the users (10/19, 52%) registered the medicines taken per day. A third of the users (6/19, 31.6%) had complete registration of their daily pill taking. There were variations in daily pills registrations among the 19 participants, with 6 (31.6%) registering during the first day and 12 (63.2%) on the second day, followed by 8 (42.1%), 11 (57.9%), 10 (52.6%), 11 (7.9%), and 12 (63.2%) during subsequent days (Table 1). The users used, on average, 1 out of 9 app features per day. The features visitation rates were highest for WhatsApp (100/270, 37.0%) and WordPress (102/270, 37.8%). WhatsApp's platform provided secure communication with a peer educator or a doctor, while WordPress contained educational materials about PrEP (Figure 6).

Table 1. Number of pill registrations per day during the first week of February 2020.

Day	Pill registrations per day (N=19), n (%)
1	6 (31.6%)
2	12 (63.2%)
3	8 (42.1%)
4	11 (57.9%)
5	10 (52.6%)
6	11 (57.9%)
7	12 (63.2%)

Figure 6. The number of App features used per day 3-10 February 2020 (N=270).



Discussion

Main Findings

This paper presents results from a participatory design approach in developing the Jichunge mobile app to improve adherence to PrEP among MSM and FSW in Tanzania. The app includes various functionalities that encourage not only adherence to PrEP but also HIV preventive behaviors. The final version of the application offers 10 unique features, as presented in Textbox 1. The development process, focusing on the relevance and design cycles of the ISR framework, ensured optimal end-user participation in the design and evaluation. MSM and FSW are stigmatized and discriminated against in the sub-Saharan region and Africa as a whole. Their behaviors are deemed illegal, limiting access to population-specific health services [26,34,36]. Therefore, innovative approaches are needed to facilitate access and adherence to health-related services such as PrEP for HIV prevention. Studies have indicated that mobile applications developed with adequate consideration of intended users' needs and preferences are more likely to be utilized [37]. Hou et al [38] also stressed the importance of user-centered involvement in developing relevant mHealth apps and ensure usability. The user participatory approach provides an opportunity to validate proposed app features and identify relevant functional features for inclusion in the final design to ensure effectiveness [27,39]. In the Jichunge app, users provided critical inputs on the design, including logo, app name, and features. The participants suggested several modifications, such as text delivery format, gamification, and improvements in the discussion forum. A user-accepted outlook (logo) may go a long way in impacting service use by merely visualizing their phone's logo. An appealing logo stresses the importance of a user-friendly app's appearance in eHealth and mHealth technologies [3].

As expected, the high usability of the application was observed on the daily medicine registration feature. Given that this was the first time these populations in Tanzania used an mHealth application, the results are promising. Moreover, they indicate the potential for future optimal usability. Therefore, an innovative interactive mHealth application may go a long way in strengthening access to health information and access to health services among these populations.

The discussion forum suggested an improvement to include virtualization of other online members, which was considered in the final design to provide a sense of community engagement [40].

Among the lessons learned during the introduction and onboarding of participants into the app was ensuring wireless internet connectivity, which reduces participants' internet costs. Moreover, to ensure that participants have a better opportunity to learn more about the app, promotional posters, app functionality fliers, and demonstration videos should be

considered alongside verbal descriptions [41,42]. Finally, in a limited resource setting, planning for mobile application intervention may be hindered by outdated versions of mobile phones, which may not be compatible with most modern applications.

Limitations

The results of this study should be interpreted in light of the following limitations. Firstly, the usability measurement was done through the number of visits to each feature by the user. However, we acknowledge that visit statistics for some features, such as educational resources, may not indicate detailed (quality) reading. It would be more informative to have measures such as design layout, comprehension of learning, and application performance.

Secondly, at both the app design and pilot stage, we had exploratory interviews with only 30 participants and conducted 2 focus group discussions involving 20 participants. Although these participants' sociodemographic characteristics did not differ from that of the user population in the city, we acknowledge that the participants may not represent the views and perspectives of the entire population of FSW and MSM. We used strategic peer-led recruitment (respondent-driven sampling) as the recommended recruitment method in stigmatized, criminalized, and hard-to-reach populations.

Thirdly, the app's reporting of drug use may be affected by a desirability bias, where people may decide to register treatment administration without taking the tablet. Similarly, there could be people who have used the tablet but did not register. As in many desirability-prone studies, confirmatory measures such as pill counts and biological markers are recommended and planned in our future trial using the Jichunge app.

Conclusions

The participatory design approach in mHealth app development helps assess and cooperate end-users' experiences and perspectives and identify potential challenges to guide future app usability. The Jichunge app accounted for the sociocultural and legal context of the target populations eliciting enthusiasm with the potential to impact adherence to PrEP for HIV prevention and general risk-free behavioral promotion.

Acknowledgments

The authors would like to thank the Research Council of Norway through the global health and vaccination (GLOBVAC) program and the European and Developing Countries Clinical Trials Partnership (EDCTP2) program supported by the European Union funding for this study. We are also indebted to the participants who provided feedback and input in developing the Jichunge application and the software company Sveve AS of Oslo, Norway, for all their support. The Research Council of Norway partly supports the PREPTA (pragmatic trial for HIV pre-exposure prophylaxis roll-out in Tanzania) project through GLOBVAC (project number 285361). The project is also part of the EDCTP2 program supported by the European Union.

Authors' Contributions

WM collected and analyzed the data and drafted the initial manuscript. EJM and KM designed the study and interpreted the data. EJM, KM, MTL, GHL, IH, HL, CM, and MK revised the manuscript, along with Neema Makyao and Emmy Metta. All authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. World Bank. Information and Communications for Development 2012: Maximizing Mobile. Washington, DC: The World Bank; 2012:116-134 URL: https://documents1.worldbank.org/curated/en/727791468337814878/pdf/722360PUB0EPI00367926B9780821389911.pdf
2. PricewaterhouseCoopers. Emerging mHealth: Paths for growth Internet. 201www.pwc.com/mhealth. 2012. URL: https://www.pwc.com/gx/en/healthcare/mhealth/assets/pwc-emerging-mhealth-full.pdf?dLDf=false [accessed 2020-05-24]
3. Anderson PL, Liu AY, Castillo-Mancilla JR, Gardner EM, Seifert SM, McHugh C, et al. Intracellular Tenofovir-Diphosphate and Emtricitabine-Triphosphate in Dried Blood Spots following Directly Observed Therapy. Antimicrob Agents Chemother 2017 Dec 21;62(1):e01710-e01717. [doi: 10.1128/aac.01710-17]
4. Ippoliti N, L'Engle K. Meet us on the phone: mobile phone programs for adolescent sexual and reproductive health in low-to-middle income countries. In: Reprod Health. Meet us on the phone: Mobile phone programs for adolescent sexual and reproductive health in low-to-middle income countries Internet. Reprod Health. BioMed Central Ltd; Jan 17, 2017:11.
5. Brown K, Williams DB, Kinchen S, et al. Status of HIV Epidemic Control Among Adolescent Girls and Young Women Aged 15–24 Years — Seven African Countries, 2015–2017. MMWR Morb Mortal Wkly Rep 2018(67):29-32 [FREE Full text] [doi: 10.15585/mmwr.mm6701a6]
6. Feroz A, Abrejo F, Ali S, Nuruddin R, Saleem S. Using mobile phones to improve young people's sexual and reproductive health in low- and middle-income countries: a systematic review protocol to identify barriers, facilitators and reported interventions. Syst Rev 2019 May 18;8(1):117 [FREE Full text] [doi: 10.1186/s13643-019-1033-5] [Medline: 31103044]

7. Stephani V, Opoku D, Quentin W. A systematic review of randomized controlled trials of mHealth interventions against non-communicable diseases in developing countries. BMC Public Health 2016 Dec 15;16:572 [FREE Full text] [doi: 10.1186/s12889-016-3226-3] [Medline: 27417513]
8. Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht U. mHealth 2.0: Experiences, Possibilities, and Perspectives. JMIR Mhealth Uhealth 2014 May;2(2):e24 [FREE Full text] [doi: 10.2196/mhealth.3328] [Medline: 25099752]
9. Shah R, Watson J, Free C. A systematic review and meta-analysis in the effectiveness of mobile phone interventions used to improve adherence to antiretroviral therapy in HIV infection. BMC Public Health 2019 Jul 9;19(1):915. [doi: 10.1186/s12889-019-6899-6]
10. Mitchell JT, LeGrand S, Hightow-Weidman LB, McKellar MS, Kashuba AD, Cottrell M, et al. Smartphone-Based Contingency Management Intervention to Improve Pre-Exposure Prophylaxis Adherence: Pilot Trial. JMIR Mhealth Uhealth 2018 Sep 10;6(9):e10456. [doi: 10.2196/10456]
11. Liu A, Vittinghoff E, Felten PV, Amico K, Anderson P, Lester R. Randomized Controlled Trial of a Mobile Health Intervention to Promote Retention and Adherence to Pre-exposure Prophylaxis among Young People at Risk for Human Immunodeficiency Virus: The EPIC Study. Clin Infect Dis 2019 May 30;68(12):2010-2017. [doi: 10.1093/cid/ciy810]
12. Ronen K, Unger J, Drake A, Perrier T, Akinyi P, Osborn L, et al. SMS messaging to improve ART adherence: perspectives of pregnant HIV-infected women in Kenya on HIV-related message content. AIDS Care 2018 Apr;30(4):500-505 [FREE Full text] [doi: 10.1080/09540121.2017.1417971] [Medline: 29254362]
13. Forrest JI, Wiens M, Kanters S, Nsanzimana S, Lester RT, Mills EJ. Mobile health applications for HIV prevention and care in Africa. Curr Opin HIV AIDS 2015;10(6):464-471. [doi: 10.1097/coh.0000000000000198]
14. Källander K, Tibenderana J, Akpogheneta O, Strachan D, Hill Z, ten Asbroek AHA, et al. Mobile health (mHealth) approaches and lessons for increased performance and retention of community health workers in low- and middle-income countries: a review. J Med Internet Res 2013 Jan 25;15(1):e17 [FREE Full text] [doi: 10.2196/jmir.2130] [Medline: 23353680]
15. WHO. Guideline on when to start antiretroviral therapyon pre-exposure prophylaxis for HIV. World Health Organization. 2014. URL: https://www.who.int [accessed 2020-05-23]
16. Van der Elst EM, Mbogua J, Operario D, Mutua G, Kuo C, Mugo P, et al. High acceptability of HIV pre-exposure prophylaxis but challenges in adherence and use: qualitative insights from a phase I trial of intermittent and daily PrEP in at-risk populations in Kenya. AIDS Behav 2013 Jul;17(6):2162-2172 [FREE Full text] [doi: 10.1007/s10461-012-0317-8] [Medline: 23080358]
17. Van Damme L, Corneli A, Ahmed K, Agot K, Lombaard J, Kapiga S, et al. Preexposure prophylaxis for HIV infection among African women. N Engl J Med 2012 Aug 2;367(5):411-422 [FREE Full text] [doi: 10.1056/NEJMoa1202614] [Medline: 22784040]
18. Hanscom B, Janes HE, Guarino PD, Huang Y, Brown ER, Chen YQ, et al. Brief report:Preventing HIV-1 Infection in Women using Oral Pre-Exposure Prophylaxis: A Meta-analysis of Current Evidence. J Acquir Immune Defic Syndr 2016;73(5):606-608. [doi: 10.1097/qai.0000000000001160]
19. McCormack S, Dunn DT, Desai M, Dolling DI, Gafos M, Gilson R, et al. Pre-exposure prophylaxis to prevent the acquisition of HIV-1 infection (PROUD): effectiveness results from the pilot phase of a pragmatic open-label randomised trial. The Lancet 2016 Jan;387(10013):53-60. [doi: 10.1016/S0140-6736(15)00056-2]
20. Wheeler DP, Fields SD, Beauchamp G, Chen YQ, Emel LM, Hightow-Weidman L, et al. Pre-exposure prophylaxis initiation and adherence among Black men who have sex with men (MSM) in three US cities: results from the HPTN 073 study. J Int AIDS Soc 2019 Feb;22(2):e25223 [FREE Full text] [doi: 10.1002/jia2.25223] [Medline: 30768776]
21. TCRA. Quarterly communications statistics Internet. Tanzania Communications Regulatory Authority. Dar es Salaam URL: https://www.tcra.go.tz/images/documents/telecommunication/TelCom_Statistics_Dec_2019.pdf [accessed 2020-06-25]
22. GlobalStat. Social Media Stats United Republic Of Tanzania. StatCounter Global Stats Internet. URL: https://gs.statcounter.com/social-media-stats/all/tanzania [accessed 2020-03-23]
23. Kushniruk A, Nøhr C. Participatory Design, User Involvement and Health IT Evaluation. Stud Health Technol Inform 2016;222:139-151. [Medline: 27198099]
24. Chhoun P, Kaplan KC, Wieten C, Jelveh I, Lienemann M, Tuot S, et al. Using participatory methods to build an mHealth intervention for female entertainment workers in Cambodia: the development of the Mobile Link project. mHealth 2019 Aug;5:24-24. [doi: 10.21037/mhealth.2019.07.02]
25. Tang T, Lim ME, Mansfield E, McLachlan A, Quan SD. Clinician user involvement in the real world: Designing an electronic tool to improve interprofessional communication and collaboration in a hospital setting. Int J Med Inform 2018 Feb;110:90-97 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.11.011] [Medline: 29331258]
26. Johnson CA. Off the Map: How HIV / AIDS programming is failing same-sex practicing people in Africa. International Gay and Lesbian Human Rights Commission, editor. New York, NY: Outright Action International; 2007. URL: http://www.iglhrc.org/files/iglhrc/otm/Off [accessed 2020-05-23]

27. Schnall R, Rojas M, Bakken S, Brown W, Carballo-Dieguez A, Carry M, et al. A user-centered model for designing consumer mobile health (mHealth) applications (apps). J Biomed Inform 2016 Apr;60:243-251 [FREE Full text] [doi: 10.1016/j.jbi.2016.02.002] [Medline: 26903153]
28. Hevner A. A three cycle view of design science research. Scand J Inf Syst 2007;4(2):1-7 [FREE Full text]
29. Martikainen S, Korpela M, Tiihonen T. User participation in healthcare IT development: a developers' viewpoint in Finland. Int J Med Inform 2014 Mar;83(3):189-200. [doi: 10.1016/j.ijmedinf.2013.12.003] [Medline: 24382475]
30. Ahmed I, Ahmad NS, Ali S, Ali S, George A, Saleem Danish H, et al. Medication Adherence Apps: Review and Content Analysis. JMIR Mhealth Uhealth 2018 Mar 16;6(3):e62 [FREE Full text] [doi: 10.2196/mhealth.6432] [Medline: 29549075]
31. Moen K, Aggleton P, Leshabari MT, Middelthon A. Gays, guys, and mchicha mwiba: same-sex relations and subjectivities in Dar es Salaam. Journal of Homosexuality 2014;61(4):511-539. [doi: 10.1080/00918369.2014.865457] [Medline: 24313863]
32. Mmbaga E, Leyna G, Leshabari M, Moen K. Early Anal Sex Experience Among Men Who Have Sex with Men in Dar Es Salaam Tanzania: Implications for HIV Prevention and Care. Arch Sex Behav 2020 Aug;49(6):2045-2055. [doi: 10.1007/s10508-019-01529-5] [Medline: 31872388]
33. Mmbaga EJ, Moen K, Leyna GH, Mpembeni R, Leshabari MT. HIV Prevalence and Associated Risk Factors Among Men Who Have Sex With Men in Dar es Salaam, Tanzania. JAIDS Journal of Acquired Immune Deficiency Syndromes 2018;77(3):243-249. [doi: 10.1097/qai.0000000000001593]
34. Mmbaga E, Moen K, Makyao N, Mpembeni R, Leshabari M. HIV and STI s among men who have sex with men in Dodoma municipality, Tanzania: a cross-sectional study. Sex Transm Infect 2017 Aug;93(5):314-319. [doi: 10.1136/sextrans-2016-052770] [Medline: 28202736]
35. App designing company. https://sveve.no/. URL: https://sveve.no/varsling.jsp?gclid=EAIaIQobChMIp8jHnv3J6QIVxeAYCh2ApwjiEAAYASAAEgKOsPD_BwE [accessed 2020-05-29]
36. Carter M. Activists arrested during attempt to highlight Uganda's neglect of HIV prevention for men who have sex with men Internet. aidsmap. URL: https://www.aidsmap.com/news/jun-2008/activists-arrested-during-attempt-highlight-ugandas-neglect-hiv-prevention-men-who [accessed 2020-05-24]
37. Brown W, Yen P, Rojas M, Schnall R. Assessment of the Health IT Usability Evaluation Model (Health-ITUEM) for evaluating mobile health (mHealth) technology. J Biomed Inform 2013 Dec;46(6):1080-1087 [FREE Full text] [doi: 10.1016/j.jbi.2013.08.001] [Medline: 23973872]
38. Hou I, Lan M, Shen S, Tsai PY, Chang KJ, Tai H, et al. The Development of a Mobile Health App for Breast Cancer Self-Management Support in Taiwan: Design Thinking Approach. JMIR Mhealth Uhealth 2020 Apr 30;8(4):e15780 [FREE Full text] [doi: 10.2196/15780] [Medline: 32352390]
39. Schnall R, Rojas M, Travers J, Brown W, Bakken S. Use of Design Science for Informing the Development of a Mobile App for Persons Living with HIV. AMIA Annu Symp Proc 2014;2014:1037-1045 [FREE Full text] [Medline: 25954413]
40. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and Promoting Effective Engagement With Digital Behavior Change Interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
41. Lefebvre RC, Tada Y, Hilfiker MA, Baur C. The Assessment of User Engagement with eHealth Content: The eHealthEngagement Scale. Journal of Computer-Mediated Communication 2010;15:666-681 [FREE Full text] [doi: 10.1111/j.1083-6101.2009.01514.x]
42. Neve MJ, Collins CE, Morgan PJ. Dropout, nonusage attrition, and pretreatment predictors of nonusage attrition in a commercial Web-based weight loss program. J Med Internet Res 2010 Dec 14;12(4):e69 [FREE Full text] [doi: 10.2196/jmir.1640] [Medline: 21156470]

Abbreviations

EDCTP2: European and Developing Countries Clinical Trials Partnership
FGD: focus group discussion
FSW: female sex workers
GLOBVAC: global health and vaccination
ISR: information system research framework
mHealth: mobile health
MSM: men who have sex with men
PrEP: pre-exposure prophylaxis

Edited by G Eysenbach; submitted 04.08.20; peer-reviewed by A Polhemus, B Guthrie; comments to author 14.12.20; revised version received 03.03.21; accepted 01.08.21; published 07.10.21.

Please cite as:
Mauka W, Mbotwa C, Moen K, Lichtwarck HO, Haaland I, Kazaura M, Leyna GH, Leshabari MT, Mmbaga EJ
Development of a Mobile Health Application for HIV Prevention Among At-Risk Populations in Urban Settings in East Africa: A Participatory Design Approach
JMIR Form Res 2021;5(10):e23204
URL: https://formative.jmir.org/2021/10/e23204
doi:10.2196/23204
PMID:34617904

Original Paper

Detecting Subclinical Social Anxiety Using Physiological Data From a Wrist-Worn Wearable: Small-Scale Feasibility Study

Ruksana Shaukat-Jali[1*], MEng; Nejra van Zalk[1*], BA, MSc, PhD; David Edward Boyle[1*], BEng, PhD

Dyson School of Design Engineering, Imperial College London, London, United Kingdom

[*]all authors contributed equally

Corresponding Author:
Nejra van Zalk, BA, MSc, PhD
Dyson School of Design Engineering
Imperial College London
25 Exhibition Road
London, SW7 2AZ
United Kingdom
Phone: 44 2083318091
Email: n.van-zalk@imperial.ac.uk

Abstract

Background: Subclinical (ie, threshold) social anxiety can greatly affect young people's lives, but existing solutions appear inadequate considering its rising prevalence. Wearable sensors may provide a novel way to detect social anxiety and result in new opportunities for monitoring and treatment, which would be greatly beneficial for persons with social anxiety, society, and health care services. Nevertheless, indicators such as skin temperature measured by wrist-worn sensors have not been used in prior work on physiological social anxiety detection.

Objective: This study aimed to investigate whether subclinical social anxiety in young adults can be detected using physiological data obtained from wearable sensors, including heart rate, skin temperature, and electrodermal activity (EDA).

Methods: Young adults (N=12) with self-reported subclinical social anxiety (measured using the widely used self-reported version of the Liebowitz Social Anxiety Scale) participated in an impromptu speech task. Physiological data were collected using an E4 Empatica wearable device. Using the preprocessed data and following a supervised machine learning approach, various classification algorithms such as Support Vector Machine, Decision Tree, Random Forest, and K-Nearest Neighbours (KNN) were used to develop models for 3 different contexts. Models were trained to differentiate (1) between baseline and socially anxious states, (2) among baseline, anticipation anxiety, and reactive anxiety states, and (3) social anxiety among individuals with social anxiety of differing severity. The predictive capability of the singular modalities was also explored in each of the 3 supervised learning experiments. The generalizability of the developed models was evaluated using 10-fold cross-validation as a performance index.

Results: With modalities combined, the developed models yielded accuracies between 97.54% and 99.48% when differentiating between baseline and socially anxious states. Models trained to differentiate among baseline, anticipation anxiety, and reactive anxiety states yielded accuracies between 95.18% and 98.10%. Furthermore, the models developed to differentiate between social anxiety experienced by individuals with anxiety of differing severity scores successfully classified with accuracies between 98.86% and 99.52%. Surprisingly, EDA was identified as the most effective singular modality when differentiating between baseline and social anxiety states, whereas ST was the most effective modality when differentiating anxiety among individuals with social anxiety of differing severity.

Conclusions: The results indicate that it is possible to accurately detect social anxiety as well as distinguish between levels of severity in young adults by leveraging physiological data collected from wearable sensors.

(JMIR Form Res 2021;5(10):e32656) doi:10.2196/32656

KEYWORDS

social anxiety; wearable sensors; physiological measurement; machine learning; young adults; mental health; mHealth; new methods; anxiety; wearable; sensor; digital phenotyping; digital biomarkers

Introduction

Background

Social anxiety is a fear of social situations in which the individual is exposed to possible scrutiny by others [1], and high levels of social anxiety are associated with a low quality of life in various domains [2,3]. Even when not clinically diagnosable (ie, subclinical or threshold social anxiety), it can greatly affect young people's lives. Fehm et al [4] showed that young adults with social anxiety who do not receive treatment are at risk of developing social anxiety disorder (SAD) and comorbid mental health problems such as depression, both of which cause further adverse life impairments [3,5]. SAD is one of the most common anxiety disorders [6]. One UK study in 2000 [7] revealed that the annual health care cost per person with SAD was £609 (US $834.59), with annual productivity losses and social security benefits adding to £1920 (US $2631.22) per person with SAD, whereas those with SAD and a comorbidity incurred even higher costs. Nevertheless, many individuals do not receive treatment owing to limited availability or lack of awareness of social anxiety among health care professionals [3,4,8]. Some may not even seek treatment owing to a fear of being negatively evaluated by health care professionals [8]. Thus, it is imperative to empower both individuals and health care professionals in early detection of social anxiety before it potentially escalates into SAD and other related problems.

Common methods for assessing social anxiety involve using subjective measures, usually in a clinical setting. Owing to the rising prevalence of social anxiety, however, it is becoming evident that traditional approaches are inadequate and unsustainable for health care services [4,5,9]. In recent years, increasing focus has been given to technological advances that might help in the early detection and subsequent intervention for anxiety-related problems. In terms of social anxiety, objective methods used to assess symptoms include monitoring physiological changes typically caused by anxiety such as an elevated heart rate (HR), increased electrodermal activity (EDA), variation in skin temperature (ST), and trembling [1,10-12].

Nevertheless, despite extensive and promising research into stress and emotion detection based on physiological indices applicable to social anxiety collected from wearable sensors (Table 1), there has been little effort to predict social anxiety particularly using this approach. This might be ascribed to the recent shift in attention toward social anxiety reported by Heimberg and Butler [13], owing to widening of the diagnostic criteria and leading to a rise in those who fit the criteria for social anxiety.

Although not without its problems, detection via wearable sensors has the potential to underpin solutions addressing the growing needs of individuals with social anxiety and complement traditional therapeutic approaches. If subclinical social anxiety could reliably and validly be detected using wearable sensors, initial treatment could subsequently transition to digital self-help solutions to aid social anxiety at earlier stages when treatment is less extensive and costly [7]. Furthermore, self-help solutions may be a more appropriate method of treatment as individuals with social anxiety often feel nervous to seek treatment in clinical settings [8]. Detecting social anxiety using evidence-based objective methods could also complement current therapeutic approaches.

Prior Work

Emotion Detection Using Machine Learning

A rise in wearable devices has further enabled researchers to investigate methods for the detection of emotion and stress states [14,15], with many studies reporting high-accuracy detection levels (Table 1). To detect emotional states using physiological data, researchers have executed data collection experiments that invoke the state to be detected, with tasks including hyperventilation and watching emotional films [16,17].

After data collection, a supervised machine learning (ML) approach is commonly used owing to the classification nature of the investigations [16-18]. In supervised ML, the training data are labeled in accordance with the correct class as the classification algorithms learn by example. Table 1 shows an overview of ML approaches focusing on emotion and stress detection. The most dominant and successful algorithm in studies involving recognition of states using physiological data is Support Vector Machine (SVM). Classifiers such as Decision Tree, Random Forest, and K-Nearest Neighbours (KNN) have also been frequently used and are reportedly effective.

Table 1. Studies on the recognition of emotion and stress states by using physiological indicators.

Study	Classification algorithms	Physiological data	Detection	Reported accuracy, %
[16]	SVM[a], Decision Tree, KNN[b], Naïve Bayes, Random Forest, Neural Network, Zero K	HR[c], ST[d], EDA[e]	Stress	65.8-100%
[18]	SVM	EDA, BVP[f], PD[g]	Stress	57.1-80%
[19]	SVM, Decision Tree, KNN, Naïve Bayes	EDA, BVP, PZT[h], EEG[i], ECG[j], EMG[k]	Emotion	17-91.3%
[20]	SVM	BVP, ST, EDA, PD	Stress	61.5-90.1%
[21]	KNN	HRV[l]	Stress	79.2-94.6%

[a]SVM: Support Vector Machine.

[b]KNN: K-Nearest Neighbours.

[c]HR: heart rate.

[d]ST: skin temperature.

[e]EDA: electrodermal activity.

[f]BVP: blood volume pulse.

[g]PD: pupillary distance.

[h]PZT: piezoelectric response.

[i]EEG: electroencephalogram.

[j]ECG: electrocardiogram.

[k]EMG: electromyography.

[l]HRV: heart rate variability.

Physiological Indicators of Social Anxiety

Classical psychological experiments commonly use impromptu public speaking tasks to elicit a social anxiety response [1,11,22,23]. These experiments are often split into stages that measure three responses: *baseline*, *anticipatory*, and *reactive* anxiety (where the nature of the speaking task is announced beforehand to provoke an anticipatory anxiety response) [1,11,23]. In conjunction, respondents are typically asked about their anxiety levels through self-reports [24]. Although self-reports are an important way of gauging individual perceptions of social anxiety, this approach is not without its problems, including a high level of subjectivity.

A more objective way to measure social anxiety is using physiological indicators. Social anxiety activates the sympathetic nervous system (SNS) [25]. HR and ST are modulated by both the parasympathetic nervous system (PNS) and SNS divisions of the autonomic nervous system (ANS), whereas EDA is modulated by the SNS alone. Therefore, EDA, HR, and ST are seen as markers of SNS activation and can be considered as potential indicators of social anxiety [26-29].

Studies investigating physiological responses to social anxiety further illustrate the potential to use EDA, HR, and ST as indicators [10-12]. Despite the potential for these indicators, however, their responses are complex, and a few studies have indicated minor differences in SNS arousal for individuals with social anxiety compared to control groups [22,23,30]. Furthermore, although ST has been explored as a social anxiety marker [12], wrist ST measurements have not been explored systematically. To our knowledge, this is the first study to explore wrist ST as an indicator of experimentally induced social anxiety.

Research Aims and Objectives

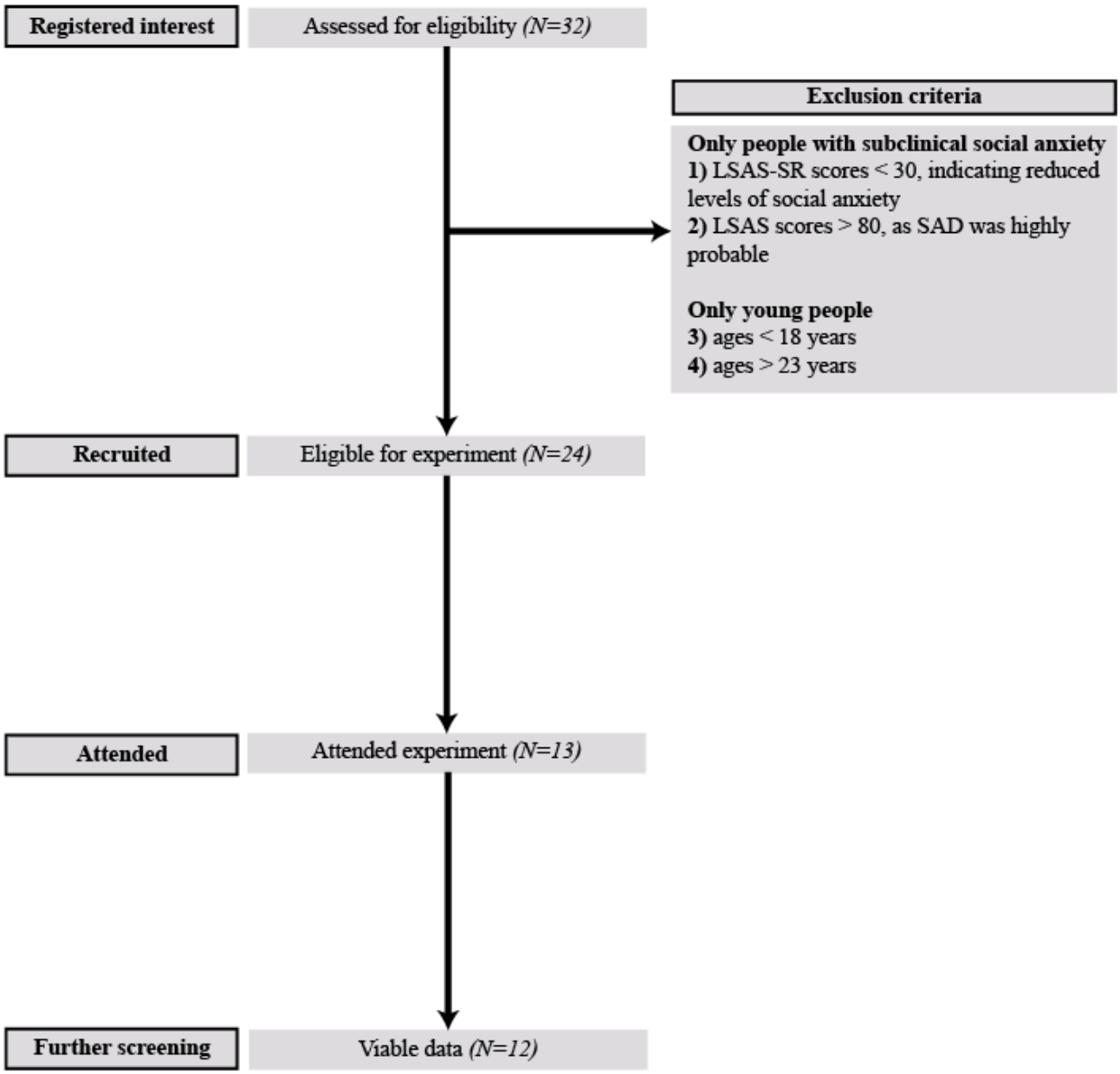
This study aimed to investigate whether social anxiety in young people with subclinical social anxiety can be detected using physiological data (based on HR, ST, and EDA) recorded from an existing multi-sensor wearable. The study aims to explore if models can be trained to differentiate (1) between baseline and socially anxious states, (2) among baseline, anticipation anxiety, and reactive anxiety states, and (3) between social anxiety among individuals with social anxiety of differing severity. This study also aims to explore the predictive capability of the singular modalities.

Methods

Recruitment

Young adult participants were recruited using posters around Imperial College London. The initial sample comprised 13 individuals who self-identified as shy or socially fearful. An exclusion criterion was created to ensure that only young adults with subclinical social anxiety were recruited, as described in Figure 1. To assess participants' social anxiety levels, the self-reported version of the Liebowitz Social Anxiety Scale (LSAS-SR) was initially used (mean 64.33, SD 13.12, range 38-80). In total, 13 individuals attended the experiment. One participant who showed up for the experiment was known to the experimenter and had their data subsequently excluded owing to likely bias. The final study sample thus comprised 12 participants (58% female; mean age, 19.75 years, SD 1.76 years; 67% Asian, 25% White, and 8% Mixed race).

Figure 1. Flow diagram explaining the study recruitment process. LSAS-SR: self-reported version of the Liebowitz Social Anxiety Scale, SAD: social anxiety disorder.



Measures

LSAS-SR

We used the self-report version of the LSAS-SR owing to its well-established validity and reliability in a large amount of previous literature [24]. The LSAS-SR allows for the classification of individuals into differing severity groups, as a higher overall LSAS-SR score is seen to correspond with greater social anxiety severity [24,31]. Furthermore, the LSAS-SR examines both affective aspects (ie, quantifying how anxious participants feel) using the fear subscale and behavioral aspects (ie, gauging to what extent they avoid various social situations) using the avoidance subscale. Each subscale consists of 24 items, with response items ranging on a 4-point scale from "none (0)" to "severe (3)" for the fear subscale, and "never (0)" to "usually (3)" for the Avoidance subscale. Prior studies indicate a high level of reliability of the LSAS-SR (Cronbach α=.95 [24]). In this study, the Cronbach α values for the fear and avoidance subscales were .69 and .69, respectively, with an overall Cronbach α of .83.

Social Phobia Screening Questionnaire

To cross-validate the LSAS-SR, we also used the Social Phobia Screening Questionnaire (SPSQ), which comprises 8 questions about how much fear individuals feel in various social situations, including speaking in front of a group of people, going to a party, and being alone with someone unfamiliar [32]. This measure has shown good validity in prior research [32]. It can be used with or without additional questions that allow an estimation of whether individuals reach the clinical cut-off for SAD and has been used in previous research to indicate subclinical social anxiety levels [33]. The response items ranged from "none (1)" and "some (2)" to "a lot (3)" (Cronbach α=.74).

Ethics

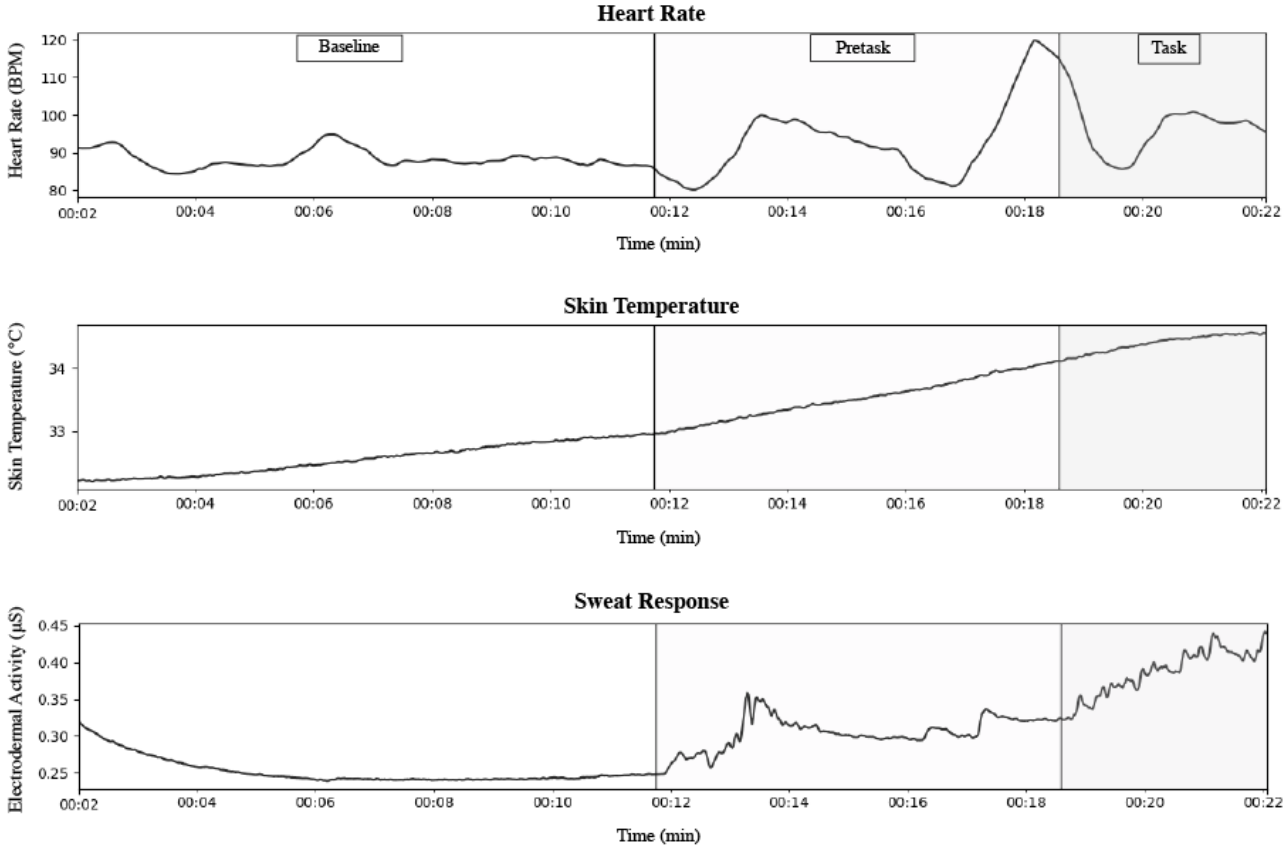
The University Ethics Committee approved all the procedures and measures used in the study. Throughout the procedure, participants were reminded that their participation was voluntary and that they could withdraw their data at any time until used for statistical analysis. The collected data were anonymized and stored in a password-protected folder.

Data Collection

The data were collected using the E4 Empatica research-grade multi-sensor wristband wearable. The device was selected as it simultaneously monitors various types of physiological data at predetermined sampling rates [34]. However, only HR, EDA, and ST data were explored in this study as they could be considered social anxiety markers [26-29]. E4 Empatica has not yet been used in many studies of this nature, although other multi-sensor wrist-worn wearables have demonstrated effectiveness [16,17].

Using the default sampling rates of the E4 [35], EDA was measured in microSiemens (μS) at 4 Hz using stainless steel electrodes positioned on the inner side of the wrist. HR was measured in beats per minute (BPM) at 1 Hz using data derived from a photoplethysmography sensor. ST was measured in °C at 4 Hz using an infrared thermophile [35]. The data were collected throughout the duration of the experiment, an example of which is shown in Figure 2. The full data set and code needed to recreate the classification models and reproduce the results, as well as functions that enable further experimentation, is available in a designated GitHub repository (Multimedia Appendix 1).

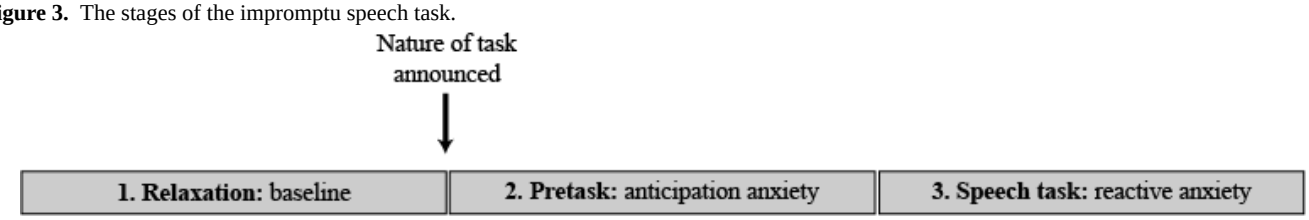
Figure 2. A participant's physiological data sample during the experimental stages. BPM: beats per minute.



Experimental Protocol

The experiments had an approximate duration of 30 minutes. Similar to previous studies [1,11,23], the experiment was split into 3 stages involving relaxation (baseline), task preparation (anticipation anxiety), and performance (reactive anxiety), as responses might differ across these stages [25]. Figure 3 illustrates the experimental stages. The timestamps for the stages were also recorded for labeling purposes. The experimental protocol is listed below.

Figure 3. The stages of the impromptu speech task.



First, the wearable was attached to the participant's wrist. The procedure commenced with a 10-minute baseline period. During this time, participants were offered magazines and ocean sounds were played to create a calming effect.

Second, the nature of the task was then announced, and the participant was given 5 minutes to prepare a 3-minute speech on a selected subject from a choice of topics chosen on the basis of their anxiety-inducing potential. These included "*Is Brexit good or bad, and why?*," "*Intelligence is not enough*," and "*The history of Western Europe until the 2000s*."

Third, a "judging" panel comprising experimenter confederates entered the room, and the participant performed the speech while being timed.

Finally, the participant was debriefed, and the wearable was removed.

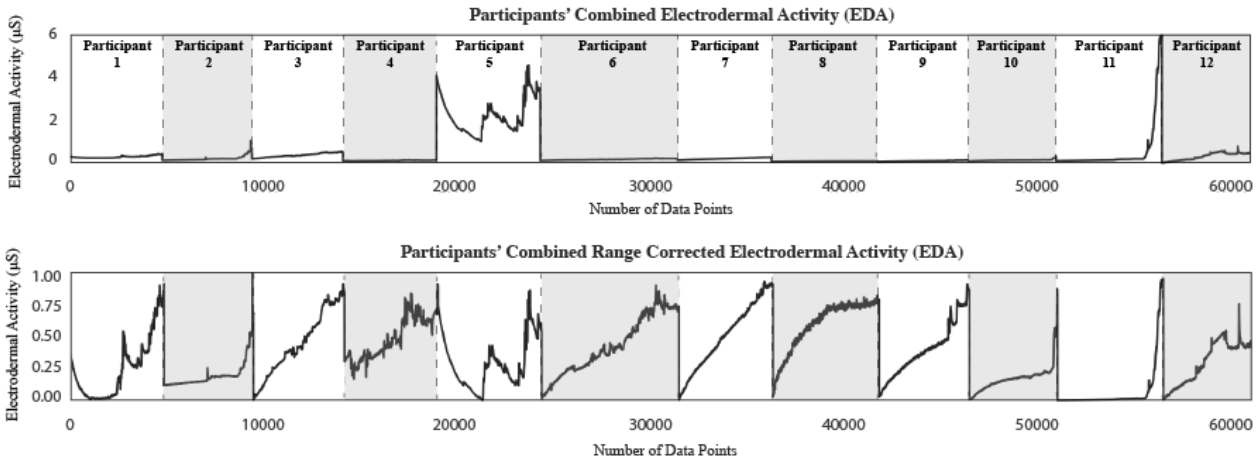
Data Preprocessing

The HR data were first upsampled to 4 Hz, similar to ST and EDA. A Moving Average Filter (Equation 1) was then applied to the data to remove noise [17] and reduce the risk of model overfitting [36,37].

Where *w* refers to window size, Input[*i*] refers to original time series signal and Output[*i*] refers to processed time series signal.

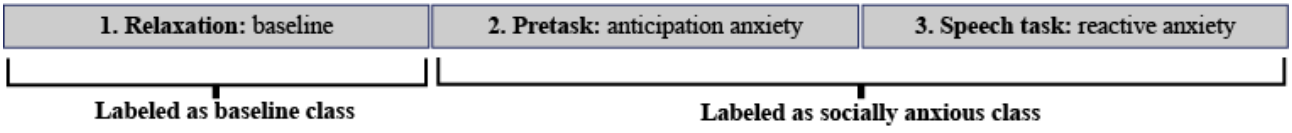
An EDA range correction method (Equation 2) was applied to each participant's EDA (*E*) data, see Figure 4 [38]. This removed inter-individual differences, particularly as physiological activation is believed to be better indicated by the variation within the EDA range rather than the range itself [39].

Figure 4. Participants' data before and after range correction.



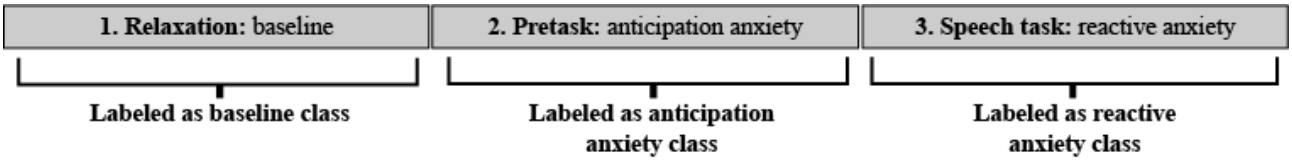
Following this, the labels were allocated on the basis of the experiment timestamps, assuming the suspected states were invoked. Classification investigation (1) examined whether models can be trained to classify baseline and socially anxious states. Therefore, the participants' data were split into the respective classes and labeled using the experiment timestamps (Figure 5).

Figure 5. Labeling arrangement for classification investigation (1).



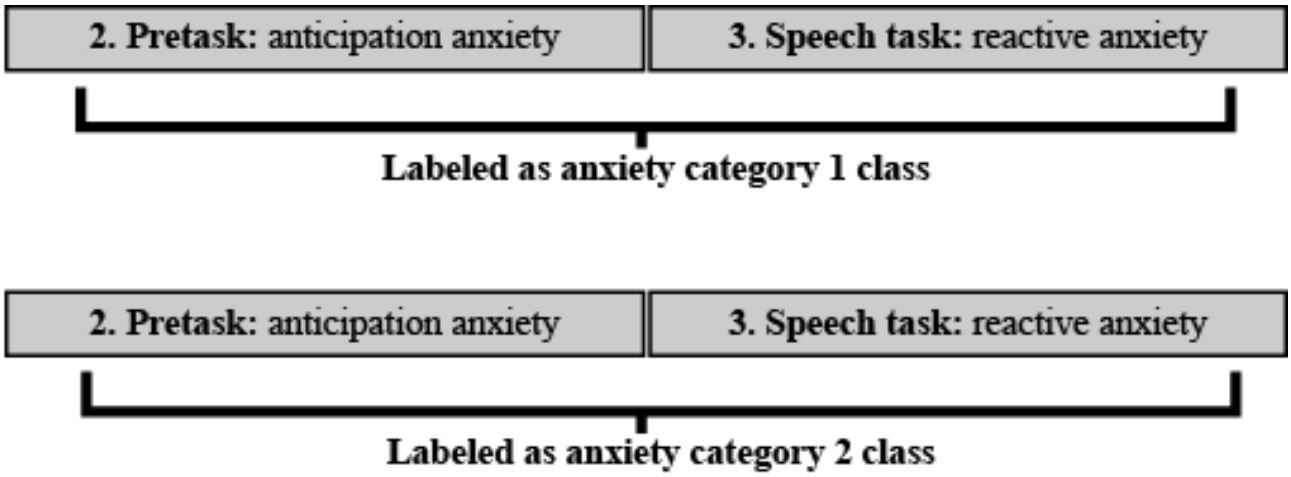
Classification investigation (2) focused on whether models can be trained to differentiate among baseline, anticipation anxiety, and reactive anxiety states. Therefore, the data were divided into the 3 respective classes using the timestamps and labeled as shown in Figure 6.

Figure 6. Labeling arrangement for classification investigation (2).



Finally, classification investigation (3) examined whether models could be trained to differentiate between anxiety experienced by individuals with differing levels of social anxiety. Therefore, the data were collected from participants within differing ranges of LSAS-SR scores, including anxiety category 1 (LSAS-SR:50-64) and anxiety category 2 (LSAS-SR:65-80) and was subsequently labeled (Figure 7).

Figure 7. Labeling arrangement for classification investigation (3).



The first 2 minutes from the baseline period were disregarded to account for acclimatization, and the recording was discarded after the task as it was not needed. All participant data were then combined.

The features were standardized to have zero mean and unit variance, which is a widely used scaling approach as algorithms such as Radial SVM assume features are centered around zero [36,40]. For each feature, the mean (μ) and the standard deviation (σ) were extracted from the raw training feature values. The training data were then standardized using equation (3), and the same transformation was applied to the test data [37].

Classification

The investigations were framed as supervised learning tasks owing to their classification nature. Four classification algorithms were explored: SVM, Random Forest, Decision Tree, and KNN. Furthermore, for classification investigation (2), a "*One Vs. Rest*" strategy was utilized as the investigation involved a multi-class data set.

The trained models were evaluated using 10-fold cross-validation. The method involves dividing the data set into k-folds with 1 fold for testing and the others for training. Confusion matrices were also utilized to calculate the average classification accuracy for each class.

Results

Study Descriptives

All study descriptives are shown in Table 2. In this sample, women had higher mean levels for all study variables than men (though the differences were nonsignificant, which is likely owing to the small sample). This is uncharacteristic, as women typically have a higher risk to develop anxiety and higher mean levels of social anxiety than men [41]. However, the self-reported LSAS scores were highly correlated with SPSQ scores (r=0.63; P=.05).

Table 2. Descriptives for all study variables by gender.

Gender	Participants, n	Score, mean (SD)
Liebowitz Social Anxiety Scale fear subscale		
Women	7	1.3095 (0.33666)
Men	5	1.5917 (0.11562)
Liebowitz Social Anxiety Scale avoidance subscale		
Women	7	1.1845 (0.34766)
Men	5	1.3583 (0.25786)
Liebowitz Social Anxiety Scale avoidance subscale overall score		
Women	7	1.2470 (0.33190)
Men	5	1.4750 (0.15548)
Social Phobia Screening Questionnaire		
Women	7	1.3469 (0.36288)
Men	5	1.7429 (0.29277)

Combined Modalities

For classification investigation (1), the yielded accuracies were between 97.54% and 99.48%, as shown in Table 3. For investigation (2), the accuracies were between 95.18% and 98.10%, as shown in Table 4. Additionally, for investigation (3) the yielded accuracies were between 98.86% and 99.52%, as shown in Table 5. In each classification investigation, Radial SVM outperformed other classifiers (Tables 3-5).

Table 3. Cross-validation results for classification investigation (1).

Classifier	Overall performance, %	Baseline state accuracy, %	Social anxiety state accuracy, %
Radial Support Vector Machine	99.48	99.40	99.52
K-Nearest Neighbours	99.08	99.12	99.05
Decision Tree	97.54	99.04	96.59
Random Forest	97.96	99.38	97.13

Table 4. Cross-validation results for classification investigation (2).

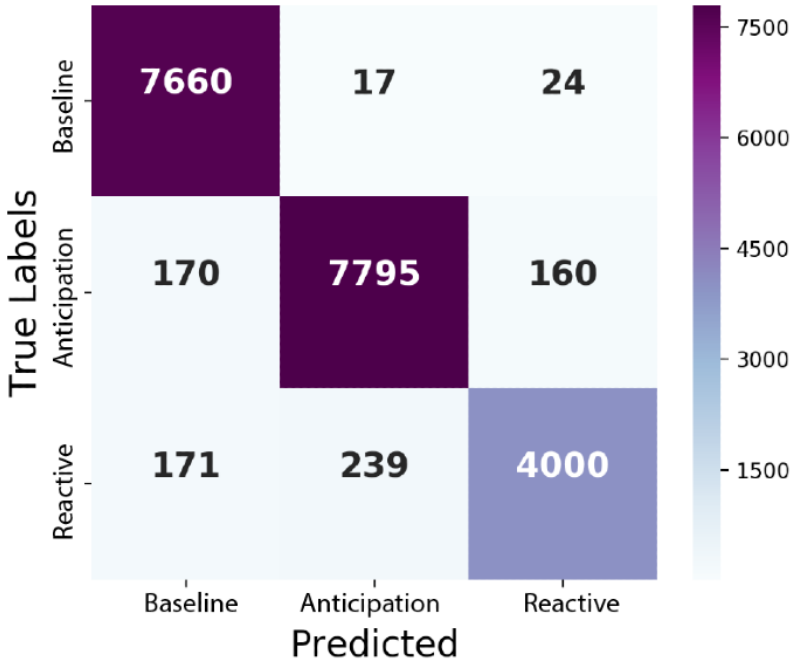
Classifier	Overall performance, %	Baseline state accuracy, %	Anticipation anxiety state accuracy, %	Reactive anxiety state accuracy, %
Radial Support Vector Machine	98.10	99.30	98.37	95.52
K-Nearest Neighbours	97.61	98.99	97.28	95.78
Decision Tree	96.63	99.39	96.86	91.36
Random Forest	95.18	99.27	95.99	85.99

Table 5. Cross-validation results for classification investigation (3).

Classifier	Overall performance, %	Anxiety category 1 accuracy, %	Anxiety category 2 accuracy, %
Radial Support Vector Machine	99.52	100	99.03
K-Nearest Neighbours	98.86	99.35	98.36
Decision Tree	99.04	100	98.09
Random Forest	99.34	100	98.70

There were common class misclassification patterns among all classifiers. For investigation (1), the models were less able to classify anxious states (Table 3). For investigation (2), reactive anxiety was misclassified the most and often mistaken for anticipation anxiety (Figure 8). Additionally, in investigation (2), the baseline class was most accurately classified, as shown in Table 4. For investigation (3), the models were not as effective at classifying anxiety category 2 (Table 5).

Figure 8. Confusion matrix from classification investigation (2) using the Decision Tree.



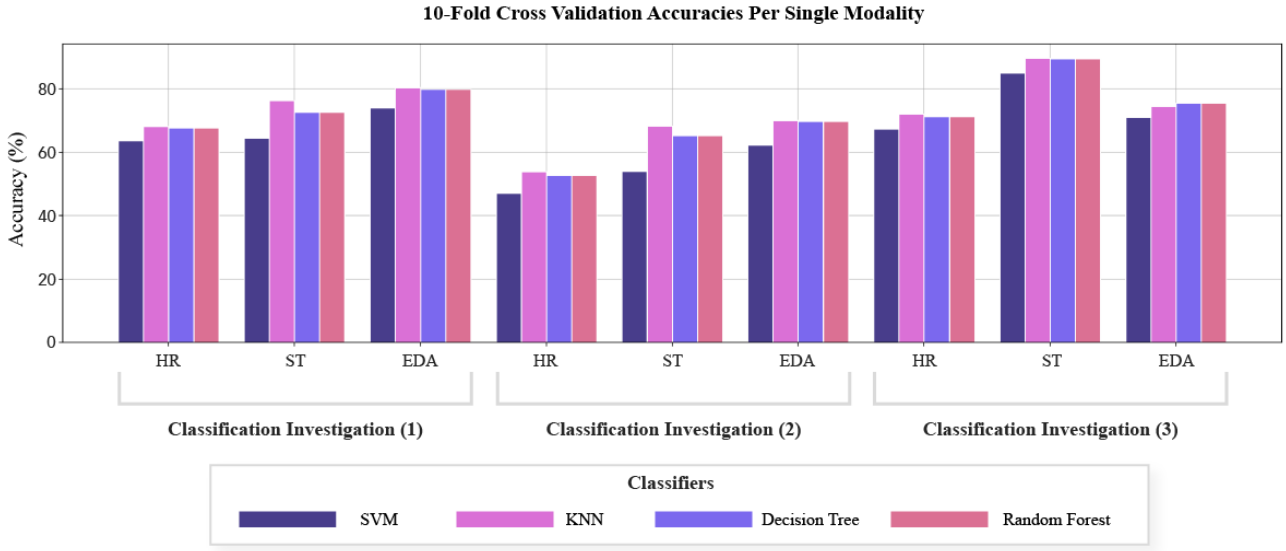
Singular Modalities

The singular modality results are shown in Table 6 and Figure 9. In classification investigation (1) EDA yielded 80.46% and was shown to have the highest predictive capability. EDA was also shown to have the highest classification accuracy of 70.02% for investigation (2), whereas ST was the most effective modality for investigation (3) with an accuracy of 89.47%. For each classification investigation, HR was observed to be the least effective modality. Furthermore, KNN generally outperformed other classifiers (Table 6).

Table 6. Highest cross-validation results per single modality.

Modality	Classifier with the highest performance	Overall performance, %
Classification investigation 1		
Heart rate	K-Nearest Neighbours	68.18
Skin temperature	K-Nearest Neighbours	76.30
Electrodermal activity	K-Nearest Neighbours	80.46
Classification investigation 2		
Heart rate	K-Nearest Neighbours	53.91
Skin temperature	K-Nearest Neighbours	68.32
Electrodermal activity	K-Nearest Neighbours	70.02
Classification investigation 3		
Heart rate	K-Nearest Neighbours	72.00
Skin temperature	K-Nearest Neighbours	89.47
Electrodermal activity	Random Forest and Decision Tree	75.66

Figure 9. Accuracies per modality. EDA: electrodermal activity, HR: heart reate, KNN:K-Nearest Neighbours, ST: skin temperature, SVM: Support Vector Machine.



Discussion

Principal Findings

Combined Modalities

This study aimed to determine if ML models could be trained to (1) classify baseline and socially anxious states, (2) differentiate among baseline, anticipation anxiety, and reactive anxiety states, and (3) classify social anxiety with differing severity levels of social anxiety. High accuracies were obtained when differentiating between baseline and socially anxious states, suggesting that it is possible to detect social anxiety using HR, ST, and EDA. These high accuracies are likely due to physiological differences between baseline and socially anxious states and have also been shown in previous research [10-12].

The models also yielded high accuracies when classifying among baseline, anticipatory, and reactive states. The classifiers' ability to differentiate between reactive and anticipatory anxiety might be due to the varying responses during these stages. It is, therefore, likely possible to detect the nature of social anxiety experienced on an individual basis.

The models also yielded high accuracies when differentiating between marked and moderate social anxiety. This demonstrates the possibility to identify social anxiety levels using physiological indices, implying that individuals with differing severity levels of social anxiety exhibit diverse physiological responses. This is in line with prior research indicating that individuals with greater social anxiety exhibit responses consistent with greater threat [10].

The results also indicated that higher modeling accuracies were yielded when all modalities were combined [42]. Research shows that models created using singular modalities may be underfit owing to lack of data [37]. This is likely because each physiological index contains varying information that enables classifiers to differentiate among certain classes, thus providing measurement granularity. Furthermore, when modalities were combined, Radial SVM outperformed the other classifiers in all investigations, which is possibly owing to the classifier's ability to formulate complex decision boundaries [37,43].

Finally, certain classes were commonly misclassified, which could be explained by class imbalances owing to the different durations of each stage during the data collection sessions. Class imbalances can cause classifiers to bias toward larger classes [44].

Singular Modalities

Each modality had varying predictive capabilities, despite the complexity of the physiological indicators used in the study. EDA was the most effective singular modality when differentiating between baseline and social anxiety states (including anticipatory and reactive states). This is possibly because EDA comprises the sum of phasic and tonic components that change following stimuli, which is likely because sweat glands responsible for EDA variation are entirely controlled by the SNS, whereas HR and ST are mediated by both the PNS and SNS [26,27,29]. Thus, EDA represents an accumulation of information that could indicate social anxiety [27].

ST was the most effective modality when differentiating between anxiety experienced by individuals with differing severity levels of social anxiety. This might be because individuals with greater social anxiety exhibit differing amounts of blood flow to the skin. This surprising finding highlights the predictive capability of ST collected around the wrist and suggests that it could be viewed as a novel social anxiety marker.

HR showed the lowest effectiveness in all investigations, which might be explained by HR being mediated by the PNS and SNS [26,27]. The comparatively low recognition accuracies may also be a result of HR being sampled at the lowest rate.

Furthermore, KNN was the most effective classifier when the modalities were singular, which is likely because KNN can formulate complex decision boundaries between classes.

Limitations

Despite these promising results, these findings are preliminary. The sample size was small, with the COVID-19 pandemic preventing further data collection. Prior to the COVID-19 pandemic, we intended to collect test data in "real-world" settings to evaluate the models' ability to detect social anxiety in practice. Instead, the models were evaluated using a subset of data from the experiment. Although this approach is often used in ML studies [16,17], it does not offer a realistic indication of model generalizability. Therefore, given the small sample size, our results need to be interpreted cautiously.

Additionally, classifiers may have been biased toward certain classes owing to the moderately differing class sizes (Multimedia Appendix 2). This may have accounted for the high accuracies but reduced model generalizability [44]. Like other studies of a similar nature (such as affect recognition studies using physiological data [45]), it was difficult to establish the ground truth of the data with respect to the presence and nature of social anxiety. Therefore, labeling was assumed to be aligned with the experimental protocol.

Furthermore, the physiological responses from the individuals could have been influenced by external factors such as caffeine and alcohol consumption [46,47], though this was not mitigated in the current study design. It is also important to note that EDA measurements can be affected by environmental conditions such as humidity and room temperature [27]. Although the experiments took place in the same room, these variables were not monitored and controlled. In sum, all of these limitations remain challenges for future research.

Comparison With Prior Work

Despite its limitations, this study has extended previous work and applications focusing on supervised machine learning in the field of physiological anxiety detection. This experiment was informed by existing study protocols, such as using an impromptu speech task, which is a cornerstone of experimental work invoking social anxiety [11,22,23]. Additionally, the study utilized the LSAS-SR measure, which is a widely used measure demonstrating good psychometric properties in previous research [24], and the social anxiety self-reports were cross-validated using another well-known indicator of subclinical social anxiety (SPSQ [32]). Overall, our findings also align with those of prior studies indicating that EDA is a "directed and undiluted" representation of the SNS [27]. Although prior work has focused on EDA as an indicator, physiological measurement from the anatomical site of the wrist had not been explored in a social anxiety context.

Conclusions

This study examined whether social anxiety could be detected in young adults using physiological data (HR, ST, and EDA) from wrist-worn sensors. The findings indicate that it is possible to detect social anxiety and its severity using this approach. Future work in this area has the potential to identify novel methods of detecting and monitoring subclinical social anxiety in young adults, which could help counteract development into SAD. As mental health provision is transitioning toward digital interventions, it is crucial that they are evidence-based and can target individuals with subclinical levels of social anxiety. The ability for future interventions to detect social anxiety before it escalates further could have great social and economic benefits for health care, society and those who experience its consequences.

Acknowledgments

We would like to thank all the individuals who volunteered to take part in this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
A Github repository was created to accompany this work. The repository contains the full dataset and code needed to recreate the classification models and reproduce the results, as well as functions that enable further experimentation.
[DOCX File , 48 KB - formative_v5i10e32656_app1.docx]

Multimedia Appendix 2
The following tables illustrate the class data distributions for each classification investigation.
[DOCX File , 49 KB - formative_v5i10e32656_app2.docx]

References

1. Heiser NA, Turner SM, Beidel DC, Roberson-Nay R. Differentiating social phobia from shyness. J Anxiety Disord 2009 May;23(4):469-476 [FREE Full text] [doi: 10.1016/j.janxdis.2008.10.002] [Medline: 19028075]
2. Eng W, Coles ME, Heimberg RG, Safren SA. Domains of life satisfaction in social anxiety disorder: relation to symptoms and response to cognitive-behavioral therapy. J Anxiety Disord 2005;19(2):143-156. [doi: 10.1016/j.janxdis.2004.01.007] [Medline: 15533701]

3. Wittchen H, Fuetsch M, Sonntag H, Müller N, Liebowitz M. Disability and quality of life in pure and comorbid social phobia--findings from a controlled study. Eur Psychiatry 1999 Jun;14(3):118-131. [doi: 10.1016/s0924-9338(99)80729-9] [Medline: 10572336]
4. Fehm L, Beesdo K, Jacobi F, Fiedler A. Social anxiety disorder above and below the diagnostic threshold: prevalence, comorbidity and impairment in the general population. Soc Psychiatry Psychiatr Epidemiol 2008 Apr;43(4):257-265. [doi: 10.1007/s00127-007-0299-4] [Medline: 18084686]
5. Spence SH, Rapee RM. The etiology of social anxiety disorder: An evidence-based model. Behav Res Ther 2016 Nov;86:50-67. [doi: 10.1016/j.brat.2016.06.007] [Medline: 27406470]
6. Kessler RC, Chiu WT, Demler O, Merikangas KR, Walters EE. Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):617-627 [FREE Full text] [doi: 10.1001/archpsyc.62.6.617] [Medline: 15939839]
7. Patel A, Knapp M, Henderson J, Baldwin D. The economic consequences of social phobia. J Affect Disord 2002 Apr;68(2-3):221-233. [doi: 10.1016/s0165-0327(00)00323-2] [Medline: 12063150]
8. National Collaborating Centre for Mental Health. Social Anxiety Disorder: Recognition, Assessment and Treatment. Leicester: British Psychological Society; 2013. URL: https://www.nice.org.uk/guidance/CG159 [accessed 2021-09-29]
9. Mccrone P, Dhanasiri S, Patel A, Knapp M, Smith SL. Paying the Price: The cost of mental health care in England to 2026. King's Fund. 2008. URL: https://www.kingsfund.org.uk/sites/default/files/Paying-the-Price-the-cost-of-mental-health-care-England-2026-McCrone-Dhanasiri-Patel-Knapp-Lawton-Smith-Kings-Fund-May-2008_0.pdf [accessed 2021-09-27]
10. Shimizu M, Seery MD, Weisbuch M, Lupien SP. Trait social anxiety and physiological activation: cardiovascular threat during social interaction. Pers Soc Psychol Bull 2011 Jan;37(1):94-106. [doi: 10.1177/0146167210391674] [Medline: 21177876]
11. Thompson T, Kaminska M, Marshall C, Van Zalk N. Evaluation of the social phobia scale and social interaction anxiety scale as assessments of performance and interaction anxiety. Psychiatry Res 2019 Mar;273:725-731. [doi: 10.1016/j.psychres.2019.01.103] [Medline: 31207859]
12. Edelmann RJ, Baker SR. Self-reported and actual physiological responses in social phobia. Br J Clin Psychol 2002 Mar;41(Pt 1):1-14. [doi: 10.1348/014466502163732] [Medline: 11931674]
13. Heimberg RG, Butler RM. Research on social anxiety disorder continues to escalate: A commentary on Asmundson and Asmundson's analysis of publication trends in the anxiety disorders. J Anxiety Disord 2018 May;56:8-10. [doi: 10.1016/j.janxdis.2018.04.001] [Medline: 29859660]
14. Peake JM, Kerr G, Sullivan JP. A Critical Review of Consumer Wearables, Mobile Applications, and Equipment for Providing Biofeedback, Monitoring Stress, and Sleep in Physically Active Populations. Front Physiol 2018;9:743 [FREE Full text] [doi: 10.3389/fphys.2018.00743] [Medline: 30002629]
15. Kanjo E, Al-Husain L, Chamberlain A. Emotions in context: examining pervasive affective sensing systems, applications, and analyses. Pers Ubiquit Comput 2015 Apr 25;19(7):1197-1212 [FREE Full text] [doi: 10.1007/s00779-015-0842-3]
16. de Arriba-Pérez F, Santos-Gago J, Caeiro-Rodríguez M, Ramos-Merino M. Study of stress detection and proposal of stress-related features using commercial-off-the-shelf wrist wearables. J Ambient Intell Human Comput 2019 Jan 8;10(12):4925-4945 [FREE Full text] [doi: 10.1007/s12652-019-01188-3]
17. Zhang Z, Song Y, Cui L, Liu X, Zhu T. Emotion recognition based on customized smart bracelet with built-in accelerometer. PeerJ 2016;4:e2258 [FREE Full text] [doi: 10.7717/peerj.2258] [Medline: 27547564]
18. Zhai J, Barreto AB, Chin C, Li C. Realization of stress detection using psychophysiological signals for improvement of human-computer interactions. 2005 Presented at: Proceedings. IEEE SoutheastCon, 2005; April 8-10, 2005; Fort Lauderdale, FL. [doi: 10.1109/secon.2005.1423280]
19. Xu Y, Hübener I, Seipp AK, Ohly S, David K. From the lab to the real-world: An investigation on the influence of human movement on Emotion Recognition using physiological signals. 2017 Presented at: 2017 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops); March 13-17, 2017; Kona, HI. [doi: 10.1109/percomw.2017.7917586]
20. Zhai J, Barreto A. Stress Detection in Computer Users Based on Digital Signal Processing of Noninvasive Physiological Variables. 2006 Presented at: 2006 International Conference of the IEEE Engineering in Medicine and Biology Society; August 30 to September 3, 2006; New York, NY. [doi: 10.1109/iembs.2006.259421]
21. Karthikeyan P, Murugappan M, Yaacob S. Analysis of Stroop Color Word Test-Based Human Stress Detection using Electrocardiography and Heart Rate Variability Signals. Arab J Sci Eng 2013 Sep 11;39(3):1835-1847 [FREE Full text] [doi: 10.1007/s13369-013-0786-8]
22. Anderson ER, Hope DA. The relationship among social phobia, objective and perceived physiological reactivity, and anxiety sensitivity in an adolescent population. J Anxiety Disord 2009 Jan;23(1):18-26 [FREE Full text] [doi: 10.1016/j.janxdis.2008.03.011] [Medline: 18436426]
23. Knight ML, Borden RJ. Autonomic and affective reactions of high and low socially-anxious individuals awaiting public performance. Psychophysiology 1979 May;16(3):209-213. [doi: 10.1111/j.1469-8986.1979.tb02979.x] [Medline: 441214]

24. Fresco D, Coles M, Heimberg R, Liebowitz M, Hami S, Stein M, et al. The Liebowitz Social Anxiety Scale: a comparison of the psychometric properties of self-report and clinician-administered formats. Psychol Med 2001 Aug;31(6):1025-1035. [doi: 10.1017/s0033291701004056] [Medline: 11513370]
25. Leary MR, Kowalski RM. Social Anxiety. New York, NY: Guilford Press; 1995:83-131.
26. Michael-Titus A, Revest P, Shortland P. The Nervous System (2nd edition). New York, NY: Churchill Livingstone; 2010:1-30.
27. Cacioppo J, Tassinary LG, Berntson GG. Handbook of Psychophysiology. Cambridge: Cambridge University Press; 2007.
28. Craske MG, Waters AM, Lindsey Bergman R, Naliboff B, Lipp OV, Negoro H, et al. Is aversive learning a marker of risk for anxiety disorders in children? Behav Res Ther 2008 Aug;46(8):954-967 [FREE Full text] [doi: 10.1016/j.brat.2008.04.011] [Medline: 18539262]
29. Koeppen BM, Stanton BA. Berne & Levy Physiology. Philadelphia, PA: Elsevier; 2018.
30. Mauss IB, Wilhelm FH, Gross JJ. Autonomic recovery and habituation in social anxiety. Psychophysiology 2003 Jul;40(4):648-653. [doi: 10.1111/1469-8986.00066] [Medline: 14570172]
31. Heimberg RG, Horner KJ, Juster HR, Safren SA, Brown EJ, Schneier FR, et al. Psychometric properties of the Liebowitz Social Anxiety Scale. Psychol Med 1999 Jan 01;29(1):199-212. [doi: 10.1017/s0033291798007879]
32. Furmark T, Tillfors M, Everz P, Marteinsdottir I, Gefvert O, Fredrikson M. Social phobia in the general population: prevalence and sociodemographic profile. Soc Psychiatry Psychiatr Epidemiol 1999 Aug;34(8):416-424. [doi: 10.1007/s001270050163] [Medline: 10501711]
33. Van Zalk N, Tillfors M, Mörtberg E. Social Anxiety-Impulsivity Subgroups and Links to Later Emotional Adjustment in Adolescence: A Latent Transition Analysis. J Early Adolesc 2020 May 19;40(9):1397-1426 [FREE Full text] [doi: 10.1177/0272431620919153]
34. E4 wristband: Real-time physiological data streaming and visualization. URL: https://www.empatica.com/research/e4/ [accessed 2021-09-27]
35. E4 wristband user's manual. URL: https://empatica.app.box.com/v/E4-User-Manual [accessed 2021-09-27]
36. Géron A. Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. Newton, MA: O'Reilly Media; 2019.
37. Dougherty G. Pattern Recognition and Classification. New York, NY: Springer; 2013.
38. Lykken DT, Rose R, Luther B, Maley M. Correcting psychophysiological measures for individual differences in range. Psychol Bull 1966 Dec;66(6):481-484. [doi: 10.1037/h0023922] [Medline: 5974620]
39. Lykken DT, Venables PH. Direct measurement of skin conductance: a proposal for standardization. Psychophysiology 1971 Sep;8(5):656-672. [doi: 10.1111/j.1469-8986.1971.tb00501.x] [Medline: 5116830]
40. Paper D. Introduction to Scikit-Learn. Hands-on Scikit-Learn for Machine Learning Applications. Berkeley, CA: Apress; 2020. URL: https://scikit-learn.org/stable/modules/preprocessing.html [accessed 2021-09-29]
41. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. Arch Gen Psychiatry 2005 Jun;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
42. Anusha A, Jose J, Preejith S, Jayaraj J, Mohanasankar S. Physiological signal based work stress detection using unobtrusive sensors. Biomed Phys Eng Express 2018 Sep 07;4(6):065001 [FREE Full text] [doi: 10.1088/2057-1976/aadbd4]
43. Kampouraki A, Manis G, Nikou C. Heartbeat time series classification with support vector machines. IEEE Trans Inf Technol Biomed 2009 Jul;13(4):512-518. [doi: 10.1109/TITB.2008.2003323] [Medline: 19273030]
44. Muller A, Guido S. Introduction to Machine Learning with Python: A Guide for Data Scientists. Newton, MA: O'Reilly Media; 2017:76-284.
45. Picard R, Vyzas E, Healey J. Toward machine emotional intelligence: analysis of affective physiological state. IEEE Trans Pattern Anal Machine Intell 2001;23(10):1175-1191 [FREE Full text] [doi: 10.1109/34.954607]
46. Barry RJ, Clarke AR, Johnstone SJ, Rushby JA. Timing of caffeine's impact on autonomic and central nervous system measures: clarification of arousal effects. Biol Psychol 2008 Mar;77(3):304-316. [doi: 10.1016/j.biopsycho.2007.11.002] [Medline: 18093716]
47. Schrieks I, Stafleu A, Kallen V, Grootjen M, Witkamp R, Hendriks H. Moderate alcohol consumption, autonomic nervous system and mood. Appetite 2013;71:485-485 [FREE Full text] [doi: 10.1016/j.appet.2013.06.061]

Abbreviations

ANS: autonomic nervous system
BPM: beats per minute
EDA: electrodermal activity
HR: heart rate
KNN: K-Nearest Neighbours
LSAS-SR: self-reported version of the Liebowitz Social Anxiety Scale
ML: machine learning

PNS: parasympathetic nervous system
SAD: social anxiety disorder
SNS: sympathetic nervous system
SPSQ: Social Phobia Screening Questionnaire
ST: skin temperature
SVM: Support Vector Machine

Edited by G Eysenbach; submitted 05.08.21; peer-reviewed by M Dechant; comments to author 26.08.21; revised version received 17.09.21; accepted 20.09.21; published 07.10.21.

Please cite as:
Shaukat-Jali R, van Zalk N, Boyle DE
Detecting Subclinical Social Anxiety Using Physiological Data From a Wrist-Worn Wearable: Small-Scale Feasibility Study
JMIR Form Res 2021;5(10):e32656
URL: https://formative.jmir.org/2021/10/e32656
doi:10.2196/32656
PMID:34617905

Original Paper

The Potential of Digital Data Collection Tools for Long-lasting Insecticide-Treated Net Mass Campaigns in Nigeria: Formative Study

Youngji Jo[1], PhD; Nathan Barthel[2], MSc; Elizabeth Stierman[2], MPH; Kathryn Clifton[2], PhD; Esther Semee Pak[3], MPH; Sonachi Ezeiru[4]; Diwe Ekweremadu[4]; Nnaemeka Onugu[4]; Zainab Ali[4]; Elijah Egwu[4]; Ochayi Akoh[4], BSc; Orkan Uzunyayla[5], MSc; Suzanne Van Hulle[2], MHS

[1]Boston Medical Center, Boston, MA, United States

[2]Catholic Relief Services, Baltimore, MD, United States

[3]Graduate Institute of International Development Studies, Geneva, Switzerland

[4]Catholic Relief Services, Abuja, Nigeria

[5]Red Rose Ödeme Sistemleri AŞ, İstanbul, Turkey

Corresponding Author:
Youngji Jo, PhD
Boston Medical Center
One Boston Medical Center Pl
Boston, MA, 02118
United States
Phone: 1 4438001626
Email: youngji1435@gmail.com

Related Article:

This is a corrected version. See correction statement: https://formative.jmir.org/2021/11/e34526

Abstract

Background: Nigeria has the world's largest malaria burden, accounting for 27% of the world's malaria cases and 23% of malaria mortality globally. This formative study describes the operational process of the mass distribution of long-lasting insecticide-treated nets (LLINs) during a campaign program in Nigeria.

Objective: This study aims to assess whether and how digital data collection and management tools can change current practices and help resolve major implementation issues.

Methods: Qualitative data on the technical features and operational processes of paper-based and information and communication technology (ICT)–based systems in the Edo and Kwara states from June 2 to 30, 2017, were collected on the basis of documented operation manuals, field observations, and informant interviews. During the LLIN campaign in Edo State, we recruited 6 local government area focal persons and monitors and documented daily review meetings during household mobilization (9 days) and net distribution (5 days) to understand the major program implementation issues associated with the following three aspects: logistic issues, technical issues, and demand creation. Each issue was categorized according to the expected degree (low, mid, and high) of change by the ICT system.

Results: The net campaign started with microplanning and training, followed by a month-long implementation process, which included household mobilization, net movement, net distribution, and end process monitoring. The ICT system can improve management and oversight issues related to data reporting and processes through user-centered interface design, built-in data quality control logic flow or algorithms, and workflow automation. These often require more than 50% of staff time and effort in the current paper-based practice. Compared with the current paper-based system, the real-time system is expected to reduce the time to payment compensation for health workers by about 20 days and produce summary campaign statistics for at least 20 to 30 days.

Conclusions: The ICT system can facilitate the measurement of population coverage beyond program coverage during an LLIN campaign with greater data reliability and timeliness, which are often compromised due to the limited workforce capacity in a paper-based system.

(JMIR Form Res 2021;5(10):e23648) doi:10.2196/23648

KEYWORDS

long-lasting insecticide-treated nets; malaria; Nigeria; information communication technology; geographic information system; supply chain management

Introduction

Background

Nigeria has the world's largest burden of malaria, accounting for 27% of the world's malaria cases and 23% of malaria mortality globally [1]. The disease continues to be a pressing public health challenge, responsible for 60% of the country's outpatient visits and 30% of hospitalizations [1]. Long-lasting insecticide-treated nets (LLINs) have played an important role in the remarkable reduction in the number of malaria cases and deaths worldwide over the past decade. Sleeping under a net has been shown to reduce all-cause mortality in children younger than 5 years by about 20% and malarial illnesses among children younger than 5 years and in pregnant women by up to 50% [2]. Studies demonstrating the effectiveness [3-6] and cost-effectiveness [7,8] have contributed to substantial success in scaling up this intervention in malaria-endemic African countries over the past decades [9]. However, more progress is required to improve sustained coverage [10] as well as access and use from an equity perspective [11].

Mobile phone technology coupled with electronic geographic information systems (GISs) has made data collection processes easier and more accurate and has become a powerful tool for managing data in the context of malaria prevention and control [12-15]. Modernized high-resolution mapping technologies and satellite imagery can help locate target populations and provide reliable estimates of populations at risk. This can also facilitate the identification of issues pertinent to health service delivery and produce data that can guide better intervention strategies so that scarce resources can be allocated in a cost-effective manner, especially in low-income countries [16]. Such systems allow government health officials in capital cities to better plan interventions in remote rural areas. This, in turn, promises a better assessment approach for trends and correlations to improve the planning, monitoring, and surveillance of public health programs in resource-limited countries.

Objectives

The use of digital data collection and management tools based on information and communication technology (ICT) can promote supply chain management [17-19] and sustainable and effective use of LLINs [20,21]; however, there is not enough evidence or guidelines that describe the specific features, processes, and consequences of how the ICT system could change the current practice. From the perspective of management decision-making, the goal of this study is to describe the detailed process and system of mass distribution LLIN campaign programs and identify major challenges and resolution approaches in the paper- and ICT-based systems. The study findings will discuss whether and how a digital data collection and management tool can potentially improve the management, monitoring, and evaluation of the program.

Methods

Study Setting

In Nigeria, Catholic Relief Services (CRS) is working in partnership with the National Malaria Elimination Program (NMEP) to distribute LLINs to prevent malaria through a program funded by the Global Fund to Fight AIDS, Tuberculosis, and Malaria [22]. Together with NMEP and state ministries of health, CRS implemented malaria LLIN replacement campaigns between April and November 2017 in 6 Southeast Nigerian states: Kwara, Ondo, Imo, Edo, Osun, and Adamawa. Among the 6 states, we collected data from Edo State—in all 18 local government areas (LGAs)—using the traditional paper-based intervention, and from Kwara State with a pilot ICT-based intervention in 1 LGA (Oyun). According to the 2016 census, Edo State consists of approximately 4 million people [23], and Kwara State consists of approximately 3 million people [24]. One LGA generally consists of approximately 16 to 30 wards, and the size of a ward varies, consisting of 500 to 8000 households per ward [25].

Research Participants

One researcher (YJ) and 2 technical staff members (OA and EE) conducted structured in-depth interviews with 10 CRS program managers and 26 local stakeholders (eg, local partner organizations, government officers from NMEP, CRS local staff) working in Edo and Kwara states. We conducted field observations of the paper-based system in Edo State. We recruited 1 LGA focal person and 1 monitor from 6 LGAs and documented daily review meetings during the household mobilization (9 days) and net distribution (5 days) campaign from June 2 to June 30, 2017. We could not conduct field observations of the ICT system in the Kwara State, as the campaign schedule had to be changed owing to delayed mobile phone procurement. Thus, we conducted informant interviews with 2 field managers in Kwara and 2 technical staff who developed the Red Rose system to understand the system characteristics, operational procedure, challenges, and resolution approaches of the ICT-based system.

Data Collection and Analyses

We reviewed the documented operation manuals and conducted informant interviews to map out operational processes and identify the technical features of paper- and ICT-based systems

(Figure 1). On the basis of the documented contents during LGA- and state-level review meetings, we categorized the major program implementation issues discussed during the meetings into the following three campaign activity components: (1) logistic issues, (2) technical issues, and (3) demand creation activity and identified typical resolution approaches in paper-based campaigns. Each issue was then reviewed and categorized into low, middle, and high levels based on the expected degree of change by the ICT system (Multimedia Appendix 1).

LLIN Campaign Structure and Processes

The campaign consisted of multiple levels of staff and a set of activities, including planning, LLIN logistics, training, demand creation, household mobilization, net distribution, and in or end process monitoring. First, during the macroquantification stage, the required number of LLINs for each LGA was determined on the basis of the working assumption of "a 1.8 person-per-net ratio" [26], a set of simple formulas for calculating administrative coverage, and data from a population-based survey or census. On the basis of the required numbers of LLINs, a logistic plan, including net procurement for each LGA, was developed. Microplanning at the state level determined the number and location of each settlement in the LGA.

Next, the training took place at the state, LGA, ward, and community levels as a cascade or training of trainers model. Through existing social and community networks, media jingles, community sensitization, and demand creation, activities were implemented to increase community awareness of the LLIN campaign and improve demand for and use of nets. Then, household mobilization took place through door-to-door visits, in advance of the net distribution, to encourage active participation in the campaigns, register households, issue net cards based on the number of household members (ie, 1 net card to be issued per 2 people but a maximum of 4 net cards to be issued per household), and provide information about collecting and using the LLINs. Household registration teams wrote information about the LGA or ward names of the household, the designated location of distribution point (DP), and the pick-up date on the net cards when they issued them to the households. Each registration team registered 24 to 35 households per day. On the scheduled dates, beneficiaries brought their net cards to their designated DP. At the DP, their net cards were exchanged for a net. Monitors randomly visited 8 to 10 households per day during the household mobilization period and 2 to 4 DPs during the net distribution period to check the net card and net distribution status during the campaign.

In parallel, the review meetings were conducted during the household mobilization activities (9 days) and net distribution activities (5-7 days) to monitor daily performance and address any challenges. The meetings were held at LGA and state levels. At the LGA level, ward or DP supervisors gathered at a meeting and reported campaign progress in their catchment areas to an LGA focal person. After the meeting, LGA focal persons and monitors gathered for a state-level meeting to report the progress and discuss major issues with state supervisors and a committee that consisted of government officers from NMEP, CRS program staff, and local partner organizations. An end process evaluation (also known as a rapid assessment) was carried out to check the net distribution status by the independent monitors 2 days after distribution in all LGAs using the two-stage cluster sampling (ie, *4-4-10 principle*: a convenient sampling of 4 wards in each LGA, 4 communities in each ward, and 10 households in each community, respectively). All the data from the process were submitted directly to the state data manager for validation and compilation. The information retrieved was used to prepare the state debriefing presentation, state campaign report, and postcampaign behavior change communication priorities.

Results

Data Flow of the LLIN Campaign

The net campaign started with microplanning and training followed by a month-long implementation, including household mobilization, net movement, net distribution, and end process monitoring (Figures 1 and 2). In paper-based systems, training attendance was checked by training attendance sheets filled by participants, and payments were manually processed based on the records (Figure 1). For paper-based registration, the household mobilization team consisted of 2 people who visited households for registration. At each household visit, one household mobilizer filled out the Household Registration Sheets (Form I-3, as presented in Figure 1) and another household mobilizer completed the Net Cards Distribution Daily Summary Sheet (Form I-4a). Once registered, households were given net cards based on the number of household members (1 net card per 2 household members, but a maximum of 4 net cards per household). At the end of the day, ward supervisors collated the net card distribution summary sheet (Form I-4b). Then, LGA supervisors completed the LGA Nets cards distribution summary sheets (Form I-4c) at the LGA and collated them into the LGA LLIN Summary Sheet (Form I-8c) at the state level at the end of the campaign.

Figure 1. Operation process and data flow of the paper-based system of the long-lasting insecticide-treated net campaign in Edo State. DP: distribution point; HH: household; ICC: inventory control card; LGA: local government area.

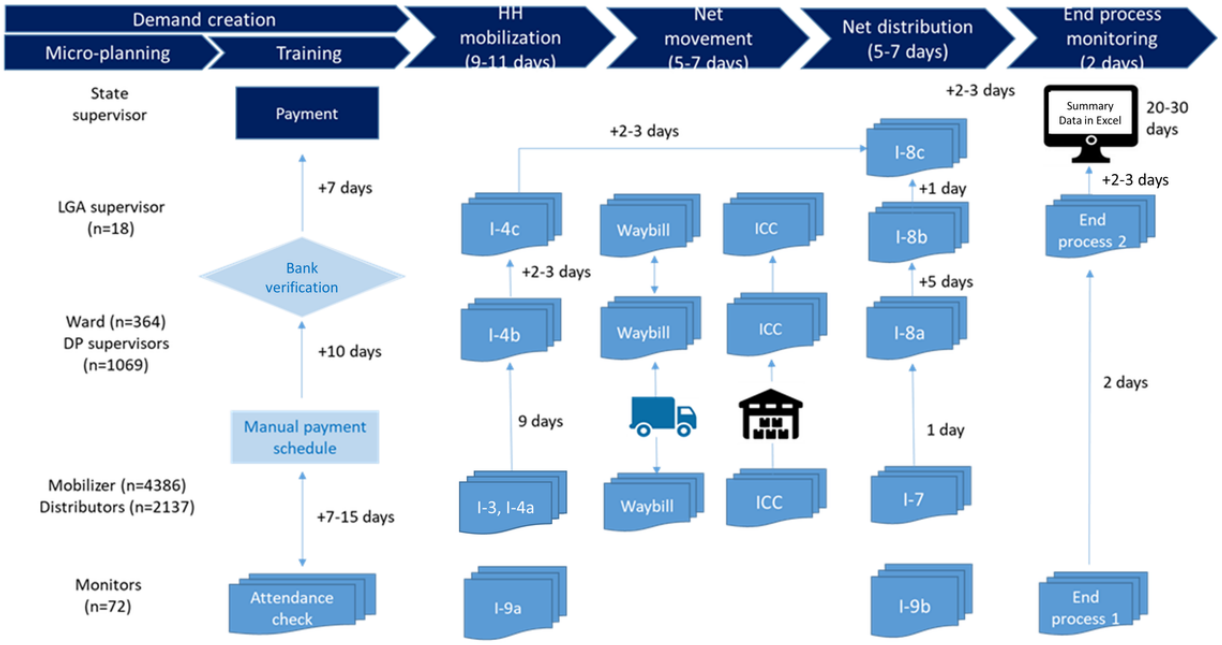
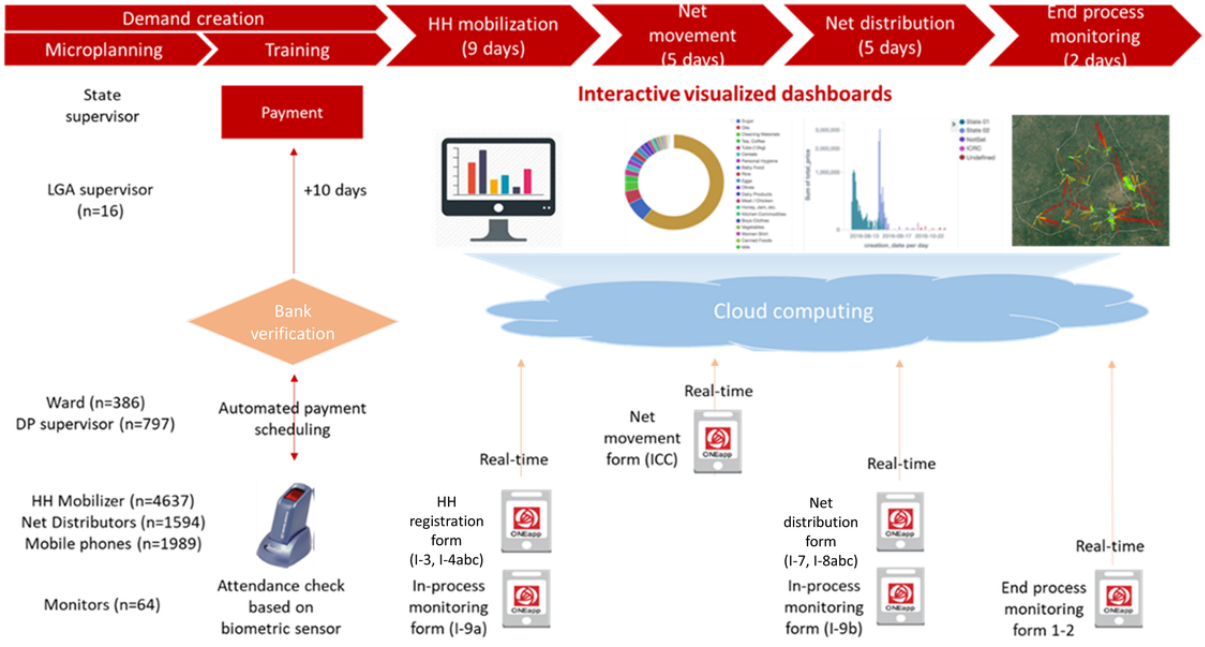


Figure 2. Operation process and data flow of the information and communication technology–based system of the long-lasting insecticide-treated net campaign in Kwara State. DP: distribution point; HH: household; ICC: inventory control card; LGA: local government area.



On the basis of the summary data at the end of the household mobilization activities, net positioning was planned for each designated DP (which takes an additional 5-7 days). In terms of net logistics, inventory control card forms were used to track the net inventory stock status by DP supervisors and were submitted to the LGA supervisor. Waybill forms were used to record net movement information such as types of vehicle, amount of nets, date of delivery, and sender or recipient of the nets. During net distribution, distributors filled the Daily Net Collection Tally Sheets (Form I-7) and DP supervisors filled out the DP net collection summary sheets (Form 1-8a) at the end of the day. Then, LGA supervisors filled out the LLIN and Net Card Aggregation Summary Sheets (Form I-8b) at the end of the net distribution and LGA LLIN summary sheets (Form I-8c) based on I-4c and I-8b forms from all LGAs.

At the end of the distribution, the DP supervisor calculated the total number of LLINs distributed over the 5-day distribution period and the 2 additional days, as well as the number of LLINs remaining in the DP storage, to the ward supervisor using the inventory control card form. Net distributors recorded these forms, and ward supervisors and LGA supervisors collated the data. Data clerks at the state level then finally entered the data into computers. In addition, independent monitors filled in in-process monitoring forms during household mobilization (Form I-9a) and net distribution (Form I-9b) periods. At the end of the net distribution, the independent monitors filled the end-process monitoring forms. Overall, these processes took approximately 20 to 30 days from the initiation of household registration activities in the current paper-based practice.

The ICT system included the use of a technology platform—Red Rose software loaded onto Android phones that can function in real time whether it is connected or disconnected to the internet network (Figure 2). The overall campaign operation processes and staffing structures were similar to those of the paper systems. The Red Rose platform was used throughout the campaign, including for tracking attendees at different training events, mobile money payments, household mobilization, net distribution, in-process monitoring, and end-process monitoring surveys. Portable power banks or wireless network adapters were provided for network connectivity. During training, biometric sensors tracked the participants' check-in or check-out times, and the participation records were automatically processed for a payment schedule. Similar to the paper system, the household mobilization team consisted of 2 people who visited households for registration. Once registered, households were given *identifiable* net cards based on the number of household members. The net cards were associated with a set of unique information, including household ID, household register ID, household phone number, mobilization date, designated DP, geocoordinate, and time data at the event of household registration and net distribution. Each net card was preprinted with an individual machine-readable quick response (QR) code, scanned using Android phones to ease tracking and prevent errors that may arise from manual entry. These QR codes were scanned by mobile phones in every transaction of the net card distribution and collection during household registration and net distribution. In this way, each net card distributed was entered into and tracked by the Red Rose platform. The number of LLINs that a household receives was automatically calculated when the number of people in the household was entered into the Android device. At each DP, a DP supervisor recorded the net stock position in the Red Rose platform. When LLINs were handed to beneficiaries, each net card was scanned, and the stock position was automatically updated. The progress of net card distribution and collection, as well as stock positions, was viewed regularly on customized and interactive dashboards. The application had built-in anomaly detection algorithms for *inactive* net cards if the net cards were not registered or assigned to different DPs. Once these data were entered by household mobilizers and net cards were scanned by net distributors, each data point was automatically synchronized based on a unique net card ID (QR code) and calculated for summary data for DP, ward, LGA, and state levels in real time.

Program Implementation Issues

As described in Table S1 in Multimedia Appendix 1, the characteristics of paper-based systems have a high dependence on individual capacity (ie, how an individual performs his or her task and the capacity of supervisors to effectively monitor activities across a wide geographic area), which are inherently exposed to several potential risks [19]. For example, during microplanning, there was a lack of reliable, up-to-date population figures of census units or village-level population, which resulted in gaps and discrepancies in household registration or net distribution planning [27]. During the training, participants' attendance checks and bank verification for payment took a long time (at least a month), which might reduce staff motivation. During the LLIN campaign, most issues were identified through staff reporting during review meetings or through oversight visits by supervisors. The reporting system also had multiple levels of reporting, with exposure to errors in data entry and requiring additional manual scrutiny. Although the system required a large volume of hard copies of reports and forms, which often require more than 50% of staff time and effort in data entry, the reporting and processing of information flow was one-way and not actionable. As a result, effective monitoring was complicated and laborious, often done merely by occasional checks (spot checks) by supervisors to monitor attendance and performance of staff; even if these were identified, there were limited mechanisms to ensure that the issues were correctly resolved. Although campaign staff were generally highly committed to their work, willing to serve the community, and had good knowledge of national guidelines, these attributes were not enough to overcome various implementation challenges, coupled with shortages of qualified workforce in resource-limited settings.

Of the identified operational issues, the ICT system mostly influenced the technical issues, in particular household mobilization and net distribution activities compared with logistic issues and demand creation issues. The ICT system has the potential to improve management and oversight through user-centered interface design, built-in data quality control logic flow or algorithms, workflow automation, and visualized interactive dashboards. The data linked to GIS information and an automated alert system could help supervisors detect issues and take immediate actions to resolve them. For example, when supervisors reviewed the satellite maps and discovered households that were not provided with net cards, they could guide them to visit the missing household. This allowed us to identify the target population more closely, without missing settlements or households. Real-time workforce performance data improved management and facilitated monitoring and supervision during household mobilization and net distribution.

Data Outputs

At the end of each phase of activities, each level's supervisor reported, collated, and produced a summary of the data in the respective catchment areas indicating the total number of households (or individuals) reached, total number of LLINs needed, and total net cards distributed. The coverage estimates were generally high, ranging between 69% (estimated program coverage of 142,635/206,717 based on population size in

Orhionmwon) and 99% (152,841/154,385 in Owan east) across 18 LGAs in Edo State (Figure S1 in Multimedia Appendix 1). The supervisors also provided DP data on the total number of net cards received, the total number of LLINs distributed, and the total number of LLINs remaining at the end of the distribution phase. These coverage measures (ie, household registration rate and net redemption rate) were compared across LGAs. In addition, end-process monitoring produced various snapshots of the overall campaign performance, based on the sampled household, in terms of net use rate, net retention rate, and net redemption rate across LGAs.

In the ICT-based system, various visual dashboards (diagrams and maps) generated from the collected data provided a comprehensive projection of the performance and coverage data according to the geographical location in close to real time (Figures 3-5). The Red Rose interactive visualization dashboard system allowed data to be tabulated, analyzed, merged, sorted, and filtered interactively to assist exploratory analyses and examine performance progress. Depending on the choice of parameters, users could decide the various visual outputs based on their interests, such as workforce performance or geographical coverage (Figure 4). The data could be accessed in real time so that any anomalies or gaps could be addressed during the campaign, not at the end of the campaign. These included workforce performance level by showing the number, location, and time of household mobilization or net distribution of health workers, percentages of household mobilization or net distribution by geographical level or location, and the service coverage level on the satellite maps (Figures 4 and 5). It also allowed additional functions, such as presenting the names of high-performance workers based on the number of household registrations (Figure S2 in Multimedia Appendix 1).

Figure 3. Interactive visualized dashboards of the information and communication technology system in Kwara State (mobilization or distribution by time and cluster).

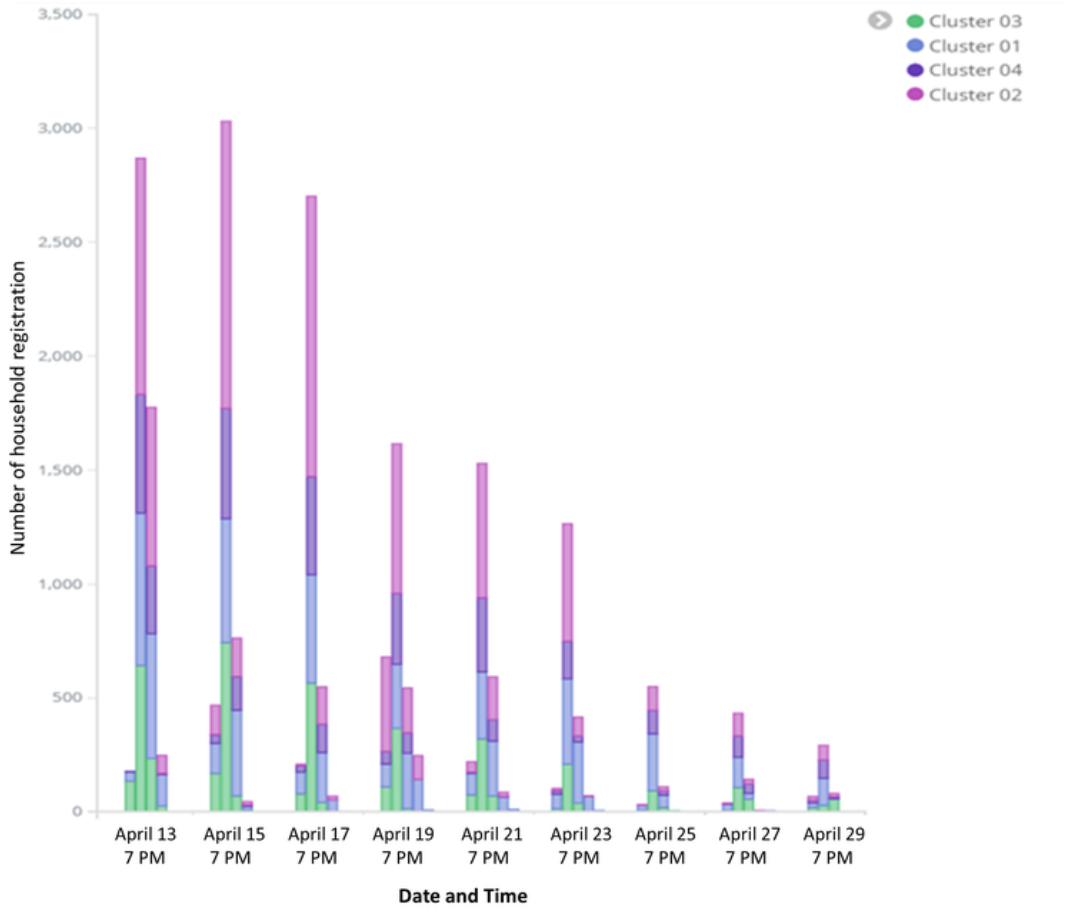


Figure 4. Interactive visualized dashboards of the information and communication technology system in Kwara State (mobilization density on the map for Oyun local government area, Kwara State).

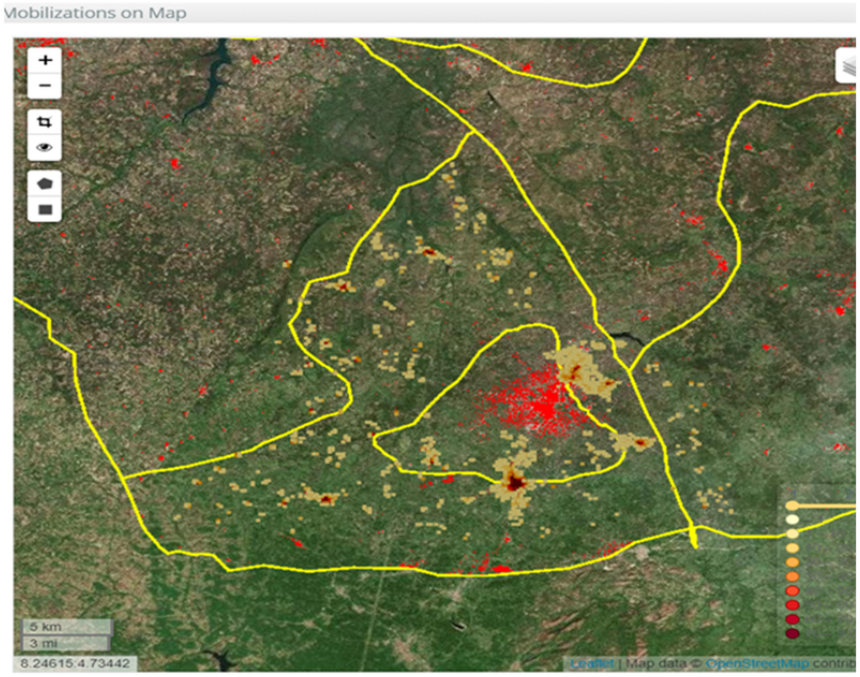
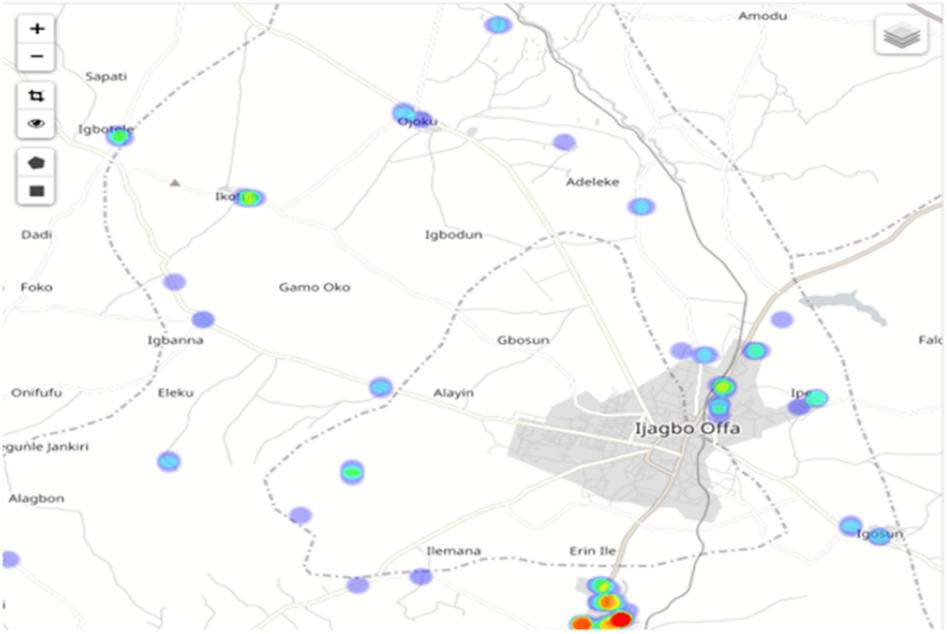


Figure 5. Interactive visualized dashboards of the information and communication technology system in Kwara State (distribution points on the map).



Discussion

Principal Findings

The LLIN campaign with paper-based systems revealed several challenges in terms of data reliability, validity, and timeliness coupled with limited workforce capacity. Although the ICT-based system would have implementation issues similar to those in the current paper-based system, it can better manage the issues related to the data reporting system and processes, which comprise significant staff time and effort in the current paper-based practice. Consequently, staff could spend more resources and time addressing issues related to LLIN supply chain management and demand creation. The information collected by the ICT system is accurate and verifiable through personal identification, geocode, and time data, which allows real-time monitoring and effective management decision-making. It can reduce any missed settlements and households based on satellite maps during the household

mobilization period and promote net card redemption during the net distribution period. Accordingly, ICT enables a system that provides a measure of accurate targets and ensures population coverage (ie, the denominator of the service coverage measure is likely the actual number of households identified by the GIS-based system and registered by staff) beyond program coverage (ie, the denominator of the service coverage measure is limited by the number of households registered by the staff) with greater confidence and transparency.

Strengths

Many previous studies have demonstrated the effectiveness and cost-effectiveness of LLIN in preventing malaria infection and reducing associated morbidity and mortality [28-30]. Mass campaigns of LLIN have thus been the primary vector control strategy for malaria-endemic countries over the past decades, with universal campaigns every 3 years, which has been the norm since 2009 [31]. Although LLIN coverage dramatically increased with mass distribution campaigns, the coverage gaps started to appear almost immediately after the campaign through net deterioration or loss of nets [32,33]. Studies also demonstrated remaining gaps in geographical heterogeneity, household net access or ownership [34], and equity gaps in terms of LLIN coverage as well as prevention effectiveness of LLIN [35,36]. The ICT system allows a verifiable transaction system through the net card that connects information of people (ie, households), place (ie, geographical location), and products (ie, nets); this integrated system can be used as push and pull mechanisms to identify and address the gaps (ie, missed households or net redemption and ensuring appropriate net use) in the service delivery process [37]. Moreover, the systematic platform and information can be also used to guide targeted strategies and improve performance for other complementary continuous distribution channels (eg, antenatal care clinics and the expanded program on immunization) even after the mass campaign as well as the next cycle of mass campaigns in 3 years, which can also benefit other population-based surveillance and intervention programs, in addition to malaria [38]. Our study described the detailed process and system of the LLIN campaign and addressed contextual barriers and facilitators to implement the campaign based on the paper- and ICT-based systems. Because establishing effectiveness or innovation itself does not guarantee its scale-up and routine use, our study provides important insights and potential research agenda of implementation science not only to the traditional paper-based practice but also to the future ICT-based program design, implementation, and evaluation [39].

Implications

When the system functions well at scale, we expect the ICT-based system to promote various supply- and demand-side factors in health systems in resource-limited settings. From the supply side, it can enhance workforce capacity and productivity by improving attendance at training, providing supervisors with better tools to monitor the activities of field staff, reducing the time required for tabulating paper forms, and automating certain processes to minimize human errors. It can also improve oversight and promote data-driven management decision-making, which allows implementation challenges to be resolved in a timely manner. On the basis of the automation of data entry and reporting, resources can be invested more in behavior change communication campaigns and to deliver nets to hard-to-reach areas for effective, sustainable, and equitable net use [17,32]. The accurate household mapping and village-level population estimates can be important data at a local level that provide accurate denominator estimates for other public health services [40]. From the demand side, spatial analysis can enable campaign staff to select accessible DPs, so people spend less time traveling, and with more efficient operations to scan net cards quickly, people can also spend less time waiting to receive the nets. Furthermore, once clients' mobile phone numbers are registered during the household registration, server systems can send SMS text messages to clients about the proper and sustained use of LLINs [20,32]. Some previous studies in sub-Saharan African countries also introduced digital health applications in LLIN campaigns for demand creation strategies by using SMS to raise awareness and eVoucher systems to monitor net distributions and reduce fraud [41]. Overall, we expect ICT to contribute to improving health outcomes through more efficient and accountable processes to ensure that all nets reach the intended users.

Recommendations

As an ICT system holistically changes the overall program process and dynamics, rather than simply expediting the current process, it requires new roles and practices of the major stakeholders [42]. First, implementing organizations need to develop new operations and training manuals and policies that define new roles and management responsibilities of staff. These may include identifying who are going to be responsible for tracking and analyzing all the data that came in through the ICT system, what was the daily set of tasks or standard checks to review the data, and what specific actions were taken as a result of reviewing these data, and in what time frame. It may also require new process or performance indicators beyond net coverage to ensure better quality and equity. Second, international and national stakeholders should foster enabling conditions and collaborations for microplanning, supply chain management, and workforce capacity based on a shared data platform [18,19]. For effective supervision using visual dashboards, investment in and installation of monitoring screens for the LGA-level meetings and computer system (or tablets) for the DP managers and improving electricity and network connectivity would be important in local centers. Furthermore, integration of the data collected from mobile phones into the national health information system is important for evaluating the progress and impact of the program at scale [40,43,44]. Finally, the technical organization should continuously promote the improvement of optimized workflow automation, a well-functioning logic flow, user-centered interface design, and interactive data dashboard. Ultimately, it is critical to ensure and foster effective use of data at each level of management capacity and policy making [45].

Limitations

This study has several limitations. The study sites (6 LGAs) in Edo State were selected on the basis of operational convenience. As we could not conduct a similar field observation in the Kwara

State due to an operational delay, alternative approaches, such as informant interviews and documentation reviews, were used to understand the ICT system by staff involved in the ICT pilot, based on their knowledge of the Red Rose platform and experience testing the system in Kwara State. Thus, although this formative study illustrates how the ICT system can potentially and positively change current practices and processes, there may be new challenges in adopting and scaling up this new system. For example, intensive learning processes may be required for staff regarding new data entry methods, workflow, and new management responsibilities. In this regard, future studies can consider observation and qualitative studies during implementation and identify challenges or unintended consequences (eg, additional training or resource requirements, system integration, stakeholder coordination) through field observations during the adaptation and scaling-up process. Future studies may also use implementation research methods, such as the normalization process theory [46] or dynamic sustainability framework [47], to understand the factors and conditions that contribute to the adoption, scale-up, and sustainability of ICT-based systems. Future studies may measure explicit quantitative or qualitative comparisons of the performance between paper and ICT systems from the aspects of the organization (ie, efficiency: time or cost savings in program implementation), health workers (ie, satisfaction or accountability: timely compensation, productivity), and community (ie, equity: equitable net distribution).

Conclusions

To conclude, our study shows that high program coverage estimates in a paper-based system may not be a sufficient measure to ensure effective and sustainable universal coverage of LLINs, given the existing gaps in data reliability and timeliness, coupled with limited workforce capacity. The systematic functions and various technical features available in the ICT system can improve the impact and quality of LLIN campaigns by reducing staff time and resources in the data reporting system and processes. Accordingly, it can facilitate the measurement of true population coverage beyond program coverage with greater confidence and transparency. The ICT system can facilitate effective service delivery, promote workforce capacity and performance, and enhance the transparency and accountability of public health systems in resource-constrained settings. Continued efforts are necessary to develop new operations and training manuals and policies and to foster enabling conditions and collaborations for supply chain management and workforce capacity. Our findings and lessons may provide useful insights for developing, implementing, and evaluating future ICT systems for malaria vector control and other public health campaigns.

Acknowledgments

The long-lasting insecticide-treated net campaign program and this research were supported by grants from the Global Fund to Fight AIDS, Tuberculosis and Malaria. The funding organizations did not participate in the design of the study; the collection, analysis, and interpretation of data; or the writing of the manuscript. The findings and conclusions in this study are those of the authors and do not necessarily represent the official position of the Catholic Relief Services, the Catholic Church, or the Global Fund. The authors are thankful for the coordination and support of the Catholic Relief Services Nigeria team (Imran Chishtie, Kingsley Godson, Gbenga Jokodola, Uchechukwu Ikhimioya, Orezi Ajugba, Selbo Gambo, Vyvyenne Chigboh, and Adebayo Adeshola) and local government area focal persons (Esther Enabulele Ebose, Pat Aigbovia, Queen N Omorogbe, Emma-Esekhile Pat, Amu Rose Ireoba, and Edith Esohe Osula) and monitors (Ochoche Grace, Ufort Anietie John, John Oteikwu Lucia, Babavwemu Sweet, Bienose Julia, and Ogunkoya Abiola) who helped to document daily review meetings in Edo State. The authors also gratefully acknowledge Issa Kawu, John Hembling, Stephen Helen, and Scott LeFevre for their review and comments on the manuscript and study findings. Finally, we would also like to thank the partner organizations in this campaign for their support —National Malaria Elimination Program in Nigeria, Red Rose team (Hakan Buyubayrak), and Society for Family Health (John Ocholi, Akimlalu Titilope, and Dangana Christopher)—who shared their experience and insights for this study.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Major program implementation issues and resolution approaches of the paper-based system and information and communication technology–based system in Nigeria.
[DOCX File , 967 KB - formative_v5i10e23648_app1.docx]

References

1. World Health Organization. 2020 Nov 30. URL: https://www.who.int/publications/i/item/9789240015791 [accessed 2021-06-23]
2. The Alliance for Malaria Prevention: A toolkit for mass distribution campaigns to increase coverage and use of long-lasting insecticide-treated nets, 2012. The Alliance for Malaria Prevention. 2012. URL: https://allianceformalariaprevention.com/wp-content/uploads/2021/03/AMP-Toolkit-2.0-English.pdf [accessed 2021-06-23]

3. Okell LC, Griffin JT, Kleinschmidt I, Hollingsworth TD, Churcher TS, White MJ, et al. The potential contribution of mass treatment to the control of Plasmodium falciparum malaria. PLoS One 2011 May 24;6(5):e20179 [FREE Full text] [doi: 10.1371/journal.pone.0020179] [Medline: 21629651]
4. Stuckey EM, Stevenson J, Galactionova K, Baidjoe AY, Bousema T, Odongo W, et al. Modeling the cost effectiveness of malaria control interventions in the highlands of western Kenya. PLoS One 2014 Oct 07;9(10):e107700 [FREE Full text] [doi: 10.1371/journal.pone.0107700] [Medline: 25290939]
5. Richards JS, Arumugam TU, Reiling L, Healer J, Hodder AN, Fowkes FJI, et al. Identification and prioritization of merozoite antigens as targets of protective human immunity to Plasmodium falciparum malaria for vaccine and biomarker development. J Immunol 2013 Jul 15;191(2):795-809 [FREE Full text] [doi: 10.4049/jimmunol.1300778] [Medline: 23776179]
6. Teklehaimanot A, Sachs JD, Curtis C. Malaria control needs mass distribution of insecticidal bednets. Lancet 2007 Jun 30;369(9580):2143-2146. [doi: 10.1016/S0140-6736(07)60951-9] [Medline: 17597200]
7. Yukich JO, Lengeler C, Tediosi F, Brown N, Mulligan J, Chavasse D, et al. Costs and consequences of large-scale vector control for malaria. Malar J 2008 Dec 17;7:258 [FREE Full text] [doi: 10.1186/1475-2875-7-258] [Medline: 19091114]
8. Value for Money in LLIN Specifications Guidance: Ensuring highest value for money in LLIN procurement by balancing cost with end-user and program benefits. USAID - Result for Development. 2012. URL: http://www.r4d.org/wp-content/uploads/ValueForMoney-in-LLIN.pdf [accessed 2021-06-23]
9. Teklehaimanot A, McCord GC, Sachs JD. Scaling up malaria control in Africa: an economic and epidemiological assessment. Am J Trop Med Hyg 2007 Dec;77(6 Suppl):138-144. [Medline: 18165486]
10. Paintain LS, Kolaczinski J, Renshaw M, Filler S, Kilian A, Webster J, et al. Sustaining fragile gains: the need to maintain coverage with long-lasting insecticidal nets for malaria control and likely implications of not doing so. PLoS One 2013 Apr 10;8(12):e83816 [FREE Full text] [doi: 10.1371/journal.pone.0083816] [Medline: 24386283]
11. Aderibigbe SA, Olatona FA, Sogunro O, Alawode G, Babatunde OA, Onipe AI, et al. Ownership and utilisation of long lasting insecticide treated nets following free distribution campaign in South West Nigeria. Pan Afr Med J 2014 Apr 11;17:263. [doi: 10.11604/pamj.2014.17.263.3927] [Medline: 25309663]
12. Sipe NG, Dale P. Challenges in using geographic information systems (GIS) to understand and control malaria in Indonesia. Malar J 2003 Nov 04;2(1):36 [FREE Full text] [doi: 10.1186/1475-2875-2-36] [Medline: 14613511]
13. Nhavoto JA, Grönlund A. Mobile technologies and geographic information systems to improve health care systems: a literature review. JMIR Mhealth Uhealth 2014 Aug 05;2(2):e21 [FREE Full text] [doi: 10.2196/mhealth.3216] [Medline: 25099368]
14. Clements AC, Reid HL, Kelly GC, Hay SI. Further shrinking the malaria map: how can geospatial science help to achieve malaria elimination? Lancet Infect Dis 2013 Aug 13;13(8):709-718. [doi: 10.1016/S1473-3099(13)70140-3] [Medline: 23886334]
15. Kelly GC, Hale E, Donald W, Batarii W, Bugoro H, Nausien J, et al. A high-resolution geospatial surveillance-response system for malaria elimination in Solomon Islands and Vanuatu. Malar J 2013 Mar 21;12(1):108 [FREE Full text] [doi: 10.1186/1475-2875-12-108] [Medline: 23514410]
16. Wangdi K, Banwell C, Gatton ML, Kelly GC, Namgay R, Clements AC. Development and evaluation of a spatial decision support system for malaria elimination in Bhutan. Malar J 2016 Mar 22;15:180 [FREE Full text] [doi: 10.1186/s12936-016-1235-4] [Medline: 27004465]
17. Zöllner C, De Allegri M, Louis VR, Yé M, Sié A, Tiendrebéogo J, et al. Insecticide-treated mosquito nets in rural Burkina Faso: assessment of coverage and equity in the wake of a universal distribution campaign. Health Policy Plan 2015 Mar 30;30(2):171-180. [doi: 10.1093/heapol/czt108] [Medline: 24463333]
18. Masaninga F, Mukumbuta N, Ndhlovu K, Hamainza B, Wamulume P, Chanda E, et al. Insecticide-treated nets mass distribution campaign: benefits and lessons in Zambia. Malar J 2018 Apr 24;17(1):173 [FREE Full text] [doi: 10.1186/s12936-018-2314-5] [Medline: 29690873]
19. Arroz JA, Candrinho B, Pedro S, Mathe G, da Silva M, Tsabete S, et al. Planning and implementation of a countrywide campaign to deliver over 16 million long-lasting insecticidal nets in Mozambique. Malar J 2018 Jul 09;17(1):254 [FREE Full text] [doi: 10.1186/s12936-018-2406-2] [Medline: 29986711]
20. Janko MM, Churcher TS, Emch ME, Meshnick SR. Strengthening long-lasting insecticidal nets effectiveness monitoring using retrospective analysis of cross-sectional, population-based surveys across sub-Saharan Africa. Sci Rep 2018 Nov 20;8(1):17110 [FREE Full text] [doi: 10.1038/s41598-018-35353-z] [Medline: 30459359]
21. Dolan CB, BenYishay A, Grépin KA, Tanner JC, Kimmel AD, Wheeler DC, et al. The impact of an insecticide treated bednet campaign on all-cause child mortality: A geospatial impact evaluation from the Democratic Republic of Congo. PLoS One 2019 Feb 22;14(2):e0212890 [FREE Full text] [doi: 10.1371/journal.pone.0212890] [Medline: 30794694]
22. Nigeria Malaria Brief: Working with the global fund to reduce the burden of malaria in Nigeria. Catholic Relief Services. URL: https://www.crs.org/sites/default/files/tools-research/malaria_brief_2018_final.pdf [accessed 2021-06-23]
23. Edo - State in Nigeria. CityPopulation. URL: https://www.citypopulation.de/php/nigeria-admin.php?adm1id=NGA012 [accessed 2021-06-23]
24. Kwara - State in Nigeria. CityPopulation. URL: https://www.citypopulation.de/php/nigeria-admin.php?adm1id=NGA024 [accessed 2021-06-23]

25. Nigeria National Population Commission. URL: https://nationalpopulation.gov.ng/ [accessed 2021-06-23]
26. Kilian A, Koenker H, Paintain L. Estimating population access to insecticide-treated nets from administrative data: correction factor is needed. Malar J 2013 Jul 26;12:259 [FREE Full text] [doi: 10.1186/1475-2875-12-259] [Medline: 23890257]
27. Augustine Uchenna A. Quality micro-planning processes - Key strategy for achieving high quality LLINs Campaign: Pilot report from Cross-River State Nigeria. Euro J Prev Med 2019 May 10;7(2):32. [doi: 10.11648/j.ejpm.20190702.11]
28. Staedke SG, Gonahasa S, Dorsey G, Kamya MR, Maiteki-Sebuguzi C, Lynd A, et al. Effect of long-lasting insecticidal nets with and without piperonyl butoxide on malaria indicators in Uganda (LLINEUP): a pragmatic, cluster-randomised trial embedded in a national LLIN distribution campaign. Lancet 2020 Apr 18;395(10232):1292-1303 [FREE Full text] [doi: 10.1016/S0140-6736(20)30214-2] [Medline: 32305094]
29. Gama E, Were V, Ouma P, Desai M, Niessen L, Buff AM, et al. Large-scale implementation of disease control programmes: a cost-effectiveness analysis of long-lasting insecticide-treated bed net distribution channels in a malaria-endemic area of western Kenya-a study protocol. BMJ Open 2016 Nov 21;6(11):e012776 [FREE Full text] [doi: 10.1136/bmjopen-2016-012776] [Medline: 27872120]
30. Worrall E, Were V, Matope A, Gama E, Olewe J, Mwambi D, et al. Coverage outcomes (effects), costs, cost-effectiveness, and equity of two combinations of long-lasting insecticidal net (LLIN) distribution channels in Kenya: a two-arm study under operational conditions. BMC Public Health 2020 Dec 07;20(1):1870 [FREE Full text] [doi: 10.1186/s12889-020-09846-4] [Medline: 33287766]
31. Koenker H. More is more: Are we delivering enough LLINs? EClinicalMedicine 2018 Jul;1:5-6 [FREE Full text] [doi: 10.1016/j.eclinm.2018.07.005] [Medline: 31193635]
32. Girond F, Madec Y, Kesteman T, Randrianarivelojosia M, Randremanana R, Randriamampionona L, et al. Evaluating effectiveness of mass and continuous long-lasting insecticidal net distributions over time in madagascar: a sentinel surveillance based epidemiological study. EClinicalMedicine 2018 Jul 18;1:62-69 [FREE Full text] [doi: 10.1016/j.eclinm.2018.07.003] [Medline: 30294720]
33. Musa JJ, Moore SJ, Moore J, Mbuba E, Mbeyela E, Kobe D, et al. Long-lasting insecticidal nets retain bio-efficacy after 5 years of storage: implications for malaria control programmes. Malar J 2020 Mar 14;19(1):110 [FREE Full text] [doi: 10.1186/s12936-020-03183-y] [Medline: 32169081]
34. Koenker H, Arnold F, Ba F, Cisse M, Diouf L, Eckert E, et al. Assessing whether universal coverage with insecticide-treated nets has been achieved: is the right indicator being used? Malar J 2018 Oct 11;17(1):355 [FREE Full text] [doi: 10.1186/s12936-018-2505-0] [Medline: 30305127]
35. Khanam F, Hossain MB, Chowdhury TR, Rahman MS, Kabir M, Naher S, et al. Exploring the gap between coverage, access, and utilization of long-lasting insecticide-treated nets (LLINs) among the households of malaria endemic districts in Bangladesh. Malar J 2018 Dec 6;17(1):455. [doi: 10.1186/s12936-018-2610-0]
36. Galactionova K, Smith TA, de Savigny D, Penny MA. State of inequality in malaria intervention coverage in sub-Saharan African countries. BMC Med 2017 Oct 18;15(1):185 [FREE Full text] [doi: 10.1186/s12916-017-0948-8] [Medline: 29041940]
37. Guidelines for managing the malaria supply chain: a companion to the logistics handbook. USAID Global Health Supply Chain Program. 2017. URL: https://www.ghsupplychain.org/guidelines-managing-malaria-supply-chain [accessed 2021-06-23]
38. Malaria Online: The digital surveillance system for Thailand malaria elimination. Division of Vector Borne Diseases, Department of Disease Control, Ministry of Public Health, Thailand. 2019. URL: https://publicadministration.un.org/unpsa/Portals/0/UNPSA_Submitted_Docs/2019/3fe4c1ba-e00b-4250-8816-f513c3b209c6/2020%20UNPSA_Malaria%20online_full%20report_27112019_111848_f62725d9-bd75-4846-a44c-8524838f4e87.pdf?ver=1441-03-30-111848-927 [accessed 2021-06-23]
39. Bauer MS, Kirchner J. Implementation science: What is it and why should I care? Psychiatry Res 2020 Jan 22;283:112376 [FREE Full text] [doi: 10.1016/j.psychres.2019.04.025] [Medline: 31036287]
40. Rosewell A, Makita L, Muscatello D, John LN, Bieb S, Hutton R, et al. Health information system strengthening and malaria elimination in Papua New Guinea. Malar J 2017 Jul 05;16(1):278 [FREE Full text] [doi: 10.1186/s12936-017-1910-0] [Medline: 28679421]
41. Krezanoski PJ. Delivering insecticide-treated nets for malaria prevention: innovative strategies. Res Rep Trop Med 2016 May 07;7:39-47 [FREE Full text] [doi: 10.2147/RRTM.S83173] [Medline: 30050338]
42. Aidenagbon G, Adeusi O, Akilah J, Umar A. Improving quality through process assessments during mass distribution campaigns of Long Lasting Insecticidal Nets (LLINs) in Nigeria: Studies from four states. IJID Online. URL: https://www.ijidonline.com/action/showPdf?pii=S1201-9712%2814%2900706-1 [accessed 2021-06-23]
43. Ashton RA, Bennett A, Al-Mafazy A, Abass AK, Msellem MI, McElroy P, et al. Use of routine health information system data to evaluate impact of malaria control interventions in Zanzibar, Tanzania from 2000 to 2015. EClinicalMedicine 2019 Jun 21;12:11-19 [FREE Full text] [doi: 10.1016/j.eclinm.2019.05.011] [Medline: 31388659]
44. Ashton RA, Bennett A, Yukich J, Bhattarai A, Keating J, Eisele TP. Methodological considerations for use of routine health information system data to evaluate malaria program impact in an era of declining malaria transmission. Am J Trop Med Hyg 2017 Sep 01;97(3_Suppl):46-57 [FREE Full text] [doi: 10.4269/ajtmh.16-0734] [Medline: 28990915]

45. Kim S, Piccinini D, Mensah E, Lynch M. Using a human-centered design approach to determine consumer preferences for long-lasting insecticidal nets in Ghana. Glob Health Sci Pract 2019 Jun 24;7(2):160-170 [FREE Full text] [doi: 10.9745/GHSP-D-18-00284] [Medline: 31249018]
46. May CR, Cummings A, Girling M, Bracher M, Mair FS, May CM, et al. Using Normalization Process Theory in feasibility studies and process evaluations of complex healthcare interventions: a systematic review. Implement Sci 2018 Dec 07;13(1):80 [FREE Full text] [doi: 10.1186/s13012-018-0758-1] [Medline: 29879986]
47. Chambers DA, Glasgow RE, Stange KC. The dynamic sustainability framework: addressing the paradox of sustainment amid ongoing change. Implement Sci 2013 Oct 02;8:117 [FREE Full text] [doi: 10.1186/1748-5908-8-117] [Medline: 24088228]

Abbreviations

CRS: Catholic Relief Services
DP: distribution point
GIS: geographic information system
ICT: information and communication technology
LGA: local government area
LLIN: long-lasting insecticide-treated net
NMEP: National Malaria Elimination Program
QR: quick response

Edited by G Eysenbach; submitted 19.08.20; peer-reviewed by L Jordan, T Turk; comments to author 11.03.21; revised version received 29.06.21; accepted 27.07.21; published 08.10.21.

Please cite as:
Jo Y, Barthel N, Stierman E, Clifton K, Pak ES, Ezeiru S, Ekweremadu D, Onugu N, Ali Z, Egwu E, Akoh O, Uzunyayla O, Van Hulle S
The Potential of Digital Data Collection Tools for Long-lasting Insecticide-Treated Net Mass Campaigns in Nigeria: Formative Study
JMIR Form Res 2021;5(10):e23648
URL: https://formative.jmir.org/2021/10/e23648
doi:10.2196/23648
PMID:34623310

Original Paper

Conducting Digital Health Care Research: Document Analysis of Challenges Experienced During Intervention Development and Feasibility Study Setup of an Internet-Administered Intervention for Parents of Children Treated for Cancer

Joanne Woodford[1], PhD; Mathilda Karlsson[1], BSc; Josefin Hagström[1], MSc; Ylva Hägg Sylvén[1], MSc; Kajsa Norbäck[1], MSc; Helena Grönqvist[1], PhD; Louise von Essen[1], PhD

Healthcare Sciences and e-Health, Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden

Corresponding Author:
Louise von Essen, PhD
Healthcare Sciences and e-Health
Department of Women's and Children's Health
Uppsala University
MTC House
Dag Hammarskjölds väg 14B
Uppsala, 752 37
Sweden
Phone: 46 704250714
Email: louise-von.essen@kbh.uu.se

Abstract

Background: The design and conduct of research to develop, test, and evaluate complex health care interventions is challenging. Although the existing literature describes key challenges associated with the design and conduct of definitive (evaluation) trials, there is a lack of information concerning specific challenges associated with the intervention development phase and setup of feasibility studies. In particular, the literature is scarce concerning the challenges associated with conducting digital health care research, such as research on internet-administered interventions and research using digital features to support the execution of study procedures (eg, recruitment, consent, retention, and data collection and management). This study is conducted in the context of the intervention development and feasibility study setup phases of an internet-administered, guided, low-intensity cognitive behavioral therapy–based intervention for parents of children previously treated for cancer.

Objective: The aim of this study is to explore the challenges experienced during the development phase of the internet-administered intervention and digital features to support the execution of the study procedures and a feasibility study setup.

Methods: To explore the key challenges experienced, we conducted a document analysis of written records from all study meetings held by the research team (meeting minutes) between June 7, 2018, and January 10, 2020, guided by a thematic analysis approach. Furthermore, discussion groups with members of the research team were held to develop a more detailed understanding of the key challenges experienced. Methods and results are reported in accordance with the relevant items from the Standards for Reporting Qualitative Research checklist.

Results: Six main themes were identified: decision-making and communication, expertise, external constraints, flexibility, planning and scheduling, and technical constraints.

Conclusions: Significant challenges were experienced during the intervention development and setup phases of the feasibility study. Implications are discussed to inform future design, conduct, and planning of internet-administered intervention development and feasibility studies, especially within the context of digital health care research.

(JMIR Form Res 2021;5(10):e26266) doi:10.2196/26266

KEYWORDS

parents; internet-administered cognitive behavior therapy; low intensity CBT; feasibility study; challenges; digital healthcare research

Introduction

Background

Clinical trials are essential to inform evidence-based health care [1]. However, clinical trials are costly and resource-intensive for both researchers and funders [2], and approximately 85% of research investment is wasted [3]. Examples of research waste include asking the wrong research questions [4], using inappropriate study designs and methods [5,6], poor and biased reporting [7], and underreporting [8]. There are a number of inefficiencies related to clinical trial conduct that can lead to research waste, especially given that the design and efficient conduct of clinical trials is challenging [9] and operationally complex [10,11]. Challenges include the creation and management of trial procedures and materials; communication with multiple stakeholders; ethical and regulatory requirements; the recruitment, training, and turnover of trial personnel; budget management; the recruitment and retention of trial participants; and data monitoring and assurance of data quality [10-14].

While numerous barriers have been identified in the successful conduct of clinical trials [12], there has been less focus on sharing experiences and lessons learned by trial teams [10]. Publications tend to focus on trial outcomes, with little reporting on trial conduct [11]. Consequently, there is scarce evidence to inform decisions concerning the management of clinical trials [1]. Furthermore, in contrast to the literature focused on the conduct of definitive (evaluation) trials, there is a lack of literature concerning the challenges of conducting intervention development research and pilot and feasibility studies, following the Medical Research Council (MRC) framework for the development and evaluation of complex interventions [15]. Indeed, there is limited guidance on how to design and conduct feasibility studies, leading to poor study design and reporting [16,17]. Given that many preparatory development [18,19] and clinical, methodological, and procedural uncertainties require testing [20] before progressing to a definitive trial, development and feasibility phases are substantial works in and of themselves, with findings often underreported [16,17].

In the context of this feasibility study (ENGAGE study: ISRCTN 57233429; ISRCTN 18404129) [21], the internet-administered low-intensity cognitive behavioral therapy (LICBT) intervention (the EJDeR intervention) is delivered on the U-CARE-portal, hereafter referred to as the Portal [22]. The Portal is designed to deliver internet-administered cognitive behavioral therapy interventions and support the execution of study procedures, for example, randomization, web-based informed consent, and data collection [23]. Digital technologies (eg, technologies using the internet) are facilitating the delivery of health care worldwide [24,25]. Indeed, internet-administered interventions have been posited as a solution to the global mental health crisis and to help overcome significant barriers to treatment access (eg, geographical and resource barriers) [26]. The promise of digital technologies to deliver mental health care interventions has been further amplified by the current COVID-19 pandemic [27], given the negative psychosocial consequences of the pandemic itself [27,28] and the ability to facilitate access to mental health care [29]. However, while the evidence base for internet-administered psychological interventions is well-established [30], many publicly available interventions delivered via digital technologies are not evidence-based [31].

Furthermore, digital technologies are being increasingly incorporated into the design and execution of health care research, for example, to facilitate recruitment, enhance retention, and collect data [24,25,32,33]. The use of digital technologies in health care research has intensified during the COVID-19 pandemic in an attempt to continue health care research in the absence of in-person contact [34]. Digital health care research is associated with reduced trial costs, improved trial efficiency [35,36], and recruitment of more diverse populations [25,34]. As such, the use of digital features to execute health care research is likely to continue to grow beyond the pandemic [34]. However, at present, challenges related to the conduct of digital health care research are less well documented [25,32] and focus on topics such as patient privacy and confidentiality, adequate infrastructure, data accuracy and integrity, and user acceptability [25,37,38]. Some challenges are beginning to be addressed by the development of new ethical and regulatory standards and increased provision of guidance for investigators in the use of digital technologies [25,37,38]. However, the literature remains in its infancy, and there is a need for researchers to publish their experiences in conducting digital health care research [25].

Context: The ENGAGE Feasibility Study

Globally, approximately 300,000 children are diagnosed with cancer each year [39], and cancer remains a leading cause of death in children worldwide [40]. Typically, parents are the primary source of support for children with cancer and report significant negative psychological [41-43] and socioeconomic impacts [44-47]. Mental health difficulties are reported after cancer treatment [42,43,48] and years after the end of treatment [42,49]. However, parents of children treated for cancer report an unmet need for psychological support [50-52]. To improve access to evidence-based psychological support, innovative solutions are being developed worldwide [53]. One such solution is the provision of guided internet-administered LICBT, which may help improve access to psychological support for parents of children treated for cancer.

Given the promise of internet-administered LICBT, we have adopted the MRC complex interventions framework [15] to develop an internet-administered LICBT intervention (the EJDeR intervention) tailored to the specific needs of parents of children previously treated for cancer [54]. Significant previous research [41,42,49,55-57] has informed the development of the EJDeR intervention alongside multiple stakeholders, including parent research partners (PRPs), clinical psychologists, software developers, and pediatric oncologists. The EJDeR intervention is delivered on the Portal and includes written, film, audio content, videoconferencing, and in-portal email guidance from an e-therapist [54].

The objectives of this study are to explore the challenges experienced during (1) the development phase of the internet-administered intervention and digital features to support the execution of the study procedures and (2) a feasibility study setup [21,58]. To explore the key challenges experienced, we

conducted a document analysis of written records from all study meetings held by the research team (meeting minutes) between June 7, 2018, and January 10, 2020, guided by a thematic analysis approach. Furthermore, discussion groups with members of the research team were held to develop a more detailed understanding of the key challenges experienced.

Methods

The methods and findings are reported in accordance with relevant items from the Standards for Reporting Qualitative Research checklist [59].

Qualitative Approach and Research Paradigm

A document analysis [60] of study meeting minutes was guided by a thematic analysis approach [61]. Document analysis was considered suitable given that it allows an examination of contextual and background information to provide historical insights into the challenges experienced by the ENGAGE research team and provides a way of tracking challenges over time [60].

Researcher Characteristics and Reflexivity

Document analysis was primarily conducted by 2 members of the research team (MK and JW). MK is a female research assistant with a Bachelor of Sport Science who joined the research team toward the end of the study setup phase and was therefore able to conduct the analysis from an *outsider* perspective. MK was trained in using thematic analysis by JW. JW is a female researcher with a PhD in psychology with experience in conducting qualitative research. JW is a coinvestigator, supervisor of the study coordinator, and research assistant in the research team and has been a member of the research team from the beginning of the study setup phase. All manuscript authors are, or have been, members of the research team.

Context

All meeting minutes were taken by the ENGAGE study coordinator, or a substitute, and meetings were held at the Department of Women and Children's Health, Uppsala University, Uppsala, Sweden. In most cases, all core members of the research team were in attendance, including the principal investigator (LVE), researchers (JW and HG), study coordinator (KN), Portal development team coordinator (YHS), and research assistants (JH and MK). Occasionally, wider research team members, such as software developers, licensed psychologists, and student interns, attended meetings. Meetings were held weekly and scheduled for 2 hours. Meeting minutes were circulated to research team members for approval after the meeting and saved on a shared folder and thus visible to all research team members.

Ethical Issues

Ethical approval was not deemed necessary as the study analyzed documentary data only.

Data Collection

All meeting minutes (N=78) from ENGAGE study meetings conducted between June 7, 2018, and January 10, 2020, were included to account for meetings concerning the development of the intervention, the development of digital features to support study execution, and the concurrent setup of the feasibility study. The number of meeting minute pages ranged from 1 to 6 and the word count range was 85-2350. Meeting minutes were designed to document (1) progress toward specific study milestones, (2) problems or challenges arising, (3) decisions, and (4) actions moving forward.

Data Analysis

Given that documents cannot be considered to be completely accurate recordings of past events [60], throughout the analysis process, MK and JW actively reflected upon the meaning of the meeting minute content and how this was related to challenges experienced by the research team. Guidelines for conducting a reflexive thematic analysis followed [61]. An inductive approach was adopted with codes and themes driven by the data [62]. The analysis of meeting minutes took place between January and May 2020. Meeting minutes were read multiple times by MK to enable familiarization with the data set as a whole. MK identified initial codes across the data set to begin organizing the data. To enhance rigor, initial codes were discussed in weekly meetings with JW. After initial coding across the data set was complete, MK sorted initial codes into initial main themes and subthemes within main themes, with continued weekly discussions with JW to establish initial consensus and refine the main themes and subthemes [61]. Next, the initial thematic map (Figure 1) was presented to the research team (HG, JH, JW, KN, LVE, MK, and YHS) for feedback in a face-to-face discussion group held on February 14, 2020 (120 minutes). During the discussion group, the initial thematic map was presented, and discussions were held concerning whether the initial main themes and subthemes reflected the main challenges experienced. One major discussion concerned whether the Portal was a main theme or an element of all the main themes.

Subsequently, a refined thematic map (Figure 2) was developed by MK with continued weekly discussions with JW and applied across the data set to ensure a good fit to the data. A second face-to-face discussion group (120 minutes) was held on March 5, 2020, to present the refined thematic map to the research team (HG, JH, JW, KN, LVE, MK, and YHS) for further discussion and feedback.

On the basis of feedback received in the second discussion group, the main themes and subthemes were further revised by MK and JW, and a final thematic map was developed with six main themes and no subthemes (Figure 3). The thematic map, alongside descriptions of each main theme, was sent to the wider research team via email for final feedback and approval, including the provision of salient examples of challenges missing from descriptors.

Figure 1. Initial thematic map. Ovals represent main themes and rectangles represent subthemes within main themes. Solid gray lines illustrate connections between main themes and subthemes. Dotted green lines illustrate where main themes and subthemes overlap with other main themes and subthemes.

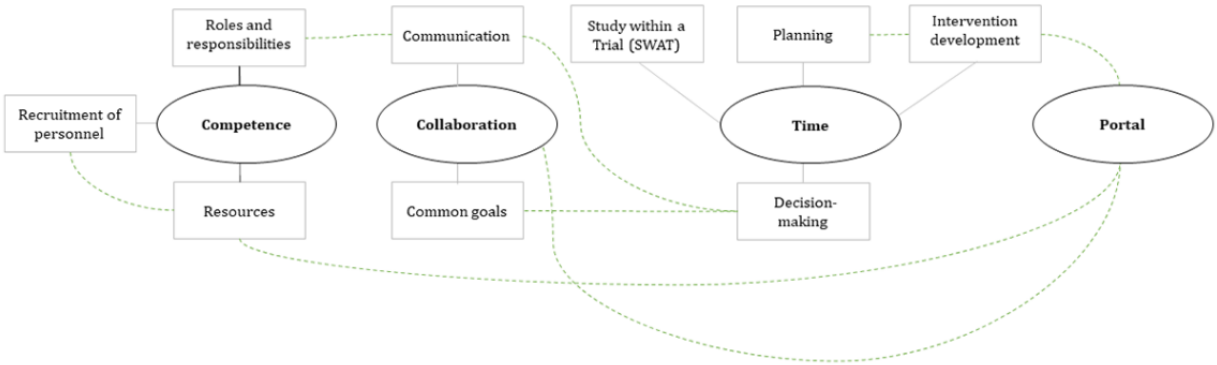


Figure 2. Refined thematic map. Ovals represent main themes and rectangles represent subthemes within each main theme. Solid lines illustrate connections between main themes and subthemes.

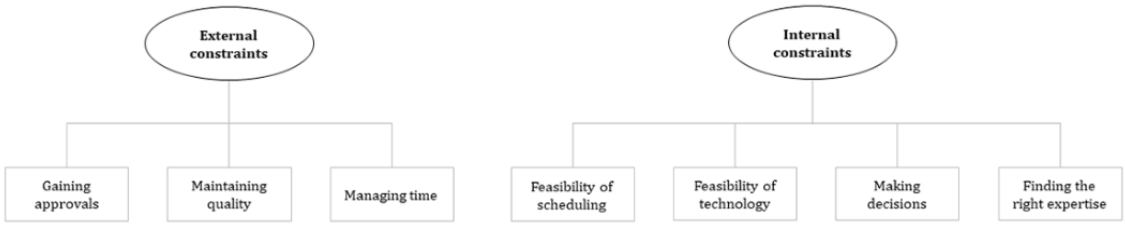


Figure 3. Final thematic map.

Trustworthiness

To enhance the trustworthiness of the analysis, a number of strategies were adopted as follows: (1) member-checking with research team members to test that the main themes and detailed descriptors were recognized [63], (2) use of thematic maps to explore the main theme and subtheme connections during the analysis process, (3) repeated returning to the whole data set to check the adequacy of each thematic map, and (4) member-checking and (5) peer debriefing with initial codes and subsequent thematic maps reviewed and discussed by MK and JW throughout the analysis process.

Results

Overview

The thematic analysis resulted in a three-step process, an initial thematic map (Figure 1), a refined thematic map (Figure 2), and a final thematic map (Figure 3). Textbox 1 presents the content of the final main themes.

Textbox 1. Final themes and descriptors of challenges.

Descriptors of challenges identified in main themes

- Decision-making and communication
  - Involving all members of the research team in making certain decisions. Considering the different preferences of research team members
  - Communicating between the research team and the portal development team concerning technical requirements, time estimation, and planning
- Expertise
  - Identifying and recruiting study personnel with the necessary expertise and experience to design and set up the feasibility study
  - External constraints
  - Uncertainties regarding time taken for tasks to be completed by contracted personnel outside of the research team and time taken for public authorities to make decisions and provide approvals
- Flexibility
  - Changing intervention requirements, elements of the study protocol, and associated study documentation
- Planning and scheduling
  - Identifying required tasks to complete
  - Estimating time to complete tasks
  - Deciding the order in which tasks should be completed, especially concerning task interdependency
- Technical constraints
  - Managing within the constraints of available technical resources and known limitations of the Portal
  - Identifying necessary technical requirements on the Portal

Decision-making and Communication

Decision-making was a significant challenge during the development of the intervention and digital features to support study execution, and the feasibility study setup phase, resulting in time inefficiencies and a delay to study start. A particular challenge was related to occasions in which the entire research team was involved in the decision-making process. Group decision-making resulted in the time taken for the principal investigator or a researcher to make a final decision being longer than was always necessary. A tension was identified concerning the desire to build and maintain a supportive team environment versus the need for research team members with more experience and leadership responsibilities to make quick decisions. This resulted in examples from the meeting minutes of the group decision-making process taking weeks, or sometimes months, before a final decision was made by the principal investigator. Group decision-making was particularly challenging when research team members had different preferences, opinions, and different levels of experience and expertise, resulting in difficulties in reaching a consensus. One example was deciding on a name for the intervention, whereby several opinion polls were made over a number of weeks before a final decision was made.

Communication was also identified as a challenge that contributed to difficulties in the decision-making process. Challenges regarding communication were particularly exacerbated in situations where research team members had different areas of expertise. A prevalent example concerned communication between researchers and the Portal development team. For example, during the development of the intervention and digital features to support study execution, the research team had many technical requirements. However, sometimes, there was a mismatch between researchers and the Portal development team in their understanding of technical requirements (eg, how interactive homework exercises were presented on the Portal). Furthermore, difficulties were identified concerning time estimation and subsequent planning. For example, the Portal team at times underestimated the length of time to deliver a technical requirement, or technical requirements were changed by the research team, subsequently impacting the study time plan. In addition, researchers sometimes held unrealistic expectations concerning delivery time for a new requirement, negatively impacting the study time plan. Difficulties with communication were more prevalent in the study setup phase. During the study setup phase, a software developer from the Portal team began to attend research team meetings, helping to avoid misunderstanding and clarifying requirements.

Expertise

Identifying personnel with the required research and clinical expertise was challenging. One challenge related to the recruitment of research team personnel with the correct competencies and experience, for example, prior experience of trial management, design and conduct of feasibility studies, and/or experience of internet-administered intervention development. For example, multiple recruitment rounds were held to recruit a postdoctoral researcher to work full-time in the study; however, no candidates with the correct expertise were

found. In addition, retention of research team personnel (eg, research assistants) was raised as a difficulty, for example, staff turnover was experienced due to short-term temporary contracts and the uncertainty of future employment being tied to study funding. As such, difficulties were experienced needing to replace temporary personnel and provide training to new research team members, taking significant time and further impacting the study time plan. This was especially problematic as it was difficult to recruit research assistants with prior experience of working in similar research environments, and thus the need for training was understandably greater.

Another challenge related to expertise concerned the further development and refinement of the intervention material. Difficulties were experienced in identifying a licensed psychologist with specific experience in writing LICBT interventions in Swedish to finalize the intervention material. Ultimately, 2 English-speaking experts in LICBT were engaged in later iterations of the intervention material after feedback from PRPs. However, this resulted in the need for translation into Swedish, back-translation into English, and cultural adaptation of the intervention material, which further delayed intervention development.

Challenges were also experienced when expectations for the required expertise changed. One example is related to illustrations for the EJDeR intervention, where illustrations were initially developed by a research team member with some experience in using design software. However, it was later decided to engage an external company with specialist experience of developing illustrations for websites and mobile devices. This decision was made to enhance the quality of the EJDeR intervention alongside considerations concerning resource allocation, as illustration design is time consuming and the skillset of the member of the research team was better used elsewhere.

External Constraints

During the development of the intervention and digital features to support study execution and feasibility study setup phases, the research team liaised with public authorities (eg, the Swedish Ethical Review Authority, the Childhood Cancer Registry, and the Swedish Tax Agency), external companies (eg, professional illustrators), and external contractors (eg, licensed psychologists and e-therapists). One challenge relates to the time taken by public authorities to make decisions and provide approval. A specific example related to the submission of an ethical amendment to the Swedish Ethical Review Authority regarding the implementation of a study within a trial [58] embedded in the ENGAGE feasibility study. The study within a trial was approved; however, the authority raised concerns regarding previously approved parts of the study design (eg, an opt-out recruitment procedure involving telephone reminders to parents who do not actively decline study participation). Raising concerns regarding a previously approved application was potentially because the Swedish government replaced regional ethical review boards with a single national review authority during the study setup phase, meaning applications and requests for amendment were no longer necessarily reviewed by the same local authority. This resulted in additional delays, with the research team needing to respond to the authority to further justify the use of an opt-out procedure. Additional delays were experienced by external companies and contractors, for example, not working at full capacity during the summer of 2019 (eg, taking a summer vacation). Another delay related to external factors was experienced when Uppsala University upgraded their servers, resulting in members of the Portal team being allocated to mitigate risks to the Portal during the upgrade and thus unable to prioritize technical requirements for the study and intervention.

Flexibility

A further challenge was related to the development of the intervention and digital features to support study execution and the feasibility study setup taking place concurrently, and the subsequent need for flexibility. For example, during the intervention development phase, multiple changes and continued improvements were made to the intervention content, language, and overall design (eg, font, color palette, logo design, professional illustrations, and layout). However, any change to language and design had a subsequent impact on other study components, for example, participant information sheets. It was perceived that there were many interrelated *moving parts* and any detail changed in the intervention or feasibility study resulted in a *snow-ball effect* on multiple other intervention and study components. A further example relates to study documentation. While refining the intervention, specific study terminology was still under development, for example, intervention name, e-therapists being referred to in the study as *parent guides*, the structure of the intervention, module, and chapter titles. During the feasibility study setup phase, study documentation was under development (eg, case report forms, data management, SMS text message and email reminder protocols and reminder content, standard operating procedures, and the study handbook). However, each terminology change resulted in revisions to all study documentation. Another example related to the study information video, which had to be rerecorded after numerous changes were made to the study information sheets based on changes to terminology used in the intervention. Research team members raised that these challenges seemed related to setting up a feasibility study when the intervention had not yet been finalized. The research team needed to be flexible and adaptable during this process and keep track of any detail and subsequent impact on other elements of the intervention and feasibility study.

Planning and Scheduling

Another challenge is related to the overall timeframe and associated planning and scheduling of tasks. For example, given the aforementioned challenges with delays in intervention development, meeting minutes documented difficulties scheduling specific events, such as training e-therapists, or presenting the intervention to PRPs for feedback. On several occasions, e-therapist training was rescheduled due to delays in intervention development; for example, finalizing the written intervention material and some technical features, such as videoconferencing, taking longer than initially anticipated to develop and test. Furthermore, time for task completion was often underestimated, which was discussed at times as being

related to expertise. This could be observed by how several research team personnel had no prior experience of working within a similar research environment, resulting in some tasks taking longer than anticipated. Planning and scheduling were also described as generally difficult given the interrelatedness of intervention development and feasibility study setup tasks. For example, any changes made to the intervention affected the entire *planning chain.*

A further complexity regarding planning and scheduling related to the different project management tools used by the research team versus the Portal team. For example, the research team used a Gantt chart, mainly focused on research-associated tasks, alongside Microsoft To Do, to enable task breakdown and allocation of tasks between research team members. However, the Portal team used Atlassian Jira Software Server Version 8.5.1, a software development tool used by agile teams to plan, track, and release software features. At times, this was experienced as a challenge as while the research team tracked overall technical feature development in the Gantt chart, this was at times out of sync with more detailed software development planning in Jira Software Server. During the study setup phase, the developer team began to report updates to the research assistant responsible for the Gantt chart each week, which improved planning. However, research team members discussed how a more integrated solution may have facilitated scheduling and planning, especially because of the detail and complexity involved in the development of new digital features for both the intervention and support study execution.

Technical Constraints

A challenge frequently mentioned in meeting minutes pertained to the availability of technical resources, such as the Portal team's time and the existing functionality of the Portal. The existing Portal did not have all the digital features required to deliver the internet-administered intervention or to support the execution of the feasibility study. Meeting minutes documented numerous new or adjusted intervention features, for example, (1) color palette change, (2) e-therapist notifications via email and SMS text message when participants send internal messages or submit chapters or homework exercises, (3) font selection, (4) intervention display to allow e-therapists to tailor the intervention, (5) carrousel feature to enhance library navigation, (6) printable PDF documents of homework exercise, (7) tab-based intervention view, and (8) videoconferencing. Meeting minutes also document new digital features required to support study execution, for example, (1) character limitation removed for SMS text messages sent via the Portal; (2) customized study registration process, including the use of recruitment ID to identify the source of recruitment; (3) individual preferences for reminders (eg, email, SMS text message, post, or telephone); (4) new home page to facilitate study sign-up and log-in for existing participants; (5) newsletter scheduling based on parents' progress in the study; (6) opt-out procedure for participants declining study participation; (7) personalization of reminders (eg, use of first name); (8) reporting features on intervention use, newsletters, reminders, and suicide alerts; and (9) study-specific technical help-texts throughout the Portal.

Some new digital feature requirements were known from study conception (eg, included in the grant application and study protocol), for example, the newsletter, individualized reminders, and opt-out procedure. However, other digital requirements, especially in relation to the intervention, were not planned and arose during the intervention development phase. During the discussion groups, it was raised that as it was very difficult to know all Portal requirements in advance, there was a need for good communication between the research team and Portal team. Some research team members discussed how their lack of prior knowledge of the Portal, especially in relation to how interventions were delivered, resulted in underestimations or differing expectations concerning the amount of technical and aesthetic changes required. Furthermore, the research team did not always have a clear understanding of how much work might be involved in developing a new technical feature or changing an existing feature. Meeting minutes also listed several occasions when new digital features, or requested changes to existing digital features, took longer than anticipated to develop and test. One reason for the difficulties in time estimations is related to the existing software architecture of the Portal. As such, the development of new digital features, or making changes to existing digital features, could have unintended or unanticipated consequences on other existing features on the Portal, and subsequently impact other studies running on the Portal.

Discussion

Principal Findings

This paper describes the challenges experienced during the development phase of the internet-administered intervention and digital features to support the execution of the study procedures and setup of the ENGAGE feasibility study. To summarize the main findings, a document analysis of meeting minutes adopting a thematic analysis approach and subsequent research team discussions resulted in six main themes as follows: (1) decision-making and communication, (2) expertise, (3) external constraints, (4) flexibility, (5) planning and scheduling, and (6) technical constraints.

Comparison With Prior Work

To provide an overview of prior work exploring challenges experienced in traditional and digital health care research and software development projects, we present each of the main themes and descriptors of challenges identified alongside similar research findings from others (Table 1). We included 10 publications [64-73] in Table 1, which have not yet been mentioned.

Table 1. Final main themes, descriptors of challenges identified, and a summary of similar research findings.

Main themes	Descriptors of challenges identified in main themes	Similar research findings from others
Decision-making and communication	• Involving all members of the research team in making certain decisions • Considering the different preferences of research team members • Communicating between the research team and Portal development team concerning technical requirements, time estimation, and planning	• Challenges in communicating between the principal investigator and the research team [9] • Cultural interference in group decision-making [64,65] • Insufficient time from the principal investigator [9] • Lack of arena for solving conflict [66] • Misunderstanding project goals [67] • Difficulties with communication between software personnel and nontechnical personnel (eg, different perceptions, knowledge, and experience) [66,68]
Expertise	• Identifying and recruiting study personnel with the necessary expertise and experience to design and set up the feasibility study	• Limited access to project personnel with the relevant expertise and knowledge [9,25,37,68] • Poor retention of personnel [69] • Time needed to train personnel [9,25]
External constraints	• Uncertainties regarding time taken for tasks to be completed by contracted personnel outside of the research team and time taken for public authorities to make decisions and provide approvals	• Time taken for ethical and regulatory approval from public authorities [69,70] • Lack of adequate regulatory and legal guidance [23,25,37,38]
Flexibility	• Changing intervention requirements, elements of the study protocol, and associated study documentation	• Adaptations needed to allow digital tools to be used in research [25,36] • Complexities of study documentation development [9] • Lack of understanding of requirement complexity [68]
Planning and scheduling	• Identifying required tasks to complete • Estimating time to complete tasks • Deciding the order in which tasks should be completed, especially concerning task interdependency	• Conflicting priorities [68] • Complexities of forecasting and planning project budgets [71] • Poor or unrealistic project planning [9,68] • Need to involve multiple stakeholders in the planning [25,37,38] • Time taken to develop high-quality documents and data collection tools [68] • Use of collaboration tools [66]
Technical constraints	• Managing within the constraints of available technical resources and known limitations of the Portal • Identifying necessary technical requirements on the Portal	• Shared an understanding of software requirements [23,66,68] • Data privacy and security [23-25,37,38,72,73] • Access to adequate infrastructure [23,25,37,38]

One interesting challenge was identified in relation to the group decision-making process affecting the study timeframe. Research concerning cultural expectations and decision-making indicates that, within the Swedish culture, there are high expectations for shared authority and decision-making between managers and personnel [64]. Generally, the workplace is less hierarchical with decisions and responsibilities shared and decisions made in larger groups [65]. However, research suggests that more unilateral or directive decision-making may be required when there are significant time pressures and critical deadlines and with more novice team members [74]. While adopting group decision-making processes may help to maintain good relationships and is more in line with cultural expectations, this was at tension with the need for more unilateral and directive decision-making, given the time critical nature of the research. Indeed, one reason for failing clinical trials may pertain to a lack of structured *business-like* trial management [12].

A related challenge concerned the difficulties in recruiting experienced study personnel. Difficulties with recruitment and inadequate training of study personnel have been identified in the literature as a key inefficiency in successful trial delivery [9]. Despite multiple recruitment attempts, it was not possible to recruit a postdoctoral researcher with the required competencies to work on the study. However, competent trial management by a dedicated trial manager responsible for day-to-day operations is essential for efficient trial conduct. Indeed, once funding is awarded, the trial manager, rather than the principal investigator and coinvestigators, has been suggested to be the most important research team member to successfully deliver a clinical trial [75]. Furthermore, it is suggested that some trials fail not due to the study design but rather problems with trial management [76,77]. In the context of the ENGAGE feasibility study, no research team member working on the study on a full-time basis had experience of similar research environments and advanced training in research methodology. Those in senior roles with advanced training and expertise in research methodology (eg, the principal investigator and coinvestigator) were key to organizing, planning, and decision-making; however, their time was spread across multiple studies and competing responsibilities. This *light* project management approach has been demonstrated to lead to team

members' expectations being unfulfilled and commitment of the team may decrease [78].

In countries such as the United Kingdom, there has been an increased emphasis on efficient clinical trials to reduce research waste, for example, by the establishment of the UK Trial Managers' Network for trial managers on academic-led noncommercial trials. To the best of our knowledge, there is no specific trial management professional career structure within academic settings in Sweden, which may account for some of the difficulties recruiting a postdoctoral researcher with the required competencies. An added complexity in the Swedish context is the employment projection law (LAS § 5a 1982:80), which stipulates that personnel can be employed on a fixed-term contract for a maximum of 2 years and must be offered a permanent position thereafter to remain in employment. However, given the time-limited nature of health care research and uncertainty concerning continued funding, this often results in research personnel only being employed for a maximum 2-year fixed-term contract. Subsequently, it is difficult to maintain experienced research staff for the duration of the trial funding periods. However, even in countries such as the United Kingdom, where trial management is more established within academic settings, barriers are experienced, such as a lack of clarity around career structures, progression, and professional status, lack of training opportunities, lack of professional accreditation, and staff retention due to lack of funding and instability of short-term contracts [11,12,79].

A further key challenge related to both the development phase of the internet-administered intervention and digital features to support the execution of study procedures and setup of the feasibility study, occurring concurrently. Complex health care intervention development and refinement involves a number of iterative and interacting stages [19]. The development of interventions using internet-based technologies is complex, and specific challenges have been identified in relation to their development (eg, iterative development life cycles, relationship between academics and developers, and characterization of intervention components and essential features) [80,81]. Furthermore, feasibility studies are complex and can involve a number of iterative phases and the testing of multiple procedural, methodological, and clinical uncertainties [17]. In the present context, a number of challenges were related to the interrelatedness of the intervention and subsequent feasibility study, for example, further development, refinement, and adaptation of the intervention resulted in multiple changes to associated documentation for the feasibility study. The development of high-quality study documentation is a difficult and lengthy process [9], and the focus on intervention development meant allocating resources away from the preparation of study documentation. As such, working on intervention development concurrent with the feasibility study setup may have added to the complexity of already complex research processes.

In addition, while a significant phased approach to intervention development had already taken place [41,42,49,54], a structured approach was not adopted regarding the technical development of the intervention. Given the complexity of the development of internet-administered interventions [80,81], it may have been beneficial to adopt a specific development model designed to inform the development of health care interventions using internet-based technologies, such as the Behavioral Intervention Technology model [82]. Adopting a more structured approach to technical development may have facilitated the communication of technical requirements to the Portal team. Indeed, challenges relating to communication between the research and Portal team may have been experienced, given that requirements were requested from those unfamiliar with software programming, and software developers are unfamiliar with research methodologies and LICBT interventions [23]. The Portal team adopts agile software development, therefore working with software requirements in short iterations [83] with a need for frequent, quick, and short-term decisions [84]. However, this approach can cause challenges when there is a need for collaborative decision-making with multiple stakeholders with varying backgrounds, expertise, and goals [66]. Barriers to efficient communication and development of shared knowledge may be related to a lack of understanding of each other's field (eg, different educational backgrounds, technical knowledge, and experience), difficulties with clients describing functional requirements, and difficulties with software developers communicating time expectations with clients in a direct manner [68]. A lack of shared understanding between software development teams and end-clients can result in unrealistic planning and frequent changes in planning [66]. In addition, transforming study procedures normally conducted in-person to digital form is a complex procedure requiring significant planning, time, and expertise [25,36]. However, involving a member of the Portal team in the weekly research team meetings significantly improved the development of shared knowledge and understanding and helped to overcome some of the aforementioned challenges. Indeed, more active engagement between software developers and the end client has been posited as a helpful strategy to facilitate communication [68].

Another noteworthy challenge is related to the constraints of the Portal. Health care interventions delivered by digital technologies have evolved rapidly over the past two decades [85]. However, this also means that technology becomes quickly outdated [86], given the technological advancements and changing end-user expectations. However, rapid technical evolution is not easily compatible with traditional research approaches such as randomized controlled trials (RCTs) [85], and translating traditional clinical trial procedures to digital form may require the complete re-engineering of trial design and processes [36]. The Portal was developed to support traditional health care research, for example, RCTs and observational studies, and multiple RCTs of internet-administered cognitive behavioral therapy interventions have been conducted on the Portal [87-90]. The Portal was originally designed to deliver two internet-administered cognitive behavioral therapy interventions [87,88]; however, over time, the Portal has been developed to support the needs of different studies and has experienced a continuous flow of technical requirements from researchers since its conception in 2010 [22,23]. However, the time taken to design and conduct traditional health care research is not in line with fast-paced technical advances [91]. Therefore, when using digital technologies for intervention delivery and the execution of study

procedures, careful planning is required as technology may change [25,32]. Multiple requests for new digital features and changes to existing features were required for both the intervention and the execution of study procedures; however, challenges were experienced relating to the legacy software architecture of the Portal. For example, new feature requests or changes to existing features can result in unanticipated negative impacts on existing features used by ongoing studies. This highlights a key challenge for research environments using internet-based technologies, especially in a context where new digital features and complex configurations are added over time [23]. Appropriate technical infrastructure (eg, having the appropriate hardware, software, and technical expertise) has been highlighted as a significant challenge in conducting successful digital health care research [38].

Limitations

Meeting minutes were written independently to the objective of the analysis, and therefore may not provide sufficient detail to reveal all the challenges experienced by the research team. As is common in document analysis, documents are often uneven in length, providing more detailed information about some topics than others [60]. As such, there is a potential that some challenges experienced were missed in the analysis as less focus was placed on them when writing meeting minutes; however, we attempted to overcome this limitation by member-checking through two group discussions. An additional limitation relates to selectivity bias [60], and it is difficult to separate the analysis from the context of the research team, for example, biases and opinions held by research team members. Attempts to limit selectivity bias were made by the analysis being primarily conducted by a new member of the research team who had not been present in any of the meetings wherein minutes were analyzed.

Conclusions

The development and feasibility testing of health care interventions using digital technologies is time- and resource-intensive. Recommendations for improving efficiency include (1) the development of networks to share good practice and training opportunities for trial staff, especially in the area of complex digital health care interventions; (2) the employment of advanced research methodology–educated, senior dedicated trial personnel who can be responsible for the day-to-day operations; (3) the completion of the intervention development phase (including technical requirements) before the feasibility study setup; and (4) the integration of members of the software development team into the research team to improve communication and develop shared knowledge and understanding. We hope that our experiences may be useful for others who are planning to conduct future research within the development and feasibility phases of the MRC complex intervention framework, especially for internet-administered interventions and research using digital features to support the execution of study procedures. Publishing challenges experienced during intervention development [19] and trial setup and conduct [11] may help to reduce future research waste, improve the quality of digital health care research, and add to the emerging literature concerning challenges experienced by integrating digital technologies into health care research [25].

Despite experiencing a number of challenges, the ENGAGE feasibility study commenced recruitment on July 3, 2020, and recruitment targets were successfully met by October 14, 2020, well within the projected 6-month recruitment period [21]. Furthermore, preliminary posttreatment follow-up data indicated that retention targets were successfully met. While delays to study start were experienced, taking a careful and detail-orientated approach to intervention development and feasibility study setup may have helped facilitate meeting recruitment and retention targets and will hopefully enhance efficiencies across subsequent phases of our planned research, for example, a definitive (evaluation) trial.

Acknowledgments

This work was supported by the Swedish Research Council (K2015-99X-20836-08-4 and 2018-02578), the Swedish Cancer Society (grants 150673 and 180589), and the Swedish Childhood Cancer Foundation (grant PR2017-0005). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The funders do not bear any responsibility for the analyses or interpretations presented here. The authors thank the 4 parent research partners (Helena Börjesson, Martin Hedqvist, Anna Mautner, and Erik Olsson) and the Portal team members outside of the ENGAGE research team (Fabian Holmberg and Ian Horne). The authors are also very grateful to ETC Communications, Stockholm, for providing the graphic design and to Rebecka Porse Schalin for designing all illustrations in the EJDeR intervention.

Authors' Contributions

Author contributions have been written in accordance with the CRediT author statement. JW contributed to the development and design of the methodology, formal analysis, writing the original draft, data visualization, supervision, and project administration. MK contributed to formal analysis, data curation, reviewing and editing the original draft, and data visualization. JH contributed to reviewing and editing the original draft, and data visualization. YHS was responsible for programming and software development and reviewing and editing the original draft. KN and HG reviewed and edited the final draft. LVE was responsible for conceptualization, methodology, provision of resources, supervision, project administration, funding acquisition, and reviewed and edited the final draft. All authors approved the final version of the manuscript before submission.

Conflicts of Interest

None declared.

References

1. Treweek S, Altman DG, Bower P, Campbell M, Chalmers I, Cotton S, et al. Making randomised trials more efficient: report of the first meeting to discuss the Trial Forge platform. Trials 2015 Jun 05;16:261 [FREE Full text] [doi: 10.1186/s13063-015-0776-0] [Medline: 26044814]
2. Moher D, Glasziou P, Chalmers I, Nasser M, Bossuyt PM, Korevaar DA, et al. Increasing value and reducing waste in biomedical research: who's listening? Lancet 2016 Apr 09;387(10027):1573-1586. [doi: 10.1016/S0140-6736(15)00307-4] [Medline: 26423180]
3. Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. Lancet 2009 Jul 04;374(9683):86-89. [doi: 10.1016/S0140-6736(09)60329-9] [Medline: 19525005]
4. Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gülmezoglu AM, et al. How to increase value and reduce waste when research priorities are set. Lancet 2014 Jan 11;383(9912):156-165. [doi: 10.1016/S0140-6736(13)62229-1] [Medline: 24411644]
5. Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D, et al. Increasing value and reducing waste in research design, conduct, and analysis. Lancet 2014 Jan 11;383(9912):166-175 [FREE Full text] [doi: 10.1016/S0140-6736(13)62227-8] [Medline: 24411645]
6. Yordanov Y, Dechartres A, Atal I, Tran V, Boutron I, Crequit P, et al. Avoidable waste of research related to outcome planning and reporting in clinical trials. BMC Med 2018 Jun 11;16(1):87 [FREE Full text] [doi: 10.1186/s12916-018-1083-x] [Medline: 29886846]
7. Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. Lancet 2014 Jan 18;383(9913):267-276. [doi: 10.1016/S0140-6736(13)62228-X] [Medline: 24411647]
8. Glasziou P, Chalmers I. Research waste is still a scandal—an essay by Paul Glasziou and Iain Chalmers. BMJ Clin Res 2018 Nov 12;363:k4645. [doi: 10.1136/bmj.k4645]
9. Duley L, Gillman A, Duggan M, Belson S, Knox J, McDonald A, et al. What are the main inefficiencies in trial conduct: a survey of UKCRC registered clinical trials units in the UK. Trials 2018 Jan 08;19(1):15 [FREE Full text] [doi: 10.1186/s13063-017-2378-5] [Medline: 29310685]
10. Arundel C, Gellatly J. Learning from OCTET - exploring the acceptability of clinical trials management methods. Trials 2018 Jul 13;19(1):378 [FREE Full text] [doi: 10.1186/s13063-018-2765-6] [Medline: 30005659]
11. Beaumont D, Arribas M, Frimley L, Balogun E, Roberts I, Shakur-Still H. Trial management: we need a cadre of high-class triallists to deliver the answers that patients need. Trials 2019 Jun 13;20(1):354 [FREE Full text] [doi: 10.1186/s13063-019-3451-z] [Medline: 31196151]
12. Farrell B, Kenyon S, Shakur H. Managing clinical trials. Trials 2010 Jul 13;11(1):78 [FREE Full text] [doi: 10.1186/1745-6215-11-78] [Medline: 20626885]
13. Goodarzynejad H, Babamahmoodi A. Project management of randomized clinical trials: a narrative review. Iran Red Crescent Med J 2015 Aug 24;17(8):e11602 [FREE Full text] [doi: 10.5812/ircmj.11602] [Medline: 26430517]
14. McCaskell DS, Molloy AJ, Childerhose L, Costigan FA, Reid JC, McCaughan M, et al. Project management lessons learned from the multicentre CYCLE pilot randomized controlled trial. Trials 2019 Aug 28;20(1):532 [FREE Full text] [doi: 10.1186/s13063-019-3634-7] [Medline: 31455384]
15. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M, Medical Research Council Guidance. Developing and evaluating complex interventions: the new Medical Research Council guidance. BMJ 2008 Sep 29;337:a1655 [FREE Full text] [doi: 10.1136/bmj.a1655] [Medline: 18824488]
16. Eldridge SM, Chan CL, Campbell MJ, Bond CM, Hopewell S, Thabane L, PAFS consensus group. CONSORT 2010 statement: extension to randomised pilot and feasibility trials. BMJ 2016 Oct 24;355:i5239 [FREE Full text] [doi: 10.1136/bmj.i5239] [Medline: 27777223]
17. Lancaster GA, Thabane L. Guidelines for reporting non-randomised pilot and feasibility studies. Pilot Feasibility Stud 2019 Oct 6;5:114 [FREE Full text] [doi: 10.1186/s40814-019-0499-1] [Medline: 31608150]
18. Bleijenberg N, de Man-van Ginkel JM, Trappenburg JC, Ettema RG, Sino CG, Heim N, et al. Increasing value and reducing waste by optimizing the development of complex interventions: enriching the development phase of the Medical Research Council (MRC) Framework. Int J Nurs Stud 2018 Mar;79:86-93 [FREE Full text] [doi: 10.1016/j.ijnurstu.2017.12.001] [Medline: 29220738]
19. O'Cathain A, Croot L, Duncan E, Rousseau N, Sworn K, Turner KM, et al. Guidance on how to develop complex interventions to improve health and healthcare. BMJ Open 2019 Aug 15;9(8):e029954 [FREE Full text] [doi: 10.1136/bmjopen-2019-029954] [Medline: 31420394]
20. Thabane L, Lancaster G. A guide to the reporting of protocols of pilot and feasibility trials. Pilot Feasibility Stud 2019 Feb 28;5:37 [FREE Full text] [doi: 10.1186/s40814-019-0423-8] [Medline: 30858987]

21. Woodford J, Wikman A, Cernvall M, Ljungman G, Romppala A, Grönqvist H, et al. Study protocol for a feasibility study of an internet-administered, guided, CBT-based, self-help intervention (ENGAGE) for parents of children previously treated for cancer. BMJ Open 2018 Jun 14;8(6):e023708 [FREE Full text] [doi: 10.1136/bmjopen-2018-023708] [Medline: 29903802]
22. Sjöström J, von Essen L, Grönqvist H. The origin and impact of ideals in eHealth research: experiences from the U-CARE research environment. JMIR Res Protoc 2014 May 23;3(2):e28 [FREE Full text] [doi: 10.2196/resprot.3202] [Medline: 24860071]
23. Grönqvist H, Olsson EM, Johansson B, Held C, Sjöström J, Lindahl Norberg A, et al. Fifteen challenges in establishing a multidisciplinary research program on eHealth research in a university setting: a case study. J Med Internet Res 2017 May 23;19(5):e173 [FREE Full text] [doi: 10.2196/jmir.7310] [Medline: 28536090]
24. Eagleson R, Altamirano-Diaz L, McInnis A, Welisch E, De Jesus S, Prapavessis H, et al. Implementation of clinical research trials using web-based and mobile devices: challenges and solutions. BMC Med Res Methodol 2017 Mar 17;17(1):43 [FREE Full text] [doi: 10.1186/s12874-017-0324-6] [Medline: 28302050]
25. Rosa C, Marsch LA, Winstanley EL, Brunner M, Campbell AN. Using digital technologies in clinical trials: current and future applications. Contemp Clin Trials 2021 Jan;100:106219. [doi: 10.1016/j.cct.2020.106219] [Medline: 33212293]
26. Wilhelm S, Weingarden H, Ladis I, Braddick V, Shin J, Jacobson NC. Cognitive-behavioral therapy in the digital age: presidential address. Behav Ther 2020 Jan;51(1):1-14 [FREE Full text] [doi: 10.1016/j.beth.2019.08.001] [Medline: 32005328]
27. Strudwick G, Sockalingam S, Kassam I, Sequeira L, Bonato S, Youssef A, et al. Digital interventions to support population mental health in Canada during the COVID-19 pandemic: rapid review. JMIR Ment Health 2021 Mar 02;8(3):e26550 [FREE Full text] [doi: 10.2196/26550] [Medline: 33650985]
28. Rauschenberg C, Schick A, Hirjak D, Seidler A, Paetzold I, Apfelbacher C, et al. Evidence synthesis of digital interventions to mitigate the negative impact of the COVID-19 pandemic on public mental health: rapid meta-review. J Med Internet Res 2021 Mar 10;23(3):e23365 [FREE Full text] [doi: 10.2196/23365] [Medline: 33606657]
29. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
30. Andersson G, Carlbring P, Rozental A. Response and remission rates in internet-based cognitive behavior therapy: an individual patient data meta-analysis. Front Psychiatry 2019 Oct 25;10:749 [FREE Full text] [doi: 10.3389/fpsyt.2019.00749] [Medline: 31708813]
31. Hwang WJ, Ha JS, Kim MJ. Research trends on mobile mental health application for general population: a scoping review. Int J Environ Res Public Health 2021 Mar 02;18(5):2459 [FREE Full text] [doi: 10.3390/ijerph18052459] [Medline: 33801537]
32. Rosa C, Campbell AN, Miele GM, Brunner M, Winstanley EL. Using e-technologies in clinical trials. Contemp Clin Trials 2015 Nov;45(Pt A):41-54 [FREE Full text] [doi: 10.1016/j.cct.2015.07.007] [Medline: 26176884]
33. Baker TB, Gustafson DH, Shah D. How can research keep up with eHealth? Ten strategies for increasing the timeliness and usefulness of eHealth research. J Med Internet Res 2014 Feb 19;16(2):e36 [FREE Full text] [doi: 10.2196/jmir.2925] [Medline: 24554442]
34. Noonan D, Simmons LA. Navigating nonessential research trials during COVID-19: the push we needed for using digital technology to increase access for rural participants? J Rural Health 2021 Jan;37(1):185-187 [FREE Full text] [doi: 10.1111/jrh.12446] [Medline: 32282959]
35. National Academies of Sciences, Engineering, and Medicine, Health and Medicine Division, Board on Health Sciences Policy, Forum on Drug Discovery, Development, and Translation. Virtual Clinical Trials: Challenges and Opportunities: Proceedings of a Workshop. Washington, DC: National Academies Press; 2019.
36. Steinhubl SR, Wolff-Hughes DL, Nilsen W, Iturriaga E, Califf RM. Digital clinical trials: creating a vision for the future. NPJ Digit Med 2019 Dec 12;2:126 [FREE Full text] [doi: 10.1038/s41746-019-0203-0] [Medline: 31872066]
37. Clinical Trials Transformation Initiative (CTTI) Program: Digital health trials. Clinical Trials Transformation Initiative. URL: https://www.ctti-clinicaltrials.org/programs/mobile-clinical-trials [accessed 2021-04-13]
38. Geoghegan C, Nido V, Bemden AB, Hallinan Z, Jordan L, Kehoe LS, et al. Learning from patient and site perspectives to develop better digital health trials: recommendations from the Clinical Trials Transformation Initiative. Contemp Clin Trials Commun 2020 Aug 13;19:100636 [FREE Full text] [doi: 10.1016/j.conctc.2020.100636] [Medline: 32913915]
39. Steliarova-Foucher E, Colombet M, Ries LA, Moreno F, Dolya A, Bray F, IICC-3 contributors. International incidence of childhood cancer, 2001-10: a population-based registry study. Lancet Oncol 2017 Jun;18(6):719-731 [FREE Full text] [doi: 10.1016/S1470-2045(17)30186-9] [Medline: 28410997]
40. International incidence of childhood cancer 3. International Agency for Research on Cancer. URL: http://iicc.iarc.fr/results/ [accessed 2020-08-18]
41. Carlsson T, Kukkola L, Ljungman L, Hovén E, von Essen L. Psychological distress in parents of children treated for cancer: an explorative study. PLoS One 2019 Jun 21;14(6):e0218860 [FREE Full text] [doi: 10.1371/journal.pone.0218860] [Medline: 31226159]

42. Ljungman L, Cernvall M, Grönqvist H, Ljótsson B, Ljungman G, von Essen L. Long-term positive and negative psychological late effects for parents of childhood cancer survivors: a systematic review. PLoS One 2014 Jul 24;9(7):e103340 [FREE Full text] [doi: 10.1371/journal.pone.0103340] [Medline: 25058607]
43. Wakefield CE, McLoone JK, Butow P, Lenthen K, Cohn RJ. Parental adjustment to the completion of their child's cancer treatment. Pediatr Blood Cancer 2011 Apr;56(4):524-531. [doi: 10.1002/pbc.22725] [Medline: 21298736]
44. Wikman A, Hovén E, Cernvall M, Ljungman G, Ljungman L, von Essen L. Parents of children diagnosed with cancer: work situation and sick leave, a five-year post end-of-treatment or a child's death follow-up study. Acta Oncol 2016;55(9-10):1152-1157. [doi: 10.3109/0284186X.2016.1167956] [Medline: 27159219]
45. Hovén E, von Essen L, Norberg AL. A longitudinal assessment of work situation, sick leave, and household income of mothers and fathers of children with cancer in Sweden. Acta Oncol 2013 Aug;52(6):1076-1085. [doi: 10.3109/0284186X.2012.760846] [Medline: 23343223]
46. Lindahl Norberg A, Montgomery SM, Bottai M, Heyman M, Hovén EI. Short-term and long-term effects of childhood cancer on income from employment and employment status: a national cohort study in Sweden. Cancer 2017 Apr 01;123(7):1238-1248 [FREE Full text] [doi: 10.1002/cncr.30436] [Medline: 27870013]
47. Öhman M, Woodford J, von Essen L. Socioeconomic consequences of parenting a child with cancer for fathers and mothers in Sweden: a population-based difference-in-difference study. Int J Cancer 2020 Dec 15;148(10):2535-2541 [FREE Full text] [doi: 10.1002/ijc.33444] [Medline: 33320976]
48. van Warmerdam J, Zabih V, Kurdyak P, Sutradhar R, Nathan PC, Gupta S. Prevalence of anxiety, depression, and posttraumatic stress disorder in parents of children with cancer: a meta-analysis. Pediatr Blood Cancer 2019 Jun;66(6):e27677. [doi: 10.1002/pbc.27677] [Medline: 30816008]
49. Ljungman L, Boger M, Ander M, Ljótsson B, Cernvall M, von Essen L, et al. Impressions that last: particularly negative and positive experiences reported by parents five years after the end of a child's successful cancer treatment or death. PLoS One 2016 Jun 7;11(6):e0157076 [FREE Full text] [doi: 10.1371/journal.pone.0157076] [Medline: 27272318]
50. Christen S, Mader L, Baenziger J, Roser K, Schindera C, Tinner EM, et al. "I wish someone had once asked me how I'm doing": disadvantages and support needs faced by parents of long-term childhood cancer survivors. Pediatr Blood Cancer 2019 Aug;66(8):e27767. [doi: 10.1002/pbc.27767] [Medline: 31090217]
51. Kukkola L, Hovén E, Cernvall M, von Essen L, Grönqvist H. Perceptions of support among Swedish parents of children after end of successful cancer treatment: a prospective, longitudinal study. Acta Oncol 2017 Dec;56(12):1705-1711. [doi: 10.1080/0284186X.2017.1374554] [Medline: 28971717]
52. Wakefield CE, McLoone J, Butow P, Lenthen K, Cohn RJ. Support after the completion of cancer treatment: Perspectives of Australian adolescents and their families. Eur J Cancer Care (Engl) 2013 Jul;22(4):530-539. [doi: 10.1111/ecc.12059] [Medline: 23730980]
53. Patel V, Saxena S, Lund C, Thornicroft G, Baingana F, Bolton P, et al. The Lancet Commission on global mental health and sustainable development. Lancet 2018 Oct 27;392(10157):1553-1598. [doi: 10.1016/S0140-6736(18)31612-X] [Medline: 30314863]
54. Woodford J, Farrand P, Hagström J, Hedenmalm L, von Essen L. Internet-administered cognitive behavioral therapy for common mental health difficulties in parents of children treated for cancer: intervention development and description study. JMIR Form Res 2021 Jul 22;5(7):e22709 [FREE Full text] [doi: 10.2196/22709] [Medline: 34142662]
55. Ljungman L, Cernvall M, Ghaderi A, Ljungman G, von Essen L, Ljótsson B. An open trial of individualized face-to-face cognitive behavior therapy for psychological distress in parents of children after end of treatment for childhood cancer including a cognitive behavioral conceptualization. PeerJ 2018 Apr 11;6:e4570 [FREE Full text] [doi: 10.7717/peerj.4570] [Medline: 29666751]
56. Wikman A, Mattsson E, von Essen L, Hovén E. Prevalence and predictors of symptoms of anxiety and depression, and comorbid symptoms of distress in parents of childhood cancer survivors and bereaved parents five years after end of treatment or a child's death. Acta Oncol 2018 Jul;57(7):950-957. [doi: 10.1080/0284186X.2018.1445286] [Medline: 29498559]
57. Woodford J, Wikman A, Einhorn K, Cernvall M, Grönqvist H, Romppala A, et al. Attitudes and preferences toward a hypothetical trial of an internet-administered psychological intervention for parents of children treated for cancer: web-based survey. JMIR Ment Health 2018 Dec 18;5(4):e10085 [FREE Full text] [doi: 10.2196/10085] [Medline: 30563814]
58. Woodford J, Norbäck K, Hagström J, Grönqvist H, Parker A, Arundel C, et al. Study within a trial (SWAT) protocol. Investigating the effect of personalised versus non-personalised study invitations on recruitment: an embedded randomised controlled recruitment trial. Contemp Clin Trials Commun 2020 Apr 24;18:100572 [FREE Full text] [doi: 10.1016/j.conctc.2020.100572] [Medline: 32420511]
59. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. Acad Med 2014 Sep;89(9):1245-1251 [FREE Full text] [doi: 10.1097/ACM.0000000000000388] [Medline: 24979285]
60. Bowen GA. Document analysis as a qualitative research method. Qualitative Research Journal 2009 Aug 03;9(2):27-40. [doi: 10.3316/qrj0902027]

61. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
62. Boyatzis R. Transforming Qualitative Information: Thematic Analysis and Code Development. Thousand Oaks, CA: Sage Publications; 1998.
63. Lietz CA, Langer CL, Furman R. Establishing trustworthiness in qualitative research in social work: implications from a study regarding spirituality. Qual Soc Work 2006 Dec 1;5(4):441-458. [doi: 10.1177/1473325006070288]
64. Auer-Rizzi W, Berry M. Business vs. Cultural frames of reference in group decision making: interactions among Austrian, Finnish, and Swedish business students. J Bus Commun 2000 Jul 01;37(3):264-288. [doi: 10.1177/002194360003700304]
65. Lämsä T. Leadership styles and decision-making in Finnish and Swedish organizations. Int J Comp Manag 2010 Jan;11(1):139-149 [FREE Full text]
66. Moe NB, Aurum A, Dybå T. Challenges of shared decision-making: a multiple case study of agile software development. Inf Softw Technol 2012 Aug;54(8):853-865. [doi: 10.1016/j.infsof.2011.11.006]
67. Ellis E, Riegel B, Hamon M, Carlson B, Jimenez S, Parkington S. The challenges of conducting clinical research: one research team's experiences. Clin Nurse Spec 2001 Nov;15(6):286-92; quiz 293. [doi: 10.1097/00002800-200111000-00013] [Medline: 11855487]
68. Ghobadi S, Mathiassen L. Perceived barriers to effective knowledge sharing in agile software teams. Info Systems J 2014 Dec 18;26(2):95-125. [doi: 10.1111/isj.12053]
69. Bonsu JM, Frasso R, Curry AE. Lessons from the field: the conduct of randomized controlled trials in Botswana. Trials 2017 Oct 27;18(1):503 [FREE Full text] [doi: 10.1186/s13063-017-2237-4] [Medline: 29078791]
70. Petrova M, Barclay S. Response to Correspondence from Kolstoe and colleagues concerning our paper entitled, Research approvals iceberg: how a 'low-key' study in England needed 89 professionals to approve it and how we can do better. BMC Med Ethics 2019 Dec 24;20(1):101 [FREE Full text] [doi: 10.1186/s12910-019-0433-3] [Medline: 31870356]
71. Martin L, Hutchens M, Hawkins C, Radnov A. How much do clinical trials cost? Nat Rev Drug Discov 2017 Jun;16(6):381-382. [doi: 10.1038/nrd.2017.70] [Medline: 28529317]
72. Kakkar AK, Sarma P, Medhi B. mHealth technologies in clinical trials: opportunities and challenges. Indian J Pharmacol 2018;50(3):105-107 [FREE Full text] [doi: 10.4103/ijp.IJP_391_18] [Medline: 30166746]
73. Kotz D, Fu K, Gunter C, Rubin A. Security for mobile and cloud frontiers in healthcare. Commun ACM 2015 Jul 23;58(8):21-23. [doi: 10.1145/2790830]
74. Sims HP, Faraj S, Yun S. When should a leader be directive or empowering? How to develop your own situational theory of leadership. Bus Horiz 2009;52(2):149-158. [doi: 10.1016/j.bushor.2008.10.002]
75. Treweek S, Littleford R. Trial management- building the evidence base for decision-making. Trials 2018 Jan 08;19(1):11 [FREE Full text] [doi: 10.1186/s13063-017-2322-8] [Medline: 29310691]
76. Campbell MK, Snowdon C, Francis D, Elbourne D, McDonald AM, Knight R, STEPS group. Recruitment to randomised trials: strategies for trial enrollment and participation study. The STEPS study. Health Technol Assess 2007 Nov;11(48):iii, ix-105 [FREE Full text] [doi: 10.3310/hta11480] [Medline: 17999843]
77. Fogel DB. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: a review. Contemp Clin Trials Commun 2018 Aug 7;11:156-164 [FREE Full text] [doi: 10.1016/j.conctc.2018.08.001] [Medline: 30112460]
78. Grant KP, Graham TS, Heberling ME. The project manager and project team involvement: implications for project leadership. J Leadersh Organ Stud 2001 Jan 1;7(4):32-42. [doi: 10.1177/107179190100700403]
79. Mitchell E, Goodman K, Hartley S, Hickey H, McDonald AM, Meadows HM, UK Trial Managers' Network Executive Group. Where do we go from here? - Opportunities and barriers to the career development of trial managers: a survey of UK-based trial management professionals. Trials 2020 May 06;21(1):384 [FREE Full text] [doi: 10.1186/s13063-020-04316-z] [Medline: 32375851]
80. Yardley L, Choudhury T, Patrick K, Michie S. Current issues and future directions for research into digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):814-815. [doi: 10.1016/j.amepre.2016.07.019] [Medline: 27745680]
81. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
82. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 5;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
83. Dybå T, Dingsøyr T. Empirical studies of agile software development: a systematic review. Inf Softw Technol 2008 Aug;50(9-10):833-859. [doi: 10.1016/j.infsof.2008.01.006]
84. Cockburn A, Highsmith J. Agile software development, the people factor. Computer 2001 Nov;34(11):131-133. [doi: 10.1109/2.963450]
85. 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 Aug 27;3(1):110 [FREE Full text] [doi: 10.1038/s41746-020-00314-2] [Medline: 34446821]

86. Patrick K, Hekler EB, Estrin D, Mohr DC, Riper H, Crane D, et al. The pace of technologic change: implications for digital health behavior intervention research. Am J Prev Med 2016 Nov;51(5):816-824. [doi: 10.1016/j.amepre.2016.05.001] [Medline: 27745681]
87. Igelström H, Hauffman A, Alfonsson S, Sjöström J, Cajander A, Johansson B. User experiences of an internet-based stepped-care intervention for individuals with cancer and concurrent symptoms of anxiety or depression (the U-CARE AdultCan trial): qualitative study. J Med Internet Res 2020 May 19;22(5):e16604 [FREE Full text] [doi: 10.2196/16604] [Medline: 32427108]
88. Norlund F, Wallin E, Olsson EM, Wallert J, Burell G, von Essen L, et al. Internet-based cognitive behavioral therapy for symptoms of depression and anxiety among patients with a recent myocardial infarction: the U-CARE heart randomized controlled trial. J Med Internet Res 2018 Mar 08;20(3):e88 [FREE Full text] [doi: 10.2196/jmir.9710] [Medline: 29519777]
89. Rondung E, Ternström E, Hildingsson I, Haines HM, Sundin O, Ekdahl J, et al. Comparing internet-based cognitive behavioral therapy with standard care for women with fear of birth: randomized controlled trial. JMIR Ment Health 2018 Aug 10;5(3):e10420 [FREE Full text] [doi: 10.2196/10420] [Medline: 30097422]
90. Sjömark J, Parling T, Jonsson M, Larsson M, Skoog Svanberg A. A longitudinal, multi-centre, superiority, randomized controlled trial of internet-based cognitive behavioural therapy (iCBT) versus treatment-as-usual (TAU) for negative experiences and posttraumatic stress following childbirth: the JUNO study protocol. BMC Pregnancy Childbirth 2018 Oct 01;18(1):387 [FREE Full text] [doi: 10.1186/s12884-018-1988-6] [Medline: 30285758]
91. Riley WT, Glasgow RE, Etheredge L, Abernethy AP. Rapid, responsive, relevant (R3) research: a call for a rapid learning health research enterprise. Clin Transl Med 2013 May 10;2(1):10 [FREE Full text] [doi: 10.1186/2001-1326-2-10] [Medline: 23663660]

Abbreviations

LICBT: low-intensity cognitive behavioral therapy
MRC: Medical Research Council
PRP: parent research partner
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 04.12.20; peer-reviewed by MDG Pimentel; comments to author 19.02.21; revised version received 16.04.21; accepted 02.05.21; published 08.10.21.

Please cite as:
Woodford J, Karlsson M, Hagström J, Hägg Sylvén Y, Norbäck K, Grönqvist H, von Essen L
Conducting Digital Health Care Research: Document Analysis of Challenges Experienced During Intervention Development and Feasibility Study Setup of an Internet-Administered Intervention for Parents of Children Treated for Cancer
JMIR Form Res 2021;5(10):e26266
URL: https://formative.jmir.org/2021/10/e26266
doi:10.2196/26266
PMID:34433524

Original Paper

Development of a Theoretically Informed Web-Based Mind-Body Wellness Intervention for Patients With Primary Biliary Cholangitis: Formative Study

Makayla Watt[1], BSc; John C Spence[1], PhD; Puneeta Tandon[1], MD, MSc, FRCP

University of Alberta, Edmonton, AB, Canada

Corresponding Author:
Puneeta Tandon, MD, MSc, FRCP
University of Alberta
130-University Campus
Edmonton, AB, T6G2X8
Canada
Phone: 1 780 492 9844
Fax: 1 780 492 9873
Email: ptandon@ualberta.ca

Abstract

Background: Mind-body interventions have the potential to positively impact the symptom burden associated with primary biliary cholangitis (PBC). Interventions are more likely to be effective if they are informed by a theoretical framework. The Behaviour Change Wheel (BCW) and the behaviour change technique taxonomy version 1 (BCTv1) provide frameworks for intervention development.

Objective: This study describes how theory has guided the development of a 12-week multicomponent mind-body wellness intervention for PBC.

Methods: The steps involved in developing the BCW intervention included specifying the target behavior; explaining barriers and facilitators using the Capability, Opportunity, Motivation, and Behaviour and the theoretical domains framework; identifying intervention functions to target explanatory domains; and selecting relevant behavior change techniques to address intervention functions. Qualitative data from patients with inflammatory bowel disease using an earlier version of the program and feedback from a PBC patient advisory team were used to guide intervention development.

Results: Barriers and facilitators to intervention participation associated with capability, opportunity, and motivation were identified. Intervention functions and behavior change techniques were identified to target each barrier and facilitator.

Conclusions: The Peace Power Pack PBC intervention was developed to help individuals with PBC manage their symptom burden. The theoretical frameworks employed in this intervention provide direction on targeting antecedents of behavior and allow standardized reporting of intervention components.

Trial Registration: ClinicalTrials.gov NCT04791527; https://clinicaltrials.gov/ct2/show/NCT04791527

(JMIR Form Res 2021;5(10):e29064) doi:10.2196/29064

KEYWORDS

liver disease; meditation; yoga; breathwork; behavior theory; COM-B model; behaviour change wheel; behaviour change taxonomy; internet; digital

Introduction

Primary biliary cholangitis (PBC) is a female predominant chronic liver disease estimated to affect between 9000 and 11,000 Canadians [1]. Despite the relatively low prevalence of PBC, global incidence and prevalence rates have been reported to be on the rise [1]. PBC is associated with symptoms including pruritus and fatigue, which can lead to social isolation and emotional dysfunction [1-3]. Fatigue, defined as a persistent state of exhaustion, inability to perform usual routines, and a decreased capacity for physical and mental work, has been reported as the most common and debilitating among these symptoms [4-7]. Individuals with PBC also commonly experience a low health-related quality of life (HRQOL), with 1 study concluding that 35% of individuals with PBC had an

impaired HRQOL compared to 6% of healthy controls [6]. Current medical therapies are ineffective at improving PBC-related symptoms or impacting quality of life [5,6,8]. Building upon the recognized need for novel interventions [6,9], our team was approached by patients and the Canadian PBC Society to develop self-care tools to manage symptom burden. Although to our knowledge, mind-body wellness interventions have never been trialed in PBC, interventions of this nature have been found to improve fatigue and HRQOL in other chronic diseases [10-12].

The use of a clear theoretical framework during the design of an intervention has been associated with increased adherence rates, and sustained changes to health-related behaviors [13-15]. The Behaviour Change Wheel (BCW), a framework synthesized from 19 individual models of behavior, has been used to guide development of several acceptable and effective theory-based interventions [16-18]. At the core of the BCW is the Capability, Opportunity, Motivation and Behaviour (COM-B) model, which describes the key antecedents to the target behavior. The BCW then outlines intervention functions that can be used to facilitate behavior change [16]. This process is further enhanced by the behavior change technique taxonomy version 1 (BCTv1), which details standardized active intervention ingredients that can be implemented to target intervention functions [19]. Optimally, theory would also extend to the evaluation of behavior change and maintenance.

This paper describes how theory has guided the development of a 12-week multicomponent mind-body wellness intervention for PBC (ClinicalTrials.gov NCT04791527) using several theoretical constructs: BCW guidelines [16], the COM-B model [16], the theoretical domains framework (TDF) [20], and the BCTv1 [19]. Development of the intervention involved the following steps, which were informed by the BCW guidelines: (1) specify the target behavior; (2) explain barriers and facilitators to the target behavior by using the COM-B model and the TDF; (3) identify intervention functions to target explanatory domains; and (4) select relevant behavior change techniques to address intervention functions.

Methods

The following sections outline the processes (methods) for each of the 4 steps of intervention development. An outline of the 4 steps of intervention development can be found in Figure 1.

Figure 1. Steps involved in intervention development. BCT: behavior change technique; COM-B: Capability, Opportunity, Motivation and Behaviour; TDF: theoretical domains framework.

Step 1: Specify the Target Behavior

The target behavior was determined through a review of the literature on adherence to behavioral health interventions, and in consultation with the Canadian PBC Society.

Step 2: Explain Barriers and Facilitators to Behavior Using the COM-B and TDF

Domains from the COM-B model and the TDF were selected to explain barriers and facilitators to the target behavior. The COM-B model outlines that for a behavior to occur, an individual must have the capability, opportunity, and motivation to perform the behavior. Capability is composed of psychological capability (knowledge), and physical capability (physical skills); opportunity is composed of physical opportunity (environmental resources) and social opportunity (cultural milieu); and motivation includes reflective motivation (evaluations, plans) and automatic motivation (emotions, impulses) [16]. As the COM-B model provides a relatively general understanding of behavior, the TDF, which outlines 14 processes involved in behavior change, is often used to provide further specification of behavioral determinants [20]. To identify barriers and facilitators driving health-related behavior, we conducted qualitative interviews with individuals who had participated in the previous iteration of the intervention carried out in a separate chronic disease group (ie, individuals with inflammatory bowel disease [IBD]) [21]. Similar to PBC, individuals with IBD experience high rates of fatigue and impaired quality of life [22,23]. These interviews were coded and thematically analyzed by 2 independent coders [21]. A COM-B characteristic and a TDF domain were then identified for each barrier and facilitator of behavior mentioned by those participants.

Step 3: Identify Intervention Functions to Target Explanatory Domains

Intervention functions were selected to address each barrier and facilitator to behavior. The BCW specifies 9 standardized intervention functions that can be used to address barriers and facilitators to behavior change [16]. The BCW guide then outlines intervention functions that are appropriate for each TDF domain [24]. The web-based nature of the program and characteristics of the target population (ie, chronic fatigue) were considered when selecting intervention functions.

Step 4: Specify Intervention Content by Selecting Relevant BCTs

Behavior change techniques were selected to allow standardized implementation of intervention functions. Following the procedure outlined by Jennings et al [13] and Tombor et al [25], BCTs were specified for each of the intervention functions identified in step 3.

Results

Results Overview

The Peace Power Pack PBC (PPP_{PBC}) intervention was co-developed with a patient advisory team from the Canadian PBC society. The web-based intervention is described in Multimedia Appendix 1. The intervention is 12 weeks in duration with each week featuring: (1) a video detailing a core practice of mindful movement (yoga, tai chi, and low-intensity exercise divided into a standing stream and a chair stream), energizing breathwork practices, and guided meditation (increasing in length from 20-30 minutes over the course of the program); (2) an introductory video describing a weekly positive psychology theme (3-5 minutes); and (3) an interactive positive psychology activity related to the theme for the week (3-5 minutes). All programming is hosted on the investigator's website [26]. Throughout the duration of the study, participants will receive standardized weekly motivational emails, weekly 10-minute motivational interviewing check-ins, and will be invited to participate in weekly group sessions with fellow participants. The following section outlines the outcomes (results) for each of the 4 steps of intervention development previously outlined.

Step 1: Specify the Target Behavior

Adherence to the video-based program at least 3 days a week was selected as the primary target behavior, with a gradual increase in the video duration over the course of the 12 weeks. Based on feedback from the Canadian PBC society, this target behavior was chosen with the intent to balance the intervention dose with likelihood of adherence. Available evidence suggests that higher levels of adherence to behavioral health interventions leads to improved outcomes in a dose-dependent manner [27]. High levels of fatigue in individuals with PBC have been associated with a decreased sense of self-efficacy for a particular behavior [28] and inability to adhere to a target could lead to further reductions in self-efficacy. To ensure participants are aware of the anticipated study commitment, the target will be advertised to participants interested in enrollment.

Step 2: Explain Barriers and Facilitators to Behavior Using the COM-B and TDF

A comprehensive list of barriers and facilitators, along with the associated COM-B and TDF domains is provided in Table 1. The most common barriers to program participation described by the individuals with IBD were difficulty fitting the program into daily routine, and finding that the movement portion of the program was not matched with the ability level [21]. Perceived facilitators to program participation included accessible presentation of content on the host website and contact with program facilitators/fellow participants. Of the 14 domains of the TDF, 9 were associated with barriers and facilitators to intervention participation: behavior regulation, physical skills, environmental context and resources, memory attention and decision processes, social influences, goals, beliefs about capabilities, beliefs about consequences, and reinforcement. The most common TDF domains were social influences (check-ins with program facilitators and other participants enhancing accountability), and behavioral regulation (fitting the program into daily routine).

Table 1. Use of behavior change techniques in developing an intervention for people living with primary biliary cholangitis.

Enabler	Barrier	COM-B[a]/TDF[b]/IF[c]	Behavior change technique	Implementation of a behavior change technique
Interactions with program facilitators enhanced accountability		• COM-B: reflective motivation • TDF: goals • IF: persuasion	• 1.5 Review behavior goal(s) • 1.6 Discrepancy between current behavior and goal • 3.1 Social support (unspecified)	• 1.5 Weekly adherence vs target adherence goal were discussed during check in • 1.6 Weekly adherence vs target adherence goal were discussed during check in • 3.1 Weekly check ins employed motivational interviewing techniques to support program adherence
Able to integrate in everyday routine	Difficulty integrating program into daily routine	• COM-B: psychological capability • TDF: behavioral regulation • IF: enablement, persuasion	• 1.4 Action planning (Future consideration • 1.6 Discrepancy between current behavior and goal • 2.2 Feedback on behavior • 15.3 Focus on past success (self-belief)	• 1.4 In week 1, participants watched an interactive video prompting them to plan their performance of the target behavior (adherence to the program at or above the set minimum adherence goal). This included committing to a personal adherence goal at or above the set minimum, and writing down (1) potential obstacles to meeting their adherence goal; and (2) actions that could be taken to avoid or overcome these obstacles. • 1.6 The host website recorded weekly participation (indicated by accessed content). At the top of the website, the user's current weekly participation was presented beside the user's adherence goal. • 2.2 The host website recorded weekly participation (indicated by accessed content). At the top of the website, the user's current weekly participation was presented. • 15.3 In week 1, participants watched an interactive video that prompted them to think about instances in which they successfully adhered to a goal.
Access to accommodations to physical activity program where needed	Insufficient access to accommodation to physical activity program	• COM-B: physical capability • TDF: skills • IF: enablement	• 4.1 Instruction on how to perform a behavior • 6.1 Demonstration of the behavior	• 4.1 Instruction for accommodations were provided • 6.1 Demonstration of accommodations were provided
Interaction with others in program associated with increased motivation		• COM-B: social opportunity • TDF: social influences • IF: persuasion, modeling	• 3.1 Social support (unspecified)	• 3.1 Participants were invited to weekly live group sessions in which they had the opportunity to participate in program practices with peers
Desire to feel better		• COM-B: reflective motivation • TDF: goals • IF: persuasion	• 5.1 Information about health consequences • 5.2 Information about emotional consequences	• 5.1 Introductory videos provided information about health consequences associated with participating in the program • 5.2 Introductory videos provided information about health consequences associated with participating in the program

Enabler	Barrier	COM-B[a]/TDF[b]/IF[c]	Behavior change technique	Implementation of a behavior change technique
	Difficult to participate when feeling unwell due to disease	• COM-B: physical capability, psychological capability • TDF: environmental context and resources • IF: environmental restructuring	• 12.1 Restructuring of the physical environment	• 12.1 Short meditations were provided that could be completed when individuals are not feeling as well • 12.1 All mindful movement was low intensity
Able to navigate website	Difficulty navigating website	• COM-B: psychological capability • TDF: memory attention and decision processes • IF: training	• 4.1 Instruction on how to perform a behavior • 6.1 Demonstration of the behavior (comparison of a behavior)	• 4.1 Individuals received an introduction to the online platform via zoom, in which the research assistant provided instruction on accessing the intervention. Written instructions were also forwarded to all participants in an email. • 6.1 Individuals received an introduction to the online platform via zoom in which the research assistant demonstrated accessing the intervention
Web-based format enhanced accessibility		• COM-B: psychological capability • TDF: environmental context and resources • IF: environmental restructuring	• 12.1 Restructuring the physical environment	• 12.1 Web-based format was maintained
	Physical movement was too difficult	• COM-B: physical capability • TDF: physical skills • IF: enablement, training, environmental restructuring	• 4.1 Instruction on how to perform a behavior • 6.1 Demonstration of the behavior (comparison of a behavior) • 12.1 Restructuring of the physical environment	• 4.1 Within each stream, the mindful movement videos featured description of how to perform each specific posture/exercise • 6.1 Within each stream, the mindful movement videos featured demonstration of how to perform each specific posture/exercise • 12.1 Two streams of mindful movement were implemented, which were differentiated by difficulty
	Physical movement was not difficult enough	• COM-B: reflective motivation • TDF: beliefs about capabilities • IF: environmental restructuring	• 12.1 Restructuring of the physical environment	• 12.1 Two streams of mindful movement were implemented, which were differentiated by difficulty
Feeling better/good after participating in intervention	Uncertain about benefit	• COM-B: reflective motivation • TDF: beliefs about consequences, reinforcement • IF: persuasion	• 5.1 Information about health consequences • 5.2 Information about emotional consequences • 9.1 Credible source	• 5.1 Introductory videos provided information about health consequences associated with participating in the program • 5.2 Introductory videos provided information about health consequences associated with participating in the program • 9.1 Introductory videos featured health care professionals discussing potential benefits associated with participating in the program

Enabler	Barrier	COM-B[a]/TDF[b]/IF[c]	Behavior change technique	Implementation of a behavior change technique
	Fear of getting injured during physical activity	• COM-B: reflective motivation • TDF: beliefs about consequences • IF: education, environmental restructuring	• 9.1 Credible source • 12.1 Restructuring of the physical environment	• 9.1 Welcome video featured a health care professional explaining that mindful movement was designed to be safe for PBC. • 12.1 Various streams of mindful movement were available, separated by difficulty. Adaptations were available within mindful movement.
Repetition in physical activity program helped build routine		• COM-B: psychological capability • TDF: memory, attention, and decisional processes • IF: enablement	• 8.3 habit formation	• 8.3 Routine varied but structure was conveyed through repetition of the same type of activity from week to week (eg, 1 day of each week was dedicated to a breath program, 1 day a flow day)

[a]COM-B: Capability, Opportunity, Motivation and Behaviour.

[b]TDF: theoretical domains framework.

[c]IF: intervention functions.

Step 3: Identify Intervention Functions to Target Explanatory Domains

The intervention functions persuasion, environmental restructuring, and education were used to target theoretical domains relating to motivation. The intervention functions persuasion, enablement, training, and environmental restructuring were selected to target theoretical domains related to capability, and the intervention functions persuasion and modelling were selected to target theoretical domains related to opportunity. See Table 1 for a full outline of the intervention functions selected for each domain**.**

Step 4: Specify Intervention Content by Selecting Relevant BCTs

The comprehensive list of selected BCTs along with a description of how they were operationalized can be found in Table 1. Examples of how selected BCTs were translated into each of the general intervention components are detailed in the following.

Implementation of BCTs Into Core Practice

To address the behavior barrier "physical movement was too difficult," the BCTs "including instructions on how to perform a behaviour," "demonstration of the behaviour," and "restructuring of the physical environment" were employed. These were operationalized by including short videos to describe and demonstrate each exercise featured in the mindful movement routines, and through restructuring the program to include participant choice between a chair versus a standing stream of mindful movement.

Implementation of BCTs Into Positive Psychology

The BCTs "action planning" and "focus on past successes" were integrated into the positive psychology portion of the program to help address the barrier "integrating the program into daily routine." Specifically, an interactive positive psychology activity at the beginning of the program was created to prompt participants to set their adherence goal, schedule their behavior, consider potential barriers and facilitators to behavior, and think about past successes with behavior change.

Implementation of BCTs Into Weekly Communications

To address the behavior facilitator "interactions with others enhances accountability," we selected the BCTs "social support (unspecified)," "review behaviour goals," and "discrepancy between current behaviour and goal." During the weekly phone check-ins, a program facilitator will implement these BCTs by providing social support through brief weekly motivational interviewing touchpoints, revisiting the participant's initial goals, and discussing weekly adherence versus initial adherence goals.

Discussion

Principal Findings

The PPP_{PBC} intervention was developed to provide individuals with PBC a tool to help better manage their symptom burden. The intervention was designed to optimize participation by enhancing a participant's physical capability (ie, enable participation in a stream of mindful movement), psychological capability (ie, enable self-regulation), automatic motivation (ie, help participants build a routine), reflective motivation (ie, building intention to participate in wellness practices), and social opportunity (ie, connect with peer models). Owing to the web-based nature of this intervention, we were not able to alter the individual's physical environment and therefore did not target physical opportunity. Capability, opportunity, and motivation were targeted through the intervention functions persuasion, education, modeling, enablement, environmental restructuring (restructuring of intervention platform), and training. Additionally, 13 BCTs from the BCT taxonomy v1 were chosen to deliver the intervention content.

Utility of a Theoretical Framework

Informing behavioral interventions by theory not only provides a means to increase the efficacy of these interventions, but also allows researchers to standardize reporting of the active ingredients of interventions through BCTs. Current guidelines for reporting behavioral interventions are largely focused on reporting intervention delivery rather than intervention content [29,30]. Consequently, few reports detail active components of existing behavioral interventions and often use different language to describe active components. This presents a barrier to evaluating and replicating aspects of interventions that effectively bring about behavioral change. Experts in behavioral medicine have reported a low level of confidence in their ability to replicate effective behavioral interventions, which is likely linked to poor reporting of these interventions [19]. The current intervention is among a small number of multicomponent behavioral interventions to report on theoretically informed intervention development in a standardized manner [30]. In addition, this is the first known mind-body intervention tailored to PBC. This report provides a basis for (1) better consensus to be reached around a standardized approach to employing behavior change theory to inform an intervention and (2) evidence to be synthesized around which BCTs are effective in the context of an intervention. Both of these factors will allow for replication of successful aspects of implementation and successful active components. Importantly, after study rollout is complete, subsequent qualitative and quantitative assessment of behavior change will be necessary to determine successful components of the intervention.

Limitations

This project is not without limitations that should be acknowledged. The qualitative feedback used to inform barriers and facilitators to participating in the intervention was provided by participants with IBD, with no large-scale data collection occurring from individuals with PBC. Given the similarity of the symptom burden experienced with IBD and PBC (eg, fatigue, depression, anxiety, stress) the barriers and facilitators provided in the interviews were deemed to be applicable to PBC. To further mitigate this limitation, we worked with an advisory team of patients with PBC to better understand how intervention design needed to be tailored to meet the specific needs of this population (eg, providing a chair stream within the mindful movement to accommodate for potential fatigue and mobility restrictions).

Conclusions

To our knowledge, the PPP_{PBC} intervention is unique in that it is a mind-body wellness program designed for individuals with PBC, and in that it has taken a structured approach to considering theory in design and evaluation. Development was informed by the BCW [16] and BCTs [19]. Application of these frameworks was guided by feedback from our patient advisory team. Further standardized reporting of complex interventions conducted in different contexts, along with subsequent assessment of behavior change, is necessary to determine how contextual variables influence the effectiveness of different BCTs.

Acknowledgments

We would like to gratefully acknowledge the Canadian PBC Society, and specifically Gail Wright and Shauna Vander Well for their invaluable guidance during the design of the Peace Power Pack program. We would also like to acknowledge the support of the Mitacs Accelerate and the Canadian PBC Society for providing funding to support M Watt's first year of graduate studies.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Description of the web-based intervention.
[DOCX File , 16 KB - formative_v5i10e29064_app1.docx]

References

1. Clinical Review Report: Obeticholic Acid (Ocaliva): (Intercept Pharmaceuticals Canada, Inc.). Ottawa (ON): Canadian Agency for Drugs and Technologies in Health 2017 Aug. [Medline: 30525349]
2. Khanna A, Hegade V, Jones D. Management of Fatigue in Primary Biliary Cholangitis. Curr Hepatology Rep 2019 May 23;18(2):127-133 [FREE Full text] [doi: 10.1007/s11901-019-00458-0]
3. Reshetnyak VI. Primary biliary cirrhosis: Clinical and laboratory criteria for its diagnosis. World J Gastroenterol 2015 Jul 07;21(25):7683-7708 [FREE Full text] [doi: 10.3748/wjg.v21.i25.7683] [Medline: 26167070]
4. Witt-Sullivan H, Heathcote J, Cauch K, Blendis L, Ghent C, Katz A, et al. The demography of primary biliary cirrhosis in Ontario, Canada. Hepatology 1990 Jul;12(1):98-105. [doi: 10.1002/hep.1840120116] [Medline: 2197212]
5. Lee JY, Danford CJ, Trivedi HD, Tapper EB, Patwardhan VR, Bonder A. Treatment of Fatigue in Primary Biliary Cholangitis: A Systematic Review and Meta-Analysis. Dig Dis Sci 2019 Aug;64(8):2338-2350. [doi: 10.1007/s10620-019-5457-5] [Medline: 30632051]

6. Mells GF, Pells G, Newton JL, Bathgate AJ, Burroughs AK, Heneghan MA, UK-PBC Consortium. Impact of primary biliary cirrhosis on perceived quality of life: the UK-PBC national study. Hepatology 2013 Jul;58(1):273-283. [doi: 10.1002/hep.26365] [Medline: 23471852]
7. Cauch-Dudek K, Abbey S, Stewart DE, Heathcote EJ. Fatigue in primary biliary cirrhosis. Gut 1998 Nov;43(5):705-710 [FREE Full text] [doi: 10.1136/gut.43.5.705] [Medline: 9824355]
8. Rudic JS, Poropat G, Krstic MN, Bjelakovic G, Gluud C. Ursodeoxycholic acid for primary biliary cirrhosis. Cochrane Database Syst Rev 2012 Dec 12;12:CD000551 [FREE Full text] [doi: 10.1002/14651858.CD000551.pub3] [Medline: 23235576]
9. Carey EJ, Ali AH, Lindor KD. Primary biliary cirrhosis. Lancet 2015 Oct 17;386(10003):1565-1575. [doi: 10.1016/S0140-6736(15)00154-3] [Medline: 26364546]
10. Shohani M, Kazemi F, Rahmati S, Azami M. The effect of yoga on the quality of life and fatigue in patients with multiple sclerosis: A systematic review and meta-analysis of randomized clinical trials. Complement Ther Clin Pract 2020 May;39:101087. [doi: 10.1016/j.ctcp.2020.101087] [Medline: 32379628]
11. Araujo RV, Fernandes AFC, Nery IS, Andrade EMLR, Nogueira LT, Azevedo FHC. Meditation effect on psychological stress level in women with breast cancer: a systematic review. Rev Esc Enferm USP 2019;53:e03529 [FREE Full text] [doi: 10.1590/S1980-220X2018031303529] [Medline: 31800820]
12. Izgu N, Gok Metin Z, Karadas C, Ozdemir L, Metinarikan N, Corapcıoglu D. Progressive Muscle Relaxation and Mindfulness Meditation on Neuropathic Pain, Fatigue, and Quality of Life in Patients With Type 2 Diabetes: A Randomized Clinical Trial. J Nurs Scholarsh 2020 Sep;52(5):476-487. [doi: 10.1111/jnu.12580] [Medline: 32536026]
13. Jennings HM, Morrison J, Akter K, Kuddus A, Ahmed N, Kumer Shaha S, et al. Developing a theory-driven contextually relevant mHealth intervention. Glob Health Action 2019;12(1):1550736 [FREE Full text] [doi: 10.1080/16549716.2018.1550736] [Medline: 31154988]
14. Cotter AP, Durant N, Agne AA, Cherrington AL. Internet interventions to support lifestyle modification for diabetes management: a systematic review of the evidence. J Diabetes Complications 2014;28(2):243-251 [FREE Full text] [doi: 10.1016/j.jdiacomp.2013.07.003] [Medline: 24332469]
15. Hagger MS, Weed M. DEBATE: Do interventions based on behavioral theory work in the real world? Int J Behav Nutr Phys Act 2019 Apr 25;16(1):36 [FREE Full text] [doi: 10.1186/s12966-019-0795-4] [Medline: 31023328]
16. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
17. Robinson E, Higgs S, Daley AJ, Jolly K, Lycett D, Lewis A, et al. Development and feasibility testing of a smart phone based attentive eating intervention. BMC Public Health 2013 Jul 09;13:639 [FREE Full text] [doi: 10.1186/1471-2458-13-639] [Medline: 23837771]
18. Mangurian C, Niu GC, Schillinger D, Newcomer JW, Dilley J, Handley MA. Utilization of the Behavior Change Wheel framework to develop a model to improve cardiometabolic screening for people with severe mental illness. Implement Sci 2017 Nov 14;12(1):134 [FREE Full text] [doi: 10.1186/s13012-017-0663-z] [Medline: 29137666]
19. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
20. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. Implement Sci 2012 Apr 24;7:37 [FREE Full text] [doi: 10.1186/1748-5908-7-37] [Medline: 22530986]
21. Watt M, Hyde A, Madsen K, Peerani F, Tandon P. A169 exploring patient perspectives on an online stress reduction based wellness intervention in patients with inflammatory bowel disease (IBD). J Can Assoc Gastroenterol 2021;4:179-180 [FREE Full text] [doi: 10.1093/jcag/gwab002.167]
22. Borren NZ, van der Woude CJ, Ananthakrishnan AN. Fatigue in IBD: epidemiology, pathophysiology and management. Nat Rev Gastroenterol Hepatol 2019 Apr;16(4):247-259. [doi: 10.1038/s41575-018-0091-9] [Medline: 30531816]
23. Knowles SR, Graff LA, Wilding H, Hewitt C, Keefer L, Mikocka-Walus A. Quality of Life in Inflammatory Bowel Disease: A Systematic Review and Meta-analyses-Part I. Inflamm Bowel Dis 2018 Mar 19;24(4):742-751. [doi: 10.1093/ibd/izx100] [Medline: 29562277]
24. Webster R, Michie S, Estcourt C, Gerressu M, Bailey JV, MenSS Trial Group. Increasing condom use in heterosexual men: development of a theory-based interactive digital intervention. Transl Behav Med 2016 Sep;6(3):418-427 [FREE Full text] [doi: 10.1007/s13142-015-0338-8] [Medline: 27528531]
25. Tombor I, Shahab L, Brown J, Crane D, Michie S, West R. Development of SmokeFree Baby: a smoking cessation smartphone app for pregnant smokers. Transl Behav Med 2016 Dec;6(4):533-545 [FREE Full text] [doi: 10.1007/s13142-016-0438-0] [Medline: 27699682]
26. The Wellness Toolbox. URL: https://wellnesstoolbox.ca/ [accessed 2021-09-10]
27. Inouye SK, Bogardus ST, Williams CS, Leo-Summers L, Agostini JV. The role of adherence on the effectiveness of nonpharmacologic interventions: evidence from the delirium prevention trial. Arch Intern Med 2003 Apr 28;163(8):958-964. [doi: 10.1001/archinte.163.8.958] [Medline: 12719206]

28. Blackburn P, Freeston M, Baker CR, Jones DEJ, Newton JL. The role of psychological factors in the fatigue of primary biliary cirrhosis. Liver Int 2007 Jun;27(5):654-661. [doi: 10.1111/j.1478-3231.2007.01500.x] [Medline: 17498251]
29. Wood CE, Hardeman W, Johnston M, Francis J, Abraham C, Michie S. Reporting behaviour change interventions: do the behaviour change technique taxonomy v1, and training in its use, improve the quality of intervention descriptions? Implement Sci 2016 Jun 07;11(1):84 [FREE Full text] [doi: 10.1186/s13012-016-0448-9] [Medline: 27268131]
30. Bourne JE, Ivanova E, Gainforth HL, Jung ME. Mapping behavior change techniques to characterize a social cognitive theory informed physical activity intervention for adults at risk of type 2 diabetes mellitus. Transl Behav Med 2020 Aug 07;10(3):705-715. [doi: 10.1093/tbm/ibz008] [Medline: 30715534]

Abbreviations

BCT: behavior change technique
BCTv1: behavior change technique taxonomy version 1
BCW: Behaviour Change Wheel
COM-B: Capability, Opportunity, Motivation and Behaviour
HRQOL: health-related quality of life
IBD: inflammatory bowel disease
PBC: primary biliary cholangitis
PPP_{PBC}: Peace Power Pack PBC
TDF: theoretical domains framework

Edited by G Eysenbach; submitted 24.03.21; peer-reviewed by F Ghezelbash, E Toki; comments to author 10.05.21; revised version received 06.06.21; accepted 11.06.21; published 08.10.21.

Please cite as:
Watt M, Spence JC, Tandon P
Development of a Theoretically Informed Web-Based Mind-Body Wellness Intervention for Patients With Primary Biliary Cholangitis: Formative Study
JMIR Form Res 2021;5(10):e29064
URL: https://formative.jmir.org/2021/10/e29064
doi:10.2196/29064
PMID:34623317

Original Paper

An mHealth App to Support Fertility Patients Navigating the World of Infertility (Infotility): Development and Usability Study

Katya Kruglova[1,2], BA; Siobhan Bernadette Laura O'Connell[1,2], MA; Shrinkhala Dawadi[1,2], MSc; Eden Noah Gelgoot[1,2], MSc; Skye A Miner[1], PhD; Stephanie Robins[1,2], MSc; Joy Schinazi[1,2], MPH; Phyllis Zelkowitz[1,3], EdD

[1]Lady Davis Institute for Medical Research, Jewish General Hospital, Montreal, QC, Canada

[2]Department of Psychiatry, Jewish General Hospital, Montreal, QC, Canada

[3]Department of Psychiatry, McGill University, Montreal, QC, Canada

Corresponding Author:
Phyllis Zelkowitz, EdD
Lady Davis Institute for Medical Research, Jewish General Hospital
Ludmer Research & Training Bldg
Montreal, QC, H3A 1A1
Canada
Phone: 1 514 340 8222
Email: phyllis.zelkowitz@mcgill.ca

Abstract

Background: The experience of infertility and its treatment engenders considerable stress and is often described as an emotional rollercoaster. A mobile health (mHealth) app may be a novel solution to address the psychoeducational and psychosocial support needs of fertility patients because of its potential to reduce stress and increase patient empowerment. There are a few fertility-related apps that provide information and support to both men and women undergoing fertility treatment; however, none have documented their development and evaluation process.

Objective: This study aims to describe the development and evaluation process of a bilingual mHealth app, *Infotility*, designed to meet the psychoeducational and psychosocial support needs of men and women undergoing fertility treatment.

Methods: To develop the *Infotility* app, we adhered to the Medical Research Council guidelines for the development and evaluation of complex interventions. First, we conducted literature reviews and needs assessment surveys of fertility patients and health care providers who informed the content and design of the app. Second, we tested the intervention with a small group of end users who provided feedback on the design and appropriateness of the app's content. Third, we evaluated the uptake and usability of the app using a pre-post study design. Finally, we updated the app's content based on participants' feedback and searched for partners to disseminate the app to the broader public.

Results: This study is the first to describe the development and evaluation process of an mHealth app for men and women undergoing fertility treatment. The app met its goal in providing fertility patients with a clinician-approved, portable resource for reliable information about medical and psychosocial aspects of infertility and its treatments and a confidential peer support forum monitored by trained peer supporters. Participants rated the engagement, functionality, information, and esthetics of the app positively, with an overall app quality mean score of 3.75 (SD 0.53) and a star rating of 3.43 (SD 0.75), with a total possible score and star rating of 5.00.

Conclusions: By documenting the systematic development and evaluation of the mHealth app for men and women undergoing fertility treatment, this paper can facilitate the replication of the study intervention and the development of similar mHealth apps.

(JMIR Form Res 2021;5(10):e28136) doi:10.2196/28136

KEYWORDS

mHealth app; mHealth development process; infertility; intervention design; mobile phone

Introduction

The Negative Psychological Consequences of Infertility

Infertility is defined as the inability to achieve pregnancy after 12 months of unprotected sexual intercourse or the inability to reproduce either as an individual or as a couple [1]. Estimates suggest that 11%-16% of Canadians experience infertility in their lifetimes [2]. Infertility is a challenging experience, with infertile individuals demonstrating higher levels of stress, anxiety, depressive symptomatology, and stigma compared with their fertile counterparts [3-5]. Although infertile women tend to feel more stigmatized by infertility and report higher levels of depression and self-blame compared with infertile men [4], men do experience physical and emotional stress, particularly after treatment failure [6]. However, even in cases of male factor infertility, the focus of treatment is usually on the woman's body, which may make men feel excluded from the fertility treatment process and suspend their own emotional needs to meet the support needs of their female partners [7]. Fertility patients may also experience stress because of the physically arduous, costly, and time-consuming nature of treatment [3]. Although infertility is stressful, many patients do not present with clinical diagnoses of psychological disorders and often do not seek formal mental health services [8-10]. Fertility patients may benefit from alternative options for support such as psychoeducational materials, practical information about treatment and test procedures, and peer support.

Reviews of psychosocial interventions for fertility patients indicate that they are effective at reducing depressive symptomatology, anxiety, and stress associated with infertility [11-13]. Research suggests that patients desire more information from their health care providers about the emotional and psychological aspects of infertility and about treatment and test procedures [14]. Patients also demonstrate an unmet need for information about accessing psychological support services [15,16] and express interest in online peer support [17]. In addition to the clinical setting, fertility patients often search for web-based health information [18]. However, existing web-based resources for fertility patients often do not meet the standards of readability and accuracy, nor do they contain information about male infertility [19].

Mobile Health Interventions

Mobile health (mHealth) is the provision of health services and information using a mobile device, such as a smartphone [20,21]. Generally, mHealth interventions have a positive effect on users and can help patients with treatment adherence and symptom monitoring [20,22]. Those interventions that include psychoeducation, online peer support, and cognitive-behavioral or mindfulness-based therapies can improve clinical outcomes among users with a variety of health conditions [23-25]. Evidence suggests that mHealth can foster behavioral changes [20,26]. mHealth apps also have the potential to increase patient involvement and feelings of control over the treatment process [27] and provide more personalized care to users. The interactive and multimedia nature of mobile devices also allows for innovation in the presentation of information. This is an important aspect of health-related interventions; visual appeal is related to patient trust, perceived ease of use, and increased understanding [28].

Almost 70% of Canadian adults aged 18 years and older report owning a smartphone, and the rates of internet use and smartphone ownership continue to increase worldwide [29]. Therefore, mHealth may be an effective way to target patient populations who traditionally experience barriers to accessing formal health care services, such as men, immigrants, ethnic minorities and people with stigmatized illnesses [30-33]. Given its potential to reduce stress, increase patient empowerment, and provide user-friendly information to a broad range of the population, an mHealth app may be a novel solution to address fertility patients' psychoeducational and psychosocial support needs.

Documenting the mHealth App Development Process

To the best of our knowledge, there are a few fertility-related apps that provide information and support to both men and women undergoing fertility treatment, but none have documented their development and evaluation process. Improved documentation of app development will allow future researchers to develop similar mHealth apps. It will also help in knowledge transfer, which may be especially useful for the more innovative aspects of mHealth, such as design principles, the presentation of information, and technical features.

Accordingly, this study describes the development process for *Infotility*, a bilingual mHealth app designed to meet the psychoeducational needs of both men and women undergoing fertility treatment. *Infotility* contains information about treatments and test procedures, financial and legal aspects of fertility treatment, fertility health and risks to fertility, mental health and wellness, and a confidential forum monitored by peer supporters. In developing this intervention, we adhered to the Medical Research Council (MRC) guidelines for the development and evaluation of complex interventions [34]. As recommended, we used both quantitative (surveys) and qualitative (focus groups or interviews) methods during the development and evaluation of the intervention. We outline the development of the *Infotility* app based on the following recommended steps from the MRC guidelines:

1. *Development of the intervention*: literature review of existing interventions, needs assessments to determine stakeholders' perspectives on the content of the intervention
2. *Feasibility and piloting*: testing procedures, estimating recruitment and retention, and determining sample size
3. *Evaluation of the intervention*: pre-post study exploring fertility patients' experiences using the app
4. *Implementation*: dissemination, surveillance and monitoring, and long-term follow-up

Methods

Development of Infotility

Existing Sources of Web-Based Information and Support for Fertility Patients

To determine whether there were existing mHealth apps that provided fertility information to both men and women, we

reviewed the Apple iTunes and Google Play stores between October and December 2016, using the keywords *fertility/fertility* and *infertility/infertilité*. The search revealed numerous menstrual tracking apps and 2 sperm home testing kit apps, *Trak:* Sperm Health and Fertility (Sandstone Diagnostics) and *YO* Sperm Analyzer (Medical Electronic Systems). In a further search of the gray literature, we found that one menstrual cycle app, *Glow*, was developing a component for men that would include fertility-related information, but this was not yet available for review [35]. Our research team did not find any English or French language mHealth apps that addressed both male and female reproductive health concerns.

As part of a project to assess fertility-related information geared specifically for men, our research team performed an analysis of fertility-related health information found on the websites of fertility clinics and major North American organizations [19]. We assessed the quality and readability of information presented on the websites of 28 Canadian fertility clinics and 13 North American organizations such as the Mayo Clinic and RESOLVE. Quality ratings were assessed using the DISCERN instrument [36]. The quality of information found on the North American organizations' websites was deemed *good*, and the quality of information on the fertility clinic websites received a grade of *fair*. Furthermore, the North American organizations' websites required an average reading level of 12.9 years of education, and the fertility clinic websites required an average level of 14.3 years, far above the grade 5-8 reading level recommended for health information material [37,38]. The results of this analysis served as a proof of concept for *Infotility*, indicating a lack of web-based sources of high-quality information about reproductive health that might be easily understood by the general population.

In parallel, our research team conducted an internet search to review how infertility blogs and forums offered social support, how they were managed, and what information and features were available to users. Some of the reviewed web-based platforms offering peer support were RESOLVE, Reddit, Association Infertilité Québec, and Fertility Matters Canada. The search showed that infertility forums posted public (openly visible) and private (member log-in required) conversations on a variety of subjects surrounding the experiences of infertility. Moderators or administrators were available to provide feedback on some conversations (or *threads*) between users. Users were required to make an anonymizing username and icon (ie, one that did not identify them), and many forums had rules related to posting, such as the prohibition of using derogatory terms or naming specific professional or clinics. Most sites also offered freely available informational content, such as their own description of fertility or pregnancy, explanations of terms or procedures used during infertility treatment, and tools for tracking ovulation. Fertility forums existed in French or English, and some, such as Reddit [39], provided content geared specifically to men.

Needs Assessment Surveys

Overview

Contextual inquiry involves assessing the needs and preferences of end users (the population that will use the mHealth app) and key stakeholders (those involved in the creation, evaluation, and distribution of the mHealth app). For *Infotility*, end users are men and women undergoing fertility treatment and the key stakeholders are patient advocates and health care providers (HCPs) in the field of reproductive health. Identifying their needs and values (and later, evaluating whether the needs and values were met) allows for improved buy-in and increased potential for app distribution. To execute a thorough contextual inquiry and use a patient-centered approach to app development, our research team undertook needs assessment surveys of fertility patients and HCPs. All survey questions were designed by our research team, including expert clinicians in the field of reproductive health. To our knowledge, this is the first study to document such a process for a fertility-related app, although mHealth interventions for other conditions have used techniques such as interviews, expert panel consultation, and focus groups to better understand their target populations [40,41].

Needs Assessment Surveys of Fertility Patients' Use of and Desire for mHealth Apps

A diverse sample of fertility patients was recruited from fertility clinics in Montreal and Toronto to complete the needs assessment survey. A total of 659 patients completed a web-based survey asking them about their experiences of, and preferences for, fertility-related information disseminated through clinical, web-based, and mobile modalities. The sociodemographic characteristics of the survey sample are given in Multimedia Appendix 1. Dawadi et al [15] provided a more detailed description of methods and measures. The results of the needs assessment survey showed that most participants had searched the internet for fertility-related information, underscoring the importance of web-based health resources in addressing the informational needs of fertility patients [42]. Moreover, most participants did not report using a fertility mobile app, but the majority were interested in using one [43]. This finding served as another proof of concept for *Infotility*, as the discrepancy between use of and interest in using a fertility app suggested a lack of publicly available, good-quality web-based resources for fertility patients.

Participants were also asked about their preferred fertility app features. The five most endorsed features were as follows: being easy to understand, including a glossary of medical terms, providing information on fertility health care coverage, providing information on reproductive health, and being free of charge [43]. These findings informed our choice of content topics and features of *Infotility*.

The needs assessment survey also examined the fertility patients' interest in online peer support and their preferences for various features of an online peer support forum [17]. The majority of the participants expressed interest in using online peer support and endorsed a monitored peer support forum that is accessible on a mobile device, allows participants to connect with peers, contains links to external resources, and is monitored by a health

professional. These results affirmed our decision to include a peer support forum in the *Infotility* app and informed the design of the forum.

Needs Assessment Surveys of Health Care Providers

To obtain the perspectives of those who have frequent contact with fertility patients in a variety of capacities, our team conducted a needs assessment survey of fertility HCPs, including physicians, nurses, mental health and wellness professionals, and administrative staff. Eligible participants were recruited at 6 sites in Montreal and Toronto, and a total of 127 participants completed a web-based survey. The survey asked fertility HCPs about features they believed a fertility mobile app should include and the types of information that fertility patients typically requested.

The majority of HCPs thought that patients would be interested in using a high-quality mobile app that provides fertility-related information and support. With respect to the app features, the ones that were most highly endorsed by HCPs included being easy to understand, providing information that promotes reproductive health, containing a glossary of medical terms, and offering links to stress reduction tools [44]. Most of the features endorsed by HCPs were also identified by the surveyed fertility patients as most desirable in a mobile fertility app.

The survey also showed that HCPs most frequently provided patients with information on tests and procedures, medications, and explanations of conditions [45]. Comparing these findings with the results of the needs assessment survey of fertility patients demonstrated potential discrepancies between the types and amount of information provided by HCPs and those that fertility patients would have liked to receive. For example, many patients wanted to receive information on insurance and regulations but were less likely to obtain this information from HCPs. On the basis of these findings, our team decided to include detailed information on fertility laws, regulations, and health care coverage in the *Infotility* app.

Literature Review, Content Development, and Expert Input

To write the content for the *Infotility* app, our research team conducted a literature review on the topics related to infertility by searching web-based scholarly databases (eg, PubMed, PsycINFO, and Medline) for scientific literature published from 2000 to 2017. We also consulted gray literature, such as articles published in popular news media (eg, The New York Times) to capture personal experiences of fertility patients.

On the basis of the literature review as well as the responses from the needs assessment surveys, our team developed a range of content categories for the *Infotility* app. For each content topic, we created brief summaries of the scholarly and gray literature and collated this information into reports. These reports were then sent to members of the team and advisory committee who specialized in the content area and fertility patients for feedback. For example, a section on the causes of male factor infertility was sent to a urologist, whereas a section on the psychosocial aspects of infertility was sent to a clinical psychologist specializing in fertility counseling. These specialists ensured the clinical relevance and accuracy of the content. The summaries, once approved, formed the basis for the app content, which was edited and presented in clear and accessible language.

The final approved content for the *Infotility* app included a variety of informational topics such as content related to reproductive health, the psychosocial challenges of infertility, and the legal and financial aspects of fertility treatment (a detailed list of content topics are included in Textbox 1).

As the app was targeted to both female and male users, our research team made a concerted effort to identify content that expressed the experiences of both men and women undergoing fertility treatment. We did so by tailoring the language and information of certain content sections (eg, sections on nutrition and exercise) to be different for men and women.

Textbox 1. Content categories and articles of the Infotility app.

How to get pregnant

- Reproduction 101
- All about eggs and ovulation
- Tracking your ovulation
- Frequently asked questions about getting pregnant

Causes and diagnoses

- Female factor infertility causes
- Risks to female fertility
- Masturbation
- Other ways to provide a sperm sample
- Causes of male infertility
- Risks to male fertility
- Am I at risk for other health problems?

Treatment options

- Ovulation induction
- Intrauterine insemination
- In vitro fertilization
- Sperm, embryo, or egg donation

Using a donor or surrogate

- When to consider using donor eggs, sperm, embryos, or surrogate
- Preparing for donation
- Pros and cons of donation
- Challenges of egg, sperm, and embryo donation
- Medical steps and additional information

Genetic testing

- How does genetic testing work?
- What are the tests?

Multiple pregnancy losses

- Causes and treatments of multiple pregnancy losses

Stopping treatment

- Why do people stop treatment?
- If you decide to stop treatment

Fertility laws and health care coverage

- Assisted Human Reproduction Act
- What expenses are covered?
- Surrogacy and gamete donation laws
- On embryos as property and human life

Your relationships

- Keeping your couple healthy
- How do I talk about infertility?

Physical well-being

- Exercise and fertility
- Nutrition for fertility
- Environmental risks to fertility

Mental well-being

- Taking care of your body and mind
- Dealing with pregnancy loss

Working with your health care team

- Choosing the clinic that's right for you
- Preparing for medical appointments
- Questions you may want to ask your doctor
- Getting a second opinion: When and why?

In addition, our research team developed a glossary of key fertility-related terms using the *International Committee for Monitoring Assisted Reproductive Technology and the World Health Organization Revised Glossary on Assisted Reproductive Technology Terminology* as a reference [46]. We then added or removed certain terms based on their presence in the *Infotility* app content and edited the glossary to ensure its readability.

To serve the bilingual population of Montreal, Quebec, the app had to be available in both English and French. Therefore, once finalized, the written content was translated into French. The reading level of the content was assessed to ensure acceptability and appropriateness for the target audience. The English content was assessed using the Flesh Kincaid grade level [47], and the French content was evaluated using a web-based tool specifically designed to assess French written information using the Gunning Fog index [48,49]. Both measures are valid and reliable indexes commonly used to assess the grade level of educational knowledge required to understand written information. The measures indicated that the English content was written at an eighth-tenth grade reading level, whereas the French content was at the 12th grade level. This is slightly above the common guidelines recommending that health information material be written at a reading level of grade 5-8 [37,38]. However, previous literature has found that people seeking fertility treatment have, on average, higher education levels than other patient populations [50,51], and our needs assessment survey found that the majority of fertility patients in our sample had a university degree or higher [42]. Taking this into consideration, the reading level of the *Infotility* app is consistent with more flexible recommendations that content be developed at a reading level 1 to 3 grades lower than the mean education level of the target population [52].

In addition to providing information on the psychosocial and medical aspects of infertility in both English and French, the *Infotility* app integrated an online peer support forum *Connect* that allowed users to confidentially post about their experiences, communicate with each other via discussion posts, and ask private questions of a peer supporter. Screenshots of the peer support forum are provided in Multimedia Appendix 2. Peer supporters were current or former fertility patients who completed a training program developed by our research team. Training involved three components: (1) reviewing a peer support manual approved by an expert on peer support manuals [16], (2) watching a training webinar containing practice questions and a discussion of ideal responses, and (3) responding to hypothetical discussion posts followed by feedback from the research team. The peer support manual explained the role of a peer supporter [53], outlined strategies for providing web-based support to participants, and provided basic information about infertility and definitions of common medical terms. Our research team recruited and trained peer supporters to communicate with users via the forum and through private messages. Each peer supporter was asked to monitor the forum between 2 and 4 hours per week. The forum was also monitored by members of the research team to ensure that there were no posts that promoted specific products or clinics, gave medical advice to another participant, or indicated that a participant was distressed or considering self-harm. The team members also monitored the responses provided by peer supporters and were available to answer their questions. A study by Grunberg et al [53] provided a full description of methods.

App Design and Operationalization

To pick a name for the app and the peer support forum, members of the research team were polled, and the suggested names were voted on. Ultimately, the name *Infotility* was chosen for the app and the name *Connect* was chosen for the forum as both are short, unique, and easy to remember. Moreover, *Infotility* reflected the main purpose of the app, that is, providing information to fertility patients, and *Connect* embodied the spirit of the forum—connecting people via the internet. As the app had to be bilingual, these names were also chosen because they had suitable French equivalents—*Infotilité* and *Connecte.*

To design a user-friendly mHealth app, members of our research team participated in a 2-day workshop organized by an independent app design company. During this workshop, we completed a number of design activities, including persona development (representation of a typical user), brand conceptualization, journey mapping of the user, and the creation

of hypothetical wireframes (screen blueprints). We then supplied the ideas generated at the workshop to a different app design company with which we collaborated in the development of the app's user experience (overall experience a user has with the app), user interface (how the app interface functions and how the user interacts with it), and information architecture (how the app's content is structured). Figures 1 and 2 show examples of the interface and architecture of the app. The process of designing *Infotility* was informed by the principles of user-centered design, which focuses on the needs of the user at all phases of development [54]. With the understanding that the amount of information available to fertility patients can be overwhelming, we ensured that app users were able to choose the information they wished to have presented to them first. The information was further divided into digestible *chunks* and was tied together with an appealing theme to keep the app cohesive. The major organizing sections of the app were *What you need to know*, focusing on informational aspects of going through infertility and its treatment, and *What you can do*, focusing on actionable items, such as exercising for fertility.

Figure 1. Dashboard of the *Infotility* app.

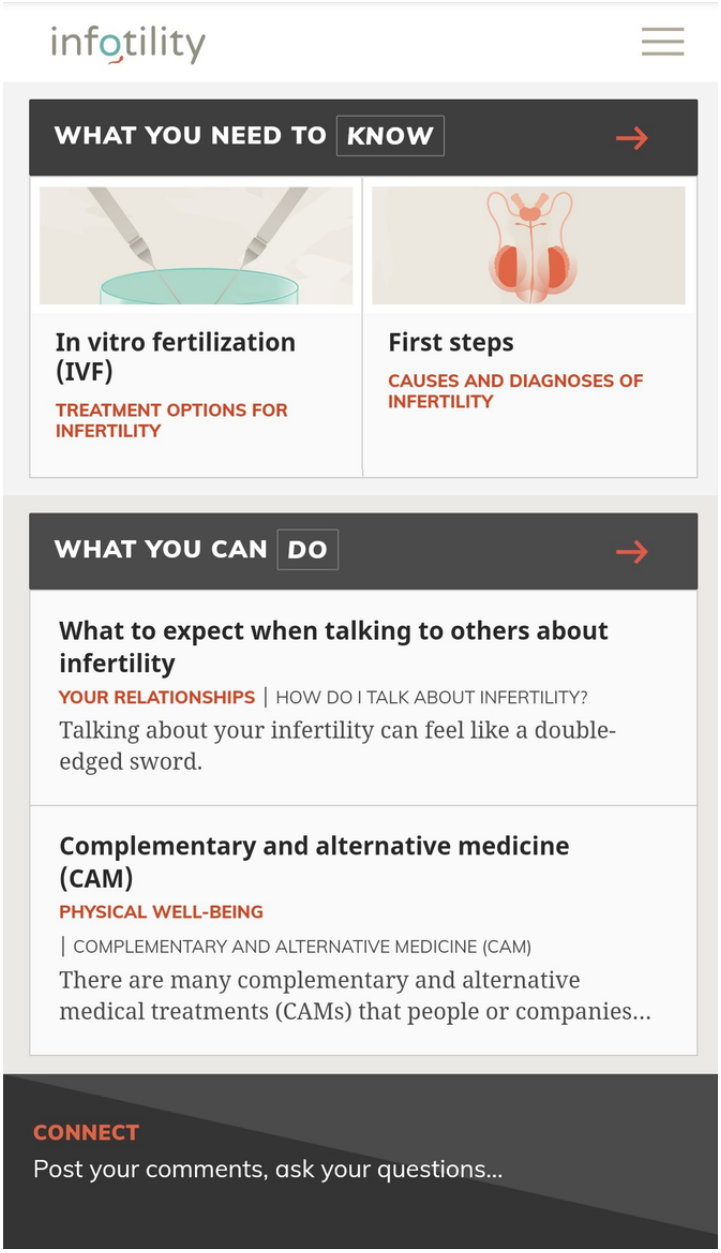
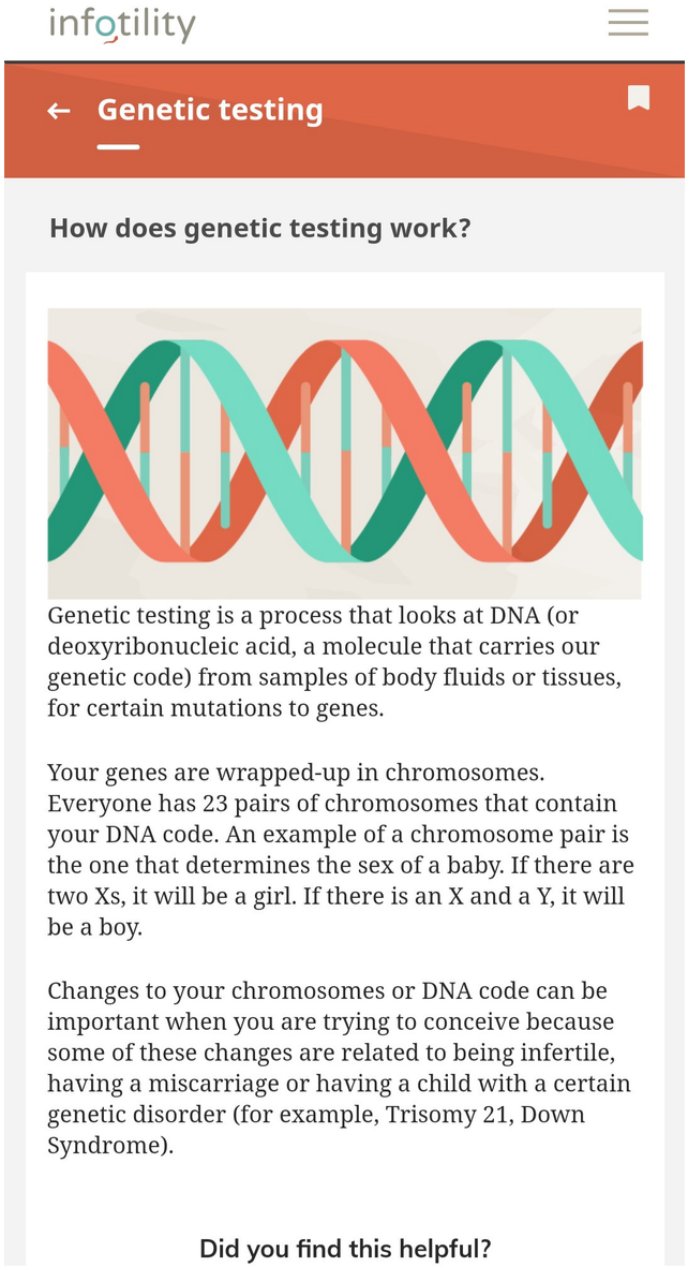


Figure 2. The article on genetic testing.



Our research team also worked with the app design company to create a theme and a mascot unique to *Infotility*. For the mascot, we chose a sperm whale. Once a color scheme and font were approved, the app design company produced graphics that either complemented or replaced the text in accordance with the guidance provided by the research team, which often included the whale mascot in different scenarios. Examples of the *Infotility* mascot and graphics are given in Multimedia Appendix 3. The intention of these graphics was to break up long blocks of text into smaller paragraphs to make the information easier to read and more appropriate for viewing on a mobile screen. In addition, although men and women had to use different log-in credentials to access the app because certain topics were sex-specific (eg, tips on how to provide a sperm sample for men), we created a feature called the *flip side*, which allowed users to view the opposite sex's app content. When a user clicked the *flip side* function, the app pages were animated to *flip* to see the opposite sex's information.

Feasibility and Piloting

An interactive prototype of the *Infotility* app was made available on a web-based platform. This prototype allowed us to discuss the flow and organization of information. Once the organization was approved, the app company consulted 2 male end users and 3 female end users who had undergone or were undergoing fertility treatment. These users provided feedback on the usability of *Infotility* and the design and appropriateness of the content. On the basis of this feedback, our research team adjusted the design, language, and tone of the app.

The sample size for the evaluation of the intervention was informed by guidelines for pilot studies of web-based interventions, which recommend at least 20 users [55]. To account for attrition and explore differences between men and women, we aimed to have at least 50 men and 50 women successfully finish the study by using the *Infotility* app for 8

weeks and completing questionnaires both before and after using the app.

Evaluation of Infotility

The primary goal of the study is to create an app that is user friendly, provides peer support, and contains reliable and easy-to-understand information about all aspects of infertility and its treatment. In the final steps of the study, we evaluated the uptake and usability of the *Infotility* app using a pre-post study design.

Participants and Procedures

Fertility patients were recruited from fertility clinics in Montreal and Toronto from October 2018 to December 2018 to test the *Infotility* app. Eligible participants met the following criteria: they were aged ≥18 years, were in a heterosexual relationship, read English and/or French, had access to the internet, and identified as male or female. In the initial stages of recruitment, we limited our inclusion criteria to only those undergoing in vitro fertilization for the first time. However, based on recommendations from our clinician partners, we expanded the inclusion criteria to recruit any fertility patients at any stage of treatment. There was great interest in the study among patients, and in the interest of being inclusive to all those who wished to participate, we did not limit the number of people who consented to participate in the study, even after our goal of 50 women and 50 men was achieved. This also allowed us to recruit a diverse sample and mitigate participant dropouts. When we felt we had a sufficient number of participants enrolled in the study to account for attrition over the 8-week study period, we stopped recruitment and determined April 30, 2019, as the study end date.

A total of 969 people (336/969, 34.6% men and 633/969, 65.3% women) were approached by recruiters. Furthermore, 661 (220/661, 33.2% men and 441/661, 66.7% women) agreed to be screened for eligibility, 505 (164/505, 32.4% men and 341/505, 67.5% women) were eligible to participate, and 387 (124/387, 32.0% men and 263/387, 67.9% women) consented to participate in the study. Before being given access to the *Infotility* app, participants were asked to complete a number of intake questionnaires measuring demographic characteristics, fertility treatments and diagnoses, fertility-related quality of life, psychological distress, and lifestyle habits, which took approximately 30 minutes to complete. Of the 387 who consented, 26% (65/250) men and 74% (185/250) women completed the intake questionnaires and visited the *Infotility* app at least once. Participants' sociodemographic characteristics are given in Multimedia Appendix 4. Participants were given access to *Infotility* for 8 weeks and could use the app as much or as little as they liked during the study period. Once using the app, participants could choose whether to access the peer support forum. After 8 weeks of using *Infotility*, of the 250 participants who used the app, 22.1% (38/172) men and 77.9% (134/172) women completed the follow-up questionnaires measuring participants' evaluations of and experiences using the app, in addition to the same measures administered at intake. Participants who completed the study were sent a Can $25 (US $32) gift card. For the purposes of this paper, data evaluating participants' experiences using the *Infotility* app, including the peer support forum *Connect*, are presented.

Measures

Google Analytics was used to collect data on how participants used the *Infotility* app during the 8-week study period. We gathered several key performance indicators to assess the frequency and patterns of app use, including the number of page views and total time spent on the app.

Participants were sent the user version of the Mobile Application Rating Scale (uMARS) to evaluate their satisfaction with the app and whether their needs and preferences were met [40]. The uMARS includes four subscales asking users to rate the app's quality of engagement, functionality, esthetics, and information [56]. Each item is measured on a Likert scale from 1-5, with higher scores representing higher quality ratings. The total scores for each subscale were generated by summing the individual items of the subscale and dividing it by the number of items in the subscale. The total app quality mean score was calculated by adding the scores from the four subscales together and dividing by four. There are four additional items of the uMARS that can be averaged to give an app *subjective* quality mean score. The research team chose the uMARS because it is a valid (α=.90) and reliable (test-retest reliability after 3 months = 0.63-0.85) measure of user satisfaction with mHealth apps and tailored toward the experiences of patients rather than professionals who work in technology or health care [56]. Three additional open-ended questions were developed by our research team to be administered after completing the uMARS, asking participants to describe (1) any topics or features that were not included in the app that they would have liked to be included, (2) what they liked best about the app, and (3) what they liked least about the app.

The Peer Support Evaluation Inventory (PSEI) was used to evaluate user satisfaction with the peer support forum. The PSEI was developed by our research team and adapted from a measure by Dennis [57]. The PSEI includes four subscales measuring supportive interactions, relationship qualities, perceived benefits, and satisfaction, with support received from the peer support forum. Each item is measured on a Likert scale from 1-4, with higher scores representing higher levels of user satisfaction with the peer support forum. The total scores for each subscale were generated by summing the individual items of the subscale and dividing by the number of items in the subscale.

Research Design and Analytic Strategy

Quantitative analyses were used to evaluate whether we achieved our goals of making the *Infotility* app useful and accessible to a sample of participants undergoing fertility treatment. Descriptive statistics of the uMARS and PSEI subscales present participants' overall ratings of the *Infotility* app and peer support forum. Bivariate analyses were used to determine whether there were any associations between the uMARS scores and app use. Qualitative responses to the three open-ended questions were read by 2 researchers, and common themes about the strengths and weaknesses of the app were identified and presented as supplementary data to the quantitative findings.

Results

Quantitative Findings

On average, participants visited approximately 34 pages and spent 22 minutes on the *Infotility* app. Participants rated the engagement, functionality, information, and esthetics of the app positively, with an overall app quality mean score of 3.75 (SD 0.53) and a star rating of 3.43 (SD 0.75), with a total possible score and star rating of 5.0. When asked whether they would recommend the app to other people who might benefit from it, 49.7% (94/189) responded "definitely" or "there were many people I would recommend this app to." When asked how many times they would use the app in the next 12 months, 5.7% (11/191) said they would use it more than 50 times, 25.6% (49/191) said 10 to 50 times, 42.9% (82/191) said 3 to 10 times, 19.3% (37/191) said 1 to 2 times, and 6.2% (12/191) said they would not use it. When asked if they would pay for the app, approximately 50.2% (95/189) said "Definitely not," and 2.1% (4/189) said "Definitely yes." A subsample of 106 *Infotility* users used the *Connect* forum. On average, *Connect* users rated the supportive interactions of peer supporters 2.91/4 (SD 0.84), the relationship qualities 2.95/4 (SD 0.49), their perceived benefits 2.92/4 (SD 0.74), and their satisfaction with support received 2.89/4 (SD 0.82; Table 1).

Table 1. App use and user ratings of the Infotility app and the Connect peer support forum (N=250).

Characteristics	Value, n (%)	Mean (SD)	Range
App use (Google Analytics)			
Total page views	250 (100)	33.93 (35.14)	1.00-270.00
Time spent on app (minutes)	250 (100)	22.07 (29.98)	0.00-193.18
uMARS[a]			
Total app quality	167 (66.8)	3.76 (0.53)	2.40-5.00
Engagement subscale	186 (74.4)	3.33 (0.64)	1.40-5.00
Functionality subscale	188 (75.5)	3.96 (0.65)	1.75-5.00
Esthetics subscale	190 (76)	3.73 (0.66)	1.67-5.00
Information subscale	174 (69.6)	3.97 (0.61)	2.00-5.00
Would you recommend this app to people who might benefit from it? (n=189)			
Not at all	9 (4.7)	N/A[b]	N/A
Very few people	23 (12.1)	N/A	N/A
Maybe	63 (33.3)	N/A	N/A
There are many people I would recommend this app to	51 (26.9)	N/A	N/A
Definitely	43 (22.7)	N/A	N/A
How many times do you think you would use this app in the next 12 months if it were relevant to you? (n= 191)			
None	12 (6.2)	N/A	N/A
1-2	37 (19.3)	N/A	N/A
3-10	82 (42.9)	N/A	N/A
10-50	49 (25.6)	N/A	N/A
>50	11 (5.7)	N/A	N/A
Would you pay for this app? (n= 189)			
Definitely not	95 (50.2)	N/A	N/A
2	36 (19.0)	N/A	N/A
3	37 (19.5)	N/A	N/A
4	17 (8.9)	N/A	N/A
Definitely yes	4 (2.1)	N/A	N/A
What is your overall (star) rating of the app? (n= 186)			
1	2 (1)	N/A	N/A
2	12 (6.4)	N/A	N/A
3	87 (46.7)	N/A	N/A
4	74 (39.7)	N/A	N/A
5	11 (5.9)	N/A	N/A
PSEI[c]			
Supportive Interactions subscale	98	2.91 (0.84)	1.00-4.00
Relationship Qualities subscale	91	2.95 (0.49)	2.00-4.00
Perceived Benefits subscale	90	2.92 (0.74)	1.00-4.00
Satisfaction With Support Received subscale	86	2.89 (0.82)	1.00-4.00

[a]uMARS: user version of the Mobile App Rating Scale.

[b]N/A: not applicable

[c]PSEI: Peer Support Evaluation Inventory.

Table 2 presents the correlations between the indicators of app use and the uMARS ratings of the app. Participants' scores on the functionality subscale were positively correlated with the number of page views (r=0.150; P=.04 for n=182) and the amount of time spent on the app (r=0.185; P=.01 for n=182). In addition, scores on the engagement subscale were correlated with the amount of time spent on the app, with a P value approaching significance, (r=0.142; P=.06 for n=180). This means that participants who felt the app functioned well and was easy to learn and navigate visited more pages and spent more time on the app, and participants who felt the app was more engaging spent more time on the app. Participants' scores on the esthetics and information subscales were not significantly correlated with app engagement.

Table 2. Pearson correlations between user ratings and engagement of the Infotility app.

uMARS[a] subscale	Total page views		Total time spent on the app	
	r	*P* value	*r*	*P* value
Engagement subscale	0.046	.54	0.142	.06
Functionality subscale	0.150[b]	.04	0.185[b]	.01
Esthetics subscale	0.096	.19	0.128	.08
Information subscale	0.051	.51	0.081	.30

[a]uMARS: user version of the Mobile App Rating Scale.

[b]Significant at α=.05.

Qualitative Feedback

Qualitative responses to the three open-ended questions administered at follow-up highlight some important findings about the strengths and limitations of the *Infotility* app. Overall, participants expressed that they appreciated the app: the "information is clear and easy to find" (Participant #03-0046), "it's a good tool to help and inform about infertility" (Participant #03-0014), and using the app was "more reassuring than googling a question" (Participant #03-0120). Some participants had suggestions for ways to improve the app, such as including more interactive features, communication with medical professionals, and informative videos:

> *An app usually should have some sort of interactive feature that brings you back, such as a fertility medication tracker, symptoms tracker, etc.* [Participant #03-0283]
>
> *It would be nice if this app was different, in the way that, it was live and interactive with specialists* [Participant #03-0248]
>
> *I think I would be happy to see short videos with the doctors of the clinic giving advices or telling a bit about their experience and statistics in infertility treatment* [Participant #03-0052]

There was also a desire for more in-depth information that was updated and customized to individual participants and personal testimonies from others:

> *No newsfeed feature. Update dashboard with new content/information based on my preferences. Why? To draw my attention, make me want to log in more often. I would see the same info each time I open the app. I would only log in if I'm looking for something in particular.* [Participant #03-0038]
>
> *There is a need for medical information that may or may not apply to you. Maybe cases? Personal stories? What happens to other people?* [Participant #03-0104]

Finally, many participants expressed that the peer support forum was their favorite part of the app and that using it reduced feelings of isolation, helped them manage stress, and provided valuable information. The *Connect* forum was perceived as confidential and safe, and peer supporters helped keep conversations respectful and on track [58].

Discussion

Principal Findings

To the best of our knowledge, this is the first study to describe the development and evaluation process of an mHealth app providing information and support to both men and women undergoing fertility treatment. The *Infotility* app was developed by our research team and was informed by extensive literature reviews and needs assessment surveys of fertility patients and HCPs. The content of *Infotility* was reviewed and approved by physicians, nurses, psychologists, and experts in the field of fertility to ensure its clinical relevance and accuracy. *Infotility* was designed by an app design company who worked alongside our team to ensure that the app was user friendly and esthetically pleasing. In addition to providing over 40 articles on the psychosocial and medical aspects of infertility, the *Infotility* app included a confidential peer support forum monitored by trained peer supporters.

In developing the *Infotility* app, our team adhered to the MRC guidelines for the development and evaluation of complex interventions, which provided a structured framework on how to approach the development, piloting, evaluation, and implementation phases of the study. By providing a complete description of all the steps of the app development process, this study can facilitate future replications of the study intervention and the development of similar mHealth apps.

Development of the Intervention

The needs assessment surveys of fertility patients and HCPs were a crucial step in identifying the informational and psychosocial needs of the end users of *Infotility*. By recruiting

a large and sociodemographically diverse sample, we were able to gain insights into fertility patients' experiences searching the internet for fertility-related information and their preferred features of a fertility app. Furthermore, obtaining the perspectives of fertility HCPs provided valuable insights into the types of information fertility patients most often requested and whether their needs were met.

The results of the needs assessment surveys guided our team throughout the entire development process of *Infotility* with respect to both its content and design. These results were especially helpful when deciding whether to include certain features on the app, which are costly and time consuming to develop. For example, the peer support forum and the detailed medical glossary were features of the *Infotility* app that took much time to develop and design. Despite this, we felt justified in including them on the app and confident in dedicating the time and resources to develop them based on the overwhelming evidence from our needs assessment surveys that these features would be beneficial to fertility patients. Furthermore, needs assessment surveys can provide insights into what you should *not* spend time and resources developing when the target audience informs you that they do not feel the need for certain features or topics. For example, we decided not to include celebrities' stories of dealing with infertility because they were rated as one of the least desired features by the surveyed fertility patients and HCPs. The results from the needs assessment surveys allowed our team to feel confident and justified throughout the entire app development process.

Feasibility and Piloting

The second stage of the MRC guidelines involves pilot-testing to determine the feasibility of complex interventions. This study assessed the feasibility of recruitment and retention of male and female participants and evaluated fertility patients' satisfaction with the *Infotility* app using the uMARS and qualitative interviews.

Recruitment of patients in waiting rooms of fertility clinics proved feasible: of 969 people approached by recruiters, 505 (52.1%; 64/505, 32.4% men and 341/505, 67.5% women) were eligible to participate and 39.9% (387/969; 124/387, 32.0% men and 263/387, 67.9% women) consented to participate in the study. However, recruitment of participants in fertility clinics might have excluded those with lower socioeconomic status who may not be able to afford fertility treatment. Nevertheless, within the limits of recruiting individuals who seek fertility treatment, we were able to obtain a sample that was diverse with respect to participants' household income (with about 30% below the median Canadian family income), ethnicity, immigrant status, and religion (Multimedia Appendix 4). Recruitment at fertility clinics might also have limited the sample size by excluding men and women who were not seeking fertility treatment but could have, however, benefited from the fertility-related information included in the app. Furthermore, this study only included heterosexual people; future studies should consider including single and nonheterosexual people to explore the experiences and psychoeducational needs of a more diverse population. In addition, the lag time between the needs assessment surveys and the launch of recruitment and data collection could have impacted the feasibility of the app intervention, as informational and support needs of end users may have evolved. Researchers planning to create similar tools should be cognizant of the inherent complexity of the development of an mHealth app and potential unanticipated delays during the process.

It is also worth noting that in line with the previous research regarding the lack of male participants in reproductive research [59], our team experienced difficulties in achieving the recruitment target of men and retaining those who agreed to participate in the study. For example, men were more likely to discontinue at some point throughout the study than women. Future studies should carefully consider the issues of recruitment and engagement of men to ensure that the psychoeducational needs of men with fertility concerns are addressed.

Evaluation of the Intervention

The evaluation of an mHealth intervention before disseminating it to a larger population or to the general public is necessary to ensure that it will be beneficial to its end users and successful in its proposed objectives. The pre-post study design proved to be effective in capturing fertility patients' experiences interacting with the *Infotility* app, including the peer support forum *Connect*. In general, participants enjoyed using *Infotility* and rated the informational side of the app and the peer support forum positively. The results showed that those who spent more time on *Infotility* were also those who rated the app higher on functionality (ie, how well the app functions and how easy it is to navigate) and engagement (ie, customization and interactivity). Our findings suggest that when developing mHealth apps, researchers and medical professionals should make the app engaging through interactive features and feedback and implement quality assurance procedures to address any technical issues that, if not resolved, may affect the level of trust among users and lead to user discontinuation [60].

The collection of both qualitative and quantitative data for the pre-post app evaluation provided nuanced and in-depth information about the app users' experiences using *Infotility* and what they felt the strengths and weaknesses of the app were. In addition, tracking participants' actual app use data through Google Analytics was highly informative and allowed our team to examine participants' self-reported data within the context of the pages they accessed on the app.

By tracking participants' actual app use data through Google Analytics, we were also able to gain insights into the amount of time participants spent on *Infotility* and whether they visited the app multiple times throughout the 8-week study period. Data showed that most participants used the app within the first 2 weeks of the study period, which might help explain the high attrition rate of the study. Although high attrition is common in studies of mHealth interventions [61], our findings suggest that using a study period of less than 8 weeks in future studies may help reduce attrition while providing sufficient information about participants' app use.

Careful analysis of participants' app use along with their qualitative and quantitative feedback is critical to improving user experiences of future versions of the mHealth app. For

example, future iterations of *Infotility* could be improved by including more interactive features that *bring you back* (eg, informative videos), more in-depth information on certain topics that were not extensively covered in the app, and summaries of current research in accessible language. Implementation of these features will help make future versions of the *Infotility* app more comprehensive and interactive, thereby potentially improving the rates of app engagement and user satisfaction.

Implementation

Finally, the evaluation of the *Infotility* app through a pre-post study design was necessary to facilitate the fourth stage of the MRC guidelines, which is implementation that includes dissemination, surveillance, and long-term follow-up.

Conducting repeated reviews of scientific literature is necessary to keep the content of *Infotility* accurate and up to date. Upon completion of the study, the entire content of the app was reviewed and edited by our research team based on feedback from the participants and by a knowledge translation consultant with expertise and personal experiences in the fertility field. Changes were made through a content management system, which is a user-friendly and accessible program that allowed our team to edit the app content in real time, without the need to outsource this task to the app company.

As this study is now complete, our team will search for partners that can maintain and disseminate *Infotility* to the broader public. We will identify potential app host partners, such as nonprofit health care organizations and health research institutes, whose strategic plan's objectives fit the mission of *Infotility*. We will then contact these organizations with a proposal and an estimated budget required to keep *Infotility* up to date, including the costs of literature and policy reviews, summarizing the information in lay terms, content translations and integration, and technical costs associated with domain registry and website hosting. On the basis of the results of the pre-post study, we can confidently conclude that fertility patients in Quebec and Ontario appreciated and enjoyed using the *Infotility* app, which will make the app more attractive to potential partners. Building a long-term partnership will help ensure that the *Infotility* app continues to provide accurate and reliable information to fertility patients.

Conclusions

Overall, the *Infotility* app succeeded in its goal to provide men and women undergoing fertility treatment with information and support through a user-friendly, credible, and single-source tool. The development and evaluation of *Infotility* highlighted the important aspects of the app creation process, which may be beneficial for researchers and medical professionals who wish to create similar mHealth apps in the future.

Acknowledgments

This study was funded by the Canadian Institute of Health Research Grant (Medical Research Council of Canada Operating Grant Program-136880).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sociodemographic characteristics of fertility patients who responded to the needs assessment survey (N=659).
[PDF File (Adobe PDF File), 160 KB - formative_v5i10e28136_app1.pdf]

Multimedia Appendix 2
Screenshots of the peer support forum Connect.
[PNG File , 1397 KB - formative_v5i10e28136_app2.png]

Multimedia Appendix 3
Screenshots of the Infotility mascot and graphics.
[PNG File , 208 KB - formative_v5i10e28136_app3.png]

Multimedia Appendix 4
Sociodemographic characteristics of study participants (N=250).
[PDF File (Adobe PDF File), 151 KB - formative_v5i10e28136_app4.pdf]

References

1. Zegers-Hochschild F, Adamson G, Dyer S, Racowsky C, de Mouzon J, Sokol R, et al. The international glossary on infertility and fertility care, 2017. Hum Reprod 2017 Sep 01;32(9):1786-1801 [FREE Full text] [doi: 10.1093/humrep/dex234] [Medline: 29117321]
2. Bushnik T, Cook JL, Yuzpe AA, Tough S, Collins J. Estimating the prevalence of infertility in Canada. Hum Reprod 2012 Jan 17;27(3):738-746. [doi: 10.1093/humrep/der465]

3. Cousineau TM, Domar AD. Psychological impact of infertility. Best Pract Res Clin Obstet Gynaecol 2007 Apr;21(2):293-308. [doi: 10.1016/j.bpobgyn.2006.12.003]
4. Fisher JR, Hammarberg K. Psychological and social aspects of infertility in men: an overview of the evidence and implications for psychologically informed clinical care and future research. Asian J Androl 2012 Jan;14(1):121-129 [FREE Full text] [doi: 10.1038/aja.2011.72] [Medline: 22179515]
5. Greil A, Slauson-Blevins K, McQuillan J. The experience of infertility: a review of recent literature. Sociol Health Illn 2010 Jan;32(1):140-162 [FREE Full text] [doi: 10.1111/j.1467-9566.2009.01213.x] [Medline: 20003036]
6. Peronace LA, Boivin J, Schmidt L. Patterns of suffering and social interactions in infertile men: 12 months after unsuccessful treatment. J Psychosom Obstet Gynaecol 2007 Jun;28(2):105-114. [doi: 10.1080/01674820701410049] [Medline: 17538818]
7. Hinton L, Miller T. Mapping men's anticipations and experiences in the reproductive realm: (in)fertility journeys. Reprod Biomed Online 2013 Sep;27(3):244-252. [doi: 10.1016/j.rbmo.2013.06.008] [Medline: 23871363]
8. Volgsten H, Skoog Svanberg A, Ekselius L, Lundkvist O, Sundström Poromaa I. Prevalence of psychiatric disorders in infertile women and men undergoing in vitro fertilization treatment. Hum Reprod 2008 Sep;23(9):2056-2063 [FREE Full text] [doi: 10.1093/humrep/den154] [Medline: 18583334]
9. Volgsten H, Schmidt L, Skoog Svanberg A, Ekselius L, Sundström Poromaa I. Psychiatric disorders in women and men up to five years after undergoing assisted reproductive technology treatment - a prospective cohort study. Hum Fertil (Camb) 2019 Dec;22(4):277-282. [doi: 10.1080/14647273.2018.1474279] [Medline: 29768933]
10. Pasch LA, Holley SR, Bleil ME, Shehab D, Katz PP, Adler NE. Addressing the needs of fertility treatment patients and their partners: are they informed of and do they receive mental health services? Fertil Steril 2016 Jul;106(1):209-15.e2. [doi: 10.1016/j.fertnstert.2016.03.006] [Medline: 27018159]
11. Boivin J. A review of psychosocial interventions in infertility. Soc Sci Med 2003 Dec;57(12):2325-2341. [doi: 10.1016/s0277-9536(03)00138-2] [Medline: 14572840]
12. Frederiksen Y, Farver-Vestergaard I, Skovgård NG, Ingerslev HJ, Zachariae R. Efficacy of psychosocial interventions for psychological and pregnancy outcomes in infertile women and men: a systematic review and meta-analysis. BMJ Open 2015 Jan 28;5(1):e006592 [FREE Full text] [doi: 10.1136/bmjopen-2014-006592] [Medline: 25631310]
13. de Liz TM, Strauss B. Differential efficacy of group and individual/couple psychotherapy with infertile patients. Hum Reprod 2005 May;20(5):1324-1332. [doi: 10.1093/humrep/deh743] [Medline: 15695317]
14. Dancet EA, Nelen WL, Sermeus W, De Leeuw L, Kremer JA, D'Hooghe TM. The patients' perspective on fertility care: a systematic review. Hum Reprod Update 2010;16(5):467-487. [doi: 10.1093/humupd/dmq004] [Medline: 20223789]
15. Dawadi S, Takefman J, Zelkowitz P. Fertility patients demonstrate an unmet need for the provision of psychological information: a cross sectional study. Patient Educ Couns 2018 Oct;101(10):1852-1858. [doi: 10.1016/j.pec.2018.06.013] [Medline: 30168420]
16. Read SC, Carrier M, Boucher M, Whitley R, Bond S, Zelkowitz P. Psychosocial services for couples in infertility treatment: what do couples really want? Patient Educ Couns 2014 Mar;94(3):390-395. [doi: 10.1016/j.pec.2013.10.025] [Medline: 24290241]
17. Grunberg PH, Dennis C, Da Costa D, Zelkowitz P. Infertility patients' need and preferences for online peer support. Reprod Biomed Soc Online 2018 Sep 15;6:80-89 [FREE Full text] [doi: 10.1016/j.rbms.2018.10.016] [Medline: 30547107]
18. Haagen E, Tuil W, Hendriks J, de Bruijn RP, Braat DM, Kremer JA. Current internet use and preferences of IVF and ICSI patients. Hum Reprod 2003 Oct;18(10):2073-2078. [doi: 10.1093/humrep/deg423] [Medline: 14507823]
19. Robins S, Barr HJ, Idelson R, Lambert S, Zelkowitz P. Online health information regarding male infertility: an evaluation of readability, suitability, and quality. Interact J Med Res 2016 Oct 21;5(4):e25. [doi: 10.2196/ijmr.6440] [Medline: 27769954]
20. Free C, Phillips G, Watson L, Galli L, Felix L, Edwards P, et al. The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. PLoS Med 2013 Jan 15;10(1):e1001363 [FREE Full text] [doi: 10.1371/journal.pmed.1001363] [Medline: 23458994]
21. Moss RJ, Süle A, Kohl S. eHealth and mHealth. Eur J Hosp Pharm 2018 Dec 21;26(1):57-58. [doi: 10.1136/ejhpharm-2018-001819]
22. Hamine S, Gerth-Guyette E, Faulx D, Green BB, Ginsburg AS. Impact of mHealth chronic disease management on treatment adherence and patient outcomes: a systematic review. J Med Internet Res 2015 Feb 24;17(2):e52. [doi: 10.2196/jmir.3951] [Medline: 25803266]
23. Berrouiguet S, Baca-García E, Brandt S, Walter M, Courtet P. Fundamentals for future mobile-health (mHealth): a systematic review of mobile phone and web-based text messaging in mental health. J Med Internet Res 2016 Jun 10;18(6):e135 [FREE Full text] [doi: 10.2196/jmir.5066] [Medline: 27287668]
24. Naslund JA, Marsch LA, McHugo GJ, Bartels SJ. Emerging mHealth and eHealth interventions for serious mental illness: a review of the literature. J Ment Health 2015 May 28;24(5):321-332 [FREE Full text] [doi: 10.3109/09638237.2015.1019054] [Medline: 26017625]
25. Donker T, Petrie K, Proudfoot J, Clarke J, Birch M, Christensen H. Smartphones for smarter delivery of mental health programs: a systematic review. J Med Internet Res 2013 Nov 15;15(11):e247 [FREE Full text] [doi: 10.2196/jmir.2791] [Medline: 24240579]

26. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
27. Martin T. Assessing mHealth: opportunities and barriers to patient engagement. J Health Care Poor Underserved 2012 Aug;23(3):935-941. [doi: 10.1353/hpu.2012.0087] [Medline: 24212144]
28. Lazard AJ, Watkins I, Mackert MS, Xie B, Stephens KK, Shalev H. Design simplicity influences patient portal use: the role of aesthetic evaluations for technology acceptance. J Am Med Inform Assoc 2016 Apr;23(e1):e157-e161 [FREE Full text] [doi: 10.1093/jamia/ocv174] [Medline: 26635314]
29. Poushter J. Smartphone ownership and internet usage continues to climb in emerging economies. Pew Research Center.: Pew Research Center; 2016 Feb 22. URL: https://www.pewresearch.org/global/2016/02/22/smartphone-ownership-and-internet-usage-continues-to-climb-in-emerging-economies/ [accessed 2021-01-15]
30. Price M, Yuen EK, Goetter EM, Herbert JD, Forman EM, Acierno R, et al. mHealth: a mechanism to deliver more accessible, more effective mental health care. Clin Psychol Psychother 2014;21(5):427-436 [FREE Full text] [doi: 10.1002/cpp.1855] [Medline: 23918764]
31. Schueller SM, Hunter JF, Figueroa C, Aguilera A. Use of digital mental health for marginalized and underserved populations. Curr Treat Options Psych 2019 Jul 5;6(3):243-255. [doi: 10.1007/s40501-019-00181-z]
32. Anderson-Lewis C, Darville G, Mercado RE, Howell S, Di Maggio S. mHealth technology use and implications in historically underserved and minority populations in the united states: systematic literature review. JMIR Mhealth Uhealth 2018 Jun 18;6(6):e128 [FREE Full text] [doi: 10.2196/mhealth.8383] [Medline: 29914860]
33. Berger M, Wagner TH, Baker LC. Internet use and stigmatized illness. Soc Sci Med 2005 Oct;61(8):1821-1827. [doi: 10.1016/j.socscimed.2005.03.025] [Medline: 16029778]
34. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. Int J Nurs Stud 2013 May;50(5):587-592. [doi: 10.1016/j.ijnurstu.2012.09.010] [Medline: 23159157]
35. Alba D. An App for Hacking Fertility Now Also Works for Men The Glow team is rounding out its app with a new component: support for male fertility. Wired. 2015 Apr 23. URL: https://www.wired.com/2015/04/glow/ [accessed 2021-01-15]
36. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111 [FREE Full text] [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
37. Mullen RJ, Duhig J, Russell A, Scarazzini L, Lievano F, Wolf MS. Best-practices for the design and development of prescription medication information: a systematic review. Patient Educ Couns 2018 Aug;101(8):1351-1367. [doi: 10.1016/j.pec.2018.03.012] [Medline: 29548600]
38. How to Create Effective Written Patient Learning Materials. Montreal, Canada: Jewish General Hospital Patient Education Network (PEN) Working Group; 2008.
39. Richard J, Badillo-Amberg I, Zelkowitz P. "So Much of This Story Could Be Me": men's use of support in online infertility discussion boards. Am J Mens Health 2017 May;11(3):663-673 [FREE Full text] [doi: 10.1177/1557988316671460] [Medline: 27702886]
40. Van Velsen L, Wentzel J, Van Gemert-Pijnen JE. Designing eHealth that matters via a multidisciplinary requirements development approach. JMIR Res Protoc 2013 Jun 24;2(1):e21 [FREE Full text] [doi: 10.2196/resprot.2547] [Medline: 23796508]
41. Vilardaga R, Rizo J, Zeng E, Kientz JA, Ries R, Otis C, et al. User-centered design of learn to quit, a smoking cessation smartphone app for people with serious mental illness. JMIR Serious Games 2018 Jan 16;6(1):e2 [FREE Full text] [doi: 10.2196/games.8881] [Medline: 29339346]
42. Brochu F, Robins S, Miner SA, Grunberg PH, Chan P, Lo K, et al. Searching the internet for infertility information: a survey of patient needs and preferences. J Med Internet Res 2019 Dec 12;21(12):e15132 [FREE Full text] [doi: 10.2196/15132] [Medline: 31829963]
43. Dawadi S, Chan P, Lo K, Mahutte N, Ouhilal S, Takefman J. Developing a fertility mHealth app: what features do patients want? In: Proceedings of the 64th annual meeting of the Canadian Fertility and Andrology Society. 2018 Presented at: 64th Annual Meeting of the Canadian Fertility and Andrology Society; Sep 13-15, 2018; Montreal, Canada.
44. Zelkowitz P. Needs assessment surveys of fertility patients and their health care providers: providing information and support through mobile solutions. 2017 Presented at: Andrology/Urology Rounds, Mount Sinai Hospital; 2017; Toronto.
45. Gelgoot EN, Kelly-Hedrick M, Miner SA, Zelkowitz P. The provision of information during fertility clinic visits: patient and provider perspectives. 2018 Presented at: 9th Annual Lady Davis Institute Scientific Retreat; 2018; Montreal, Canada.
46. Zegers-Hochschild F, Adamson GD, de Mouzon J, Ishihara O, Mansour R, Nygren K, International Committee for Monitoring Assisted Reproductive Technology, World Health Organization. International Committee for Monitoring Assisted Reproductive Technology (ICMART) and the World Health Organization (WHO) revised glossary of ART terminology, 2009. Fertil Steril 2009 Nov 15;92(5):1520-1524. [doi: 10.1016/j.fertnstert.2009.09.009] [Medline: 19828144]

47. Kincaid J, Fishburne R, Rogers R, Chissom B. Derivation of new readability formulas (automated readability index, FOG count, and Flesch Reading Ease Formula) for Navy enlisted personnel. Research Branch Report 1975:8-75. [doi: 10.21236/ada006655]
48. Gunning R. The fog index after twenty years. J Bus Commun 1969 Jan 01;6(2):3-13. [doi: 10.1177/002194366900600202]
49. Aarts JW, van Empel IW, Boivin J, Nelen WL, Kremer JA, Verhaak CM. Relationship between quality of life and distress in infertility: a validation study of the Dutch FertiQoL. Hum Reprod 2011 May;26(5):1112-1118. [doi: 10.1093/humrep/der051] [Medline: 21372046]
50. Jain T. Socioeconomic and racial disparities among infertility patients seeking care. Fertil Steril 2006 Apr;85(4):876-881. [doi: 10.1016/j.fertnstert.2005.07.1338] [Medline: 16580368]
51. Mott JM, Stanley MA, Street RL, Grady RH, Teng EJ. Increasing engagement in evidence-based PTSD treatment through shared decision-making: a pilot study. Mil Med 2014 Feb;179(2):143-149. [doi: 10.7205/MILMED-D-13-00363] [Medline: 24491609]
52. French KS, Larrabee JH. Relationships among educational material readability, client literacy, perceived beneficence, and perceived quality. J Nurs Care Qual 1999 Aug;13(6):68-82. [doi: 10.1097/00001786-199908000-00008] [Medline: 10476626]
53. Grunberg PH, Dennis C, Da Costa D, Gagné K, Idelson R, Zelkowitz P. Development and evaluation of an online infertility peer supporter training program. Patient Educ Couns 2020 May;103(5):1005-1012. [doi: 10.1016/j.pec.2019.11.019] [Medline: 31761526]
54. User centered design. Interaction Design Foundation. URL: https://www.interaction-design.org/literature/topics/user-centered-design [accessed 2021-07-14]
55. U.S. Department of Health and Human Services. Recruiting usability test participants. usability.gov. URL: https://www.usability.gov/how-to-and-tools/methods/recruiting-usability-test-participants.html [accessed 2021-01-15]
56. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the User Version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
57. Dennis C. Peer support within a health care context: a concept analysis. Int J Nurs Stud 2003 Mar;40(3):321-332. [doi: 10.1016/s0020-7489(02)00092-5] [Medline: 12605954]
58. O'Connell SB, Gelgoot EN, Grunberg PH, Schinazi J, Da Costa D, Dennis C, et al. 'I felt less alone knowing I could contribute to the forum': psychological distress and use of an online infertility peer support forum. Health Psychol Behav Med 2021 Feb 11;9(1):128-148 [FREE Full text] [doi: 10.1080/21642850.2021.1884556] [Medline: 34104553]
59. Culley L, Hudson N, Lohan M. Where are all the men? The marginalization of men in social scientific research on infertility. Reprod Biomed Online 2013 Sep;27(3):225-235. [doi: 10.1016/j.rbmo.2013.06.009] [Medline: 23871364]
60. 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 Oct 16;11:84 [FREE Full text] [doi: 10.1186/s13098-019-0480-4] [Medline: 31636719]
61. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. J Consult Clin Psychol 2020 Jan;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]

Abbreviations

HCP: health care provider
mHealth: mobile health
MRC: Medical Research Council
PSEI: Peer Support Evaluation Inventory
uMARS: user version of the Mobile Application Rating Scale

Edited by G Eysenbach; submitted 23.02.21; peer-reviewed by H Holzer, A Sible Prabhakar; comments to author 02.07.21; revised version received 20.07.21; accepted 01.08.21; published 12.10.21.

Please cite as:
Kruglova K, O'Connell SBL, Dawadi S, Gelgoot EN, Miner SA, Robins S, Schinazi J, Zelkowitz P
An mHealth App to Support Fertility Patients Navigating the World of Infertility (Infotility): Development and Usability Study
JMIR Form Res 2021;5(10):e28136
URL: https://formative.jmir.org/2021/10/e28136
doi:10.2196/28136
PMID:34636741

Original Paper

The Use of Task Shifting to Improve Treatment Engagement in an Internet-Based Mindfulness Intervention Among Chinese University Students: Randomized Controlled Trial

Marcus Rodriguez[1,2], PhD; Tory A Eisenlohr-Moul[3], PhD; Jared Weisman[4,5], BA; M Zachary Rosenthal[6], PhD

[1]Department of Psychology, Pitzer College, Claremont, CA, United States

[2]Boston Child Study Center, Los Angeles, CA, United States

[3]Neuropsychiatric Institute, Department of Psychiatry, University of Illinois at Chicago College of Medicine, Chicago, IL, United States

[4]Pitzer College, Claremont, CA, United States

[5]MCR Labs, LLC, Framingham, MA, United States

[6]Department of Psychology & Neuroscience, Duke University, Durham, NC, United States

Corresponding Author:
Jared Weisman, BA
Pitzer College
1050 N Mills Ave
BN 205
Claremont, CA, 91711
United States
Phone: 1 9784605088
Email: jweisman@pitzer.edu

Abstract

Background: Traditional in-person psychotherapies are incapable of addressing global mental health needs. Use of computer-based interventions is one promising solution for closing the gap between the amount of global mental health treatment needed and received.

Objective: Although many meta-analyses have provided evidence supporting the efficacy of self-guided, computer-based interventions, most report low rates of treatment engagement (eg, high attrition and low adherence). The aim of this study is to investigate the efficacy of an adjunctive treatment component that uses task shifting, wherein mental health care is provided by nonspecialist peer counselors to enhance engagement in an internet-based, self-directed, evidence-based mindfulness intervention among Chinese university students.

Methods: From 3 universities across China, 54 students who reported at least mild stress, anxiety, or depression were randomly assigned to a 4-week internet-based mindfulness intervention (MIND) or to the intervention plus peer counselor support (MIND+), respectively. *Be Mindful* delivers all the elements of mindfulness-based cognitive therapy in an internet-based, 4-week course. Participants completed daily monitoring of mindfulness practice and mood, as well as baseline and posttreatment self-reported levels of depression, anxiety, stress, and trait mindfulness. We screened 56 volunteer peer counselor candidates who had no former training in the delivery of mental health services. Of these, 10 were invited to participate in a day-long training, and 4 were selected. Peer counselors were instructed to provide 6 brief (15-20 minute) *sessions* each week, to help encouraging participants to complete the internet-based intervention. Peer counselors received weekly web-based group supervision.

Results: For both conditions, participation in the internet-based intervention was associated with significant improvements in mindfulness and mental health outcomes. The pre-post effect sizes (Cohen *d*) for mindfulness, depression, anxiety, and stress were 0.55, 0.95, 0.89, and 1.13, respectively. Participants assigned to the MIND+ (vs MIND) condition demonstrated significantly less attrition and more adherence, as indicated by a greater likelihood of completing posttreatment assessments (16/27, 59% vs 7/27, 26%; χ^2_1=6.1; *P*=.01) and a higher percentage of course completion (72.6/100, 72.6% vs 50.7/100, 50.7%; t_{52}=2.10; *P*=.04), respectively. No significant between-group differences in daily frequency and duration of mindfulness practice were observed. Multilevel logistic growth models showed that MIND+ participants reported significantly greater pre-post improvements in daily stress ratings (interaction estimate 0.39, SE 0.18; t_{317}=2.29; *P*=.02) and depression (interaction estimate 0.38, SE 0.16; t_{330}=2.37; *P*=.02) than those in the MIND condition.

Conclusions: This study provides new insights into effective ways of leveraging technology and task shifting to implement large-scale mental health initiatives that are financially feasible, easily transportable, and quickly scalable in low-resource settings. The findings suggest that volunteer peer counselors receiving low-cost, low-intensity training and supervision may significantly improve participants' indices of treatment engagement and mental health outcomes in an internet-based mindfulness intervention among Chinese university students.

(JMIR Form Res 2021;5(10):e25772) doi:10.2196/25772

KEYWORDS

mindfulness; mental health; social support; internet-based intervention; treatment outcome; university students; smartphone; mobile phone

Introduction

Background

Approximately 1 in 5 adults in the United States experiences mental illness annually [1], with young adults (aged 18-25 years) reporting the highest prevalence. In economic terms, these problems cost the United States more than US $193.2 billion each year, both in direct (eg, treatment) and indirect (ie, productivity loss at workplace, school, and home) expenses [2]. Globally, 3 out of 4 people report that they prefer therapy to psychopharmacology; however, despite the existence of empirically supported behavioral interventions targeting serious mental illness conditions, nearly two-third of US adults with these conditions do not receive services [3,4]. In short, there is a large gap between treatment needed and treatment received for mental disorders, and the mental health treatment gap is larger in low- and middle-income countries (LMICs) than in higher-income countries [5]. In China, more than 1 in 4 university students reported being depressed; moreover, although 4 out of 10 people worldwide suffer from a psychiatric illness at some point in their lives, nearly half of the world's population lives in a country with less than 1 psychiatrist per 100,000 people [6,7]. The reasons for these treatment gaps are multifaceted and include factors such as stigma, perceived helpfulness of treatments, and convenience. However, insufficient human resources and mental health infrastructure are arguably the greatest contributors.

Traditional psychotherapy training models (ie, 2-7 years of graduate school) and psychotherapies (one-on-one and in-person sessions) are incapable of singularly closing the global mental health gap in LMICs. Fundamentally, new approaches are needed to increase access to effective mental health care in an economic, feasible, and scalable manner to address global needs [8], and computer- and web-based interventions are promising solutions to closing the large gap between mental health treatment required and treatment received. In addition to addressing concerns of physical accessibility to treatment providers, web-based interventions accessed privately from the household allow individuals to avoid the perceived stigma associated with seeking mental health services [9]. Although not every country prioritizes the enhancement of their mental health infrastructure, most countries prioritize the development of their telecommunications infrastructure.

Mindfulness- and Acceptance-Based Interventions

Mindfulness- and acceptance-based interventions have been successfully used to address clinical dysfunction across a range of physical and mental health disorders. Specifically, systematic reviews and meta-analyses have demonstrated mindfulness and acceptance to be beneficial in treating physical health conditions, such as chronic pain [10], as well as mental health problems including depression [11,12], substance use disorders [13], eating disorders [14], anxiety [12,15], stress [16], and general psychological health [17]. A recent meta-analysis explored the efficacy of mindfulness-based interventions delivered through the internet [18]. Overall pre-post, between-group effects (Hedge *g*) were reported for stress (g=0.51), depression (g=0.29), anxiety (g=0.22), well-being (g=0.23), and mindfulness (g=0.32), all with nominal statistical significance thresholds of P<.50 [18]. Overall, the results of this meta-analysis provide promising initial evidence for the efficacy of mindfulness-based interventions delivered through the internet.

Existing Barriers to Treatment

Although technology-based mindfulness interventions can be effective, treatment engagement is a key barrier to the implementation of these approaches. High attrition and low adherence rates are commonly observed in research and practice. Nonadherence can diminish the effectiveness of interventions [19-21]. For example, although *PTSD Coach*, an app developed and used by the Department of Veteran's Affairs, was downloaded more than 150,000 times, fewer than 15% of the people used it within 1 week [22].

Adherence is especially relevant in mindfulness training because regular practice is thought to be essential for developing mindfulness skills. In their meta-analytic review of web-based mindfulness interventions, Spijkerman et al [18] reported adherence rates between 39.5% and 92%, when adherence was defined as completion of all sessions. However, adherence rates were only reported for 33% (5/15) of the studies that were included in the review.

Although many web-based mindfulness interventions have been shown to be effective, about half of the published studies did not report treatment engagement outcomes, and of those that did, most reported low engagement (high attrition and low adherence). Researchers have repeatedly noted that in spite of the effectiveness of web-based interventions, it is a consistent challenge to reliably measure the amount and quality of mindfulness practice with which people engage [23,24]. Moreover, previous approaches to improve the scale and scope

of mental health care systems in LMICs have largely failed because of budgetary constraints, lack of cost estimates, and logistical breakdowns, inhibiting the development of suitable infrastructure [25]. Novel, feasible, and scalable approaches are needed to not only improve treatment implementation but also treatment engagement. As an emerging platform for the provision of mental health interventions, web-based interventions need more quantitative data to evaluate factors associated with treatment efficacy, as well as the feasibility and acceptability of these approaches. Treatment attrition and adherence rates are two important indices of treatment feasibility that can be assessed in mobile interventions and correlated with factors that can be used to promote patient retention. Given the axiom that the development of viable mindfulness skills requires discipline and regular practice, further research is needed to explore the existing barriers to adherence to develop strategies for enhancing treatment engagement.

Existing Research

Previous research indicates that providing therapist support has a positive influence on adherence and enhances the effectiveness of web-based psychological interventions [26-28]. Spijkerman et al [18] reported larger effect sizes for mindfulness and stress for web-based mindfulness interventions with therapist guidance (g=0.89 and g=0.43, respectively), than for those interventions without guidance (g=0.19 and g=0.22, respectively). Although offering therapist guidance may potentially improve adherence and treatment outcomes, it is costly and may restrict the scalability of the intervention, particularly in LMICs. To overcome these barriers, some web-based interventions provide automated support, which has preliminary evidence supporting its efficacy [29-31]. For example, Oinas-Kukkonen and Harjumaa [32] suggest that system design and automated support may even be as effective as human support.

Task Sharing

Task sharing or *task shifting*, wherein mental health care is provided by nonspecialist peer counselors (eg, nurses, clergymen, teachers, community leaders), has recently been investigated as a promising strategy to overcome human resource shortages in LMICs. In this model of care, peer counselors receive training, supervision, and oversight from mental health care professionals (eg, psychiatrists, psychologists, and clinical social workers). Task shifting has been shown to be effective in the treatment of mental health problems [33,34], and in enhancing treatment engagement (eg, HIV medication compliance [35], child and maternal health care [36], and noncommunicable disease management [including depression] [37]). Thus, a task-shifting model shows promise as an alternative strategy to therapist guidance or peer support for enhancing treatment engagement and outcomes in self-guided, web-based mindfulness interventions.

Objectives

The aim of this study is to examine whether an adjunctive, task-shifting component (MIND+) enhances treatment engagement in a mindfulness intervention for stress and depression among Chinese undergraduate and graduate students. Individuals were randomly assigned to a brief (4-week), self-guided, web-based, mindfulness intervention (MIND), or the intervention plus support from nonspecialist peer counselors (MIND+). Peer counselors were instructed to engage in brief (15-20 minutes) weekly meetings with MIND+ participants via text or phone call during the course of treatment, with the intention of supporting and encouraging participants to complete the internet-based intervention. It was hypothesized that at posttreatment, participants randomly assigned to MIND+ (vs those assigned to MIND) would show (1) less attrition (higher completion rates of assessment), (2) greater adherence (higher percentage of course completion), (3) greater reductions in stress and depression levels, and (4) greater increases in mindfulness.

Methods

Participant Overview

Participants were 54 currently enrolled university students (undergraduate, master's, and doctoral programs) from 36 universities across China. Their mean age was 23.5 years (SD 3.17), and 74% (40/54) identified as female. Out of the 54 participants, 29 (54%) were master's students, 21 (39%) were undergraduate students, and 4 (7%) were doctoral students. All participants reported having passed an English proficiency test: 11% (6/54) reported passing the College English Test (CET; level 4); 72% (39/54), the CET (level 6); and 17% (9/54), the Test of English as a Foreign Language. All participants denied currently receiving formal mental health treatment; 80% (43/54) reported no history of mental health treatment, 11% (6/54) reported formerly receiving therapy, 4% (2/54) reported formerly receiving medication, and 6% (3/54) reported formerly receiving both treatment and medication. Out of the 54 participants, 47 (87%) reported no previous mindfulness training, and 3 (6%) reported practicing mindfulness meditation in the past year.

Participant Recruitment

Participants were recruited via WeChat blogs, student club listservs, and university websites' listing of available jobs and research opportunities. Interested individuals completed a web-based screening assessment. Eligible students were contacted by the study coordinator, who conducted phone interviews and orientation to the study procedures. Students who provided proof of student status and emergency contact information received a link to the baseline assessment measures. Students who completed the web-based baseline assessment measures were randomly assigned to a brief, 4-week internet-based mindfulness intervention (MIND), or to the intervention plus peer counselor support (MIND+). The inclusion and exclusion criteria have been provided in Textbox 1.

Textbox 1. Participant inclusion and exclusion criteria.

Inclusion criteria

- Is currently enrolled in a university in China (undergraduate, graduate, or doctoral)
- Has a smartphone and regular access to the internet
- Demonstrates the ability to read and understand Mandarin
- Reports passing at least College English Test (level 4)
- Experiences at least mild depression and anxiety

Exclusion criteria

- Is aged <18 years
- Does not provide proof of current student status and emergency contact
- Currently experiences manic or psychotic symptoms
- Expresses suicidal or homicidal ideation during the intake phone interview

Peer Counselor Overview

Overview

Peer counselors included 4 currently enrolled female students at 3 different universities in Beijing. At the time of recruitment, their mean age was 27.5 years (SD 6.8), including 1 undergraduate (psychology), 1 master's (business), and 2 doctoral (nursing) students. None of the peer counselors reported formal training or experience in mindfulness practice or the provision of mental health services. All participants reported having passed at least the CET (level 6).

Peer Counselor Recruitment

Web-based advertisements were posted on university research, student club, and mindfulness listservs. A total of 56 candidates responded to the web-based survey, expressing interest in participating in the study as peer counselors. Those who met the inclusion criteria were contacted via telephone to screen for the exclusion criteria and confirm their understanding of the study and willingness to participate in the in-person training and orientation. Volunteer peer counselor candidates who met all the inclusion criteria were invited to the in-person training and orientation. After this training, participants were contacted via telephone to once again assess their willingness to engage in the study. We selected 4 individuals as peer counselors based on their English proficiency, reported level of enthusiasm for the project, and the researchers' assessment of their nonspecific factors. Each individual was given access to the internet-based intervention and a 6-week period to complete the course. After completing the course, peer counselors were paired with study participants who were randomized to the MIND+ group. The inclusion and exclusion criteria for peer counselors have been illustrated in Textbox 2.

Textbox 2. Peer counselor inclusion and exclusion criteria.

Inclusion criteria

- Is currently enrolled in a university in Beijing (undergraduate, graduate, or doctoral)
- Has a smartphone and regular access to the internet
- Demonstrates the ability to read and communicate in Mandarin and English
- Is willing to provide brief (15-20 minute) peer-support chats per week per participant
- Is willing to participate in web-based group supervision for 1 hour per week
- Is willing to complete the internet-based mindfulness intervention

Exclusion criteria

- Is aged <18 years
- Reports previous or current format training in mindfulness or psychotherapy
- Reports current treatment (psychotherapy or medication) for a mental health problem
- Is unable to attend the day-long, in-person training in Beijing

Peer Counselor Training

The in-person training took place for 8 hours in Beijing. All lectures and discussions were conducted in Mandarin. Peer counselor candidates listened to lectures on topics related to peer counseling and the current research project. The candidates were given opportunities to practice using the skills in dyads

and to receive coaching and feedback from the first author and research assistants.

Training was didactic and experimental and included (1) mindfulness theory and practice (2 hours), (2) orientation to the study and role of a peer counselor (1.5 hours), ethics, confidentiality, and mandated reporting (30 minutes), (4) lunch break and personal introductions (1 hour), (5) fundamentals of counseling listening skills (30 minutes), (6) validation techniques (1.5 hours), and (7) motivational interviewing (1 hour).

Peer Counselor Supervision Meetings

Weekly group supervision was attended by the research coordinator (MR), 2 research assistants, and the 4 peer counselors. Meetings were conducted in Mandarin and via a videoconferencing software (Zoom; Zoom Video Communications) after peer counselors were matched with their first participant. The structure of the supervision meetings was modeled after the elements of dialectical behavior therapy consultation team meetings [38]. The meetings began with a brief mindfulness practice and a discussion of the observations. Team members took turns leading the mindfulness practice. Next, issues were addressed according to the following hierarchy: (1) life-threatening behaviors or concerns, (2) therapy interfering behaviors, and (3) quality of life–related issues. Team members presented consultation questions and supported each other using peer counseling techniques (eg, validation and motivational interviewing), in an effort to enhance capabilities and motivation. Supervision was framed for peer counselors for both clinical consultation and peer support. The study coordinator (MR) provided 5- to 10-minute didactic lessons on the common challenges faced by peer counselors. Peer counselors also used a group chat on their mobile devices to provide each other with ongoing updates and support. They were offered the opportunity to schedule additional individual supervision from the study coordinator on an as-needed basis, or in case of a participant emergency.

The Be Mindful Internet-Based Intervention

The *Be Mindful* course is an internet-based mindfulness training program produced by Wellmind Media, with support from the UK-based charity Mental Health Foundation. *Be Mindful* delivers all the elements of mindfulness-based cognitive therapy in an internet-based course that can be completed in 4 weeks. It can be accessed through their website [39], where its development and design are fully detailed. To date, 7 peer-reviewed papers have been published reporting study results based on the *Be Mindful* course (for full details, visit the website [40]). The course can be accessed on a computer, laptop, tablet, or smartphone. The course is self-guided, that is there is no contact with mindfulness teachers or other course participants. An overview of the course content is presented in Table 1.

When this study was conducted in May 2018, the *Be Mindful* course website reported that over 20,000 people had taken the course since 2011. The participants in this study were able to complete the course for free. Project staff received technical support and administrative access to randomize participants, track progress (eg, participant log-ins, module completion, and date of completion), and download data. All materials on the course were translated into Mandarin, including the videos, audio recordings, and homework assignments. Each week, the research coordinators emailed materials to the participants after they were ready to progress to the next chapter in the course. Furthermore, 6 weeks after initial enrollment, participants received an email thanking them for their participation and a link inviting them to complete the posttreatment assessment. If they did not complete the survey within 1 week, they were contacted via text message, WeChat, and email over the course of the next week. Participants were paid upon completion of the posttreatment questionnaire packet.

Table 1. *Be Mindful* course content and assignments by week.

Chapter and title	Content	Materials and homework assignments
Before; Getting Started	Orientation to the course and format	3 videos (>6 min total); assignments: stress assessment, reflection on goals, and motivation for practicing
Week 1; Stepping out of Automatic Pilot	Introduction to the concept of mindfulness	4 videos (>12 min total); 1 audio file (30 min); assignments: events diary, body scan, routine activity, and mindful meal
Week 2; Reconnecting with Body and Breath	Awareness of thoughts and feelings	3 videos (>17 min total); 2 audio files (19 min total); assignments: difficult thoughts checklist, event awareness, mindful movement, and mindful breathing
Week 3; Working with Difficulties	Acknowledging difficult thoughts and emotions without judgment or attachment	3 videos (>9 min total); 1 audio file (22 min); assignments: stress awareness, sitting meditation, and breathing space
Week 4; Mindfulness in Daily Life	Awareness of (1) personal patterns, (2) associations to changes in mind and body, and (3) stress indicators	3 videos (>11 min total); assignments: list of four helpful and unhelpful strategies, activity awareness, breathing space, and chosen practice
After; Going Forward	Reflecting on lessons learned	3 videos (>5 min total); assignments: stress assessment, letter to yourself, and review additional resources

Institutional Review Board Approval, Consent, and Compensation

Participants were compensated for completing the baseline questionnaire packet, posttreatment questionnaires, and for responding to each daily assessment; the total amount that the participants could make from this course was approximately US $28. This study received institutional review board approval from the Psychology Research Ethics Committee at the Beijing Institute of Technology, and all participants electronically signed a digital informed consent form.

The MIND+ Condition

Participants randomized to the MIND+ condition (n=27) completed the same procedures as those in the MIND condition (n=27). However, those in the MIND+ condition were informed by the study coordinator via email that they were paired with a peer counselor who would provide them support and encouragement. Participants were instructed to contact their peer counselor within a week to schedule a time to chat. Peer counselors were instructed to contact their participants if they did not hear from them within 5 days. Peer counselors were encouraged to provide brief (15-20 minutes) weekly meetings to support and encourage participants in their completion of the internet-based intervention.

Daily Reporting

Daily assessments were completed using the Qualtrics software. Participants rated their state mindfulness and mood (stress, depression, and happiness) on a 5-point Likert scale (1=*very low*, 5=*very high*). Daily questionnaires also assessed participants' self-reported frequency and duration (in minutes) of mindfulness practice the previous day. Links to questionnaires expired within 4 hours if not completed.

Self-report Questionnaires

A self-report questionnaire packet was completed at screening, baseline, postintervention, and at the 1-month follow-up after the end of the intervention.

The Demographic Data Survey-Modified is a self-report measure used to obtain demographic information (gender, age, university, year in school, and field of study), as well as self-report data about the patient's English proficiency, meditation experience (previous training and current practice), psychiatric diagnostic and treatment history, and emergency contact information.

The 7-item Generalized Anxiety Disorder (GAD-7) questionnaire is a 7-item measure of the severity of anxiety symptoms in the last 2 weeks [41,42]. The Cronbach α reliability coefficient for the GAD-7 in this sample was .87.

The Patient Health Questionnaire-9 (PHQ-9) is a 9-item measure of the severity of depression symptoms in the last 2 weeks [43,44]. The Cronbach α reliability coefficient for the PHQ-9 in this sample was .85.

The Five-Factor Mindfulness Questionnaire (FFMQ), originally developed by Baer, is a 39-item measure of trait mindfulness that is organized into 5 subscales (Observing, Describing, Nonjudging of inner experience, Nonreactivity to Inner Experience, and Acting with Awareness), with 7 or 8 items in each subscale [45,46]. Only the full-scale FFMQ score was used in the analyses. The Cronbach α reliability coefficients for the total scale at baseline and posttreatment were .72 and .89, respectively.

The Depression Anxiety Stress Scale is a 21-item measure comprising 3 subscales (Depression, Anxiety, and Stress) of 7 items each, which provide indices of depression [47]. The Cronbach α reliability coefficients for depression, anxiety, and stress in this sample were .82, .74, and .77, respectively.

Perceived Stress Scale (PSS) is a 14-item measure of perceived stress in the last month [48]. The Cronbach α reliability coefficient for the PSS in this sample was .893.

Statistical Analyses

All analyses were conducted in SAS (version 9.4, SAS Institute).

Coding of Time

To capture any nonlinear changes across the study, the phases were coded as follows: *early study* (days 1-11), *mid study* (days 12-23), and *late study* (days 24-35).

Retention

First, a chi-square analysis was used to test the hypothesis that participants randomly assigned to MIND+ (vs those assigned to MIND) would show less attrition as indicated by higher completion (vs noncompletion) rates of posttreatment assessment (yes or no).

Adherence

Second, independent samples, two-tailed *t* tests were used to test the hypotheses that participants randomly assigned to MIND+ (vs those assigned to MIND) would show greater program adherence, as indicated by more frequent use of the course (higher number of total log-ins) and a higher percentage of the course completed.

Psychosocial Outcomes

Three multilevel models (identical to those used for the number of minutes of mindfulness practice mentioned above) were used to test the hypothesis that the randomization to the MIND+ condition would result in a greater increase in mindfulness across the trial, and greater decrease in depression and stress levels across the trial. Time was also alternatively defined by examining contrasts of the beginning of the study (days 1-11) with both the middle (days 12-23) and end (days 24-35) of the study.

Model fitting was accomplished using the −2 log likelihood model to determine model fit. Random slopes were retained when this improved the model fit. Person-standardized daily values (today's value minus overall person mean, divided by overall person SD) were used for graphical depictions of continuous outcomes to depict only the within-person changes in the outcome across the study, consistent with multilevel modeling results.

Effect Size

The size of group differences in each outcome, or change over time in each outcome, was estimated using Cohen *d*. For

multilevel modeling outcomes, Cohen *d* was calculated for the group means of difference between scores on days in the early study phase and the late study phase.

Randomization

The randomization sequence was sourced through random.org [49], an automated, web-based randomization service that generates randomness using atmospheric noise. Using random.org, two 25-person blocks were used to randomize the participants into 2 equally-sized groups.

Results

Participant Flow and Descriptive Statistics

Figure 1 shows a flow chart illustrating participant flow from recruitment to study completion. Table 2 provides descriptive statistics for demographics and key study variables for both the total sample and for each condition.

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

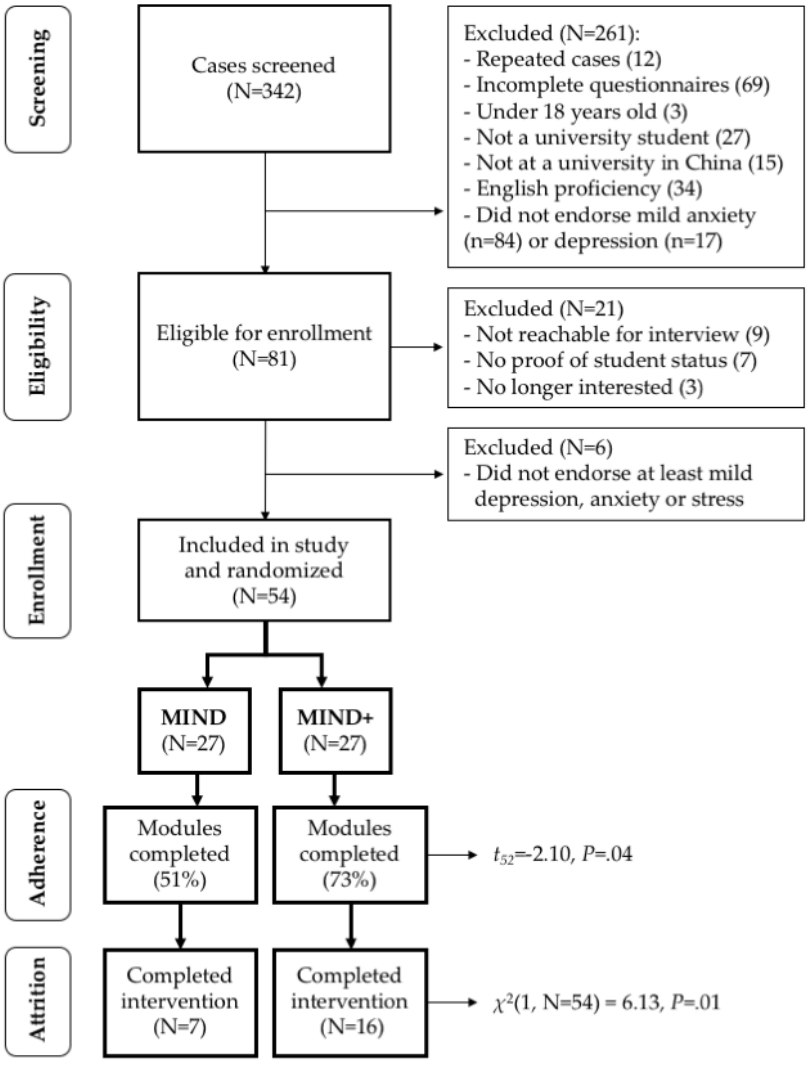


Table 2. Descriptive statistics in the full sample and by condition (N=54).

Variable	Total sample (N=54)	MIND (n=27)	MIND+ (n=27)	Comparisons
Sex (female), n (%)	40 (74)	18 (67)	22 (81)	X^2_1=1.5; *P*=.21; ϕ^a=0.17
Age (years), mean (SD)	23.53 (3.17)	23.77 (3.72)	23.29 (2.56)	t_{52}=0.55; *P*=.58; Cohen *d*=0.15
Mindfulness experience, n (%)				X^2_2=3.2; *P*=.20; Cramer V=0.24
Previously practiced	4 (7)	2 (7)	2 (7)	
Currently practice	3 (6)	3 (11)	0	
Never practiced	47 (87)	22 (81)	25 (93)	
English language competence, n (%)				X^2_2=4.6; *P*=.10; Cramer V=0.29
CET[b] level 4	7 (13)	6 (22)	1 (4)	
CET level 6	38 (70)	16 (59)	22 (81)	
TOEFL[c] or IELTS[d]	9 (17)	5 (19)	4 (15)	
Patient Health Questionnaire-9, mean (SD)				
Pretreatment	10.63 (5.0)	9.70 (4.7)	11.56 (5.3)	t_{52}=−1.38; *P*=.17; Cohen *d*=−0.37
Posttreatment	7.11 (4.8)	7.42 (5.4)	6.78 (4.2)	t_{45}=0.45; *P*=.65; Cohen *d*=0.13
Pre-to-post change	−3.23 (5.29)	−2.38 (5.08)	−4.13 (5.46)	t_{45}=1.14; *P*=.26; Cohen *d*=0.33
7-item General Anxiety Disorder, mean (SD)				
Pretreatment	8.65 (4.1)	8.33 (4.0)	8.96 (4.2)	t_{52}=−0.57; *P*=.57; Cohen *d*=−0.15
Posttreatment	5.96 (4.2)	5.83 (3.3)	6.09 (5.1)	t_{45}=−0.21; *P*=.83; Cohen *d*=−0.06
Pre-to-post change	−2.77 (3.97)	−2.62 (3.69)	−2.91 (4.32)	t_{45}=−0.25; *P*=.80; Cohen *d*=−0.07
DASS-21[e] depression, mean (SD)				
Pretreatment	6.45 (3.8)	5.73 (2.8)	7.15 (4.6)	t_{52}=−1.39; *P*=.17; Cohen *d*=−0.38
Posttreatment	3.96 (3.0)	3.79 (3.0)	4.13 (3.1)	t_{45}=−0.38; *P*=.70; Cohen *d*=−0.11
Pre-to-post change	−2.40 (3.84)	−2.62 (3.69)	−2.87 (4.32)	t_{45}=0.83; *P*=.41; Cohen *d*=0.24
DASS-21 anxiety, mean (SD)				
Pretreatment	6.38 (3.2)	6.19 (3.1)	6.56 (3.4)	t_{52}=−0.42; *P*=.57; Cohen *d*=−0.11
Posttreatment	4.72 (2.7)	4.79 (3.0)	4.65 (2.3)	t_{45}=0.18; *P*=.86; Cohen *d*=0.05
Pre-to-post change	−1.47 (3.44)	−1.29 (3.26)	−1.65 (3.22)	t_{45}=0.36; *P*=.72; Cohen *d*=0.10
DASS-21 stress, mean (SD)				
Pretreatment	8.91 (3.4)	8.31 (3.5)	9.48 (3.3)	t_{52}=−1.27; *P*=.21; Cohen *d*=−0.35
Posttreatment	6.70 (3.6)	6.33 (2.9)	7.09 (4.3)	t_{45}=−0.71; *P*=.48; Cohen *d*=−0.21
Pre-to-post change	−2.17 (3.79)	−2.0 (3.78)	−2.34 (3.86)	t_{45}=0.31; *P*=.76; Cohen *d*=0.09
Five-Factor Mindfulness Questionnaire, mean (SD)				
Pretreatment	115.46 (10.3)	116.27 (8.9)	114.65 (11.6)	t_{52}=−0.57; *P*=.57; Cohen *d*=0.16
Posttreatment	119.96 (14.8)	121.63 (12.8)	118.22 (16.7)	t_{45}=0.79; *P*=.44; Cohen *d*=0.23

Variable	Total sample (N=54)	MIND (n=27)	MIND+ (n=27)	Comparisons
Pre-to-post change	4.87 (13.18)	6.04 (9.84)	3.59 (16.21)	t_{44}=0.63; *P*=.53; Cohen *d*=0.18
Completed course, n (%)	23 (43)	7 (26)	16 (59)	X^2_1=6.1; *P*=.01; ϕ=0.34
Percentage of course completion, mean (SD)	61.66 (39.37)	50.74 (37.40)	72.59 (38.88)	t_{52}=2.10; *P*=.04; Cohen *d*=−0.57
Number of log-ins, mean (SD)	15.72 (12.35)	13.92 (9.30)	17.51 (12.39)	t_{52}=−1.07; *P*=.28; Cohen *d*=−0.33

[a]ϕ=phi coefficient (ie, mean square contingency coefficient).

[b]CET: College English Test.

[c]TOEFL: Test of English as a Foreign Language.

[d]IELTS: International English Language Testing System.

[e]DASS-21: Depression Anxiety and Stress Scale.

Hypotheses and Data

Hypothesis 1 predicted that participants randomly assigned to MIND+ (vs those assigned to MIND) would show less attrition, as indicated by a greater likelihood of completing the posttreatment assessment (as a dichotomous, between-person variable). A chi-square analysis comparing dichotomous condition assignment (MIND vs MIND+) and posttreatment assessment (completed vs not completed) revealed a greater number of completers in the MIND+ condition (χ^2_1=6.1; *P*=.01).

Hypothesis 2 predicted that participants randomly assigned to MIND+ (vs MIND) would show greater program adherence, as indicated by a higher percentage of course completion as a continuous, between-person variable. An independent samples *t* test indicated a higher mean percentage of course completion in the MIND+ condition than in the MIND condition (mean 61.66%, SD 39.37; mean difference 21.85, 95% CI 42.69-1.010; t_{52}=2.10; *P*=.04).

Hypothesis 3 predicted that participants randomly assigned to MIND+ (vs those assigned to MIND) would show more robust improvements in stress, depression, and mindfulness levels across the trial (as continuous, daily within-person variables). The results of the multilevel models testing this hypothesis are presented in Tables 3-14. Consistent with successful randomization, there were no condition effects on the baseline levels of stress, depression, or mindfulness. Both stress and depression decreased linearly as the number of study days increased, and this linear effect did not differ by condition. There were no linear effects of study days on daily mindfulness, and this effect did not differ by condition. For the daily outcomes of stress and depression, those randomized to MIND+ demonstrated a significantly greater decline from study phase 1 to phase 3. Graphs depicting daily outcomes (person-standardized) over time and phase are presented in Figures 2-4.

Table 3. Covariance parameters for the interactive effect of condition and time (study day) predicting daily self-reported stress.

Parameter	Estimate (SE)	Z value[a]	*P* value
Intercept	0.513 (0.192)	2.68	.004
Covariance (I,S)[b]	−0.006 (0.005)	−1.18	.24
Study day	0.000 (0.000)	2.20	.01
Residual (VC[c])	0.556 (0.029)	19.19	<.001

[a]The Z value represents the test value of the z distribution on which statistical significance is determined for this analysis.

[b]Covariance between the random parameters for intercept and slope in the multilevel model.

[c]VC: variance component (method for structuring the covariance matrix).

Table 4. Fixed effects for the interactive effect of condition and time (study day) predicting daily self-reported stress.

Effect	Estimate (SE)	*t* test (*df*)	*P* value
Intercept	2.998 (0.185)	16.24 (34.1)	<.001
Condition[a]	−0.279 (0.268)	−1.04 (35.2)	.30
Study day	−0.022 (0.006)	−3.90 (39.4)	<.001
Condition X study day	0.009 (0.008)	1.08 (40.3)	.29

[a]Condition is coded as a dichotomous variable, where 0=MIND only and 1=MIND+.

Table 5. Covariance parameters for the interactive effect of condition and time (study phase) predicting daily self-reported stress.

Parameter	Estimate (SE)	Z value[a]	*P* value
Intercept	0.460 (0.112)	4.12	<.001
Study phase (3 vs 1)	0.340 (0.036)	9.55	<.001
Residual (VC[b])	0.601 (0.031)	19.53	<.001

[a]The Z value represents the test value of the z distribution on which statistical significance is determined for this analysis.
[b]VC: variance component (method for structuring the covariance matrix).

Table 6. Fixed effects for the interactive effect of condition and time (study phase) predicting daily self-reported stress.

Effect	Estimate (SE)	*t* test (*df*)	*P* value
Intercept	2.580 (0.167)	15.48 (62.4)	<.001
Condition[a]	0.118 (0.230)	0.51 (62.2)	.61
Study phase (2 vs 1)	−0.268 (0.111)	−2.42 (360)	.02
Study phase (3 vs 1)	−0.285 (0.112)	−2.54 (272)	.01
Condition X phase (2 vs 1)	0.093 (0.151)	0.62 (344)	.54
Condition X phase (3 vs 1)	−0.234 (0.113)	−2.07 (266)	.04

[a]Condition is coded as a dichotomous variable, where 0=MIND only and 1=MIND+.

Table 7. Covariance parameters for the interactive effect of condition and time (study day) predicting daily self-reported depression.

Parameter	Estimate (SE)	Z value[a]	*P* value
Intercept	0.566 (0.191)	2.97	.002
Covariance (I,S)[b]	−0.007 (0.005)	−1.49	.14
Study day	0.000 (0.000)	2.19	.01
Residual (VC[c])	0.505 (0.027)	18.92	<.001

[a]The Z value represents the test value of the z distribution on which statistical significance is determined for this analysis.
[b]Covariance between the random parameters for intercept and slope in the multilevel model.
[c]VC: variance component (method for structuring the covariance matrix).

Table 8. Fixed effects for the interactive effect of condition and time (study day) predicting daily self-reported depression.

Effect	Estimate (SE)	*t* test (*df*)	*P* value
Intercept	2.439 (0.189)	12.94 (38.1)	<.001
Condition[a]	−0.160 (0.273)	−0.59 (39.1)	.56
Study day	−0.013 (0.005)	−2.40 (39.2)	.02
Condition X study day	0.008 (0.008)	1.01 (40.1)	.32

[a]Condition is coded as a dichotomous variable, where 0=MIND only and 1=MIND+.

Table 9. Covariance parameters for the interactive effect of condition and time (study phase) predicting daily self-reported depression.

Parameter	Estimate (SE)	Z value[a]	*P* value
Intercept	0.422 (0.104)	4.08	<.001
Study phase (3 vs 1)	0.359 (0.035)	10.29	<.001
Residual (VC[b])	0.544 (0.028)	19.30	<.001

[a]The Z value represents the test value of the z distribution on which statistical significance is determined for this analysis.
[b]VC: variance component (method for structuring the covariance matrix).

Table 10. Fixed effects for the interactive effect of condition and time (study phase) predicting daily self-reported depression.

Effect	Estimate (SE)	*t* test (*df*)	*P* value
Intercept	2.164 (0.160)	13.55 (61.5)	<.001
Condition[a]	0.117 (0.221)	0.53 (61.3)	.60
Study phase (2 vs 1)	0.025 (0.106)	0.23 (362)	.82
Study phase (3 vs 1)	−0.063 (0.108)	−0.58 (271)	.56
Condition X phase (2 vs 1)	−0.162 (0.145)	−1.12 (347)	.26
Condition X phase (3 vs 1)	−0.280 (0.128)	−2.19 (265)	.03

[a]Condition is coded as a dichotomous variable, where 0=MIND only and 1=MIND+.

Table 11. Covariance parameters for the interactive effect of condition and time (study day) predicting daily self-reported mindfulness.

Parameter	Estimate (SE)	Z value[a]	*P* value
Intercept	0.395 (0.134)	2.95	.002
Covariance (I,S)[b]	−0.006 (0.004)	−1.64	.10
Study day	0.000 (0.000)	2.49	.006
Residual (VC[c])	0.526 (0.025)	21.09	<.001

[a]The Z value represents the test value of the z distribution on which statistical significance is determined for this analysis.
[b]Covariance between the random parameters for intercept and slope in the multilevel model.
[c]VC: variance component (method for structuring the covariance matrix).

Table 12. Fixed effects for the interactive effect of condition and time (study day) predicting daily self-reported mindfulness.

Effect	Estimate (SE)	*t* test (*df*)	*P* value
Intercept	2.561 (0.164)	15.61 (40.5)	<.001
Condition[a]	0.007 (0.237)	0.03 (42.0)	.98
Study day	0.002 (0.005)	0.30 (41.2)	.76
Condition X study day	0.007 (0.008)	0.93 (42.1)	.36

[a]Condition is coded as a dichotomous variable, where 0=MIND only and 1=MIND+.

Table 13. Covariance parameters for the interactive effect of condition and time (study phase) predicting daily self-reported mindfulness.

Parameter	Estimate (SE)	Z value[a]	*P* value
Intercept	0.300 (0.073)	4.13	<.001
Study phase (3 vs 1)	0.211 (0.037)	5.75	<.001
Residual (VC[b])	0.565 (0.027)	21.26	<.001

[a]The Z value represents the test value of the z distribution on which statistical significance is determined for this analysis.
[b]VC: variance component (method for structuring the covariance matrix).

Table 14. Fixed effects for the interactive effect of condition and time (study phase) predicting daily self-reported mindfulness.

Effect	Estimate (SE)	*t* test (*df*)	*P* value
Intercept	2.732 (0.139)	19.69 (69.6)	<.001
Condition[a]	−0.142 (0.191)	−0.74 (69.3)	.46
Study phase (2 vs 1)	0.023 (0.100)	0.23 (405.0)	.82
Study phase (3 vs 1)	0.123 (0.099)	1.24 (324.0)	.22
Condition X phase (2 vs 1)	−0.084 (0.136)	−0.62 (383.0)	.54
Condition X phase (3 vs 1)	−0.053 (0.136)	−0.39 (314.0)	.69

[a]Condition is coded as a dichotomous variable, where 0=MIND only and 1=MIND+.

Figure 2. Daily (A) and phase (B) means for outcome daily self-reported stress.

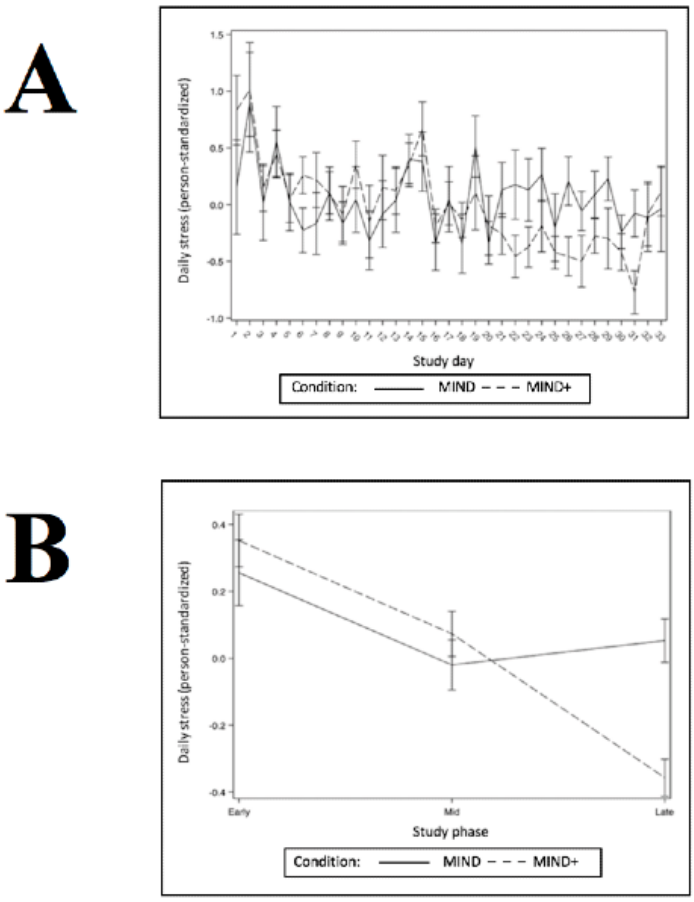


Figure 3. Daily (A) and phase (B) means for outcome daily self-reported depression.

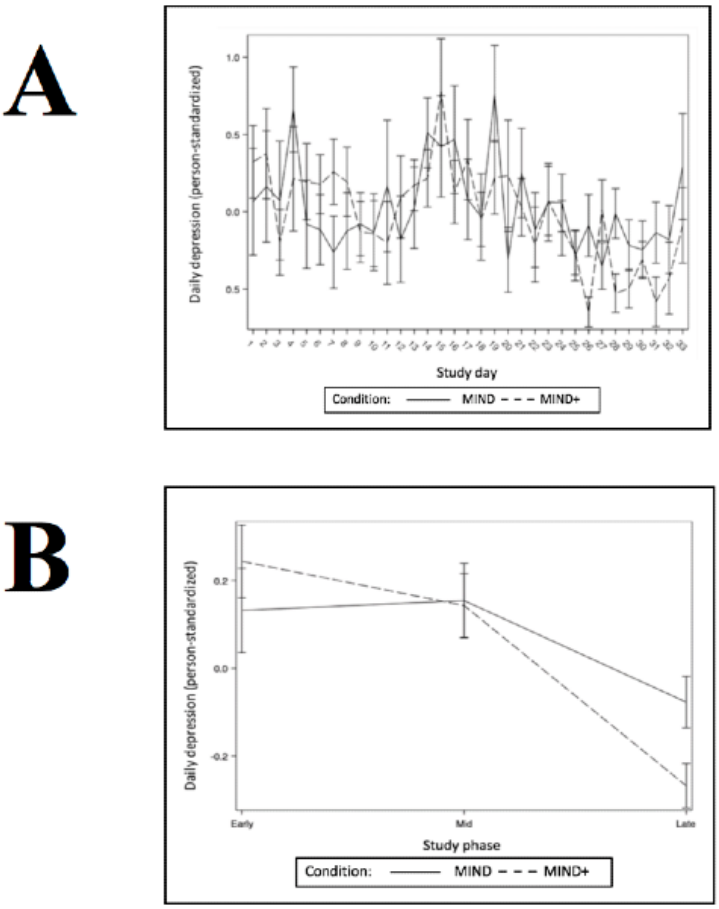
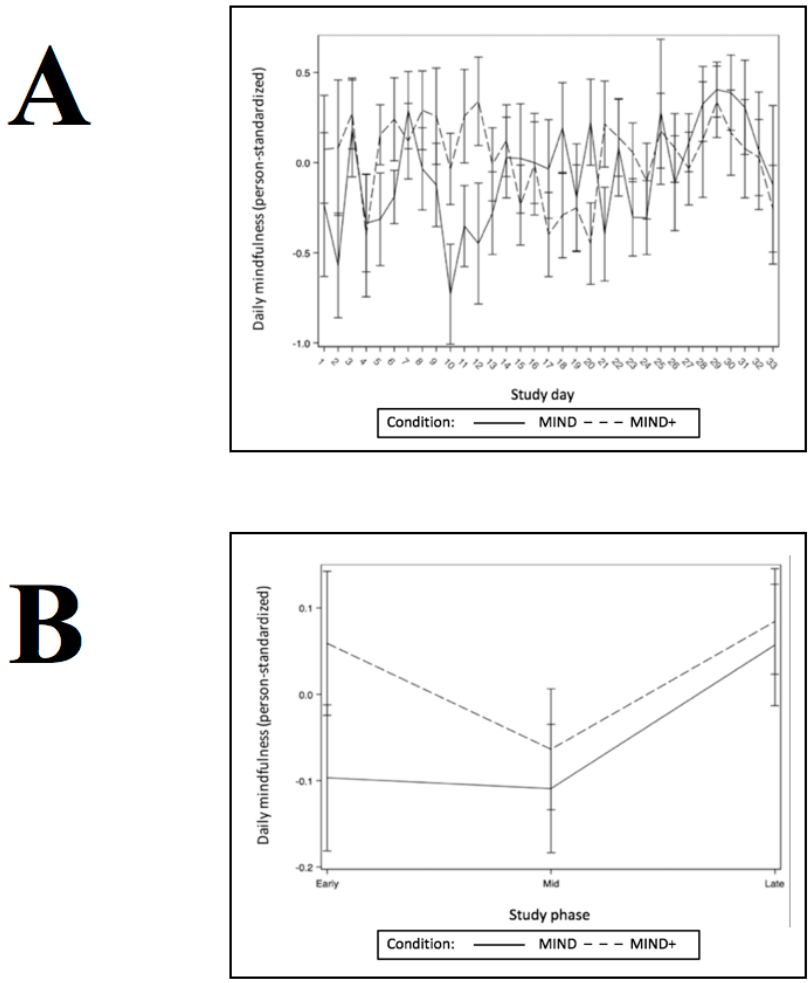


Figure 4. Daily (A) and phase (B) means for outcome daily self-reported mindfulness.



Discussion

Principal Findings

The aim of this study was to investigate the efficacy of an adjunctive treatment component that uses task-shifting (ie, nonspecialist peer counselors) to enhance engagement in a self-directed, web-based mindfulness intervention for stress and depression among Chinese undergraduate and graduate students.

The results indicated that participants assigned to the MIND+ (vs those assigned to the MIND) condition showed significantly less attrition and more adherence, as indicated by a greater likelihood of completing posttreatment assessments and a higher percentage of course completion, respectively. In addition, individuals in the MIND+ condition reported significant improvements in daily ratings of stress and depression levels across the trial compared with individuals in the MIND condition. These findings suggest that volunteer peer counselors receiving brief training and weekly supervision may significantly improve participants' indices of treatment engagement and mental health outcomes in an internet-based mindfulness intervention among college and graduate students in China.

Unique Contributions of This Study

This study makes several unique contributions to the literature. First, an internet-based platform was used to deliver a mindfulness intervention in a sample of individuals from a non-Western LMIC. There have been no publications of results from randomized controlled trials investigating the efficacy of a self-guided, web-based, mindfulness intervention in China.

Second, this study uses a task-shifting informed approach aimed at increasing retention and adherence to an existing evidence-based intervention. There have been no publications of results from randomized controlled trials investigating the efficacy of an adjunctive, web-based, peer support intervention component intended to enhance treatment engagement in a self-guided, web-based, mental health intervention in an LMIC. Furthermore, only 3 studies have been published on task shifting in mental health services of any type in China [50-52]. Accordingly, the results of this study advance the literature in this area, providing promise in the use of task shifting to improve mental health outcomes in China, the world's most populous country.

Third, this study explored the effects of a very low-intensity, low-cost task-shifting intervention. Nonspecialist peer

counselors received only 1 day of in-person training with ongoing web-based, group supervision once per week. Furthermore, each participant in the MIND+ group only received a mean of 4.69 peer counseling *sessions* (SD 2.04; range 0-7), lasting an average total of 120.85 minutes (SD 66.53; range 12-345) per MIND+ participant. All peer counselors maintained full-time student status *and* part-time jobs throughout the course of volunteering for this study. Peer counselors reported completion of responsibilities for this study (including supervision, peer counseling, and documentation) and required, on average, less than 2 hours per week, when supporting 2 patients at the same time.

Fourth, this study contributes new insights into the selection, supervision, and evaluation practices for task-shifting initiatives. According to a recent systematic review of 137 studies from 48 countries employing task shifting to deliver evidence-based mental health services in LMICs, fewer than 1 in 5 studies reported providing supervision on a weekly or biweekly basis [53]. In this study, nonspecialist peer counselors were evaluated for knowledge (questionnaires), application (group supervision discussions), competence (role-plays during training and in group supervision), and quality of care (including patients' evaluations, pre to postsession changes in participants' self-reported mood and motivation, and 2 *session* audio recordings).

Completion Rates in This Study Versus Those in Comparators

Completion rates in this study (7/27, 26% for MIND and 16/27, 59% for MIND+) appear to be lower than those in previous studies. In comparison, Querstret et al [54] reported 73.7% (87/118) of participants completing the web-based program. Krusche et al [55] reported 29.39% (1497/5094) of participants completing the course *and* follow-up assessments; however, only 11% (3/27) of MIND participants and 15% (4/27) of MIND+ participants completed both.

One explanation for the lower completion rates in this study is that participants completed the course in their second language. Of the 43 participants who completed the posttreatment assessment, when asked the degree to which language was a barrier in completing the course, 16 (37%) indicated *not at all*, and 17 (40%) reported *a little*. However, 9% (4/43) of participants indicated that language was *very much* a barrier to completion. It is also possible that language was more of a barrier for the 11 participants who did not complete the posttreatment assessment. However, follow-up analyses did not reveal an association between language proficiency and course completion or engagement. Specifically, follow-up analyses revealed that self-reported baseline English proficiency and posttreatment perception of language as a barrier to completion were not significant predictors of treatment completion.

Another explanation for this difference is that participants in the studies by Krusche et al [55] and Querset et al [54] were paying US $60-$90 for participation; therefore, they were perhaps more motivated. Before the start of the study, 26% (14/54) participants reported being *very* motivated to learn and practice mindfulness, and 44% (24/54) reported being *extremely* motivated. However, across both groups, baseline self-reported energy to learn and practice mindfulness was predictive of failure to activate the *Be Mindful* course (r=0.288; P=.04) and total login count (r=0.300; P=.03).

Participants assigned to the MIND+ (vs those assigned to the MIND) condition showed significantly less attrition, as indicated by a greater likelihood of completing the posttreatment assessments in the internet-based course. MIND+ participants also demonstrated a nonsignificant trend toward lower rates of nonuse attrition (P=.05), defined as never responding to the daily assessment or not responding for at least the last 3 weeks of the study. These findings suggest that a low-cost, low-intensity task-shifting component is promising as a feasible and scalable approach for enhancing retention in web-based evidence-based treatments. Previous research suggests that personalizing the contact (eg, provision of therapist name and photo vs a virtual therapist or personalized vs standardized messaging) is associated with lower rates of treatment termination [28,56]. Apart from strategies that involve human support, future research can also continue to explore automated strategies to provide more personalized treatment experience in self-guided, internet-based mindfulness interventions.

Unique Findings of This Study

Participants assigned to the MIND+ (vs those assigned to MIND) condition showed greater program adherence, as indicated by a higher percentage of the course completed. However, there were no between-group differences in attrition, as indicated by (1) more frequent log-ins to the course, (2) a less robust decrease in daily self-reports of mindfulness practice, or (3) a less robust decrease in daily self-reports of minutes of mindfulness practiced over the course of the treatment. Overall, these data suggest that the MIND+ task-shifting component increased participants' likelihood of completing the program but not necessarily their likelihood to be more actively engaged in the program (ie, more frequent log-ins) or to report higher frequency or duration of mindfulness practice.

It is worth noting that participants in this study presented with mean baseline PSS (stress), GAD-7 (anxiety), and PHQ-9 (depression) scores of 23.27 (SD 4.28), 9.90 (SD 3.98), and 11.31 (SD 5.06), respectively. These means are higher than the scores provided in published population norms for the PSS (between 11.9 and 14.7) [57], GAD-7 (between 2.7 and 3.8) [58,59], and PHQ-9 (approximately 3.3) [42]. Instead, the participants in this study would, on average, be considered *highly stressed* [59,60], *moderately anxious* [41,61], and *moderately depressed* [42].

Although participants were randomized to study conditions, the MIND+ group participants reported significantly higher mean baseline PSS scores than those of the MIND group participants. PSS was the only baseline measure with significant between-group differences in this study. However, it is possible that this difference in stress helps in explaining why MIND+ participants completed more modules but did not report more frequent and longer-lasting mindfulness practice than MIND participants. On the other hand, it is possible that MIND+ participants reported more stress because they were assigned to the condition with a peer counselor, and they felt more pressure to complete the course. Existing research suggests that

positive, high-quality social support can enhance resilience to stress and reduce depressive symptomology and medical morbidity and mortality [62,63]. Nevertheless, future research could explore the mechanisms by which peer support enhances adherence, including factors such as social desirability, expectation, and compliance.

Enrollment in this study was associated with a significant increase in reported trait mindfulness, as indicated by the FFMQ scores. This was true across both groups, with no significant between-group differences. Further analyses should explore whether changes in mindfulness mediate the effects of interventions on depression, anxiety, and stress. The results of Querstret et al [54] showed that although the intervention worked to increase levels of the *describing* and *nonjudging* facets of mindfulness, only Acting with Awareness mediated the effects of the intervention on mental health outcomes.

This study improves upon previous studies of daily practice in relation to the *Be Mindful* internet-based course. Krusche et al [55,64] asked participants to rate their mindfulness activities once per week using a high (*every day or most days*), medium (*sometimes*), and low (*rarely or never*) scale [64]. In the 2013 study, 51.65% (141/273) of participants reported practicing *sometimes*, and in the 2012 study, 55% (55/100) of participants reported practicing *sometimes*. This study improves upon this approach by collecting daily data and having participants record the number of minutes practiced per day. However, future research would benefit from collecting separate data for formal and informal mindfulness practice, or by collecting data in real time (eg, asking participants to report immediately before and after mindfulness practice).

Significance of These Findings

The results of this study indicate that participation in the internet-based intervention was associated with significant improvements in pre to posttreatment stress outcomes. The pre-post effect size (Cohen *d*) for stress among the completers was 1.13. This was equivalent to those reported in previous studies of the *Be Mindful* course [55], web-based interventions [65,66], and in-person mindfulness interventions [17,67,68].

These findings are significant because stress has been shown to be associated with a wide range of physical and mental health problems [69,70], including autoimmune diseases, depression, substance abuse, and suicidal behavior. Research suggests that the prevalence of stress among college students, particularly Chinese students [71], is increasing [72]. Students from Confucian Asian countries (eg, Japan, Korea, and China) report higher levels of stress, anxiety, and self-doubt than students from European regions [73]. Moreover, compared with their Korean and Japanese counterparts, Chinese college students reported the highest number of stressors and the highest levels of stress, along with passive and ineffective coping [74,75].

The pre-post effect sizes (Cohen *d*) for anxiety (GAD-7) and depression (PHQ-9) among completers were 0.89 and 0.95, respectively, which are equivalent to those reported in previous studies of the *Be Mindful* course and in-person mindfulness interventions [17,55,66].

In addition, individuals in the MIND+ condition reported significant improvements in daily ratings of stress and depression across the trial compared with individuals in the MIND condition. These findings suggest that volunteer peer counselors receiving brief training and weekly supervision may significantly improve participants' indices of treatment engagement and mental health outcomes in an internet-based mindfulness intervention among college and graduate students in China. It is worth noting that these differences between groups were not linear across the course of the study. The benefits of assignment to the MIND+ group appear late during the treatment, that is, between phases 2 and 3. In the middle of the study, MIND+ participants did not report less anxiety or depression, and they did not report practicing more than the MIND-only participants. Therefore, one explanation for the benefit of the program was weekly contact with peer counselors. Another explanation is that they received more content during the intervention. The effect size of the internet-based course on stress, depression, and anxiety scores suggests that this treatment is effective for Chinese students, regardless of whether they have contact with peers or a therapist.

MIND+ participants did not report significant improvements in daily ratings of state mindfulness across the trial compared with participants in the MIND-only condition. Instead, there was a main effect of treatment on improvements on daily mindfulness ratings. Similarly, there were no between-group differences in pre-post FFMQ scores, although there was a moderate main effect of mindfulness (FFMQ) among completers (Cohen d=0.55). These outcomes suggest that the effect of having peer support did not increase reported mindfulness as measured in daily assessments or in pre-post measures. Furthermore, it is unlikely that the effect of practice and change in trait mindfulness did not mediate the change in daily reports of stress and depression over the course of the treatment.

Study Limitations

This study has several limitations. First, the sample was small, achieving 80% power to detect only moderately-sized group differences (ie, Cohen d=0.78 or larger). The sample also consisted of a nonclinical sample of English-speaking university students. Before these findings can be generalized, this research should be replicated among participants using the internet-based intervention in their native language and among larger, more diverse samples. Related to this, the significant effects described here would not survive correction for the number of tests performed in this study; nonetheless, we believe that it is important to share our findings with the field. Second, there were insufficient follow-up data to be able to analyze whether the effects of treatment were sustained over time. Data related to peer counselors' communication with participants were not included in these analyses. It is possible that the frequency and duration of peer counseling influences the effect of treatment conditions on engagement and mental health outcomes. Future studies should explore and detect the possible dosage effects. Moreover, future research would benefit from having more objective indicators of study, practice, and mindfulness meditation. For example, one patient might report practicing mindfulness once a day, but it could be a 45-minute body scan, and another could report practicing 35 times because they could

have noticed their thoughts or body sensations that many times throughout the day. Finally, the exclusion of candidates who expressed suicidal ideation might have led to an underestimation of the overall treatment effect, because the cohort most severely affected by depression was not considered in the study.

Conclusions

This study provides preliminary support for the effectiveness of a 4-week, internet-based mindfulness course for the reduction of self-reported symptoms of stress, depression, and anxiety among English-speaking university students in China. The effects were compared with those reported in other mindfulness courses delivered on the web and in-person. Furthermore, these results highlight the potential of leveraging task shifting to enhance treatment engagement in self-guided evidence-based treatments. The combination of these approaches may represent a financially feasible, easily transportable, and quickly scalable way to provide mental health services in low-resource settings.

Conflicts of Interest

MZR is a Scientific Advisor for the Misophonia Research Fund, The Real Odin, and BehaVR.

Multimedia Appendix 1
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1253 KB - formative_v5i10e25772_app1.pdf]

References

1. Merikangas KR, He J, 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 Oct;49(10):980-989 [FREE Full text] [doi: 10.1016/j.jaac.2010.05.017] [Medline: 20855043]
2. Insel TR. Assessing the economic costs of serious mental illness. Am J Psychiatry 2008 Jun;165(6):663-665. [doi: 10.1176/appi.ajp.2008.08030366] [Medline: 18519528]
3. McHugh RK, Whitton SW, Peckham AD, Welge JA, Otto MW. Patient preference for psychological vs pharmacologic treatment of psychiatric disorders: a meta-analytic review. J Clin Psychiatry 2013 Jun;74(6):595-602 [FREE Full text] [doi: 10.4088/JCP.12r07757] [Medline: 23842011]
4. Results from the 2014 National Survey on Drug Use and Health. Substance Abuse and Mental Health Services Administration. 2015 Sep. URL: https://www.samhsa.gov/data/sites/default/files/NSDUH-DetTabs2014/NSDUH-DetTabs2014.pdf [accessed 2021-09-13]
5. Saxena S, Thornicroft G, Knapp M, Whiteford H. Resources for mental health: scarcity, inequity, and inefficiency. Lancet 2007 Sep 8;370(9590):878-889. [doi: 10.1016/S0140-6736(07)61239-2] [Medline: 17804062]
6. Gao L, Xie Y, Jia C, Wang W. Prevalence of depression among Chinese university students: a systematic review and meta-analysis. Sci Rep 2020 Sep 28;10(1):15897 [FREE Full text] [doi: 10.1038/s41598-020-72998-1] [Medline: 32985593]
7. Mental Health Atlas. Geneva, Switzerland: World Health Organization; 2014.
8. Kazdin AE, Blase SL. Rebooting psychotherapy research and practice to reduce the burden of mental illness. Perspect Psychol Sci 2011 Jan;6(1):21-37. [doi: 10.1177/1745691610393527] [Medline: 26162113]
9. Fairburn CG, Patel V. The impact of digital technology on psychological treatments and their dissemination. Behav Res Ther 2017 Jan;88:19-25 [FREE Full text] [doi: 10.1016/j.brat.2016.08.012] [Medline: 28110672]
10. Veehof MM, Oskam M, Schreurs KM, Bohlmeijer ET. Acceptance-based interventions for the treatment of chronic pain: a systematic review and meta-analysis. Pain 2011 Mar;152(3):533-542 [FREE Full text] [doi: 10.1016/j.pain.2010.11.002] [Medline: 21251756]
11. Piet J, Hougaard E. The effect of mindfulness-based cognitive therapy for prevention of relapse in recurrent major depressive disorder: a systematic review and meta-analysis. Clin Psychol Rev 2011 Aug;31(6):1032-1040. [doi: 10.1016/j.cpr.2011.05.002] [Medline: 21802618]
12. Hofmann SG, Sawyer AT, Witt AA, Oh D. The effect of mindfulness-based therapy on anxiety and depression: a meta-analytic review. J Consult Clin Psychol 2010 Apr;78(2):169-183 [FREE Full text] [doi: 10.1037/a0018555] [Medline: 20350028]
13. Zgierska A, Marcus MT. Mindfulness-based therapies for substance use disorders: part 2. Subst Abus 2010 Apr;31(2):77-78 [FREE Full text] [doi: 10.1080/08897071003641248] [Medline: 20408058]
14. Wanden-Berghe RG, Sanz-Valero J, Wanden-Berghe C. The application of mindfulness to eating disorders treatment: a systematic review. Eat Disord 2011;19(1):34-48. [doi: 10.1080/10640266.2011.533604] [Medline: 21181578]
15. Vøllestad J, Nielsen MB, Nielsen GH. Mindfulness- and acceptance-based interventions for anxiety disorders: a systematic review and meta-analysis. Br J Clin Psychol 2012 Sep;51(3):239-260. [doi: 10.1111/j.2044-8260.2011.02024.x] [Medline: 22803933]
16. Chiesa A, Serretti A. Mindfulness-based stress reduction for stress management in healthy people: a review and meta-analysis. J Altern Complement Med 2009 May;15(5):593-600. [doi: 10.1089/acm.2008.0495] [Medline: 19432513]
17. Keng S, Smoski MJ, Robins CJ. Effects of mindfulness on psychological health: a review of empirical studies. Clin Psychol Rev 2011 Aug;31(6):1041-1056 [FREE Full text] [doi: 10.1016/j.cpr.2011.04.006] [Medline: 21802619]

18. Spijkerman MP, Pots WT, Bohlmeijer ET. Effectiveness of online mindfulness-based interventions in improving mental health: a review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016 Apr;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]
19. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
20. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
21. Wangberg SC, Bergmo TS, Johnsen JK. Adherence in Internet-based interventions. Patient Prefer Adherence 2008 Feb 02;2:57-65 [FREE Full text] [Medline: 19920945]
22. Owen JE, Jaworski BK, Kuhn E, Makin-Byrd KN, Ramsey KM, Hoffman JE. mHealth in the wild: using novel data to examine the reach, use, and impact of PTSD coach. JMIR Ment Health 2015 Mar 25;2(1):e7 [FREE Full text] [doi: 10.2196/mental.3935] [Medline: 26543913]
23. Del Re AC, Flückiger C, Goldberg SB, Hoyt WT. Monitoring mindfulness practice quality: an important consideration in mindfulness practice. Psychother Res 2013;23(1):54-66. [doi: 10.1080/10503307.2012.729275] [Medline: 23046287]
24. Cavanagh K, Churchard A, O'Hanlon P, Mundy T, Votolato P, Jones F, et al. A randomised controlled trial of a brief online mindfulness-based intervention in a non-clinical population: replication and extension. Mindfulness (N Y) 2018;9(4):1191-1205 [FREE Full text] [doi: 10.1007/s12671-017-0856-1] [Medline: 30100934]
25. Jacob KS. Repackaging mental health programs in low- and middle-income countries. Indian J Psychiatry 2011 Jul;53(3):195-198 [FREE Full text] [doi: 10.4103/0019-5545.86798] [Medline: 22135435]
26. Andersson G, Cuijpers P. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. Cogn Behav Ther 2009;38(4):196-205. [doi: 10.1080/16506070903318960] [Medline: 20183695]
27. Richards D, Richardson T. Computer-based psychological treatments for depression: a systematic review and meta-analysis. Clin Psychol Rev 2012 Jun;32(4):329-342. [doi: 10.1016/j.cpr.2012.02.004] [Medline: 22466510]
28. Spek V, Cuijpers P, Nyklícek I, Riper H, Keyzer J, Pop V. Internet-based cognitive behaviour therapy for symptoms of depression and anxiety: a meta-analysis. Psychol Med 2007 Mar;37(3):319-328. [doi: 10.1017/S0033291706008944] [Medline: 17112400]
29. Furmark T, Carlbring P, Hedman E, Sonnenstein A, Clevberger P, Bohman B, et al. Guided and unguided self-help for social anxiety disorder: randomised controlled trial. Br J Psychiatry 2009 Nov;195(5):440-447. [doi: 10.1192/bjp.bp.108.060996] [Medline: 19880935]
30. Morgan AJ, Jorm AF, Mackinnon AJ. Email-based promotion of self-help for subthreshold depression: mood memos randomised controlled trial. Br J Psychiatry 2012 May;200(5):412-418 [FREE Full text] [doi: 10.1192/bjp.bp.111.101394] [Medline: 22442102]
31. Titov N, Andrews G, Davies M, McIntyre K, Robinson E, Solley K. Internet treatment for depression: a randomized controlled trial comparing clinician vs. technician assistance. PLoS One 2010 Jun 08;5(6):e10939 [FREE Full text] [doi: 10.1371/journal.pone.0010939] [Medline: 20544030]
32. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009 Mar;24(1). [doi: 10.17705/1CAIS.02428]
33. Eaton J, McCay L, Semrau M, Chatterjee S, Baingana F, Araya R, et al. Scale up of services for mental health in low-income and middle-income countries. Lancet 2011 Oct 29;378(9802):1592-1603. [doi: 10.1016/S0140-6736(11)60891-X] [Medline: 22008429]
34. Padmanathan P, De Silva MJ. The acceptability and feasibility of task-sharing for mental healthcare in low and middle income countries: a systematic review. Soc Sci Med 2013 Nov;97:82-86. [doi: 10.1016/j.socscimed.2013.08.004] [Medline: 24161092]
35. Mdege ND, Chindove S, Ali S. The effectiveness and cost implications of task-shifting in the delivery of antiretroviral therapy to HIV-infected patients: a systematic review. Health Policy Plan 2013 May;28(3):223-236. [doi: 10.1093/heapol/czs058] [Medline: 22738755]
36. Dawson AJ, Buchan J, Duffield C, Homer CS, Wijewardena K. Task shifting and sharing in maternal and reproductive health in low-income countries: a narrative synthesis of current evidence. Health Policy Plan 2014 May;29(3):396-408. [doi: 10.1093/heapol/czt026] [Medline: 23656700]
37. Joshi R, Alim M, Kengne AP, Jan S, Maulik PK, Peiris D, et al. Task shifting for non-communicable disease management in low and middle income countries--a systematic review. PLoS One 2014 Aug 14;9(8):e103754 [FREE Full text] [doi: 10.1371/journal.pone.0103754] [Medline: 25121789]
38. Linehan M. Cognitive-Behavioral Treatment of Borderline Personality Disorder. New York: Guilford Press; 1993.
39. Be Mindful. URL: https://www.bemindfulonline.com/ [accessed 2021-05-01]
40. Proven effective by world-class published research. Wellmind Health. URL: https://www.wellmindhealth.com/clinical-studies [accessed 2021-09-23]
41. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]

42. He X, Li C, Qian J, Cui H, Wu W. Reliability and validity of a generalized anxiety disorder scale in general hospital outpatients. Shanghai Arch Psychiatry 2010 Jan;22(4):200-203. [doi: 10.3969/j.issn.1002-0829.2010.04.002]
43. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
44. Liu S, Yeh Z, Huang H, Sun F, Tjung J, Hwang L, et al. Validation of Patient Health Questionnaire for depression screening among primary care patients in Taiwan. Compr Psychiatry 2011;52(1):96-101. [doi: 10.1016/j.comppsych.2010.04.013] [Medline: 21111406]
45. Baer RA. Mindfulness training as a clinical intervention: a conceptual and empirical review. Clin Psychol Sci Pract 2003;10(2):125-143. [doi: 10.1093/clipsy.bpg015]
46. Deng Y, Liu X, Rodriguez MA, Xia C. The five facet mindfulness questionnaire: psychometric properties of the Chinese version. Mindfulness 2011 Mar 15;2(2):123-128. [doi: 10.1007/s12671-011-0050-9]
47. Wang K, Shi H, Geng F, Zou L, Tan S, Wang Y, et al. Cross-cultural validation of the Depression Anxiety Stress Scale-21 in China. Psychol Assess 2016 May;28(5):e88-100. [doi: 10.1037/pas0000207] [Medline: 26619091]
48. Cohen S, Janicki-Deverts D. Who's stressed? Distributions of psychological stress in the United States in probability samples from 1983, 2006, and 2009. J Appl Soc Psychol 2012;42(6):1320-1334. [doi: 10.1111/j.1559-1816.2012.00900.x]
49. random.org. URL: https://www.random.org/ [accessed 2021-09-23]
50. Gao Y, Pan B, Sun W, Wu H, Wang J, Wang L. Anxiety symptoms among Chinese nurses and the associated factors: a cross sectional study. BMC Psychiatry 2012 Sep 14;12:141 [FREE Full text] [doi: 10.1186/1471-244X-12-141] [Medline: 22978466]
51. Kaufman JA, Zeng W, Wang L, Zhang Y. Community-based mental health counseling for children orphaned by AIDS in China. AIDS Care 2013;25(4):430-437. [doi: 10.1080/09540121.2012.712661] [Medline: 22880692]
52. Tiwari A, Fong DY, Yuen KH, Yuk H, Pang P, Humphreys J, et al. Effect of an advocacy intervention on mental health in Chinese women survivors of intimate partner violence: a randomized controlled trial. JAMA 2010 Aug 04;304(5):536-543. [doi: 10.1001/jama.2010.1052] [Medline: 20682933]
53. Singla DR, Kohrt BA, Murray LK, Anand A, Chorpita BF, Patel V. Psychological treatments for the world: lessons from low- and middle-income countries. Annu Rev Clin Psychol 2017 May 08;13:149-181 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032816-045217] [Medline: 28482687]
54. Querstret D, Cropley M, Fife-Schaw C. Internet-based instructor-led mindfulness for work-related rumination, fatigue, and sleep: assessing facets of mindfulness as mechanisms of change. A randomized waitlist control trial. J Occup Health Psychol 2017 Apr;22(2):153-169. [doi: 10.1037/ocp0000028] [Medline: 27054503]
55. Krusche A, Cyhlarova E, Williams JM. Mindfulness online: an evaluation of the feasibility of a web-based mindfulness course for stress, anxiety and depression. BMJ Open 2013 Nov 29;3(11):e003498 [FREE Full text] [doi: 10.1136/bmjopen-2013-003498] [Medline: 24293203]
56. Palmqvist B, Carlbring P, Andersson G. Internet-delivered treatments with or without therapist input: does the therapist factor have implications for efficacy and cost? Expert Rev Pharmacoecon Outcomes Res 2007 Jun;7(3):291-297. [doi: 10.1586/14737167.7.3.291] [Medline: 20528315]
57. Cohen S. Perceived stress in a probability sample of the United States. In: The Social Psychology of Health. Thousand Oaks, CA: Sage Publications; 1988.
58. Löwe B, Decker O, Müller S, Brähler E, Schellberg D, Herzog W, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care 2008 Mar;46(3):266-274. [doi: 10.1097/MLR.0b013e318160d093] [Medline: 18388841]
59. Löwe B, Spitzer RL, Williams JB, Mussell M, Schellberg D, Kroenke K. Depression, anxiety and somatization in primary care: syndrome overlap and functional impairment. Gen Hosp Psychiatry 2008;30(3):191-199. [doi: 10.1016/j.genhosppsych.2008.01.001] [Medline: 18433651]
60. Shapiro SL, Carlson LE, Astin JA, Freedman B. Mechanisms of mindfulness. J Clin Psychol 2006 Mar;62(3):373-386. [doi: 10.1002/jclp.20237] [Medline: 16385481]
61. Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med 2007 Mar 06;146(5):317-325. [doi: 10.7326/0003-4819-146-5-200703060-00004] [Medline: 17339617]
62. Ozbay F, Johnson DC, Dimoulas E, Morgan CA, Charney D, Southwick S. Social support and resilience to stress: from neurobiology to clinical practice. Psychiatry (Edgmont) 2007 May;4(5):35-40 [FREE Full text] [Medline: 20806028]
63. Aneshensel CS, Stone JD. Stress and depression: a test of the buffering model of social support. Arch Gen Psychiatry 1982 Dec;39(12):1392-1396. [doi: 10.1001/archpsyc.1982.04290120028005] [Medline: 7149899]
64. Krusche A, Cyhlarova E, King S, Williams JM. Mindfulness online: a preliminary evaluation of the feasibility of a web-based mindfulness course and the impact on stress. BMJ Open 2012 May 21;2(3):e000803 [FREE Full text] [doi: 10.1136/bmjopen-2011-000803] [Medline: 22614170]
65. Morledge TJ, Allexandre D, Fox E, Fu AZ, Higashi MK, Kruzikas DT, et al. Feasibility of an online mindfulness program for stress management--a randomized, controlled trial. Ann Behav Med 2013 Oct;46(2):137-148 [FREE Full text] [doi: 10.1007/s12160-013-9490-x] [Medline: 23632913]

66. Wolever RQ, Bobinet KJ, McCabe K, Mackenzie ER, Fekete E, Kusnick CA, et al. Effective and viable mind-body stress reduction in the workplace: a randomized controlled trial. J Occup Health Psychol 2012 Apr;17(2):246-258. [doi: 10.1037/a0027278] [Medline: 22352291]
67. Dobkin PL. Mindfulness-based stress reduction: what processes are at work? Complement Ther Clin Pract 2008 Feb;14(1):8-16. [doi: 10.1016/j.ctcp.2007.09.004] [Medline: 18243937]
68. Piet J, Würtzen H, Zachariae R. The effect of mindfulness-based therapy on symptoms of anxiety and depression in adult cancer patients and survivors: a systematic review and meta-analysis. J Consult Clin Psychol 2012 Dec;80(6):1007-1020. [doi: 10.1037/a0028329] [Medline: 22563637]
69. Schneiderman N, Ironson G, Siegel SD. Stress and health: psychological, behavioral, and biological determinants. Annu Rev Clin Psychol 2005;1:607-628 [FREE Full text] [doi: 10.1146/annurev.clinpsy.1.102803.144141] [Medline: 17716101]
70. Richardson S, Shaffer JA, Falzon L, Krupka D, Davidson KW, Edmondson D. Meta-analysis of perceived stress and its association with incident coronary heart disease. Am J Cardiol 2012 Dec 15;110(12):1711-1716 [FREE Full text] [doi: 10.1016/j.amjcard.2012.08.004] [Medline: 22975465]
71. Lin S. Occupational stress research: a review. Psychol Sci Shanghai 2003;26(3):494-497.
72. Robotham D, Julian C. Stress and the higher education student: a critical review of the literature. J Further High Educ 2006 May;30(2):107-117. [doi: 10.1080/03098770600617513]
73. Stankov L. Unforgiving Confucian culture: a breeding ground for high academic achievement, test anxiety and self-doubt? Learn Individ Differ 2010 Dec;20(6):555-563. [doi: 10.1016/j.lindif.2010.05.003]
74. Kim KI, Won H, Liu X, Liu P, Kitanishi K. Students' stress in China, Japan and Korea: a transcultural study. Int J Soc Psychiatry 1997;43(2):87-94. [doi: 10.1177/002076409704300202] [Medline: 9252822]
75. Zhang X, Wang H, Xia Y, Liu X, Jung E. Stress, coping and suicide ideation in Chinese college students. J Adolesc 2012 Jun;35(3):683-690. [doi: 10.1016/j.adolescence.2011.10.003] [Medline: 22032975]

Abbreviations

CET: College English Test
FFMQ: Five-Factor Mindfulness Questionnaire
GAD-7: 7-item Generalized Anxiety Disorder
LMIC: low- and middle-income country
PHQ-9: Patient Health Questionnaire-9
PSS: Perceived Stress Scale

Edited by G Eysenbach; submitted 14.11.20; peer-reviewed by P Dattathreya, E Feofanova; comments to author 21.11.20; revised version received 08.03.21; accepted 01.08.21; published 13.10.21.

Please cite as:
Rodriguez M, Eisenlohr-Moul TA, Weisman J, Rosenthal MZ
The Use of Task Shifting to Improve Treatment Engagement in an Internet-Based Mindfulness Intervention Among Chinese University Students: Randomized Controlled Trial
JMIR Form Res 2021;5(10):e25772
URL: https://formative.jmir.org/2021/10/e25772
doi:10.2196/25772
PMID:34643532

Original Paper

Need for Cognition Among Users of Self-Monitoring Systems for Physical Activity: Survey Study

Kirsi Halttu[1], MSc; Harri Oinas-Kukkonen[1], MSc, PhD

Oulu Advanced Research on Service and Information Systems Research Unit, Faculty of Information Technology and Electrical Engineering, University of Oulu, Oulu, Finland

Corresponding Author:
Kirsi Halttu, MSc
Oulu Advanced Research on Service and Information Systems Research Unit
Faculty of Information Technology and Electrical Engineering
University of Oulu
P.O. Box 3000
Oulu, FI-90014
Finland
Phone: 358 458601190
Email: kirsi.halttu@oulu.fi

Abstract

Background: Need for cognition (NFC) is among the most studied personality traits in psychology. Despite its apparent relevance for engaging with technology and the use of information, it has not been studied in the context of self-monitoring systems and wearables for health. This study is the first to explore the relationship between NFC and commercial self-monitoring systems among healthy users.

Objective: This study aims to explore the effect of NFC levels on the selection of self-monitoring systems and evaluation of system features of self-monitoring and feedback, as well as perceived credibility and perceived persuasiveness. We also assessed perceived behavior change in the form of self-reported activity after adopting the system.

Methods: Survey data were collected in October 2019 among university students and personnel. The invitation to respond to the questionnaire was addressed to those who had used a digital system to monitor their physical activity for at least two months. The web-based questionnaire comprised the following 3 parts: details of system use, partially randomly ordered theoretical measurement items, and user demographics. The data were analyzed using structural equation modeling. The effect of NFC was assessed both as 3 groups (low, moderate, and high) and as a continuous moderator variable.

Results: In all, 238 valid responses to the questionnaire were obtained. Individuals with high NFC reported all tested system features with statistically significantly higher scores. The NFC also had some effect on system selection. Hypothesized relationships with perceived credibility gained support in a different way for individuals with low and high NFC; for those with low NFC, credibility increased the persuasiveness of the system, but this effect was absent among individuals with high NFC. For users with high NFC, credibility was related to feedback and self-monitoring and perhaps continuously evaluated during prolonged use instead of being a static system property. Furthermore, the relationship between perceived persuasiveness and self-reported activity after adopting the system had a large effect size (Cohen f^2=0.355) for individuals with high NFC, a small effect size for individuals with moderate NFC (Cohen f^2=0.107), and a nonsignificant path (*P*=.16) for those with low NFC. We also detected a moderating effect of NFC in two paths on perceived persuasiveness but only among women. Our research model explained 59.2%, 63.9%, and 47.3% of the variance in perceived persuasiveness of the system among individuals with low, moderate, and high NFC, respectively.

Conclusions: The system choices of individuals seem to reflect their intrinsic motivations to engage with rich data, and commercial systems might themselves be a tailoring strategy. Important characteristics of the system, such as perceived credibility, have different roles depending on the NFC levels. Our data demonstrate that NFC as a trait that differentiates information processing has several implications for the selection, design, and tailoring of self-monitoring systems.

(JMIR Form Res 2021;5(10):e23968) doi:10.2196/23968

KEYWORDS

self-monitoring; wearables; physical activity tracking; mHealth; need for cognition; persuasive design; tailoring; user research; mobile phone

Introduction

Theoretical Background

Monitoring oneself in terms of behavior is one of three subfunctions in self-regulation models [1] and is elementary to behavior change. Self-monitoring is a behavior change technique (BCT) that refers to keeping a record of specified behavior related to a target behavior change domain or outcome of behaviors, such as weight loss [2]. Increasing numbers of individuals are turning to their smartphones and additional sensing devices such as wearables to collect data on their own behavior and to facilitate their personal self-regulation needs. Self-monitoring improves self-awareness; increases self-knowledge; and makes habitual, often unconscious, behavioral patterns more visible. Although many benefits of self-monitoring can also be attained with nondigital approaches, technology has increased its potential [3,4]. Currently, self-monitoring is the most common BCT among mobile apps designed to support physical activity [5] and health behavior change [6] and in health interventions [7], and it is the core function of activity trackers [8,9].

BCTs derived from self-regulation theory [10] and control theory [11,12] are generally considered effective. Self-monitoring combined with other elements of self-regulation, such as prompting intention formation and goal setting, providing feedback, and reviewing goals, is considered the most effective technique for achieving more physical activity and healthier eating [13]. Many self-monitoring systems have implemented theory- and evidence-based techniques [14] and present a viable and useful approach to support self-regulation toward healthier behaviors [15].

Self-monitoring is also a persuasive feature because of its ability to influence thinking and is part of the persuasive systems design (PSD) model [16]. Self-monitoring systems are, indeed, persuasive systems implementing persuasion, “a deliberate attempt to change attitudes or behaviors or both” [17]. A general theory of attitude change, the elaboration likelihood model (ELM) [18,19], is a cognitively oriented model of persuasion that explains how an influence process may affect attitude and behavior change. The ELM is based on the concept of message elaboration through central or peripheral routes that represent different levels and types of information processing: when presented with a message, individuals may process information through careful consideration (central route) or more automatically (peripheral route), relying on effortful analysis instead of simple decision rules [20]. Any piece of information, regardless of whether it is a rigorously designed motivational message or a small detail in the implementation of software or the environment of the persuasion event, can change the elaboration mode in a situation. This *multiple roles* notion suggests that situational factors also affect the likelihood and extent of elaboration, and any variable can influence it by serving as an argument, cue, determinant of the extent of elaboration itself, or source of bias [21,22].

The personal relevance of the argument is one of the strongest variables exerting an effect on the motivation to elaborate. Petty and Cacioppo [23] regard this construct as the personal meaning and intrinsic importance of an issue that people expect “to have significant consequences for their own lives” [24]. Behavior change approaches that aim to increase the personal relevance of information provided, thereby creating a more optimal environment for persuasion [25,26], are tailoring and personalization strategies. The effectiveness of these strategies is presumably based on increased involvement in, and engagement with, the subject matter [27], both possible outcomes of increased personal relevance [28]. Tailoring might use any part of the system, but most often, informational content is tailored to contain more relevant information for particular groups of users. To summarize, tailoring improves the fit between the user and the system, and it usually focuses on motivation to elaborate the information provided.

Tailoring is based on the assumption that target audiences differ in terms of the selected tailoring trait. There are rather stable individual differences in the intrinsic motivation to engage in extensive thinking and enjoy effortful cognitive activities, such as the need for cognition (NFC) [29]. The implications of the extensive studies on this personality trait are that individuals with high NFC have stronger information-seeking habits [30] and that they are in general more influenced by argument quality than peripheral cues [31]. They are also more motivated to process messages that they perceive as complex [32] and are more easily persuaded using cognition-based messages [33]. There are some preliminary indications that NFC affects how individuals interact with, and use, software, which implies that information processing and NFC itself affect both behaviors and actions. For example, those with high NFC use adaptive user interface features more frequently [34] and prefer personalized content and choose more preference-matched offers compared with individuals with low NFC [35]. These examples from existing studies support the relevance of NFC to interactive systems such as self-monitoring systems that provide considerable amounts of information to process.

One might expect that individuals who adopt and continue to use self-monitoring systems are inherently interested in information and are able to base their attitudes and behaviors accordingly. However, research has not addressed these assumptions, and it is not known how NFC is associated with the selection of systems or the evaluation of system features. In this study, we aim to fill this gap by exploring how individual differences in NFC influence the evaluations of self-monitoring system features in commercial systems for physical activity. Our survey sample comprised individuals who use these systems volitionally and have selected their systems themselves. In this setting, we examine whether users with different levels of NFC select their systems similarly or evaluate self-monitoring features

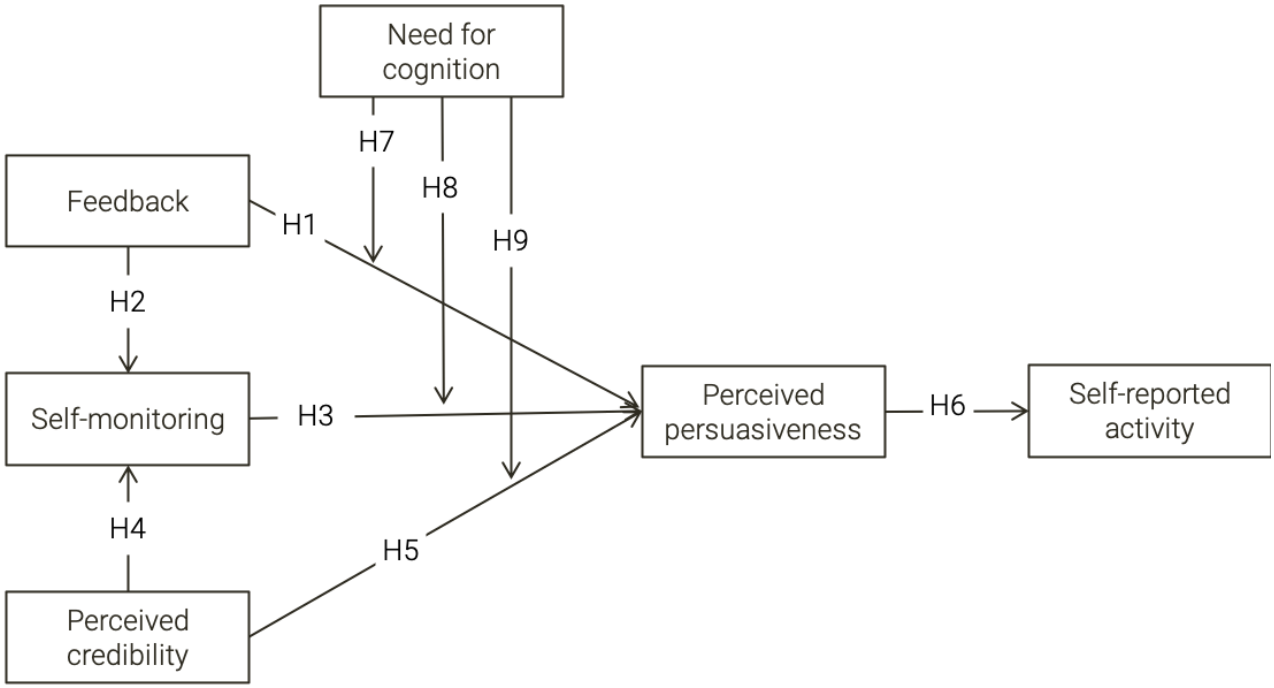
differently and whether this information is relevant for future tailoring approaches.

Model Development and Hypotheses

Research Model

We built a research model on the relationships among elements based on the PSD model [16]. These relationships have been validated in several studies using both web-based and mobile apps. All the relationships among the constructs, represented by the arrows (Figure 1), are assumed to be positive.

Figure 1. Research model and hypotheses.



Feedback and Perceived Persuasiveness

Feedback on behavior or on outcomes of behaviors can be considered *good* if it is delivered at the right time, is personally relevant, and is actionable [36]. On the basis of these assumptions, our feedback construct was created to measure these general functional qualities of feedback features rather than the actual content in terms of argument quality or wording. The connection between elaboration and persuasion refers to any change that results from exposure to communication [18], that is, the feedback the system provides to the user. Feedback features, regardless of being status information with numbers or words or instructions on how to reach the goal, form the arguments for persuasive attempts. Therefore, we expect the following relationship between feedback and subjective persuasiveness evaluation:

H1: Feedback functionality is positively related to the perceived persuasiveness of the system.

Feedback and Self-Monitoring

Feedback is an essential part of self-monitoring activity. It provides information for both setting realistic goals and evaluating goal attainment [10]. Feedback features can be implemented in many forms, and actual feedback can be provided through numbers, messages, and graphics. Some wearables also have sound, vibration, or lights that communicate the user's current status. Feedback is often obtained by taking quick glances at a device [37], but some devices do not have screens at all and provide feedback only through a mobile app (eg, basic pedometers, some activity wristbands, and smart rings). In any case, looking at data does not necessarily mean that they have been understood and reflected on [38]. According to the study by Michie et al [2], feedback and self-monitoring are categorized as different techniques, and both can target either behaviors or outcomes of behaviors. We followed this approach by developing our own construct for both techniques but aimed to keep them as general principles without specifically focusing on behaviors or their outcomes. This approach enables the evaluation of several systems by using the same constructs and is also congruent with system design practice: both self-monitoring and feedback can be designed and implemented in many different ways, and either might be dysfunctional in terms of the persuasiveness and behavior change support of the system.

Self-monitoring combined with feedback or other self-regulation– or control-theory–driven techniques such as giving feedback is considered effective for physical activity support [13]. The positive relationship between dialog support and the primary task features of the PSD model—feedback and self-monitoring features developed from these constructs of the model—has also been reported elsewhere [38-41]. We set the following hypothesis to evaluate the theorized and previously reported relationship between feedback and self-monitoring:

H2: Feedback functionality is positively related to self-monitoring features.

Self-Monitoring and Perceived Persuasiveness

Self-monitoring is a BCT where an individual monitors and records their own behavior (eg, number of steps per day) or outcomes of behaviors (eg, weight; Michie et al [2]). Self-monitoring is a functionality that has been shown to affect continuance intention through its perceived usefulness [15] and potential to disrupt unwanted habits [42,43]. Its ability to evoke self-adjustment actions when information about the monitored factor is presented has its roots in a phenomenon known as reactivity, referring to changes in behavior that occur because of merely being monitored by someone else. The reactivity effects of monitoring seem to be similar when a person is self-monitoring [44]. However, self-monitoring activity itself is not perceived as strongly motivating in cases where users have not chosen to start it based on personal motivations [45]. In addition, the notoriously high abandonment rates of self-monitoring tools imply that mere self-monitoring does not always increase the persuasiveness or desirability of a device. These findings imply that the persuasive nature and reactivity of self-monitoring might be based on certain pre-existing motivations to understand oneself or on behavior-change goals [46] or that the practicalities of actual self-tracking are often too difficult to maintain in the long term.

However, self-monitoring is often referred to as a persuasive feature [17,47] and is included as one of the primary task support features in the PSD model [16]. Previous research has shown that primary task support features increase the perceived persuasiveness of a system [39,40]. Because of the personal relevance of the self-monitoring data and self-selected nature of continued use of the system, we expect that self-monitoring features do influence the persuasiveness of the system. Hence, we set the following hypothesis:

H3: Self-monitoring is positively related to the user's perceived persuasiveness evaluation.

Perceived Credibility and Self-Monitoring

The credibility of the system is the core of the process. Users must trust the system to provide accurate information in terms of progressing toward, and reaching, their goals, as well as understanding their current status. High credibility is mandatory for a system that instructs or advises its users, reports measurements, and provides information and analysis. All these characteristics are typical of self-monitoring systems. If users trust the system, they are more likely to consider the advice provided and follow the recommendations. Therefore, the act of self-monitoring is highly dependent on the perceived credibility of the system. Alternatively, the system might be abandoned. The perceived credibility construct refers to trustworthiness, believability, and reliability, all relevant for the subjective evaluation of credibility. Therefore, we hypothesize as follows:

H4: The perceived credibility of the system has a positive influence on self-monitoring features.

Perceived Credibility and Perceived Persuasiveness

In persuasive communication, source credibility refers to the effects of the credibility of the message source. The nature of source credibility is perceived, that is, it refers to a subjective evaluation of several source-related context variables that have been shown to alter persuasion outcomes, for example, in child obesity campaigns [48]. However, the literature usually denotes source credibility as comprising expertise and trustworthiness [49-52]. A credible source is usually more persuasive and influential and is evaluated more favorably [53]. Credibility refers to an object of interest (the system in this case) and the trust of the message recipient: whether (or not) the recipient trusts the system [54]. The trustworthiness and precision of the information provided by the device are important in terms of its effectiveness and overall evaluations of the system [55]. Although trust in a service provider is a somewhat larger phenomenon in the context of self-monitoring systems that often entail both device and software, the findings in the field of health information systems indicate that credibility forms through interaction with the system or service [56] and transforms into trust with prolonged use [57]. Issues that might endanger the credibility of the system include poor style and errors [54], which signal a lack of expertise and a decrease in trust.

Our construct queries the trustworthiness of the system, believability of the measurements, and expertise as impressions that the system was built by professionals. The relationship between perceived credibility and perceived persuasiveness has previously been reported as significant and positive [39,40] in terms of the PSD model constructs that we have used as the baseline for our items. On the basis of existing studies on how the credibility of systems is developed and how self-monitoring systems should perform, we offer the following hypothesis:

H5: Perceived credibility has a positive relationship with perceived persuasiveness.

Perceived Persuasiveness and Subjective Activity Change

The study by Lehto [58] defines perceived persuasiveness as "individuals' favorable impressions toward the [behavior change support systems]." Therefore, it has an attitudinal component. The construct is created to measure subjective rather than objective persuasiveness, which is a variable that requires measuring actual behavior or attitude changes. The measurement items of the construct encourage thinking about the system's influence, for example, by inviting metacognitive processing, which has been connected to high message elaboration conditions [59]. Previous studies have reported that the perceived persuasiveness of a system has a positive influence on continuance intention, intention to use, and intention to adopt health behavior change support systems [39,40]. We hypothesize that the salience of these self-evaluative thoughts positively influences self-reported activity:

H6: Perceived persuasiveness has a positive influence on self-reported level of activity after adopting a self-monitoring system.

NFC as Moderator

Personal Relevance of Overall Feedback

In the context of the volitional use of self-selected systems used for self-monitoring, we have assumed that the personal relevance of overall feedback is rather high. The data collected are always

personal behavioral data, and the insights gained through the information that the data provide are similar to personalized or tailored messages because they are based on the user's own measurement data. However, not all information is of the same relevance, depending on an individual's goals, current situation, or more long-standing lifestyles.

Feedback and Perceived Persuasiveness

The feedback construct was created to measure the general functional qualities of feedback features instead of their actual content in terms of argument quality or wording. Previously, quantified types of motivation to use self-monitoring systems had been connected to both affective and informational feedback [60]. In the context of wearable trackers and smartphone apps with sensors, we consider the content of the feedback to be cognitive in nature as well. Cognition-based messages have been identified as more persuasive and attention-enhancing among individuals with high NFC than among individuals who have a high need for affect [33] and individuals with low NFC. This effect has also recently been demonstrated using neuroscientific methods [61]. Feedback can also be categorized as an assessment type of information, which has recently been linked to individuals with high NFC [62]. Therefore, we expect that NFC moderates the relationship between feedback and perceived persuasiveness, leading to the following hypothesis:

H7: NFC moderates the relationship between feedback and perceived persuasiveness. The positive influence of feedback is stronger for individuals with high NFC because of the informative nature of feedback in self-monitoring systems.

Self-Monitoring and Perceived Persuasiveness

The tendency demonstrated by NFC is heavily geared toward cognitive functions. NFC correlates positively with objectivity and refers to a tendency to rely on empirical information and rational consideration [63]. Messages based on cognition were more persuasive than affective messages among individuals with high NFC compared with individuals with low NFC; NFC also influenced receptivity to cognition-based messages but not to affect-based messages [33]. These results imply that the informative nature of a system's self-tracked, sensor-based information that supports self-monitoring activities should appeal especially to individuals with high NFC. NFC has also been connected to intrinsic motivation, and it predicts intrinsic enjoyment and both self-perceived and behavioral motivation to engage in cognitive, effortful elaboration [64]. We hypothesize that NFC has a similarly positive influence on self-monitoring activity that provides data, information, and the opportunity to engage in conscious processing in terms of information related to the self:

H8: NFC moderates the relationship between self-monitoring and perceived persuasiveness.

Perceived Credibility and Perceived Persuasiveness

The role of source credibility, often in the form of source expertise, in persuasion is crucial in ELM and NFC research. It is considered a peripheral cue, and therefore individuals with low NFC should be more susceptible to processing such information. However, according to the multiple roles postulate, a variable such as credibility "...can have the same impact on judgements by different processes in different situations" [21]. Therefore, the credibility indicators of the system can also be processed in both ways. For example, the social influence strategy of authority [65], which is often referred to as a peripheral credibility cue, might also be processed through the central route among individuals with high NFC [66]. In general, credible sources are more persuasive; however, studies have indicated that peripheral cues are considered less meaningful when the argument is strong and personally relevant. In our research setting, we consider arguments to be mainly strong and personally relevant, which might partly suppress the effect of credibility. To evaluate the relevance of the aforementioned assumptions, we set the following hypothesis to address the role of NFC:

H9: NFC moderates the relationship between perceived credibility and perceived persuasiveness.

Methods

Measurement Instrument

We adapted scales that have been validated in several previous studies. Feedback, self-monitoring, perceived credibility, and perceived persuasiveness are all based on existing constructs developed to evaluate a PSD model [16]. The first two of these constructs were modified slightly to provide more details regarding self-monitoring system features. Therefore, they have also been named according to the PSD model principle instead of according to the category from which they are derived (dialog support and primary task support). Self-reported activity is a single-item measure developed for this study to query the subjective evaluation of perceived change in physical activity. To measure NFC, we used a shortened version of the NFC scale assessed in the study by Chiesi et al [67]. A detailed list of the measurement items and their exact wordings as used in this study are presented in Multimedia Appendix 1 [29,39-41,67].

Data Collection

The questionnaire was implemented with the Webropol survey and reporting tool and sent to both students and employees of the University of Oulu, Finland, using mailing lists. The introductory text for the survey indicated that the respondents should be using a digital system that enables them to monitor their physical activity but that it does not have to be for that purpose only. We targeted users who had used the system for at least two months, which was also mentioned in the survey invitation. There were no other requirements or restrictions for participation, and the survey was fully anonymous.

The questionnaire consisted of 3 parts. The first part queried which systems the respondent used and how long they had used them to monitor their activity. The question regarding the estimation of current physical activity level compared with activity level before using the system was presented before the second part, which presented the theoretical measurement items. The final part collected respondent demographics. A detailed list of measurement items is presented in Multimedia Appendix 1. In terms of measurement items, 2 constructs were included on each page of the survey, and the pages were in the same fixed order for all respondents. The items for the 2 constructs on each

page were randomly ordered. All the constructs were measured on a 7-point Likert scale, and the single-item measure for the subjective evaluation of the increase in physical activity was a scale of 5 statements that used typical 5-point Likert wordings.

During the 1-week survey period in October 2019, 261 responses were received. All the questions were mandatory; therefore, no data were missing. We removed cases that reported the use of systems that did not match our criteria (self-monitoring of physical activity). In addition, outlier responses were removed if they had a high likelihood of being faulty, such as when all the questions on a page (consisting of 2 constructs) had the same extreme value and the other values did not reflect similar evaluations of the system. It should be noted that we did not remove outliers based only on the values themselves, but also on highly inconsistent responses in general. The final sample consisted of 238 valid responses.

The choice of analysis method was structural equation modeling (SEM), namely partial least squares SEM (PLS-SEM). PLS-SEM is especially suitable for explorative research with non–normally distributed data [68,69]. It is also able to analyze complex models with a relatively small sample size [69]. However, for our model, the minimum sample size was achieved for the full sample and all subgroups that we derived from the sample. According to the study by Hair et al [69], the minimum sample should be 10 times the largest number of structural paths directed at a particular latent construct in the structural model, which in our model was 30. The collected data were analyzed using SPSS software (version 26.0; IBM Corp) and SmartPLS software (version 3.0; SmartPLS GmbH) [70].

Results

Demographic Information and Use of Systems

The final sample consisted of 238 valid responses. The basic demographics of the participants are presented in Table 1. The respondents were primarily women. Of the 238 respondents, 144 (60.5%) were aged below 30 years, 180 (75.6%) had an undergraduate or higher degree, and 99 (41.6%) had used a monitoring system for more than 2 years.

Altogether, the respondents used more than 20 different services, and 29.8% (71/238) provided the name of the specific tracker they used in addition to a mobile app (adding a tracker was not mandatory because many services can be used without an additional tracker). A few respondents used several devices such as the Ōura Ring (Ōura Health Oy), Vivofit (Garmin Ltd), or Polar M400 (Polar Electro) for activity tracking and a specific sports watch for measuring training sessions. The most used system was the Polar Flow app (92 users), which is used to support the use of Polar activity trackers and heart rate monitors. Many of these apps can be used with several different trackers, as is the case for apps such as Polar Flow (Polar Electro), Suunto (Amer Sports Oyj), Garmin (Garmin Ltd), and Fitbit (Fitbit LLC). Multimedia Appendix 2 illustrates the basic self-monitoring and feedback features of the 8 most common systems.

Table 1. Sample characteristics (N=238).

Variable and category	Participant, n (%)
Gender	
Women	137 (57.6)
Men	100 (42)
Other	1 (0.4)
Age groups (years)	
19-29	144 (60.5)
30-39	48 (20.2)
40-49	27 (11.3)
≥50	19 (8)
Service use time (months)	
2-6	41 (17.2)
6-12	40 (16.8)
12-24	58 (24.4)
24-36	41 (17.2)
>36	58 (24.4)
Service (app name)	
Polar Flow	92 (38.7)
Apple Health	19 (8)
Sports Tracker	19 (8)
Suunto app	17 (7.1)
Ōura	16 (6.7)
Garmin Connect	14 (5.9)
Fitbit	13 (5.5)
Samsung Health	10 (4.2)
Education	
High school diploma	47 (19.7)
Vocational degree	11 (4.6)
Bachelor's degree	82 (34.5)
Master's degree	68 (28.6)
Doctor's degree	30 (12.9)

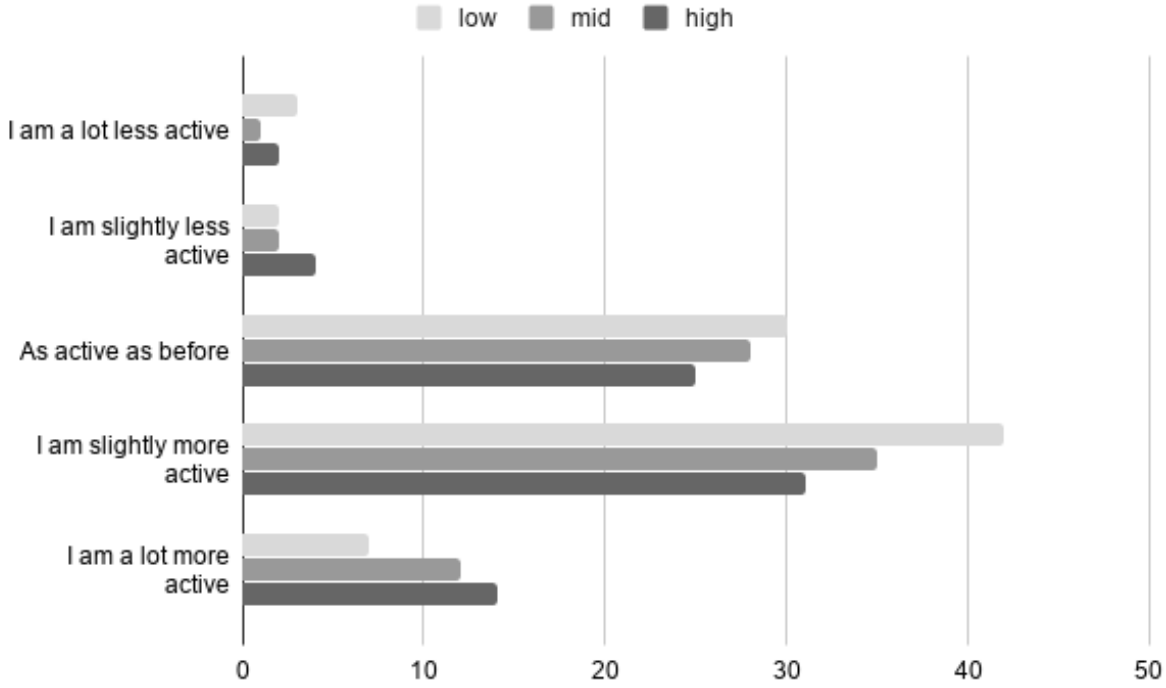
Analysis of Distributions and Group Characteristics

In this section, we present the results of the statistical analysis of the collected data. First, we analyzed whether our results differed in relation to the NFC scores or other groups possibly relevant for this study. To study the distribution among the levels of NFC scores, we divided our sample into 3 proportions—low, moderate, and high NFC—using visual binning. The upper limit of the cut point was included in the group, and this resulted in groups with 84, 78, and 76 individuals, respectively. This three-group approach is partly aligned with the recommendation for system design and message tailoring provided in the study by Nikoloudakis et al [71]. This recommendation, however, prefers categorization using SD because it correctly identifies the nature of the measurement (most individuals fall in the moderate-level category), but our sample was too small for such a grouping.

We also tested whether our theoretical measures were distributed equally across other groups such as the self-monitoring systems that the respondents used, duration of system use, age group, education, or gender. There were significantly different distributions among the NFC subgroups, genders, and systems used. The Kruskal-Wallis test showed that the NFC subgrouping significantly affected how individuals responded in terms of self-monitoring (Kruskal–Wallis H_2=7.576; P=.02), feedback (Kruskal–Wallis H_2=8.639; P=.01), perceived credibility (Kruskal–Wallis H_2=17.463; P<.001), and perceived persuasiveness (Kruskal–Wallis H_2=16.786; P<.001). In all the aforementioned tests, individuals with high NFC reported

significantly higher mean scores than those in the low or moderate subgroups. For self-monitoring and feedback, pairwise comparison revealed a significant difference between the low- and high-NFC groups (P=.048 and P=.01, respectively). Perceived credibility and perceived persuasiveness differed in pairwise comparisons between both low and high (P<.001 and P=.001, respectively) and moderate and high (P=.02 and P=.001, respectively). The distributions of NFC were not significantly different for the self-reported activity levels. The frequencies of responses in the NFC subgroups are presented in Figure 2. Overall, 58.8% (140/238) of the respondents perceived that their activity was higher than before they started using the system.

Figure 2. Self-reported activity after adopting the system in the need for cognition groups.



Women evaluated their systems significantly higher in terms of perceived persuasiveness (Kruskal–Wallis H_2=7.485; P=.02). The duration of system use, age group, and education all exhibited the same distributions across the subgroups. Table 2 presents the basic statistics of the full sample and the groups that showed differences, namely, the gender and NFC subgroups.

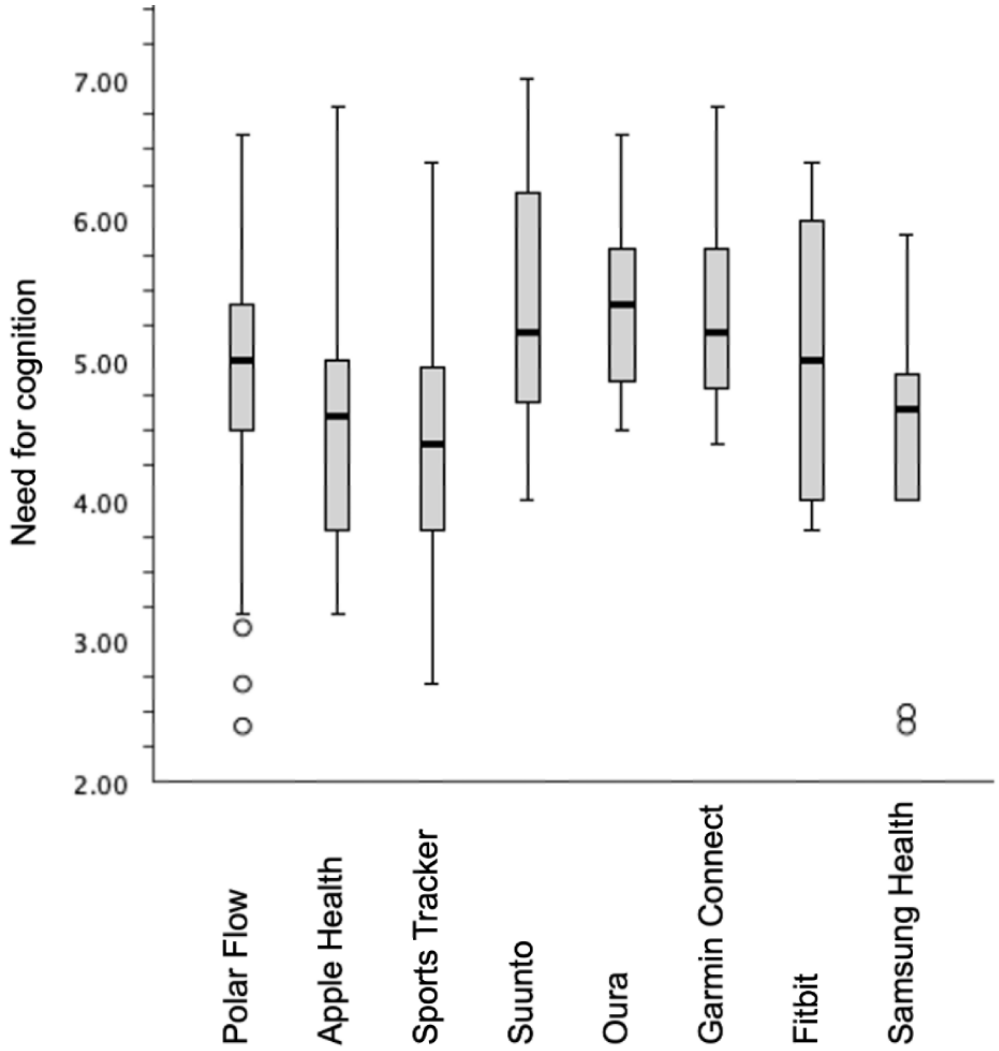
Table 2. Means (SDs) for variables per full sample and subgroups.

Variables	Value, mean (SD)					
	Full sample (N=238)	Women (n=137)	Men (n=100)	NFC[a] group		
				Low (n=76)	Moderate (n=78)	High (n=84)
Self-monitoring	4.96 (1.20)	5.01 (1.24)	4.90 (1.15)	4.83 (1.17)	4.78 (1.26)	5.29 (1.11)
Feedback	4.83 (1.24)	4.83 (1.28)	4.82 (1.19)	4.65 (1.17)	4.69 (1.25)	5.17 (1.25)
Perceived credibility	5.33 (0.97)	5.35 (1.03)	5.29 (0.88)	5.06 (0.97)	5.29 (0.89)	5.66 (0.95)
Perceived persuasiveness	5.31 (1.16)	5.47 (1.13)	5.09 (1.18)	5.10 (1.17)	5.12 (1.17)	5.72 (1.04)
NFC	4.94 (0.91)	4.93 (0.96)	4.95 (0.84)	4.00 (0.57)	4.97 (0.15)	5.95 (0.45)
Self-reported activity	3.65 (0.85)	3.80 (0.76)	3.45 (0.94)	3.57 (0.83)	3.71 (0.81)	3.67 (0.93)

[a]NFC: need for cognition.

The distributions of measured theoretical constructs (self-monitoring, feedback, perceived credibility, and perceived persuasiveness) were the same across all 8 self-monitoring systems used. However, the distribution of NFC scores differed across the 8 systems (H_7=21.709; P=.003). NFC differed significantly in one pairwise comparison, with the difference between Sports Tracker (Sports Tracking Technologies) and Ōura Ring users (P=.02) in terms of adjusted, Bonferroni-corrected values. Figure 3 shows the means and distributions of the NFC scores among the 8 most used systems. Sports Tracker users scored the lowest (mean and overall). Their NFC levels were similar to those of the users of the platform services Apple Health and Samsung Health. Dedicated health trackers (Polar, Suunto, Ōura, Garmin Connect, and Fitbit) scored marginally higher overall.

Figure 3. Distributions of need for cognition scores among users of top eight self-monitoring systems.



Structural Equation Modeling

Measurement Model

PLS-SEM analysis comprises 2 steps. First, the measurement model is assessed by analyzing the relationship of each indicator with its corresponding construct. All our constructs were reflective. Internal consistency reliability was measured using Cronbach α and composite reliability, as presented in Table 3. Composite reliability varies between 0 and 1, with higher values indicating higher levels of reliability. It is generally interpreted in the same way as Cronbach α, and values between 0.70 and 0.90 can be regarded as satisfactory [72]. In our data, composite reliability was acceptable for all the constructs after we followed the instructions provided in the study by Hair et al [73] and removed some items. The full set of our original and final measurement items is presented in Multimedia Appendix 1, with references to relevant studies and item loadings.

Table 3. Properties of latent variables, including Cronbach α, composite reliability, and average variance extracted.

Variables	Cronbach α	CR[a]	AVE[b]	NFC[c]	CRED[d]	PEPE[e]	FEEDB[f]	SELFM[g]
NFC	.845	0.889	0.616	*0.785* [h]	—[i]	—	—	—
CRED	.862	0.901	0.645	0.241	*0.803*	—	—	—
PEPE	.862	0.906	0.706	0.237	0.645	*0.840*	—	—
FEEDB	.805	0.885	0.720	0.185	0.521	0.563	*0.848*	—
SELFM	.773	0.868	0.687	0.168	0.589	0.657	0.670	*0.829*

[a]CR: composite reliability.

[b]AVE: average variance extracted.

[c]NFC: need for cognition.

[d]CRED: perceived credibility.

[e]PEPE: perceived persuasiveness.

[f]FEEDB: feedback.

[g]SELFM: self-monitoring.

[h]The variables in italics show the square root of average variance extracted and interconstruct correlations.

[i]Not applicable.

Discriminant validity was assessed using the Fornell-Larcker criterion: the values in italics in Table 3 showing the square root of average variance extracted should be higher than the interconstruct correlations. We omitted one item each from the self-monitoring and feedback constructs to tackle a heterotrait-monotrait ratio that was initially above 0.9. These omissions resulted in a satisfactory heterotrait-monotrait ratio level. The loading of each item of the constructs and the cross-loadings that validate the discriminant validity of our constructs are presented in Multimedia Appendix 3.

Structural Model and Hypothesis Testing

The second part of PLS-SEM, evaluation of the structural model, represents the underlying theory of the path model and allows for the determination of how well the empirical data collected supports the theory-derived hypotheses. The key results are obtained by defining the path coefficients and explained variances (R^2 values). We followed the recommendation provided in the study by Hair et al [73] and opted not to use the goodness-of-fit criterion for PLS-SEM. We used the complete bootstrapping method with 5000 resamples and parallel processing with no sign changes. The CI method was the two-tailed bias-corrected and accelerated bootstrap (default setting). For the moderation hypotheses, we generated interaction terms with NFC as a continuous moderator and used the product indicator calculation method and mean-centered product term generation as recommended in the study by Hair et al [73].

The results of the hypothesis testing for the research model are presented in Table 4. All the basic hypotheses were supported by the full sample. Effect size (Cohen f^2) was considered large if it was above 0.35. Nonsignificant (below 0.02) effect sizes have been removed from the table. There was no significant relationship between feedback features and perceived persuasiveness of systems for men, which was the only hypothesis not supported for both genders. Analysis of the 3 NFC groups revealed that feedback had a positive influence on perceived persuasiveness only for individuals with high NFC. Perceived credibility had a positive effect on perceived persuasiveness in all the groups except the one with individuals with high NFC, with a large effect size for the moderate-NFC group. Finally, perceived persuasiveness was positively related to self-reported activity evaluations for all groups except the one with individuals with low NFC, with a large effect size only for individuals with high NFC. To summarize, of the 6 basic hypotheses analyzed, 3 indicated significant differences among the NFC groups. The relationship between feedback and self-monitoring was the most stable path, with all group hypotheses supported with $P<.001$. The respondents also reported their perceived physical activity level compared with their activity level before using the system. There was a significant relationship between perceived persuasiveness and self-reported activity for all groups, except for individuals with low NFC. However, the relationship had a large effect size only for individuals with high NFC.

Table 4. Results of hypothesis testing for the full sample and representative groups (N=238).

Hypothesis and group	Path	*t* test (*df*)	*P* value	Cohen f^2	Support
H1: Feedback functionality is positively related to the perceived persuasiveness of the system					
Full sample	0.140	2.188 (8)	.04	0.023	Yes
Women	0.158	1.931 (8)	.05	0.028	Yes
Men	0.161	1.677 (8)	.13	0.032	No
Low NFC[a]	0.018	0.162 (8)	.90	—[b]	No
Moderate NFC	0.159	1.628 (8)	.36	0.045	No
High NFC	0.218	2.047 (8)	.03	0.049	Yes
H2: Feedback functionality is positively related to self-monitoring features					
Full sample	0.499	8.200 (8)	<.001	0.385	Yes
Women	0.503	5.811 (8)	<.001	0.421	Yes
Men	0.502	7.057 (8)	<.001	0.356	Yes
Low NFC	0.662	8.735 (8)	<.001	0.844	Yes
Moderate NFC	0.447	3.669 (8)	<.001	0.277	Yes
High NFC	0.414	4.494 (8)	<.001	0.283	Yes
H3: Self-monitoring is positively related to the user's perceived persuasiveness evaluation					
Full sample	0.348	4.494 (8)	<.001	0.125	Yes
Women	0.234	2.361 (8)	.02	0.054	Yes
Men	0.478	4.325 (8)	<.001	0.271	Yes
Low NFC	0.406	3.065 (8)	.003	0.146	Yes
Moderate NFC	0.354	3.486 (8)	<.001	0.205	Yes
High NFC	0.400	2.631 (8)	.01	0.127	Yes
H4: The perceived credibility of the system has a positive influence on self-monitoring features					
Full sample	0.329	5.013 (8)	<.001	0.167	Yes
Women	0.348	3.862 (8)	<.001	0.202	Yes
Men	0.306	3.564 (8)	<.001	0.133	Yes
Low NFC	0.212	2.368 (8)	.02	0.087	Yes
Moderate NFC	0.306	2.363 (8)	.02	0.123	Yes
High NFC	0.451	4.619 (8)	<.001	0.336	Yes
H5: Perceived credibility has a positive relationship with perceived persuasiveness					
Full sample	0.352	4.870 (8)	<.001	0.168	Yes
Women	0.446	5.620 (8)	<.001	0.279	Yes
Men	0.220	1.848 (8)	.07	0.070	Yes
Low NFC	0.443	3.596 (8)	<.001	0.312	Yes
Moderate NFC	0.476	6.023 (8)	<.001	0.475	Yes
High NFC	0.132	0.921 (8)	.36	0.019	No
H6: Perceived persuasiveness has a positive influence on self-reported level of activity after adopting a self-monitoring system					
Full sample	0.335	5.484 (8)	<.001	0.127	Yes
Women	0.260	3.331 (8)	.001	0.072	Yes
Men	0.372	4.026 (8)	<.001	0.162	Yes
Low NFC	0.182	1.420 (8)	.16	0.034	No
Moderate NFC	0.311	2.745 (8)	.006	0.107	Yes

Hypothesis and group	Path	*t* test (*df*)	*P* value	Cohen f^2	Support
High NFC	0.511	6.429 (8)	<.001	0.355	Yes
H7: NFC moderates the relationship between feedback and perceived persuasiveness. The positive influence of feedback is stronger for individuals with high NFC because of the informative nature of feedback in self-monitoring systems					
Full sample	–0.048	0.851 (8)	.40	0.025	No
Women	–0.066	1.421 (8)	.16	0.052	No
Men	0.027	0.360 (8)	.72	0.007	No
H8: NFC moderates the relationship between self-monitoring and perceived persuasiveness					
Full sample	–0.040	0.822 (8)	.41	0.016	No
Women	–0.083	2.875 (8)	.004	0.074	Yes
Men	0.047	0.634 (8)	.53	0.015	No
H9: NFC moderates the relationship between perceived credibility and perceived persuasiveness					
Full sample	–0.090	1.363 (8)	.17	0.051	No
Women	–0.098	2.077 (8)	.04	0.080	Yes
Men	0.126	0.871 (8)	.38	0.124	No

[a]NFC: need for cognition.

[b]Nonsignificant values are omitted.

Two of the three moderation hypotheses were supported, but only for women. Moderator path coefficients were negative for women and positive for men. Simple slopes were drawn for the supported hypotheses for women. In terms of perceived persuasiveness of the system, self-monitoring features made no difference to those with high NFC (Figure 4); however, for women with low NFC, high self-monitoring increased the perceived persuasiveness of the system. For both women with low NFC and women with high NFC, high perceived credibility of the system increased the perceived persuasiveness of the system (Figure 5). However, this effect was stronger for individuals with low NFC, although it reached levels of persuasiveness similar to those among women with high NFC when the credibility was high.

Figure 4. Simple slope showing the need for cognition moderation effect for self-monitoring's influence for perceived persuasiveness among women (hypothesis H8). NFC: need for cognition.

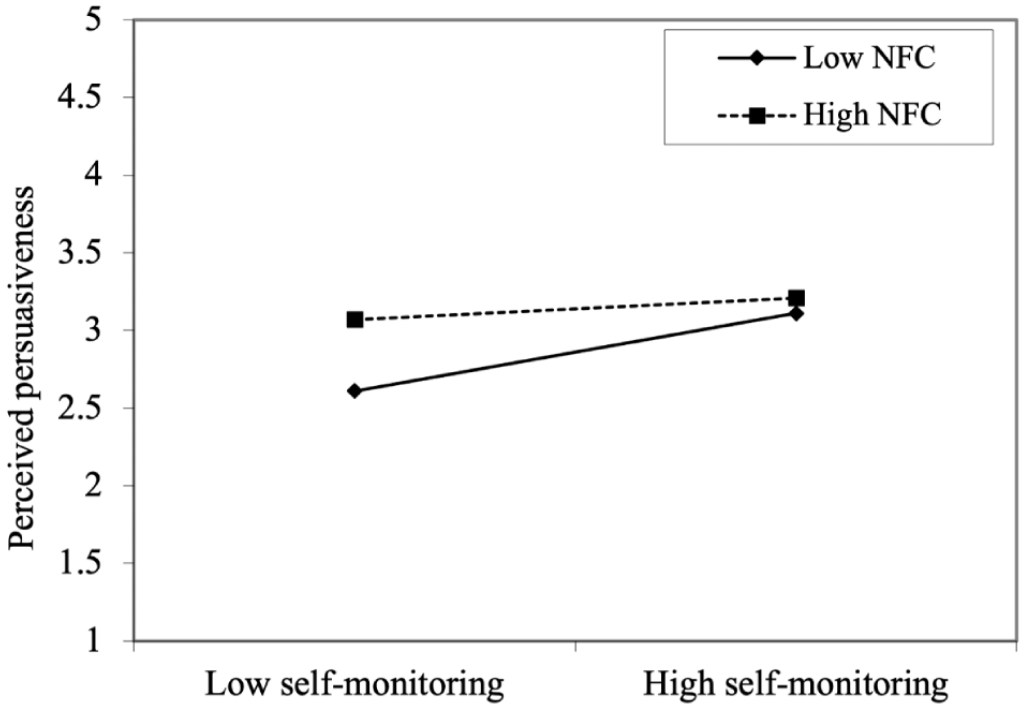
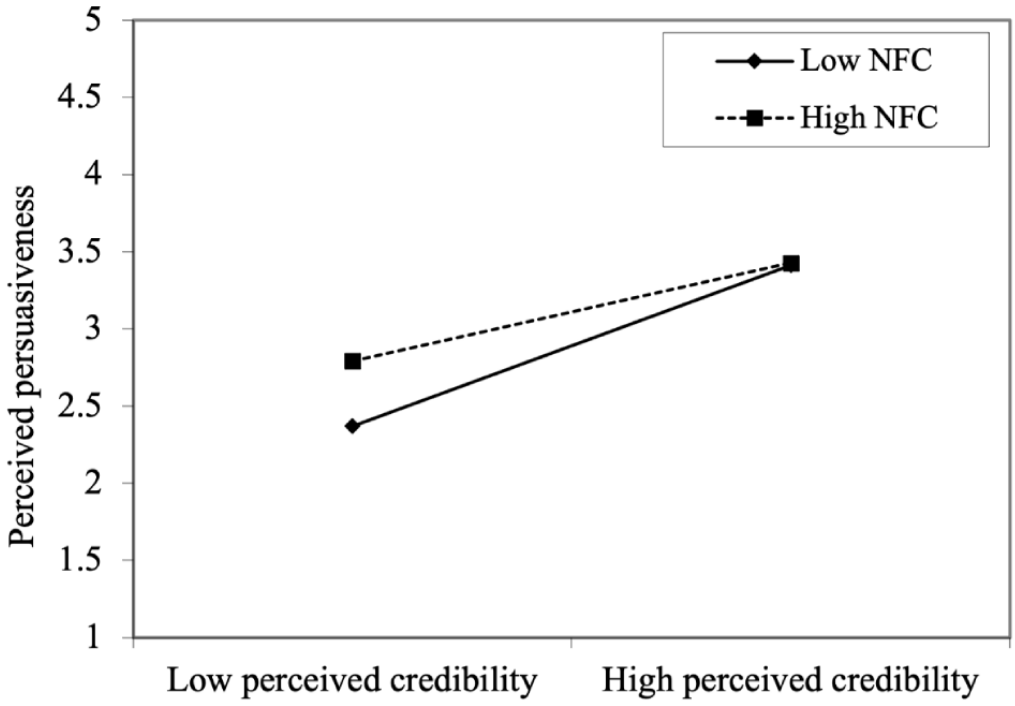


Figure 5. Simple slope showing the need for cognition moderation effect for perceived credibility's influence for perceived persuasiveness among women (hypothesis H9). NFC: need for cognition.



Analysis of R^2 values (Table 5) indicated that, altogether, feedback, self-monitoring, and perceived credibility explained 54.4% of the variance in the perceived persuasiveness of the system. This value peaked among individuals with moderate NFC, where self-monitoring and perceived credibility explained approximately 64% of the perceived persuasiveness construct. For individuals with high NFC, feedback and self-monitoring explained 47.3% of perceived persuasiveness. Variance in the self-monitoring construct was best explained for individuals with low NFC (63.8%) and least for the moderate-NFC group (41.1%). Finally, the single-item construct, self-reported activity, was significant only for men with high NFC and the high NFC group. Overall, both exogenous constructs (perceived persuasiveness and self-monitoring) revealed more differences within the NFC groups than between the genders.

Table 5. Variances explained as percentages (R^2; N=238).

Construct	Variance (%)					
	Full sample, variance	Women, variance	Men, variance	NFC[a] group, variance		
				Low	Moderate	High
Perceived persuasiveness	54.4	54.5	59.4	59.2	63.9	47.3
Self-monitoring	52.8	54.7	51.8	63.8	41.1	58.2
Self-reported activity	11.3	6.8[b]	14	3.8[b]	9.7[b]	26.2

[a]NFC: need for cognition.

[b]Nonsignificant value.

Discussion

Principal Findings

This study explored the relationship between the individual trait of NFC and commercial self-monitoring systems. In addition, it aims to understand whether this trait shows promise in designing systems for self-monitoring. We found that the respondents evaluated all studied features significantly differently based on their level of NFC, and, to some extent, there were differences in their system choices. Perceived credibility did not contribute to persuasiveness for individuals with high NFC; for these individuals, high persuasiveness was constructed based on self-monitoring and feedback. For individuals with low and moderate NFC, perceived credibility also influenced the persuasiveness of the system. There was no relationship between perceived persuasiveness and self-reported activity level among individuals with low NFC, but the effect was large for individuals with high NFC. In addition, it seems that the individuals' system choices reflected their intrinsic motivations to engage with rich sources of data, exemplified by an overall lower level of NFC among users of smartwatch systems and a higher level of NFC among dedicated fitness device users.

SEM revealed several differences among the NFC groups. Our study reports that feedback features increased the perceived persuasiveness of the system only for individuals with high NFC, and the effect was small. Previous research indicated that NFC is a trait that might motivate individuals to seek feedback. The recent study by Vaughan-Johnston and Jacobson [62] reported 2 studies that focused on preferences for different types of self-relevant feedback. NFC was related to an assessment type of feedback, and this relationship was stronger for individuals with high NFC. In our study, the relationship between feedback and self-monitoring features, in contrast, was strong among all NFC subgroups, and the effect size for individuals with low NFC (0.844) was exceptionally large. These results are in line with previous research suggesting that individuals with low NFC tend to take feedback more literally [74], without extensive elaboration; therefore, feedback and self-monitoring features form the strongest connection among those with low NFC. On the basis of this insight, it seems that individuals with high NFC do not accept feedback as is but use it as a data source to elaborate on the issue.

There was a prominent difference among the NFC groups in terms of the role of perceived credibility in forming the persuasiveness of the system. This relationship was strong in both the low- and moderate-NFC groups but nonexistent in the high-NFC group. This is in line with previous research that states that individuals with low NFC are more susceptible to persuasion using peripheral cue variables such as the 2 dimensions of credibility: expertise and trustworthiness [53,75,76]. Overall, the influence of perceived credibility on the users' evaluation of a primary task type of feature, in this case self-monitoring, was stronger for individuals with high NFC, implying that the overall perception of source credibility for individuals with high NFC might be mediated through their evaluation of self-monitoring tasks rather than through more general credibility cues such as trustworthiness or source credibility.

Perceived persuasiveness is usually explained by the direct effect of self-monitoring, which lends support for its innate persuasive nature and ability to evoke goal-directed actions. This was strongest for individuals with moderate NFC and weakest for individuals with high NFC, for whom the direct effect of perceived credibility was lacking. Individuals with high NFC seemed to derive persuasiveness also from direct interaction with the feedback features instead of more peripheral source credibility variables. In general, both credibility and persuasiveness evaluations were high in the full sample, indicating positive attitudes toward these systems in general. In addition, these scores were significantly higher among individuals with high NFC, which is to be expected based on the nature of the NFC trait. Considering the length of time our respondents had used these systems, prolonged use itself suggests user satisfaction. The repetitive nature of the interaction with the system might also contribute to high persuasiveness because (moderate) repetition has been identified as elaboration-enhancing and persuasive [77-80]. Repeating weak arguments would most likely not support long-term engagement with the system; therefore, our findings might suggest that, overall, users of these systems perceive the messages that the systems deliver as strong arguments, which increases the perceived persuasiveness in prolonged use. However, perceived behavior change, measured by self-reported activity after adopting the system, is not related to perceived persuasiveness among individuals with low NFC.

Although NFC is considered gender-neutral [30], our moderation hypotheses resulted in significant moderation only for women. To our knowledge, gendered moderation has not been previously reported. There were no significant differences in the systems the women used or in NFC values, but the women evaluated their systems as significantly more persuasive than the men. For women with low NFC, the evaluation of self-monitoring clearly increased the persuasiveness of the system such that a high rate of use of the self-monitoring feature resulted in a more persuasive system. There was no similar effect for women with high NFC, who rated persuasiveness very similarly, regardless of how they perceived self-monitoring. The same finding was present also for hypothesis H3. Altogether, self-monitoring and credibility features increased the perceived persuasiveness for women with low NFC to the same degree as for their high-NFC counterparts, which could imply that these features can improve attitudinal responses toward these systems.

Comparison With Prior Work

There is very little previous research to compare our findings with, in terms of NFC and use of self-monitoring technologies. The only study available [81] used a short German version of an NFC scale with 4 items [82]. The German version is based on the original 34-item scale [29], and we used a 10-item scale [67] that discarded 2 of the items in the German scale. One item included in the German scale was removed from our SEM analysis because of low loading, leaving our study to share only one item with the study by Attig et al [81]. Our study is, to our knowledge, the first to report a significant relationship between NFC and commercial self-monitoring tools. When we compared our study with previous research on tailoring in the health domain, we found that a few studies had used three-item questionnaires [83,84]. All these 3 items were included in our questionnaire, which showed good internal consistency, but 2 of them were discarded from the SEM analysis because of insufficient loadings.

Self-monitoring systems that are designed specifically for self-tracking and that collect several different metrics and present a complex set of indicators to the user to reflect upon (in this study, examples of such systems are Polar, Suunto, and Ōura) seem to be favored by users with high NFC. Exploring new or different systems can itself be considered a cognitively engaging activity [85], and a tendency to seek support and solutions for personal health issues from sources such as measuring devices fits well with previous research on individuals with high NFC. Users of self-monitoring apps without a dedicated fitness or activity tracker, exemplified by Apple Health, Sports Tracker, and Samsung Health, scored generally lower in terms of NFC. This might be a result of self-tracking mainstreaming through smartwatches and mobile sensors. Many individuals end up tracking their activity without a specific aim to do so because the functionalities are available through multipurpose devices. The wide range of NFC scores

for Fitbit, which has tracking devices for both low and high ends of technical and functional requirements, also supports the idea that NFC might function at the system selection level, and tailoring system features and content might not be the most obvious tailoring targets. In terms of volitionally used self-selected systems and the information they present to users, the subjective nature of complexity [32] poses interesting questions regarding preferred systems, especially because individuals with high NFC restrain themselves from engaging in effortful mental work when a message is perceived to be simple.

There is surprisingly little published research on NFC as an individual difference affecting system use. We identified differences in system evaluations in terms of NFC in every software feature construct that we measured. This implies that differences in subjective evaluations of features might indeed be a facet that allows access to differences in information processing among NFC groups. There are several possible explanations for why individuals with high NFC seem to evaluate their systems more favorably in terms of persuasive features. Previous research has indicated that individuals with high NFC are less supportive of punitive measures [86], and the study suggests that this might be due to their tendency to form complex attributions for human behavior. This would be based on their innate motivation to engage in effortful thinking about the causes of human behavior in general. This could apply to the complexity of human behavior around physical activity. Perhaps individuals with high NFC do not expect simple and actionable insights to guide their behaviors, as individuals with low NFC do, but instead value data and raw information that they can use in their own thinking process. The extensive amount of information provided by these systems may also allow for greater task complexity, which individuals with high NFC prefer [32]. If self-monitoring systems feed their enjoyment in working out their own behavior with data, it might be intrinsically motivating and persuasive for them.

Alternatively, or in addition, individuals with high NFC might also have a significant amount of affinity for technology interaction [85]. The data that the systems provide smoothen their system evaluations, and individuals with high NFC accept more discrepancies or shortcomings from their systems. Individuals with high NFC might also form stronger engagement with the device concerned because of their interest in data: the more a monitoring system is used, the more data are collected for further elaboration. Positive evaluations of the system might also be consistent with a high amount of system use, but our data or findings do not address this issue because we did not measure system use frequency. To summarize our findings, higher levels of NFC were associated with higher system evaluations; for perceived credibility and persuasiveness, this trend is so strong that it separates the high-NFC group from both lower-NFC groups. This finding cannot be explained merely by differences among the systems because there was only one pair of systems (Sports Tracker vs Ōura) that differed from each other in terms of the NFC levels of the users.

Practical Implications

NFC has been studied widely for several decades, and its basic characteristics are well established. However, in the context of systems design, it is usually considered a trait for tailoring or personalization. Our results, based on systems that individuals with different levels of NFC had selected themselves, imply that targeting NFC might not be feasible for all types of systems. This is because the selection procedure might have excluded some systems from certain levels of NFC. For example, simple ones with basic functions might not be perceived as worth a further look by individuals with high NFC. Therefore, there is no need to tailor these systems for individuals with high NFC. This selection process might be influenced by brand image and communication, peers, peer reviews, and social media, which are not always controllable by those in charge of product development. However, practitioners might want to consider whether their marketing and product information communication fits with a desire to engage with problem solving and an interest in rich data. Similarly, products that seem to persuade users with high NFC might be equally useful and interesting to individuals with lower levels of NFC when accompanied by strong credibility cues and clearly designed self-monitoring features. For example, the very strong relationship between feedback and self-monitoring among individuals with low NFC implies that they might appreciate actionable and unambiguous support from the system and wish not to rely on their own cognitive work. The extensive communication of advanced product characteristics might not resonate well with their goals, and they may then choose other systems.

Our results also indicate that regardless of the NFC level, perceived credibility is important for the persuasiveness of these systems. It is, however, constituted slightly differently based on the level of NFC. For individuals with high NFC, its impact is transferred through self-monitoring activity and feedback features and might be more akin to a continuous evaluative process to validate the arguments that the system presents case by case. For individuals with low NFC, strong credibility might be more akin to an indication that the system is overall a good choice and should be used because its impact mainly targets overall persuasiveness instead of self-monitoring activity itself. Although our work does not allow elaboration likelihood or strength considerations, the results may be an indication of the use of source credibility as a cue for peripheral processing or as a variable whose validity is evaluated in daily activities through more effortful processing.

Limitations and Directions for Future Research

Our results should be interpreted against the background of some limitations. Our sample was collected in a university setting using only English. Although university students in Finland, including native Finnish students, are generally fluent in English, this might have biased our sample toward those with higher-than-average NFC because (high) NFC is positively correlated with verbal information processing [30]. We aimed to decrease this bias by including the upper limit value in the NFC groups that we formed with visual binning. This resulted in relatively larger groups of individuals with moderate and high NFC.

The self-selection of systems also has implications for the generalizability of our findings. Most likely, our sample comprised individuals who are interested in health technology and, overall, have higher-than-average satisfaction with their systems. Our findings are therefore relevant for natural use settings but less so for studies where users cannot choose their systems freely or use them fully voluntarily. Consequently, we consider that our work advances the field of self-help tools used outside of the health care sector and without moderation by an outside party. The levels of NFC in the general population in Finland and in other commercial system users would help to address the amount of self-selection bias in our study and guide future data collection. It is also noteworthy that we did not measure self-selection but assumed that individuals more often select and buy these devices themselves. In addition, our participants were considered healthy individuals because health status information was not collected. However, a limitation of our data is that these issues were not addressed in the questionnaire or recruitment. Future studies might focus on individuals with different chronic conditions to understand if NFC has a relationship with some health issues.

In self-monitoring systems, feedback features are often embedded in self-monitoring activity itself, for example, as a chart that is updated according to progress during the day. Therefore, it is possible that the self-monitoring construct we used also embeds some of the influence of feedback, although the constructs are independent. Although we phrased the questions in a way that would reflect more prominent feedback features, it is possible that the otherwise theoretically sound approach to defining feedback and self-monitoring as separate BCTs is not fully applicable to self-monitoring systems. Future research should develop new scales to measure theory-based characteristics of feedback and self-monitoring in relation to both behaviors and outcomes of behaviors to shed additional light on the relationship among these BCTs [2] in actual, implemented systems.

In addition, the use of several different types of physical activity self-monitoring systems might have caused additional heterogeneity in our results. We did not find statistically significant differences among the top 8 systems in terms of theory-driven system features, but this might be due to the low numbers of users in several systems. Although these types of commercial systems are currently rather similar in their feature sets [38] and we measured only the basic features, some trends might have been partly caused by the details of different systems. Future studies should seek to replicate our findings with single self-selected system users or compare different software implementations in behavior change features. Such studies could also go deeper into implementation details and analyze where the actual differences among NFC levels, and especially among genders in terms of NFC, arise. For example, previous studies have suggested that the differences may lie in visual perception [87], or they may be due to the differences in elaboration style or interaction with, and perception of, the features themselves [88,89].

Conclusions

Our study reported insights into a widely studied personality trait, the NFC, among users of commercial self-monitoring tools. In contrast to most of the research in the field of persuasion that focuses on this trait for attitude or behavior change, we used it to understand how individuals choose and use systems in a natural use environment. NFC does affect both the selection and use of systems, but the nature of the findings indicates that extensive aims to tailor content based on this trait might not always be a feasible approach. Instead, some features can also enable individuals with low NFC to benefit from these systems. However, the availability of different commercial systems might itself be a tool for tailoring, and individuals choose systems based on their innate characteristics.

Regardless of the apparent relevance of NFC to engaging with personally relevant data collection with self-monitoring tools, this paper is, to our knowledge, the first to report the role of NFC among self-monitoring users of wearables. Our data demonstrate that NFC as a trait that differentiates information processing has several implications for the selection, design, and tailoring of self-monitoring systems and their use in health interventions.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Survey instrument.
[PDF File (Adobe PDF File), 158 KB - formative_v5i10e23968_app1.pdf]

Multimedia Appendix 2
Short introductions of the eight most used self-monitoring systems of this study.
[PDF File (Adobe PDF File), 430 KB - formative_v5i10e23968_app2.pdf]

Multimedia Appendix 3
Cross-loadings for measurement items of the study.
[PDF File (Adobe PDF File), 126 KB - formative_v5i10e23968_app3.pdf]

References

1. Bandura A. The primacy of self-regulation in health promotion. Appl Psychol 2005 Apr;54(2):245-254. [doi: 10.1111/j.1464-0597.2005.00208.x]
2. Michie S, Ashford S, Sniehotta FF, Dombrowski SU, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
3. Jiang J, Cameron A. IT-enabled self-monitoring for chronic disease self-management: an interdisciplinary review. MIS Q 2020;44(1b):451-508. [doi: 10.25300/misq/2020/15108]
4. Turner-McGrievy GM, Beets MW, Moore JB, Kaczynski AT, Barr-Anderson DJ, Tate DF. Comparison of traditional versus mobile app self-monitoring of physical activity and dietary intake among overweight adults participating in an mHealth weight loss program. J Am Med Inform Assoc 2013 May 1;20(3):513-518 [FREE Full text] [doi: 10.1136/amiajnl-2012-001510] [Medline: 23429637]
5. Middelweerd A, Mollee JS, van der Wal CN, Brug J, te Velde SJ. Apps to promote physical activity among adults: a review and content analysis. Int J Behav Nutr Phys Act 2014 Jul 25;11:97 [FREE Full text] [doi: 10.1186/s12966-014-0097-9] [Medline: 25059981]
6. Zhao J, Freeman B, Li M. Can mobile phone apps influence people's health behavior change? An evidence review. J Med Internet Res 2016 Oct 31;18(11):e287 [FREE Full text] [doi: 10.2196/jmir.5692] [Medline: 27806926]
7. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
8. Lyons EJ, Lewis ZH, Mayrsohn BG, Rowland JL. Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis. J Med Internet Res 2014 Aug 15;16(8):e192 [FREE Full text] [doi: 10.2196/jmir.3469] [Medline: 25131661]
9. Mercer K, Li M, Giangregorio L, Burns C, Grindrod K. Behavior change techniques present in wearable activity trackers: a critical analysis. JMIR Mhealth Uhealth 2016 Apr 27;4(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4461] [Medline: 27122452]
10. Bandura A. Social cognitive theory of self-regulation. Organ Behav Hum Decis Process 1991 Dec;50(2):248-287. [doi: 10.1016/0749-5978(91)90022-L]
11. Carver CS, Scheier MF. Control theory: a useful conceptual framework for personality-social, clinical, and health psychology. Psychol Bull 1982 Jul;92(1):111-135. [doi: 10.1037/0033-2909.92.1.111] [Medline: 7134324]
12. Vinet L, Zhedanov A. A 'missing' family of classical orthogonal polynomials. J Phys A Math Theor 2011 Jan 31;44:085201 [FREE Full text] [doi: 10.1088/1751-8113/44/8/085201]
13. Michie S, Abraham C, Whittington C, McAteer J, Gupta S. Effective techniques in healthy eating and physical activity interventions: a meta-regression. Health Psychol 2009 Nov;28(6):690-701. [doi: 10.1037/a0016136] [Medline: 19916637]
14. Stiglbauer B, Weber S, Batinic B. Does your health really benefit from using a self-tracking device? Evidence from a longitudinal randomized control trial. Comput Human Behav 2019 May;94:131-139. [doi: 10.1016/j.chb.2019.01.018]
15. Huang G, Ren Y. Linking technological functions of fitness mobile apps with continuance usage among Chinese users: moderating role of exercise self-efficacy. Comput Human Behav 2020 Feb;103:151-160. [doi: 10.1016/j.chb.2019.09.013]
16. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009;24(1):-. [doi: 10.17705/1cais.02428]
17. Fogg B. Persuasive Technology : Using Computers to Change What We Think and Do. Amsterdam: Morgan Kaufmann Publishers; 2003.
18. Petty R, Cacioppo J. Communication and Persuasion: Central and Peripheral Routes to Attitude Change. In: Schlüsselwerke der Medienwirkungsforschung. Wiesbaden: Springer VS; 2016.
19. Petty RE, Cacioppo JT. The elaboration likelihood model of persuasion. Adv Exp Soc Psychol 1986;19:123-205. [doi: 10.1016/S0065-2601(08)60214-2]
20. Fiske ST, Taylor SE. Social Cognition: From Brains to Culture. Los Angeles: SAGE Publications; 2013.
21. Petty RE, Heesacker M, Hughes JN. The elaboration likelihood model: implications for the practice of school psychology. J Sch Psychol 1997;35(2):107-136. [doi: 10.1016/s0022-4405(97)00003-4]
22. Petty RE, Briñol P. A multiprocess approach to social influence. In: Kenrick DT, Goldstein SL, editors. Six Degrees of Social Influence: Science, Application, and the Psychology of Robert Cialdini. New York, NY: Oxford University Press; 2012.
23. Cacioppo JT, Petty RE, Kao CF, Rodriguez R. Central and peripheral routes to persuasion: an individual difference perspective. J Pers Soc Psychol 1986 Nov;51(5):1032-1043. [doi: 10.1037/0022-3514.51.5.1032]
24. Apsler R, Sears DO. Warning, personal involvement, and attitude change. J Pers Soc Psychol 1968 Jun;9(2):162-166. [doi: 10.1037/h0021248] [Medline: 5662739]
25. Kreuter MW, Wray RJ. Tailored and targeted health communication: strategies for enhancing information relevance. Am J Health Behav 2003;27 Suppl 3:S227-S232. [doi: 10.5993/ajhb.27.1.s3.6] [Medline: 14672383]

26. Rimer BK, Kreuter MW. Advancing tailored health communication: a persuasion and message effects perspective. J Communication 2006 Aug 04;56(s1):S184-S201. [doi: 10.1111/j.1460-2466.2006.00289.x]
27. Cortese J, Lustria ML. Can tailoring increase elaboration of health messages delivered via an adaptive educational site on adolescent sexual health and decision making? J Am Soc Inf Sci Tec 2012 Jun 21;63(8):1567-1580. [doi: 10.1002/asi.22700]
28. Lustria ML, Cortese J, Gerend MA, Schmitt K, Kung YM, McLaughlin C. A model of tailoring effects: a randomized controlled trial examining the mechanisms of tailoring in a web-based STD screening intervention. Health Psychol 2016 Nov;35(11):1214-1224. [doi: 10.1037/hea0000399] [Medline: 27441869]
29. Cacioppo JT, Petty RE. The need for cognition. J Pers Soc Psychol 1982 Jan;42(1):116-131. [doi: 10.1037/0022-3514.42.1.116]
30. Cacioppo JT, Petty RE, Feinstein JA, Jarvis WB. Dispositional differences in cognitive motivation: the life and times of individuals varying in need for cognition. Psychol Bull 1996;119(2):197-253. [doi: 10.1037/0033-2909.119.2.197]
31. Haugtvedt C, Petty R, Cacioppo J, Steidley T. Personality and ad effectiveness: exploring the utility of need for cognition. Adv Consum Res 1988;15:209-212.
32. See YH, Petty RE, Evans LM. The impact of perceived message complexity and need for cognition on information processing and attitudes. J Res Pers 2009 Oct;43(5):880-889. [doi: 10.1016/j.jrp.2009.04.006]
33. Haddock G, Maio GR, Arnold K, Huskinson T. Should persuasion be affective or cognitive? The moderating effects of need for affect and need for cognition. Pers Soc Psychol Bull 2008 Jun;34(6):769-778. [doi: 10.1177/0146167208314871] [Medline: 18344496]
34. Gajos K, Chauncey K. The influence of personality traits and cognitive load on the use of adaptive user interfaces. In: Proceedings of the 22nd International Conference on Intelligent User Interfaces. 2017 Presented at: 22nd International Conference on Intelligent User Interfaces; Mar 13-16, 2017; Limassol Cyprus. [doi: 10.1145/3025171.3025192]
35. Ho SY, Davern MJ, Tam KY. Personalization and choice behavior: the role of personality traits. Adv Databases Inf Syst 2008 Oct 31;39(4):31-47. [doi: 10.1145/1453794.1453800]
36. Schembre SM, Liao Y, Robertson MC, Dunton GF, Kerr J, Haffey ME, et al. Just-in-time feedback in diet and physical activity interventions: systematic review and practical design framework. J Med Internet Res 2018 Mar 22;20(3):e106 [FREE Full text] [doi: 10.2196/jmir.8701] [Medline: 29567638]
37. Gouveia R, Pereira F, Karapanos E, Munson S, Hassenzahl M. Exploring the design space of glanceable feedback for physical activity trackers. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2016 Presented at: UbiComp '16: 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing; Sep 12-16, 2016; Heidelberg Germany. [doi: 10.1145/2971648.2971754]
38. Halttu K, Oinas-Kukkonen H. Persuading to reflect: role of reflection and insight in persuasive systems design for physical health. Hum Comput Interact 2017 Jan 27;32(5-6):381-412. [doi: 10.1080/07370024.2017.1283227]
39. Lehto T, Oinas-Kukkonen H, Drozd F. Factors affecting perceived persuasiveness of a behavior change support system. In: Proceedings of the Thirty Third International Conference on Information Systems. 2012 Presented at: Thirty Third International Conference on Information Systems,; Dec 16-19, 2012; Orlando, FL, USA. [doi: 10.1007/978-3-642-31037-9_14]
40. Lehto T, Oinas-Kukkonen H, Pätiälä T, Saarelma O. Consumers' perceptions of a virtual health check: an empirical investigation. In: Proceedings of the 20th European Conference on Information Systems. 2012 Presented at: Proceedings of the 20th European Conference on Information Systems; 2012; Barcelona, Spain p. 29-40. [doi: 10.1007/978-3-642-32850-3_3]
41. Lehto T, Oinas-Kukkonen H. Explaining and predicting perceived effectiveness and use continuance intention of a behaviour change support system for weight loss. Behav Inf Technol 2015;34(2):176-189. [doi: 10.1080/0144929X.2013.866162]
42. Quinn JM, Pascoe A, Wood W, Neal DT. Can't control yourself? Monitor those bad habits. Pers Soc Psychol Bull 2010 Apr;36(4):499-511. [doi: 10.1177/0146167209360665] [Medline: 20363904]
43. Hermsen S, Frost J, Renes RJ, Kerkhof P. Using feedback through digital technology to disrupt and change habitual behavior: a critical review of current literature. Comput Human Behav 2016 Apr;57:61-74. [doi: 10.1016/j.chb.2015.12.023]
44. Nelson RO, Hayes SC. Theoretical explanations for reactivity in self-monitoring. Behav Modif 1981 Jan 01;5(1):3-14. [doi: 10.1177/014544558151001]
45. Rapp A, Cena F. Personal informatics for everyday life: how users without prior self-tracking experience engage with personal data. Int J Hum Comput Stud 2016 Oct;94:1-17. [doi: 10.1016/j.ijhcs.2016.05.006]
46. Korotitsch WJ, Nelson-Gray RO. An overview of self-monitoring research in assessment and treatment. Psychol Assess 1999;11(4):415-425. [doi: 10.1037/1040-3590.11.4.415]
47. Orji R, Vassileva J, Mandryk RL. Modeling the efficacy of persuasive strategies for different gamer types in serious games for health. User Model User Adap Inter 2014;24(5):453-498. [doi: 10.1007/s11257-014-9149-8]
48. Meitz TG, Ort A, Kalch A, Zipfel S, Zurstiege G. Source does matter: contextual effects on online media-embedded health campaigns against childhood obesity. Comput Human Behav 2016 Jul;60:565-574. [doi: 10.1016/j.chb.2016.02.067]
49. Tormala ZL, Petty RE. Source credibility and attitude certainty: a metacognitive analysis of resistance to persuasion. J Consum Psychol 2008 Jan 25;14(4):427-442. [doi: 10.1207/s15327663jcp1404_11]

50. Hovland C, Weiss W. The influence of source credibility on communication effectiveness. Public Opin Q 1951 Jan 01;15(4):635-650. [doi: 10.1086/266350]
51. Ohanian R. Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. J Advert 1990;19(3):39-52. [doi: 10.1080/00913367.1990.10673191]
52. Tormala ZL, Clarkson JJ. Source trustworthiness and information processing in multiple message situations: a contextual analysis. Soc Cogn 2008 Jun;26(3):357-367. [doi: 10.1521/soco.2008.26.3.357]
53. Pornpitakpan C. The persuasiveness of source credibility: a critical review of five decades' evidence. J Appl Social Pyschol 2004 Feb;34(2):243-281. [doi: 10.1111/j.1559-1816.2004.tb02547.x]
54. Everard A, Galletta DF. How presentation flaws affect perceived site quality, trust, and intention to purchase from an online store. J Manag Inf Syst 2014 Dec 09;22(3):56-95. [doi: 10.2753/MIS0742-1222220303]
55. Benbunan-Fich R. User satisfaction with wearables. AIS Trans Hum Comput Interac 2020 Mar 31;12(1):1-27. [doi: 10.17705/1thci.00126]
56. Wathen CN, Burkell J. Believe it or not: factors influencing credibility on the Web. J Am Soc Inf Sci 2002;53(2):134-144. [doi: 10.1002/asi.10016]
57. Shin D, Lee S, Hwang Y. How do credibility and utility play in the user experience of health informatics services? Comput Human Behav 2017 Feb;67:292-302. [doi: 10.1016/j.chb.2016.11.007]
58. The importance of persuasive systems design in enhancing consumers' perceptions and adoption of health behavior change support systems. University of Oulu. 2013. URL: http://jultika.oulu.fi/files/isbn9789526201450.pdf [accessed 2021-09-13]
59. Tormala ZL, Petty RE. Resistance to persuasion and attitude certainty: the moderating role of elaboration. Pers Soc Psychol Bull 2004 Nov;30(11):1446-1457. [doi: 10.1177/0146167204264251] [Medline: 15448308]
60. Hassan L, Dias A, Hamari J. How motivational feedback increases user's benefits and continued use: a study on gamification, quantified-self and social networking. Int J Inf Manag 2019 Jun;46:151-162. [doi: 10.1016/j.ijinfomgt.2018.12.004]
61. Aquino A, Alparone FR, Pagliaro S, Haddock G, Maio GR, Perrucci MG, et al. Sense or sensibility? The neuro-functional basis of the structural matching effect in persuasion. Cogn Affect Behav Neurosci 2020 Jun;20(3):536-550. [doi: 10.3758/s13415-020-00784-7] [Medline: 32185623]
62. Vaughan-Johnston TI, Jacobson JA. "Need" personality constructs and preferences for different types of self-relevant feedback. Pers Individ Dif 2020 Feb 01;154:109671. [doi: 10.1016/j.paid.2019.109671]
63. Leary M, Shepperd J, McNeil M, Jenkins T, Barnes B. Objectivism in information utilization: theory and measurement. J Pers Assess 1986;50(1):32-43. [doi: 10.1207/s15327752jpa5001_5] [Medline: 16367446]
64. Thompson EP, Chaiken S, Hazlewood JD. Need for cognition and desire for control as moderators of extrinsic reward effects: a person x situation approach to the study of intrinsic motivation. J Pers Soc Psychol 1993 Jun;64(6):987-999. [doi: 10.1037//0022-3514.64.6.987] [Medline: 8326474]
65. Cialdini R. Influence: Science and Practice 4th Edition. New York: Harper Collins; 2001.
66. Whittler TE, Manolis C. Toward understanding persuasion expressions: the activation of attitudes. Psychol Mark 2015 Jul 07;32(8):874-890. [doi: 10.1002/mar.20824]
67. Chiesi F, Morsanyi K, Donati MA, Primi C. Applying item response theory to develop a shortened version of the need for cognition scale. Adv Cogn Psychol 2018 Sep 30;14(3):75-86 [FREE Full text] [doi: 10.5709/acp-0240-z] [Medline: 32337000]
68. Hair JF, Risher JJ, Sarstedt M, Ringle CM. When to use and how to report the results of PLS-SEM. Eur Bus Rev 2019 Jan 14;31(1):2-24. [doi: 10.1108/ebr-11-2018-0203]
69. Hair JF, Ringle CM, Sarstedt M. PLS-SEM: indeed a silver bullet. J Mark Theory Pract 2014 Dec 08;19(2):139-152. [doi: 10.2753/mtp1069-6679190202]
70. Ringle C, Wende S, Will A. SmartPLS GmbH. Software Smart PLS. 2015. URL: https://www.smartpls.com/ [accessed 2021-09-13]
71. Nikoloudakis IA, Crutzen R, Rebar AL, Vandelanotte C, Quester P, Dry M, et al. Can you elaborate on that? Addressing participants' need for cognition in computer-tailored health behavior interventions. Health Psychol Rev 2018 Dec;12(4):437-452. [doi: 10.1080/17437199.2018.1525571] [Medline: 30221580]
72. Nunnally J, Bernstein I. Psychometric Theory 3rd Edition. New York, NY: McGraw-Hill; 1994.
73. Hair J, Hult G, Ringle C, Sarstedt M. A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM) 1st edition. Thousand Oaks: Sage; 2014.
74. Dickhäuser O, Reinhard M, Diener C, Bertrams A. How need for cognition affects the processing of achievement-related information. Learn Individ Differ 2009 Jun;19(2):283-287. [doi: 10.1016/j.lindif.2008.11.003]
75. Polydoropoulou A, Pagoni I, Tsirimpa A. Ready for mobility as a service? Insights from stakeholders and end-users. Travel Behav Soc 2020 Oct;21:295-306. [doi: 10.1016/j.tbs.2018.11.003]
76. Zhang Y, Buda R. Moderating effects of need for cognition on responses to positively versus negatively framed advertising messages. J Advert 1999;28(2):1-15. [doi: 10.1080/00913367.1999.10673580]
77. Suka M, Yamauchi T, Yanagisawa H. Persuasive messages can be more effective when repeated: a comparative survey assessing a message to seek help for depression among japanese adults. Patient Educ Couns 2020 Apr;103(4):811-818. [doi: 10.1016/j.pec.2019.11.008] [Medline: 31761527]

78. Cacioppo JT, Petty RE. Effects of message repetition and position on cognitive response, recall, and persuasion. J Pers Soc Psychol 1979;37(1):97-109. [doi: 10.1037/0022-3514.37.1.97]
79. Cacioppo JT, Petty RE. Effects of message repetition on argument processing, recall, and persuasion. Basic Appl Soc Psych 1989;10(1):3-12. [doi: 10.1207/s15324834basp1001_2]
80. Shi JJ, Smith SW. The effects of fear appeal message repetition on perceived threat, perceived efficacy, and behavioral intention in the extended parallel process model. Health Commun 2016;31(3):275-286. [doi: 10.1080/10410236.2014.948145] [Medline: 26305152]
81. Attig C, Karp A, Franke T. User diversity in the motivation for wearable activity tracking: a predictor for usage intensity? In: Proceedings of the 20th Congress of the International Ergonomics Association. 2018 Presented at: 20th Congress of the International Ergonomics Association; Aug 26-30, 2018; Florence, Italy. [doi: 10.1007/978-3-319-96077-7_46]
82. Beißert H, Köhler M, Rempel M, Beierlein C. Eine deutschsprachige kurzskala zur messung des konstrukts need for cognition [a german short-scale for measuring the construct need for cognition]. GESIS- Work Pap Pap 2014;32:-. [doi: 10.1037/t75189-000]
83. Williams-Piehota P, Pizarro J, Navarro Silvera SA, Mowad L, Salovey P. Need for cognition and message complexity in motivating fruit and vegetable intake among callers to the cancer information service. Health Commun 2006;19(1):75-84. [doi: 10.1207/s15327027hc1901_8] [Medline: 16519594]
84. Steward WT, Schneider TR, Pizarro J, Salovey P. Need for cognition moderates responses to framed smoking-cessation messages. J Appl Social Pyschol 2006 Jun 31;33(12):2439-2464. [doi: 10.1111/j.1559-1816.2003.tb02775.x]
85. Franke T, Attig C, Wessel D. A personal resource for technology interaction: development and validation of the affinity for technology interaction (ATI) scale. Int J Hum Comput Interact 2018 Mar 30;35(6):456-467. [doi: 10.1080/10447318.2018.1456150]
86. Sargent MJ. Less thought, more punishment: need for cognition predicts support for punitive responses to crime. Pers Soc Psychol Bull 2004 Nov;30(11):1485-1493. [doi: 10.1177/0146167204264481] [Medline: 15448311]
87. Fleischhauer M, Miller R, Enge S, Albrecht T. Need for cognition relates to low-level visual performance in a metacontrast masking paradigm. J Res Pers 2014 Feb;48:45-50. [doi: 10.1016/j.jrp.2013.09.007]
88. Millecamp M, Verbert K, Naveed S, Ziegler J. To explain or not to explain: the effects of personal characteristics when explaining music recommendations. In: Proceedings of the 24th International Conference on Intelligent User Interfaces. 2019 Presented at: 24th International Conference on Intelligent User Interfaces; Mar 17-20, 2019; Marina del Ray California. [doi: 10.1145/3301275.3302313]
89. Santos CP, Hutchinson K, Khan V, Markopoulos P. Profiling personality traits with games. ACM Trans Interact Intell Syst 2019 Apr;9(2-3):1-30. [doi: 10.1145/3230738]

Abbreviations

BCT: behavior change technique
ELM: elaboration likelihood model
NFC: need for cognition
PLS-SEM: partial least squares structural equation modeling
PSD: persuasive systems design
SEM: structural equation modeling

Edited by G Eysenbach; submitted 01.09.20; peer-reviewed by O Rivera, C Hoving; comments to author 18.11.20; revised version received 16.01.21; accepted 23.08.21; published 14.10.21.

Please cite as:
Halttu K, Oinas-Kukkonen H
Need for Cognition Among Users of Self-Monitoring Systems for Physical Activity: Survey Study
JMIR Form Res 2021;5(10):e23968
URL: https://formative.jmir.org/2021/10/e23968
doi:10.2196/23968
PMID:34647894

Original Paper

Acceptability, Engagement, and Effects of a Mobile Digital Intervention to Support Mental Health for Young Adults Transitioning to College: Pilot Randomized Controlled Trial

Brian Suffoletto[1], MS, MD; Tina Goldstein[2], PhD; Dawn Gotkiewicz[2], MD; Emily Gotkiewicz[3], BS; Brandie George[2], MA; David Brent[2], MD

[1]Department of Emergency Medicine, Stanford University, Palo Alto, CA, United States

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

[3]University of Delaware, Newark, DE, United States

Corresponding Author:
Brian Suffoletto, MS, MD
Department of Emergency Medicine
Stanford University
900 Welch Road, Suite 350
Palo Alto, CA, 94304
United States
Phone: 1 412 901 6892
Email: suffbp@stanford.edu

Abstract

Background: The transition from high school to college can exacerbate mental health problems in young adults yet barriers prevent seamless mental health care. Existing digital support tools show promise but are not yet designed to optimize engagement or implementation.

Objective: The goal of the research was to test acceptability and effects of an automated digital Mobile Support Tool for Mental Health (MoST-MH) for young adults transitioning to college.

Methods: Youths aged 18 years and older with a current mental health diagnosis preparing to transition to college (n=52; 85% female [45/52], 91% White [48/52]) were recruited from a primary care (n=31) and a mental health clinic (n=21). Participants were randomized 2:1 to either receive MoST-MH (n=34) or enhanced Usual Care (eUC; n=18). MoST-MH included periodic text message and web-based check-ins of emotional health, stressors, negative impacts, and self-efficacy that informed tailored self-care support messages. Both eUC and MoST-MH participants received links to a library of psychoeducational videos and were asked to complete web-based versions of the Mental Health Self-Efficacy Scale (MHSES), College Counseling Center Assessment of Psychological Symptoms (CCAPS), and Client Service Receipt Inventory for Mental Health (C-SRI) monthly for 3 months and the Post-Study System Usability Scale (PSSUQ) at 3-months.

Results: MoST-MH participants were sent a median of 5 (range 3 to 10) text message check-in prompts over the 3-month study period and 100% were completed; participants were sent a median of 2 (range 1 to 8) web-based check-in prompts among which 78% (43/55) were completed. PSSUQ scores indicate high usability (mean score 2.0). Results from the completer analysis demonstrated reductions in mental health symptoms over time and significant between-group effects of MoST-MH compared to eUC on depressive symptom severity (d=0.36, 95% CI 0.08 to 0.64). No significant differences in mental health self-efficacy or mental health health care use were observed.

Conclusions: In this pilot trial, we found preliminary evidence that MoST-MH was engaged with at high rates and found to be highly usable and reduced depression symptoms relative to eUC among youth with mental health disorders transitioning to college. Findings were measured during the COVID-19 pandemic, and the study was not powered to detect differences in outcomes between groups; therefore, further testing is needed.

Trial Registration: ClinicalTrials.gov NCT04560075; https://clinicaltrials.gov/ct2/show/NCT04560075

(JMIR Form Res 2021;5(10):e32271) doi:10.2196/32271

KEYWORDS

college; mental health; self-management; digital intervention; mHealth

Introduction

Mental health disorders are common among young adults. Internationally, 20% of college students meet criteria for a mental health diagnosis, and 83% of these individuals have onset prior to college matriculation [1]. Nationally, 31% of US college students report having a mental health diagnosis [2], and rates of mental health disorders in young adults have been increasing [3]. Mental health disorders can have significant negative impact on young adults including lower grade point averages and higher rates of college attrition [4]. Mental health disorders can also put a young adult at greater risk for physical health problems [5] and suicide [6].

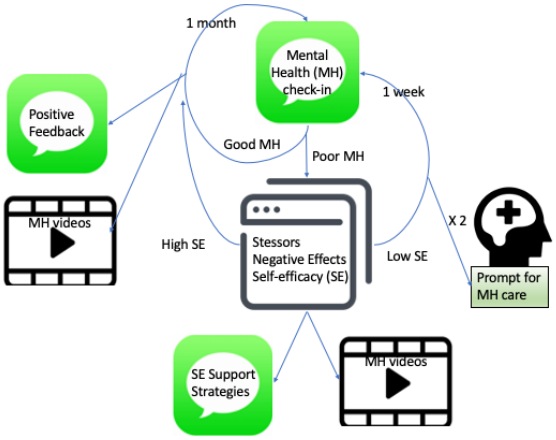
The transition from high school to college is a critical period where mental health support is lacking yet urgently needed. New stresses related to academics, finances, and relationships are heightened upon college initiation [7]. These stressors can worsen mental health symptoms and precipitate suicidal ideation [8]. Concurrently, the transition from pediatric to adult health services that frequently occurs at this juncture means that young adults often leave the providers with whom they have built longitudinal trusting relationships. Compounding this are stigma [9], scheduling difficulties [10], and lack of in-person resources to accommodate the mental health needs of all students [11,12]. For these reasons, novel approaches to supporting college students with mental health self-management are urgently needed.

Digital technology could assist young adults who are transitioning to college by providing rapid and efficient access to mental health support. In a systematic review of 19 studies (n=11 randomized controlled trials [RCTs]) of digital mental health interventions for college students, findings suggest that they can be effective for improving depression, anxiety, and psychological well-being [13]. Another recent systematic review focused on mobile mental health interventions for college students found that they reduce psychological symptomatology associated with stress, depression, anxiety, and general student mental health [14]. Despite these promising findings, most of these interventions addressed only a single mental health disorder. Also, many of these interventions incorporated some form of human support, making broad and cost-efficient implementation problematic. Finally, it is widely understood that, to date, longitudinal engagement with mental health digital interventions has been poor [15], limiting durability of effects.

To address these barriers, we developed an automated Mobile Support Tool for Mental Health (MoST-MH). MoST-MH was iteratively designed and refined by a multidisciplinary team with expertise in psychology, psychiatry, primary care, and digital interventions with integral feedback from a college student ambassador. MoST-MH was intended to provide support independent of mental health diagnosis type (ie, transdiagnostic) and designed to minimize the burden of intensive digital interactions using a stepwise algorithm which adapts frequency of interaction to the needs of the youth. MoST-MH is an ecological momentary intervention [16] in that it provides support in the context of a young adult's current state and needs. Specifically, MoST-MH incorporated periodic text message mental health check-ins, triggering web-based check-ins (when mental health was rated low) to understand stressors, negative effects, and self-efficacy, which informed self-efficacy support strategies and prompted links to psychoeducational videos focused on college and mental health. Figure 1 outlines the design of MoST-MH.

In this paper, we present pilot randomized trial findings of MoST-MH where we examined intervention engagement, acceptability, and estimates of effects of MoST-MH compared to enhanced usual care (eUC). We hypothesized that youth would engage with MoST-MH at high rates over the first 3 months of college and that they would report high levels of usability. We also hypothesized that youth who received MoST-MH, as compared with youth who receive eUC, would report greater mental health self-efficacy, lower symptom severity, and higher rates of follow-through with mental health care at 3 months.

Figure 1. Mobile Support Tool for Mental Health design diagram.



Methods

Study Design

We conducted a pilot randomized trial among youth with a current mental health disorder or recent mental health care preparing to transition to college. Design and a priori hypotheses were registered at ClinicalTrials.gov (NCT04560075). As this was a pilot study, we were not powered to detect significant differences in mental health outcomes between groups. All participants completed written informed consent. Study investigators and outcome assessors were blinded to allocation to treatment arms. All procedures were approved by the institutional review board at the University of Pittsburgh.

Participants

Participants were recruited from one primary care (n=31) and one mental health clinic (n=21) in Pittsburgh, PA, from August to October 2020. We chose to recruit from health care sites because we view the ultimate implementation to be initiated by care providers who are able to identify individuals with mental health needs prior to leaving for college. The youth's care providers identified potentially eligible youth and asked the youth about interest in participating in the study; interested youth were texted or emailed a web link that provided information about the study. If they were interested, they contacted the research team via telephone, where enrollment criteria were confirmed. Inclusion criteria included age 18 years or older, current mental health diagnosis documented in their electronic medical record or received mental health services within 3 months per self-, parent-, or clinician-report, graduated high school, plan to attend college or higher education within 6 weeks, and own a personal mobile phone with text messaging. We excluded non-English–speaking individuals given that intervention materials were in English only.

Randomization

We used block randomization whereby two-thirds of participants were randomly assigned to receive MoST-MH and one-third to receive eUC. Blocks balanced the groups based on recruitment site. Random assignment allocation occurred following completion of baseline assessments.

MoST-MH Intervention

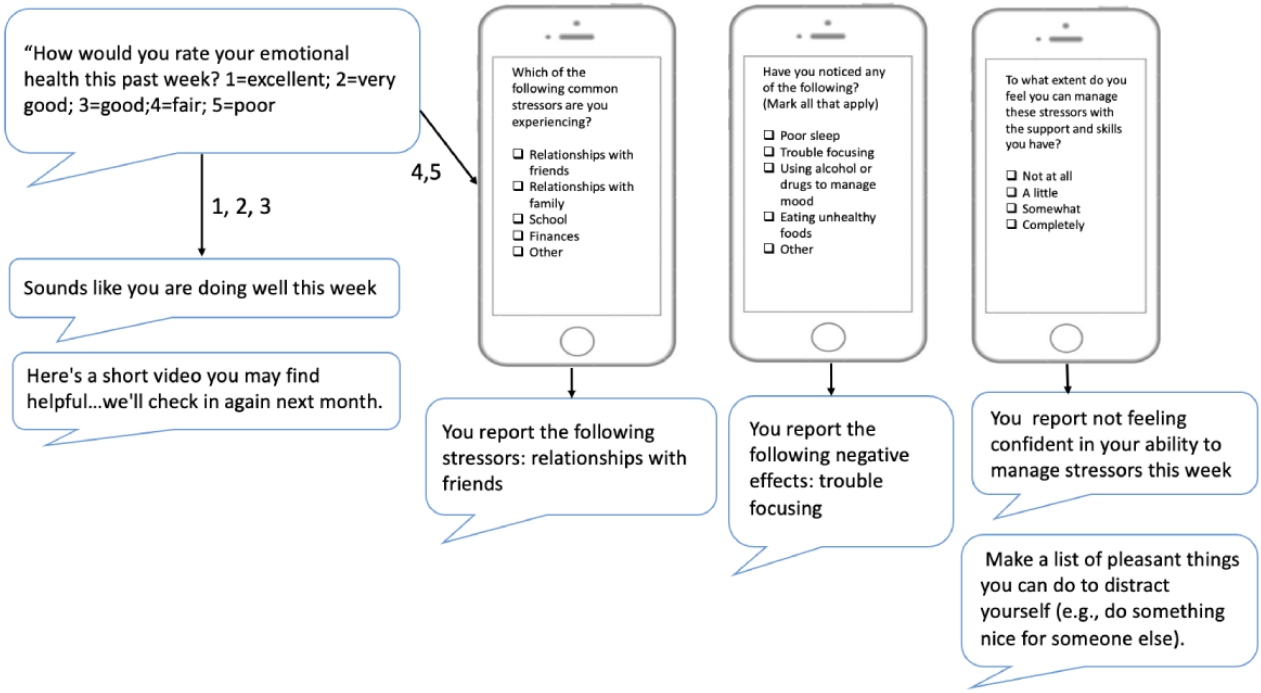
MoST-MH aimed to enhance an individual's mental health by raising awareness of current symptoms, stressors, and impact of stressors and boosting self-efficacy by prompting evidence-based strategies taken from positive psychology, cognitive behavioral therapy, and dialectical behavioral therapy. MoST-MH used text messaging as the primary communication modality given its ubiquity and preferential role for communication among youth [17] as well as its proven effectiveness to deliver other forms of health support [18]. MoST-MH incorporated web-based check-ins when mental health was rated as suboptimal to collect more detailed information to help guide tailoring of support and because lengthy checklists would have been cumbersome using text messaging. The MoST-MH intervention software was run by the Office of Academic Computing at the University of Pittsburgh Medical Center.

Upon allocation, MoST-MH participants were prompted to text a unique keyword to a study phone number to initiate the program. Once program was initiated, participants received a series of welcome messages describing what to expect over the intervention period and ways to reduce breach of privacy (eg, "Welcome to MoST-MH. Over the next 3 months we'll be checking in by text message. Set up a password on your phone and erase messages you do not want anyone to see after reading them"). Participants were instructed that they could drop out of the MoST-MH program at any time by texting Quit.

Starting on the day of enrollment, MoST-MH participants received mental health check-ins via text message: "How would you rate your emotional health this past week?" If they replied excellent, very good or good, they received a positive feedback text message and link to video library. The brief 2-minute videos were created by the study team and included psychoeducation about mental health self-care during college. If they replied fair or poor, they were sent a link to complete a web-based check-in. When the web link opened, a page displayed a checklist of common stressors [7] and negative effects. The student was asked a self-efficacy question: "To what extent do you feel you can manage your stressors and negative effects with supports and skills you have?" If they reported high self-efficacy (completely), they received positive feedback and a web link to a library of mental health videos and the program was timed to check in with them in a month.

If they reported low self-efficacy (somewhat, a little, or not at all), they received a text message from a skills library and the link to the videos and were asked if it was ok to check in next week. On subsequent MoST-MH check-ins, their reports of stressors and negative effects were compared to the prior assessment and feedback incorporated relative improvement or unresolved stressors/effects. If the ability to self-manage stressors or negative effects was still reported as suboptimal, the individual was prompted to consider making an appointment for seeking mental health care: "Your doctor or another health professional can help. Would you be willing to reach out to them to set up an appointment?" If they were willing, they were provided with a link to resources to assist. Throughout all program queries, missing responses were reprompted once only. To ensure safety, if an individual reported poor mental health and low self-efficacy 2 weeks in a row, they were prompted to seek formal mental health care. Figure 2 demonstrates a sample communication exchange.

Figure 2. Mobile Support Tool for Mental Health example exchange.



Enhanced Usual Care

The eUC participants received a web link to a library of the same psychoeducational videos provided to the MoST-MH group. eUC received no text message or web-based mental health check-ins.

Measures

MoST-MH engagement was assessed using text messaging and web-based responses. Mental health symptom severity and health care use were assessed using self-report measures collected monthly for 3 months. Each monthly assessment battery was estimated to take 15 minutes to complete and were completed on a smartphone, laptop, tablet, or desktop. Participants in both groups were sent text message reminders every 3 days up to 3 times prompting them to complete their web-based follow-up assessment batteries. Participants were eligible to receive a total of $100 for participation in the study, including $20 for completing each monthly assessment battery. Participants were not compensated for completing text messages nor web-based assessments.

MoST-MH usability was measured via the Post-Study System Usability Scale (PSSUQ) [19] at 3 months only. The PSSUQ includes 19 items, each rated on a 7-point Likert-type scale ranging from 1 (strongly agree) to 7 (strongly disagree). The psychometric factors of the PSSUQ are overall usability, system usefulness, information quality, and interface quality. The lower the score (to a limit of one), the higher the perceived usability.

Mental health self-efficacy was measured using the 6-item self-report Mental Health Self-Efficacy Scale (MHSES) [20], which asks participants to rate each statement on a 10-point Likert scale ranging from 1 (not at all confident) to 10 (totally confident) whereby higher scores indicate higher self-efficacy: "On an average day in the next month, how confident are you that… (1) You can keep your stress, anxiety, or depression from interfering with the things that you want to do? (2) You can do the different tasks and activities needed to manage your stress, anxiety, or depression so as to reduce your need to see a doctor? (3) You can do things other than just taking medicine to reduce how much your stress, anxiety, or depression affects your everyday life? (4) You can make your days at least moderately enjoyable? (5) You will have moderate amounts of time where you do not experience stress, anxiety, or depression? (6) You will be able to effectively manage any stress, anxiety, or depression that you do experience?"

Symptom severity was measured using the College Counseling Center Assessment of Psychological Symptoms (CCAPS) [21] which has 62 items with 8 distinct subscales of psychological symptoms for college students: (a) depression (13 items), (b) generalized anxiety (9 items), (c) social anxiety (7 items), (d) academic distress (5 items), (e) eating concerns (9 items), (f) family distress (6 items), (g) hostility (7 items), and (h) substance use (6 items). Items are scored on a 5-point Likert scale from 0 (not at all like me) to 4 (extremely like me), whereby higher scores indicate higher symptom severity.

Mental health treatment use was measured using the brief self-report Client Service Receipt Inventory for Mental Health (C-SRI) [22] including outpatient, inpatient, and medication management services.

Data Analysis

To determine whether any significant differences between groups existed at baseline, independent *t* tests were conducted on continuous baseline variables (ie, age, MHSES, CCAPS), and chi-square analyses were conducted on categorical or nominal variables (ie, gender, race, ethnicity, college plans and living situation, site of recruitment, C-SRI mental health care). We tested the hypothesis that youth would engage with MoST-MH at high rates (>80% response rate) by calculating text message and web check-in completions within and between

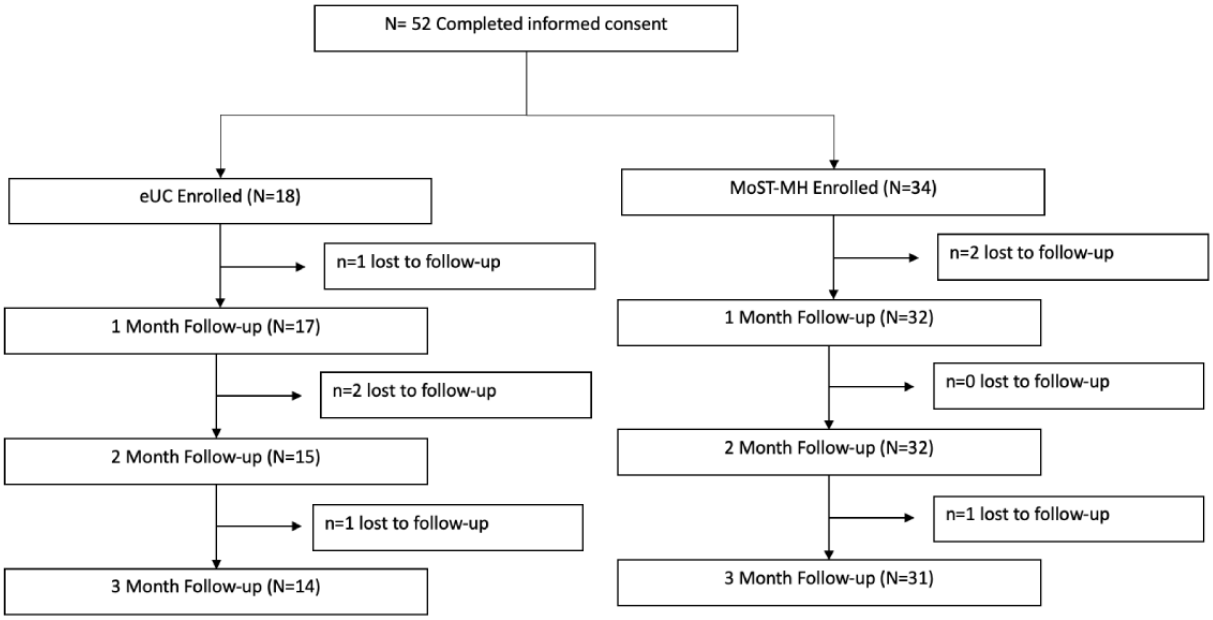
individuals. We tested the hypothesis that youth would report high levels of usability with MoST-MH (mean PSSUQ rating ≤2) by computing PSSUQ ratings at 3-month follow-up. We explored the effect of MoST-MH as compared with eUC on mental health self-efficacy (MHSES), symptom severity (CCAPS), and mental health care services use (C-SRI) using mixed effect (general estimating equation [GEE]) models. Mixed effects models using GEE are recommended for analysis of repeated-measures data and can properly account for missing data [23]. To understand for whom the intervention may work better or worse, we explored associations between patient factors (sex, race, planned college attendance, baseline CCAPS scores) and engagement, usability, and mental health outcomes using univariate GEE models. Primary analyses were conducted using listwise deletion. For sensitivity analyses, we conducted intention-to-treat analyses using multiple imputation procedures where missing CCAPS outcome data were assumed to be missing at random. A simulated dataset with 20 imputed outcomes was generated. All analyses were conducted using Stata (version 15.0, StataCorp LLC).

Results

Overview

Figure 3 shows the participant flow throughout the study. A total of 98 youths were referred to the study, 73 were reached for screening, and 52 completed informed consent. The resultant sample was randomized via computer algorithm to receive either instructions to initiate MoST-MH (n=34) or eUC (n=18) after completion of online questionnaires at baseline. A total of 49/52 (94%) of participants completed 1-month follow-up assessment batteries, 47/52 (90%) completed 2-month follow-up assessment batteries, and 45/52 (87%) completed 3-month follow-up assessment batteries. There were no differences in attrition by sex, race, college living plans, or treatment arm.

Figure 3. Consolidated Standards of Reporting Trials diagram. eUC: enhanced usual care; MoST-MH: Mobile Support Tool for Mental Health.



Participants

Table 1 shows the demographics of the sample of MoST-MH and eUC participants. Most identified as female (45/52, 85%) and White (48/52, 91%). The majority of the sample (45/52, 84%) was planning on attending a 4-year college, although with a wide variety of plans for college living situation. There was a higher percentage of White youths and individuals planning on living on campus in the MoST-MH arm compared to eUC. For that reason, we included race and planned college living situation (in addition to sex) as covariates in all models.

Table 1. Participant characteristics.

Characteristic	eUC[a] (n=18)	MoST-MH[b] (n=34)	*P* value
Age (years), mean (SD)	18.7 (0.48)	18.7 (0.42)	.47
Sex, female, n (%)	18 (100)	27 (79)	.08
Race, n (%)			.04
White	15 (83)	33 (97)	—[c]
Black	0	1 (3)	—
More than one	3 (17)	0	—
Hispanic, n (%)	1 (6)	2 (6)	.53
College plans, n (%)			.48
4 year	15 (83)	30 (88)	—
Community	2 (11)	1 (3)	—
Professional/trade	1 (6)	1 (3)	—
Other	0	2 (6)	—
Plans for living, n (%)			.02
On campus, dorms (roommate)	6 (33)	25 (74)	—
On campus, dorms (by self)	3 (17)	2 (6)	—
Off campus	3 (17)	1 (3)	—
At parents	6 (33)	6 (18)	—
Site of recruitment, n (%)			.66
Primary care	10 (56)	21 (62)	—
Mental Health clinic	8 (44)	13 (38)	—

[a]eUC: enhanced usual care.

[b]MoST-MH: Mobile Support Tool for Mental Health.

[c]Not applicable.

MoST-MH Engagement

Participants were sent a median of 5 (range 3-10) text message check-ins over the study period (depending on their risk level), and 100% of the text message queries were completed. The 21 participants who reported poor MH via text message at least once over the study period received a median of 2 (range 1 to 8) web-based check-ins. Of the 55 times when a web check-in was prompted, 43 (78%) were completed. We did not find that sex, race, or college living plans were significantly associated with web check-ins. No MoST-MH participants dropped out (ie, texted Quit).

The median number of stressors reported per check-in was 2 (range 0 to 5); the median number of negative effects reported per check-in was 3 (range 0 to 8). There were higher mean stressors reported in female compared to male participants (beta=0.49, 95% CI 0.21to 0.76). The most common stressors were related to school and finances; the most common negative effects were feeling worn out and low motivation. Table 2 shows the percentage of check-ins with a given stressor and negative effects reported. Self-management self-efficacy was rated as high 9% (4/43) of the time and low 91% (39/43) of the time. We did not find that patient factors (ie, sex, race, college plans) were significantly associated with self-efficacy. Higher baseline anxiety scores were associated with lower self-efficacy (beta=–0.32, 95% CI –0.55 to –0.08).

Table 2. Frequency of reported stressors and negative effects for Mobile Support Tool for Mental Health participants (n=34).

Characteristic	Value, n (%)
Stressors	
School	22 (51)
Finances	16 (37)
Relationships with friends/roommates	12 (28)
Relationships with family	11 (26)
Romantic relationships	9 (21)
Other	3 (7)
Negative effects	
Feeling worn out	28 (65)
Low motivation	24 (56)
Trouble focusing on schoolwork	22 (51)
Poor sleep	20 (47)
Unhealthy eating habits	18 (42)
Feeling like you are overreacting	14 (33)
Feeling like you don't have people to talk to	13 (30)
Using alcohol or drugs to manage emotions	5 (12)
Other	1 (2)

Intervention Usability

Using the PSSUQ completed at the 3-month follow-up, overall mean usability score was 2.0 (SD 1.6). For subscale rating, mean score for system usefulness was 1.9 (SD 1.7), information quality was 2.2 (SD 1.5), and interface quality was 1.9 (SD 1.7). We did not find that patient factors (ie, sex, race, college plans) or baseline mental health scores were significantly associated with usability.

Mental Health Self-Efficacy

Using the MHSES, there were no significant improvements in either treatment arm over time for mental health self-efficacy as measured. In GEE analysis, there were no significant time by treatment effects of MoST-MH compared to eUC. We did not find that patient factors (ie, sex, race, college plans) or baseline mental health scores were significantly associated with self-efficacy.

Mental Health Symptom Severity

Using the CCAPS, in the MoST-MH arm, mental health symptom severity was reduced from baseline to 3 months in all subscales save substance abuse. In eUCs, reduced symptoms over 3 months occurred for general anxiety, family distress, and hostility only. Table 3 and Figure 4 show the mean scores on CCAPS subscales across treatment and time. In GEE analysis, there was a significant time effect such that at 1 month, depression scores were lower than baseline (beta=–.28, 95% CI –0.53 to –0.04) and time × treatment effect such that MoST-MH had lower depression scores relative to eUC by 3 months (beta=–0.34, 95% CI –0.67 to –0.03). In sensitivity analyses with imputed CCAPS outcome data, no significant effects of treatment were seen.

Table 3. Mental health symptoms over time by treatment.

CCAPS[a] subscale	BL[b] (eUC[c] n=18, MoST-MH[d] n=34), mean (SD)	1 m (eUC n=17, MoST-MH n=32), mean (SD)	2 m (eUC n=15, MoST-MH n=32), mean (SD)	3 m (eUC n=12, MoST-MH n=27), mean (SD)
Depression				
eUC	1.81 (0.99)	1.57 (1.00)	1.85 (1.02)	1.83 (1.11)
MoST-MH	1.43 (0.95)	1.36 (0.96)	1.24 (0.94)	1.12 (0.94)
General anxiety				
eUC	2.22 (0.95)	2.02 (1.07)	2.24 (0.99)	1.96 (0.93)
MoST-MH	1.82 (1.05)	1.82 (1.05)	1.70 (1.15)	1.42 (0.99)
Social anxiety				
eUC	2.37 (0.93)	2.15 (0.95)	2.39 (1.02)	2.39 (0.90)
MoST-MH	2.11 (0.93)	2.06 (1.07)	1.92 (1.07)	1.81 (1.03)
Academic distress				
eUC	1.90 (0.88)	1.76 (0.96)	2.13 (0.90)	2.12 (1.11)
MoST-MH	1.42 (1.02)	1.45 (1.06)	1.39 (1.06)	1.31 (0.98)
Eating concerns				
eUC	1.51 (1.06)	1.61 (1.11)	1.72 (1.13)	1.55 (0.97)
MoST-MH	1.37 (0.89)	1.22 (1.03)	1.21 (0.97)	1.01 (0.86)
Family distress				
eUC	1.58 (0.81)	1.41 (1.00)	1.13 (0.83)	1.10 (0.71)
MoST-MH	1.28 (0.85)	1.13 (0.96)	1.18 (0.82)	0.93 (0.63)
Hostility				
eUC	1.38 (0.61)	1.26 (0.81)	1.39 (0.86)	1.1 (0.73)
MoST-MH	1.00 (0.83)	0.93 (0.80)	0.75 (0.82)	0.82 (0.89)
Substance abuse				
eUC	0.33 (0.72)	0.32 (0.54)	0.21 (0.38)	0.39 (0.76)
MoST-MH	0.33 (0.60)	0.43 (0.72)	0.38 (0.60)	0.40 (0.58)

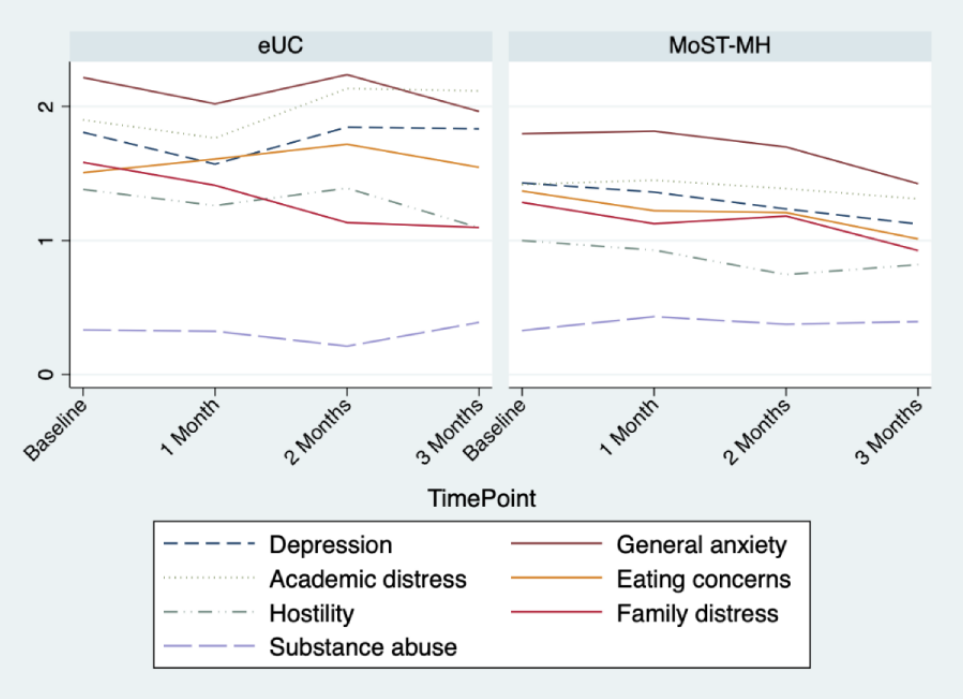
[a]CCAPS: College Counseling Center Assessment of Psychological Symptoms.

[b]BL: baseline.

[c]eUC: enhanced usual care.

[d]MoST-MH: Mobile Support Tool for Mental Health.

Figure 4. Change over time in College Counseling Center Assessment of Psychological Symptoms subscales. eUC: enhanced usual care; MoST-MH: Mobile Support Tool for Mental Health.



Mental Health Care Services

In the month prior to enrollment, 56% of MoST-MH and 67% of eUC participants had received any mental health care. Over the first 3 months of enrollment, 52% of MoST-MH and 65% of eUC participants received any mental health care. Table 4 shows the percentage of participants who received inpatient, outpatient, and primary care for mental health in the prior month at each assessment point. In GEE analysis, there were no significant time, treatment, or time by treatment effect of MoST-MH compared to eUC.

Table 4. Mental health care over time by treatment.

Type of care	BL[a] (eUC[b] n=18, MoST-MH[c] n=34), mean (SD)	M1 (eUC n=17, MoST-MH n=32), mean (SD)	M2 (eUC n=15, MoST-MH n=32), mean (SD)	M3 (eUC n=13, MoST-MH n=31), mean (SD)
Any mental health care				
eUC	12 (67)	11 (65)	11 (73)	7 (54)
MoST-MH	19 (56)	17 (53)	16 (50)	15 (48)
Inpatient				
eUC	2 (11)	0	1 (7)	0
MoST-MH	2 (6)	0	0	0
Outpatient				
eUC	9 (50)	11 (65)	10 (67)	7 (54)
MoST-MH	13 (38)	15 (47)	12 (38)	12 (39)
Primary care				
eUC	4 (22)	1 (6)	3 (20)	1 (8)
MoST-MH	9 (26)	4 (13)	6 (19)	5 (16)

[a]BL: baseline.

[b]eUC: enhanced usual care.

[c]MoST-MH: Mobile Support Tool for Mental Health.

Discussion

Principal Findings

In this pilot trial, we found evidence that, for young adults with mental health diagnoses transitioning to college, the MoST-MH intervention was engaged with at high rates, had high usability ratings, and may have reduced depression symptoms over time relative to the eUC group. Together, these findings provide initial support for an automated digital intervention incorporating periodic text message mental health check-ins that trigger web-based check-ins to understand stressors, negative effects, and self-efficacy and then provide self-efficacy support strategies.

Compared to other mental health digital interventions, we found good engagement:100% of text message check-ins and 78% of web-based check-ins were completed over 3 months, and no participant dropped out (ie, texted Quit). In contrast, among 28 digital mental health interventions for college students included in a prior systematic review [13], the average engagement rate was 56%. Complementing our finding high MoST-MH engagement, we found high usability ratings, suggesting the ease of use of text messages and web-based interfaces as well as the brief clear nature of support messaging.

We speculate that the high engagement and usability findings for MoST-MH can be attributed to several key design features. First, the study was introduced by a care provider, increasing trust in the intervention. Second, we designed MoST-MH with key input from a college student ambassador. Third, using text messaging to conduct most communication limited the amount of extra steps traditionally involved with apps or web pages, thus increasing ease of interaction. Fourth, the MoST-MH intervention adapted over time so that if an individual was not in need of support, the check-ins were stepped down (ie, occurred monthly), reducing the burden on individuals who seemed to not need help at that time.

The finding of reduced mental health symptoms in MoST-MH compared with eUC suggests that the intervention may successfully support self-management of mental health symptoms and that mood regulation may be a key mechanism. This is encouraging given the simplicity of the intervention and lack of any intensive cognitive behavioral treatment components. However, despite an intended aim of MoST-MH to enhance an individual's self-efficacy, we did not find evidence to support this as a mechanism of action. Future studies of MoST-MH will be necessary to identify mechanisms of effects.

The use of mental health care decreased slightly in both treatment arms over time, yet it was still occurring in approximately half of all participants any given month. Also, there were no apparent signals of either increased or decreased mental health care use in the MoST-MH arm compared to eUC. On one hand, this is discouraging given that the MoST-MH intervention prompted many participants to reach out for mental health care when they reported low self-efficacy 2 weeks in a row. On the other hand, it may be that the MoST-MH intervention promoted confidence in self-management not reflected in the self-efficacy scales. Future studies are needed to identify how MoST-MH may modify cognitions around mental health care seeking.

Limitations and Strengths

Several limitations should be discussed. First, we recruited mostly female White youths, therefore findings may not be valid in men or racial or ethnic minorities. Second, we did not follow participants for more than 3 months, limiting understanding of durability and prolonged engagement. Third, our cohort was recruited entirely during the COVID-19 pandemic. As such, the experience and stresses related to the transition to college and the college experience were atypical. Fourth, our sample is transdiagnostic, and we did not assess for specific mental health diagnoses or collect measures on type and other current treatments being received for mental health diagnoses, which may have been key moderators of effectiveness.

We note several strengths of our study and intervention design. First, we recruited individuals with any mental health diagnosis, thus rendering our intervention and findings broadly relevant to young adults transitioning to college (transdiagnostic). Second, we compared the MoST-MH intervention to eUC instead of a waitlist control, isolating effects of the digital interactions from those of attention and the psychoeducational videos. Third, we achieved high follow-up rates (86% at 3 months), reducing the likelihood of biased outcome analyses. Fourth, we measured and reported engagement through detailed analysis of reports through both text messages and web check-ins, notably absent in much prior digital MH intervention research [13]. Fifth, MoST-MH was completely automated, allowing low-cost scalability. Although human interaction has been identified as an important component of digital mental health interventions [24,25] and the majority of text message interventions for adolescent mental health and substance abuse involve some human communication [26], it is not feasible in the current reimbursement landscape to expect mental health care to fund these personnel. Digital intervention science should focus on identifying features to optimize human-computer interaction.

Conclusions

We found preliminary evidence in support of an automated digital mental health intervention using periodic check-ins to tailor self-management support for youth with mental health disorders transitioning to college. This study is timely, as there is an urgent need for evidence-based mental health support programs for youth in transition to college that are used longitudinally and can be scaled easily. A program like MOST-MH, if found to be effective at reducing mental health symptoms and improving psychological functioning in a larger trial, could fill a needed gap in supporting youth in their mental health.

Acknowledgments

We would like to thank Jack Doman and the Office of Academic Computing at Western Psychiatric Hospital and Clinic for all their hard work and creativity in building the text message program.

Authors' Contributions

This work was supported by grant 1P50MH115838 from the National Institutes of Health (principal investigator: DB).

Conflicts of Interest

TG received research funding from the National Institute of Mental Health, American Foundation for Suicide Prevention, and University Of Pittsburgh Clinical and Translational Science Institute and royalties from Guilford Press.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 97 KB - formative_v5i10e32271_app1.pdf]

References

1. Auerbach RP, Alonso J, Axinn WG, Cuijpers P, Ebert DD, Green JG, et al. Mental disorders among college students in the World Health Organization World Mental Health Surveys. Psychol Med 2016 Oct;46(14):2955-2970 [FREE Full text] [doi: 10.1017/S0033291716001665] [Medline: 27484622]
2. National college health assessment: executive summary report. Silver Spring: American College Health Association; 2019. URL: https://www.acha.org/documents/ncha/NCHA-II_Fall_2018_Undergraduate_Reference_Group_Executive_Summary.pdf [accessed 2021-02-03]
3. Duffy ME, Twenge JM, Joiner TE. Trends in mood and anxiety symptoms and suicide-related outcomes among U.S. undergraduates, 2007-2018: evidence from two national surveys. J Adolesc Health 2019 Nov;65(5):590-598. [doi: 10.1016/j.jadohealth.2019.04.033] [Medline: 31279724]
4. Keyes CLM, Eisenberg D, Perry GS, Dube SR, Kroenke K, Dhingra SS. The relationship of level of positive mental health with current mental disorders in predicting suicidal behavior and academic impairment in college students. J Am Coll Health 2012 Feb;60(2):126-133. [doi: 10.1080/07448481.2011.608393] [Medline: 22316409]
5. Goodwin GM. Depression and associated physical diseases and symptoms. Dialogues Clin Neurosci 2006;8(2):259-265 [FREE Full text] [Medline: 16889110]
6. Liu CH, Stevens C, Wong SH, Yasui M, Chen JA. The prevalence and predictors of mental health diagnoses and suicide among U.S. college students: implications for addressing disparities in service use. Depress Anxiety 2019 Jan 06;36(1):8-17 [FREE Full text] [doi: 10.1002/da.22830] [Medline: 30188598]
7. Karyotaki E, Cuijpers P, Albor Y, Alonso J, Auerbach RP, Bantjes J, et al. Sources of stress and their associations with mental disorders among college students: results of the World Health Organization World Mental Health Surveys International College Student Initiative. Front Psychol 2020 Jul 30;11:1759 [FREE Full text] [doi: 10.3389/fpsyg.2020.01759] [Medline: 32849042]
8. Stewart JG, Shields GS, Esposito EC, Cosby EA, Allen NB, Slavich GM, et al. Life stress and suicide in adolescents. J Abnorm Child Psychol 2019 Oct 26;47(10):1707-1722 [FREE Full text] [doi: 10.1007/s10802-019-00534-5] [Medline: 31028559]
9. Eisenberg D, Downs MF, Golberstein E, Zivin K. Stigma and help seeking for mental health among college students. Med Care Res Rev 2009 Oct 19;66(5):522-541. [doi: 10.1177/1077558709335173] [Medline: 19454625]
10. Xiao H, Carney DM, Youn SJ, Janis RA, Castonguay LG, Hayes JA, et al. Are we in crisis? National mental health and treatment trends in college counseling centers. Psychol Serv 2017 Nov;14(4):407-415. [doi: 10.1037/ser0000130] [Medline: 29120199]
11. Salaheddin K, Mason B. Identifying barriers to mental health help-seeking among young adults in the UK: a cross-sectional survey. Br J Gen Pract 2016 Sep 29;66(651):e686-e692 [FREE Full text] [doi: 10.3399/bjgp16x687313]
12. Herman S, Archambeau OG, Deliramich AN, Kim BSK, Chiu PH, Frueh BC. Depressive symptoms and mental health treatment in an ethnoracially diverse college student sample. J Am Coll Health 2011 Nov;59(8):715-720 [FREE Full text] [doi: 10.1080/07448481.2010.529625] [Medline: 21950252]
13. Lattie EG, Adkins EC, Winquist N, Stiles-Shields C, Wafford QE, Graham AK. Digital mental health interventions for depression, anxiety, and enhancement of psychological well-being among college students: systematic review. J Med Internet Res 2019 Jul 22;21(7):e12869 [FREE Full text] [doi: 10.2196/12869] [Medline: 31333198]
14. Oliveira C, Pereira A, Vagos P, Nóbrega C, Gonçalves J, Afonso B. Effectiveness of mobile app-based psychological interventions for college students: a systematic review of the literature. Front Psychol 2021 May 11;12:647606 [FREE Full text] [doi: 10.3389/fpsyg.2021.647606] [Medline: 34045994]

15. Torous J, Nicholas J, Larsen ME, Firth J, Christensen H. Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements. Evid Based Ment Health 2018 Aug 05;21(3):116-119. [doi: 10.1136/eb-2018-102891] [Medline: 29871870]
16. Heron KE, Smyth JM. Ecological momentary interventions: incorporating mobile technology into psychosocial and health behaviour treatments. Br J Health Psychol 2010 Feb;15(Pt 1):1-39 [FREE Full text] [doi: 10.1348/135910709X466063] [Medline: 19646331]
17. Lenhart A. Teens, smartphones and texting. Washington: Pew Internet and American Life Project; 2012. URL: https://www-pewinternet-org.laneproxy.stanford.edu/wp-content/uploads/sites/9/media/Files/Reports/2012/PIP_Teens_Smartphones_and_Texting.pdf [accessed 2020-12-04]
18. Hall AK, Cole-Lewis H, Bernhardt JM. Mobile text messaging for health: a systematic review of reviews. Annu Rev Public Health 2015 Mar 18;36:393-415 [FREE Full text] [doi: 10.1146/annurev-publhealth-031914-122855] [Medline: 25785892]
19. Lewis JR. IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. Int J Human Comput Interact 1995 Jan;7(1):57-78. [doi: 10.1080/10447319509526110]
20. Clarke J, Proudfoot J, Birch M, Whitton AE, Parker G, Manicavasagar V, et al. Effects of mental health self-efficacy on outcomes of a mobile phone and web intervention for mild-to-moderate depression, anxiety and stress: secondary analysis of a randomised controlled trial. BMC Psychiatry 2014;14:272 [FREE Full text] [doi: 10.1186/s12888-014-0272-1] [Medline: 25252853]
21. McAleavey AA, Nordberg SS, Hayes JA, Castonguay LG, Locke BD, Lockard AJ. Clinical validity of the counseling center assessment of psychological symptoms-62 (CCAPS-62): further evaluation and clinical applications. J Couns Psychol 2012 Oct;59(4):575-590. [doi: 10.1037/a0029855] [Medline: 22946980]
22. Beecham J, Knapp M. Costing psychiatric interventions. In: Thornicroft G, editor. Measuring Mental Health Needs. Bideford: Edward Gaskell Publishers; 2001:200-224.
23. Hardin J, Hilbe J. Generalized Estimating Equations. Boca Raton: Chapman and Hall/CRC Press; 2003.
24. Rickwood D, Bradford S. The role of self-help in the treatment of mild anxiety disorders in young people: an evidence-based review. Psychol Res Behav Manag 2012;5:25-36 [FREE Full text] [doi: 10.2147/PRBM.S23357] [Medline: 22427736]
25. Garrido S, Millington C, Cheers D, Boydell K, Schubert E, Meade T, et al. What works and what doesn't work? A systematic review of digital mental health interventions for depression and anxiety in young people. Front Psychiatry 2019 Nov 13;10:759 [FREE Full text] [doi: 10.3389/fpsyt.2019.00759] [Medline: 31798468]
26. MacDougall S, Jerrott S, Clark S, Campbell LA, Murphy A, Wozney L. Text message interventions in adolescent mental health and addiction services: scoping review. JMIR Ment Health 2021 Jan 08;8(1):e16508 [FREE Full text] [doi: 10.2196/16508] [Medline: 33416504]

Abbreviations

CCAPS: College Counseling Center Assessment of Psychological Symptoms
C-SRI: Client Service Receipt Inventory for Mental Health
eUC: enhanced usual care
GEE: general estimating equation
MHSES: Mental Health Self-Efficacy Scale
MoST-MH: Mobile Support Tool for Mental Health
PSSUQ: Post-Study System Usability Scale
RCT: randomized controlled trial

Edited by G Eysenbach; submitted 20.07.21; peer-reviewed by I Shubina, A Tannoubi; comments to author 30.07.21; revised version received 03.08.21; accepted 03.08.21; published 14.10.21.

Please cite as:
Suffoletto B, Goldstein T, Gotkiewicz D, Gotkiewicz E, George B, Brent D
Acceptability, Engagement, and Effects of a Mobile Digital Intervention to Support Mental Health for Young Adults Transitioning to College: Pilot Randomized Controlled Trial
JMIR Form Res 2021;5(10):e32271
URL: https://formative.jmir.org/2021/10/e32271
doi:10.2196/32271
PMID:34647893

Original Paper

The Accuracy of Tidal Volume Measured With a Smart Shirt During Tasks of Daily Living in Healthy Subjects: Cross-sectional Study

Denise Mannée[1], MSc; Frans de Jongh[2], PhD; Hanneke van Helvoort[1], PhD

[1]Department of Pulmonary Disease, Radboud University Medical Centre, Nijmegen, Netherlands

[2]Department of Engineering Fluid Dynamics, University of Twente, Enschede, Netherlands

Corresponding Author:
Denise Mannée, MSc
Department of Pulmonary Disease
Radboud University Medical Centre
Geert Grooteplein Zuid 10
Nijmegen, 6525 GA
Netherlands
Phone: 31 24 361 1111
Email: denise.mannee@radboudumc.nl

Abstract

Background: The Hexoskin is a smart shirt that can take continuous and objective measurements and could be part of a potential telemonitoring system.

Objective: The aim of this study was to determine the accuracy of the calibrated Hexoskin in measuring tidal volumes (TVs) in comparison to spirometry during various tasks.

Methods: In a cross-sectional study, the TV of 15 healthy subjects was measured while performing seven tasks using spirometry and the Hexoskin. These tasks were performed during two sessions; between sessions, all equipment was removed. A one-time spirometer-based calibration per task was determined in session 1 and applied to the corresponding task in both sessions. Bland-Altman analysis was used to determine the agreement between TV that was measured with the Hexoskin and that measured with spirometry. A priori, we determined that the bias had to be less than ±5%, with limits of agreement (LOA) of less than ±15%. Lung volumes were measured and had to have LOA of less than ±0.150 L.

Results: In the first session, all tasks had a median bias within the criteria (±0.6%). In the second session, biases were ±8.9%; only two tasks met the criteria. In both sessions, LOA were within the criteria in six out of seven tasks (±14.7%). LOA of lung volumes were greater than 0.150 L.

Conclusions: The Hexoskin was able to correctly measure TV in healthy subjects during various tasks. However, after reapplication of the equipment, calibration factors were not able to be reused to obtain results within the determined boundaries.

Trial Registration: Netherlands Trial Register NL6934; https://www.trialregister.nl/trial/6934

(JMIR Form Res 2021;5(10):e30916) doi:10.2196/30916

KEYWORDS

telemonitoring; Hexoskin smart shirt; smart textiles; textile sensors; accuracy; healthy subjects; calibration; tidal volume

Introduction

Various pulmonary diseases are characterized by exacerbations, which are defined as acute and subacute increases in respiratory symptoms beyond normal variations. These exacerbations may warrant change in regular therapy, and sometimes hospitalization is necessary. Many patients have difficulties perceiving acute exacerbation of chronic obstructive pulmonary disease (AE-COPD) symptoms in a timely manner and, therefore, are not likely to seek consultation for additional therapy [1]. To decrease the unwanted effects of unreported AE-COPD, insight should be gained in the early changes in respiratory parameters and symptoms in the prodrome phase of AE-COPD. Telemonitoring parameters related to breathing provide these insights and could lead to a reduction in exacerbations and hospitalizations [2-4]. Current telemonitoring methods have the following disadvantages: (1) subjective evaluation of symptoms could increase awareness, causing overresponsiveness [5]; (2) measurement of forced expiratory volume in the first second of expiration (FEV_1) alone does not fully capture the symptoms

and the patient's perception of their health status [6]; and (3) lung function parameters are measured at rest and, therefore, poorly represent the daily problems with physical activities [7]. A potential method without these disadvantages, where measurements can be taken continuously and objectively, is the use of a Hexoskin smart shirt (Carré Technologies). The shirt can be worn by a patient and connected to a "smart device." The Hexoskin contains respiratory inductance plethysmography (RIP), which consists of a thoracic and abdominal belt. In addition to the RIP sensors, the Hexoskin also contains three electrocardiogram sensors and an acceleration sensor. Various lung parameters can be derived from the signals, such as minute ventilation, respiratory rate, tidal volume (TV), and inspiratory and expiratory times.

The validity of the lung parameters measured by the Hexoskin has been studied by multiple researchers in the last several years. In these studies, it has been shown that the uncalibrated signals obtained with the Hexoskin follow the clinical standard to a good extent [8-10]. There are no studies showing the accuracy of the Hexoskin in measuring spirometer-calibrated TV; however, other studies with another smart shirt, LifeShirt, using RIP technology were found. The no-longer available LifeShirt is a "snuggly" fit garment containing two RIP belts, electrocardiogram sensors, an accelerometer, and a pulse oximeter [11]. Clarenbach et al [11] found no significant bias, and limits of agreement (LOA) were within 8% for TV with qualitative diagnostics calibration. Brüllmann et al [12] used task-specific qualitative diagnostics calibration in lying, standing, and sitting. The bias was zero in all activities, and LOA ranged from 7% to 14%. Hollier et al [13] found a bias of 5 mL (1%) and LOA of 20% in healthy controls. They used fixed-volume calibration while the subject was in the sitting position. The calibration parameters for calculating real volumes were strongly dependent on the position of the belts and on the posture and activity of the subject [12,14,15].

Ideally, the Hexoskin should be able to measure TV accurately during various tasks of daily living, and a patient should be able to remove and reapply the Hexoskin without repeating the calibration procedure. The main focus of our study is, therefore, to determine the accuracy of the calibrated Hexoskin to measure TV in comparison to a spirometer in healthy subjects, before burdening patients with advanced lung disease. As a secondary goal, we investigated whether a one-time spirometry-based calibration could be reapplied after removal and reapplication of the Hexoskin. Moreover, we investigated the possibility of using the Hexoskin to measure functional lung volumes, such as slow vital capacity (SVC), forced vital capacity (FVC), and FEV_1. If the accuracy for functional lung volumes is good, this would make it possible to combine all measurements during home monitoring into one device. Lastly, we questioned the comfort of our subjects while wearing the Hexoskin.

Methods

Subjects

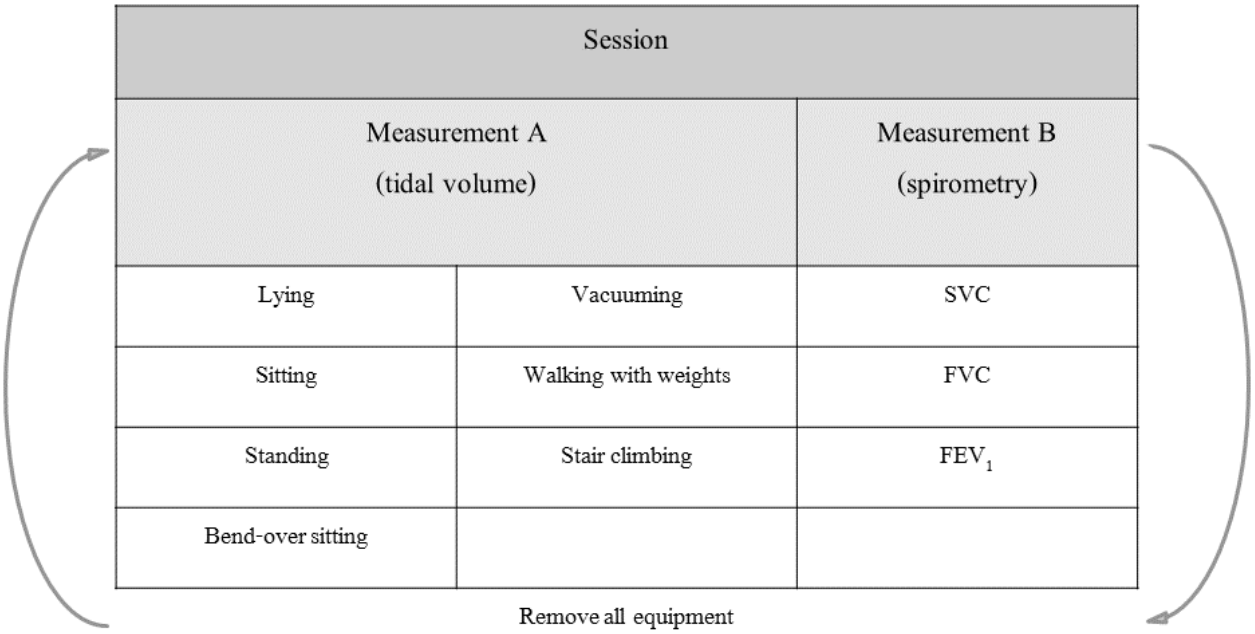
The Hexoskin study was performed between October 2018 and February 2019, as part of a multicenter, cross-sectional study. The study was registered with the Netherlands Trial Register (NL6934) and was approved by the Twente Medical Ethics Committee. A total of 15 subjects without known pulmonary or cardiac disease and with normal pulmonary function (ie, FEV_1>80% of predicted), between the ages of 18 and 80 years, were recruited at the University of Twente, Enschede, the Netherlands. Subjects were excluded when they were (1) not able to perform the tests, (2) not able to fit into one of the available Hexoskin sizes, (3) had a pacemaker or implantable cardioverter defibrillator, or (4) were unable to read and understand the Dutch language. All subjects gave written informed consent prior to participation.

Study Design

Overview

The circumferences of the thorax and abdomen of a subject, measured according to Carré Technologies' instructions, were used to select the most proper-sized Hexoskin. The RIP sensors in the Hexoskin measure fluctuations in the cross-sectional area of the chest wall based on changes in inductance, reflecting ventilation of the person wearing the Hexoskin. All subjects wore the Hexoskin for a minimum of 5 minutes before starting with the measurement to get used to the Hexoskin. After this period, they performed two sessions of measurements; see Figure 1 for an overview. Each session contained two measurements: measurement A and measurement B. At the end of session 1, all equipment was removed, including the Hexoskin. Within an hour, all equipment was reattached, and all measurements were repeated in session 2. Afterwards, subjects answered two questions describing their experience with the measurements.

Figure 1. Overview of session and measurements. After a session, all equipment was removed and reapplied. FEV_1: forced expiratory volume in the first second of expiration; FVC: forced vital capacity; SVC: slow vital capacity.



Measurement A

During measurement A, seven tasks of daily living were executed. The tasks constituted of 5 minutes each of lying, sitting, standing, bend-over sitting, vacuum cleaning, walking with weights (5 kg), and stair climbing (15 steps up and down). The latter four tasks were based on tasks that cause patients with pulmonary disease to complain of dyspnea in their daily lives [16]. The activities were performed in a pace determined by the subjects. During measurement A, respiration was recorded with the Hexoskin and a calibrated flow sensor from the Oxycon Mobile (Vyaire Medical).

Raw Hexoskin data were obtained from the online database of Carré Technologies. The data were processed with an algorithm made by the author in MATLAB (version 2018a; MathWorks). The data were filtered with a Savitzky-Golay filter to decrease high-frequency noise and movement artefacts. Thoracic and abdominal signals were summed to obtain one signal. Cross-correlation was used to correct for any time delay between the Hexoskin and the spirometer. All data were split into seven pieces, representing the seven tasks. The algorithm determined all end-expiratory and end-inspiratory lung volumes, which were used to calculate TV. TV measured with the Hexoskin (TV_{HX}) was calibrated based on TV measured with the spirometer (TV_{SPIRO}). All TV values in a task were used to determine calibration factors *a* and *b* with least squares linear regression (see equation 1). This resulted in 15×7 sets of calibration factors.

After calibration, an averaging window of 5 TVs was applied to all data to perform further analysis. The seven calibration functions from session 1 were applied to the seven tasks of session 2 to determine accuracy after removal of the Hexoskin.

A Bland-Altman analysis was used to determine the agreement between TV_{HX} and TV_{SPIRO}. The difference (diffTV) between the methods was represented as a percentage of TV_{SPIRO}:

For each task and for each individual, a bias and LOA were derived, and the mean or median bias and LOA for all subjects was determined per task. The definition of good agreement between the methods was established a priori. The median or mean bias had to be between –5% and 5%, and the LOA had to be between –15% and 15%. These criteria were established by clinicians and their opinions on accuracy in monitoring of TVs, based on their experience with cardiopulmonary exercise testing in the hospital.

Measurement B

During measurement B, subjects performed spirometry maneuvers according to the European Respiratory Society and American Thoracic Society guidelines [17] for lung function measurements. Lung function was synchronously recorded with the Hexoskin and a pneumotachograph, the Vyntus SPIRO (Vyaire Medical). Lung parameters measured were SVC, FVC, and FEV_1.

The raw data from measurement B were processed in the same manner as those from measurement A, as seen in the Measurement A section. The calibration function of the “sitting” task in measurement A was used to calibrate the Hexoskin signal to obtain SVC, FVC, and FEV_1.

To calculate the bias and LOA per session, the SVC, FVC, and FEV_1 for all subjects were used in Bland-Altman analysis. The bias should be close to zero, and LOA should not exceed ±0.150 L. These criteria were based on the European Respiratory Society and American Thoracic Society statements on lung function measurement [17]. The coefficient of variation was calculated per subject for SVC and FVC by dividing the SD by the mean.

Questions on Experience

After the measurements were conducted, we asked all subjects to answer two questions on their experience with the smart shirt. Responses to both questions were answered using a scale from 1 to 10. The first question was "What did you think of the fit of the shirt?"; a response of 1 represented "very loose," 5 represented "perfect fit," and 10 represented "very tight." The second question was "How comfortable were you whilst wearing the shirt?"; a response of 1 represented "not comfortable," 5 represented "slightly comfortable," and 10 represented "very comfortable."

Statistical Analysis

Baseline characteristics are described by mean (SD) or median (IQR), depending on normality. Normality was assessed by a Shapiro-Wilk test.

Results

Overview

A total of 15 healthy subjects were included in this study; no subjects were excluded. Baseline characteristics are presented in Table 1.

Table 1. Baseline characteristics of the participants.

Characteristic	Value (N=15)
Age (years), median (IQR)	28 (25-51)
Sex, n (%)	
Male	8 (53)
Female	7 (47)
Height (cm), median (IQR)	180 (173-188)
Weight (kg), median (IQR)	70 (65-83)
Circumference of thorax (cm), median (IQR)	87 (79-92)
Circumference of abdomen (cm), median (IQR)	85 (76-89)
Hexoskin size, n (%)	
Extra small	3 (20)
Small	2 (13)
Medium	5 (33)
Large	4 (27)
Extra large	1 (7)

Measurement A

In both sessions, out of 15 participants, individuals had to be excluded from further analysis due to artefacts in Hexoskin data regarding the following tasks: bend-over sitting (n=1, 7%), vacuuming (n=3, 20%), walking with weights (n=3, 20%), and stair climbing (n=4, 27%). Data from vacuuming and walking with weights tasks had to be removed from measurements for both subjects 1 and 2. Moreover, subject 2 data also showed large artefacts in the stair climbing task. In total, 50% of the removed data were explained by the removal of data from measurements for two individuals. Regarding the Hexoskin data, TV could not be distinguished from the artefacts and, therefore, we were unable to match TV between the two methods. If more than 50% of the data within a task were unmatchable, they were removed from further analysis. Figure 2 shows an example of the excluded data.

TV was normally distributed. We listed the tasks with the lowest and highest mean TV. In session 1, the mean TV_{SPIRO} ranged from 0.67 (SD 0.14) L for bend-over sitting to 1.57 (SD 0.48) L for stair climbing. The mean TV_{HX} ranged from 0.67 (SD 0.13) L for bend-over sitting to 1.57 (SD 0.46) L for stair climbing. In session 2, the mean TV_{SPIRO} ranged from 0.61 (SD 0.12) L for bend-over sitting to 1.63 (SD 0.49) L for stair climbing. The mean TV_{HX} ranged from 0.61 (SD 0.21) L for lying to 1.69 (SD 0.44) L for stair climbing.

With Bland-Altman analysis, the bias and LOA were calculated comparing TV_{SPIRO} and TV_{HX} in each session; see Figure 3 for an example. The biases found in subjects were nonnormally distributed. The median bias (median LOA) for all subjects per task are given in Table 2. In session 1, all tasks, except for vacuuming, met the criteria for the bias and LOA. In session 2, all tasks, except for vacuuming, met the criteria for LOA. Two tasks met the criteria for the bias.

Figure 2. Example of data excluded from further analysis because of extensive artefacts. In the upper plot, 60 seconds of the spirometry data are displayed. In the lower plot, the corresponding Hexoskin data are displayed. The blue stars represent the end-expiratory and end-inspiratory levels used for tidal volume calculation. As can be seen from both panels, in more than 50% of the measurements, it is impossible to find corresponding peaks between the spirometer and the Hexoskin.

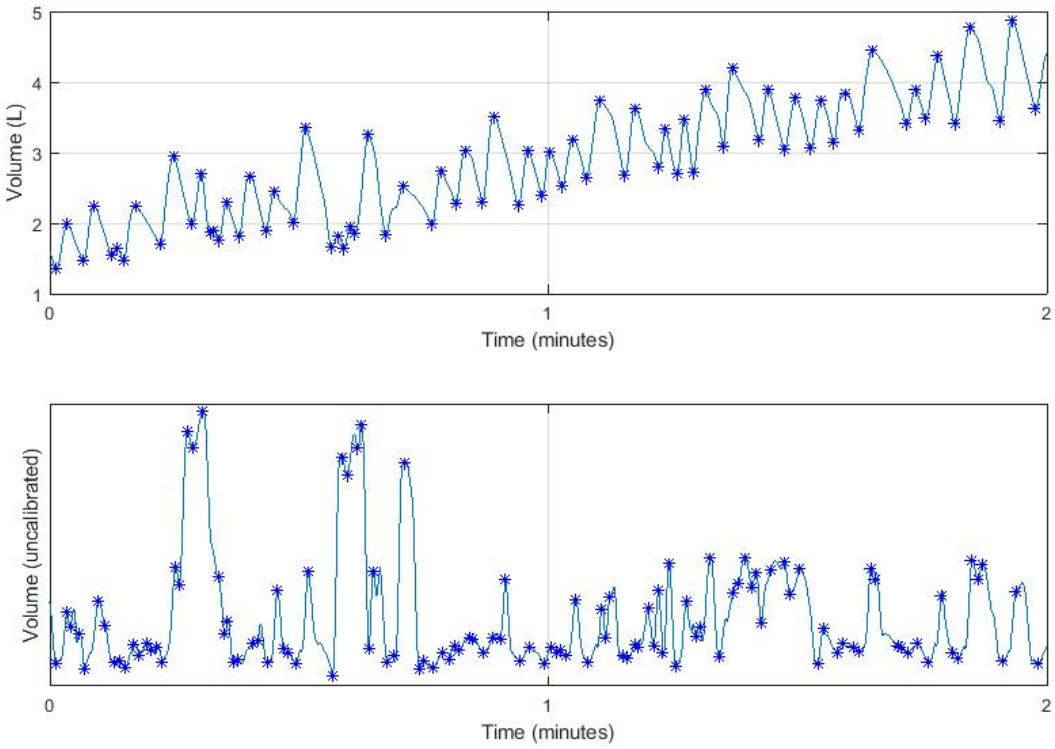


Figure 3. Example of a Bland-Altman plot of tidal volume (TV) of the task "lying" for subject 6. TV_{HX}: tidal volume measured with the Hexoskin; TV_{SPIRO}: tidal volume measured by spirometry.

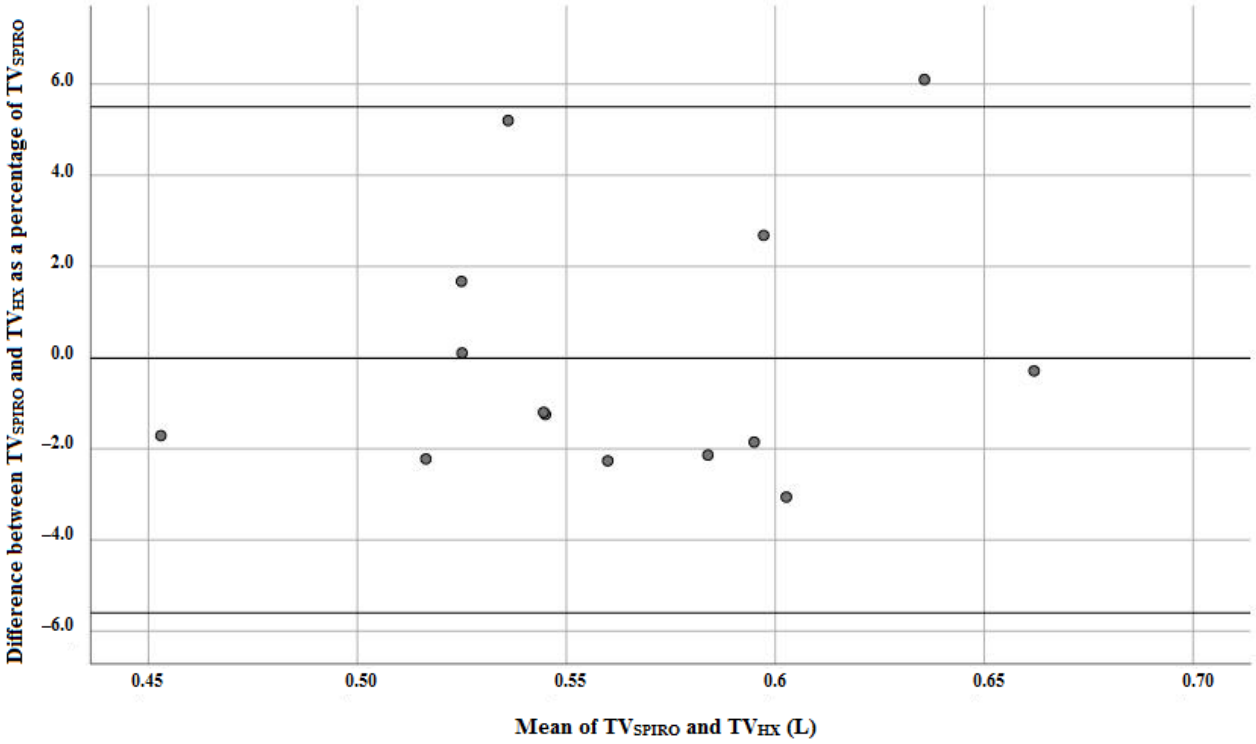


Table 2. Bland-Altman analysis was used to calculate the percentage bias (limits of agreement [LOA]) between tidal volume measured with spirometry and tidal volume measured with the Hexoskin.

Session	Task, median % bias (median LOA)						
	Lying (N=15)	Sitting (n=14)	Standing (N=15)	Bend-over sitting (N=15)	Vacuuming (n=12)	Walking with weights (n=12)	Stair walking (n=13)
Session 1	0.0 (5.6)	–0.2 (12.7)	–0.2 (12.0)	–0.4 (10.8)	–0.6 (16.7)	–0.3 (11.1)	–0.5 (12.6)
Session 2	7.5 (6.7)	–6.4 (14.7)	–8.5 (13.0)	3.3 (10.6)	–8.9 (16.4)	–7.3 (12.0)	–4.5 (15.0)

Measurement B

Due to excessive movement artefacts, as a result of breathing maneuvers, data from 5 subjects in the Hexoskin group had to be excluded from further analysis in both sessions. All lung parameters were normally distributed. A box plot of all functional lung parameters can be found in Figure 4.

With Bland-Altman analysis, the bias and LOA were calculated by comparing the various lung function parameters measured with the spirometer and the Hexoskin. The bias (LOA) of the SVC, FVC, and FEV_1 for each session can be found in Table 3. All Bland-Altman plots have similar features; see Figure 5 for Bland-Altman plots of SVC. Trends were found in all Bland-Altman plots. None of the parameters met the criteria for accuracy, and Hexoskin data showed large deviations from spirometry data. The coefficient of variation was smaller than 10% for SVC and FVC measured with both methods.

Figure 4. Box plots of all lung function parameters (n=30). FEV_1: forced expiratory volume in the first second of expiration; FVC: forced vital capacity; HX: Hexoskin; SPIRO: spirometer; SVC: slow vital capacity.

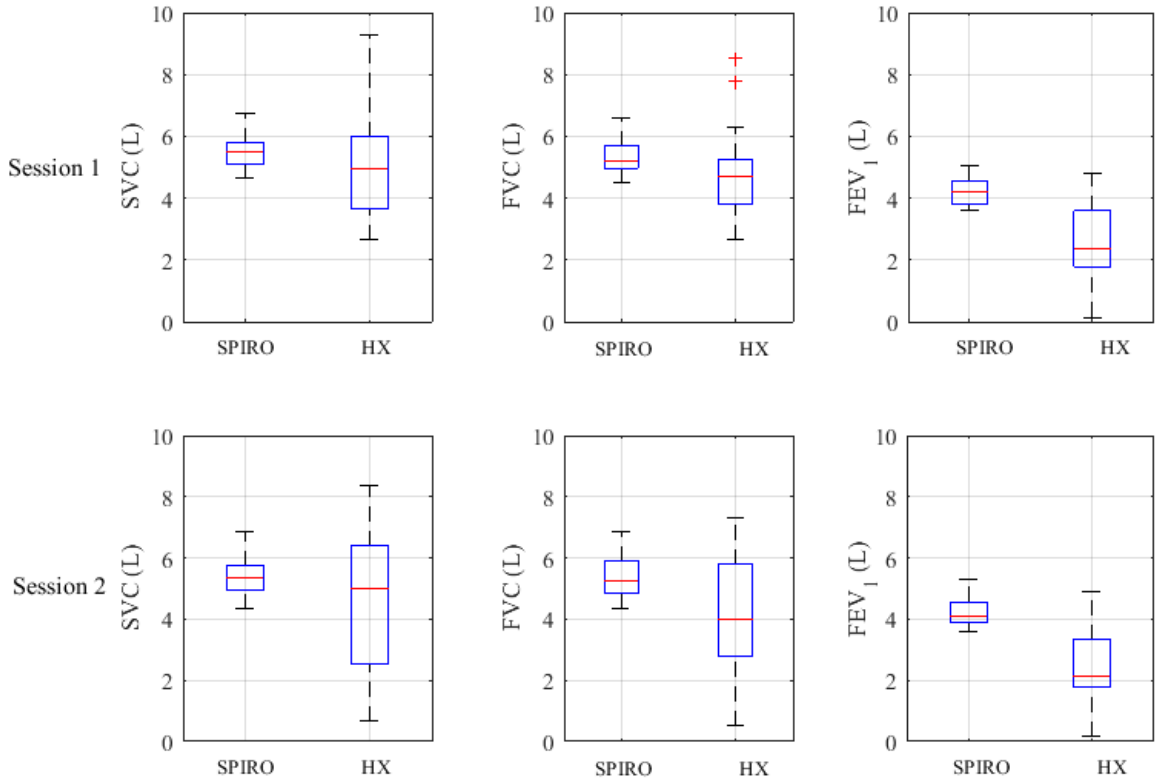
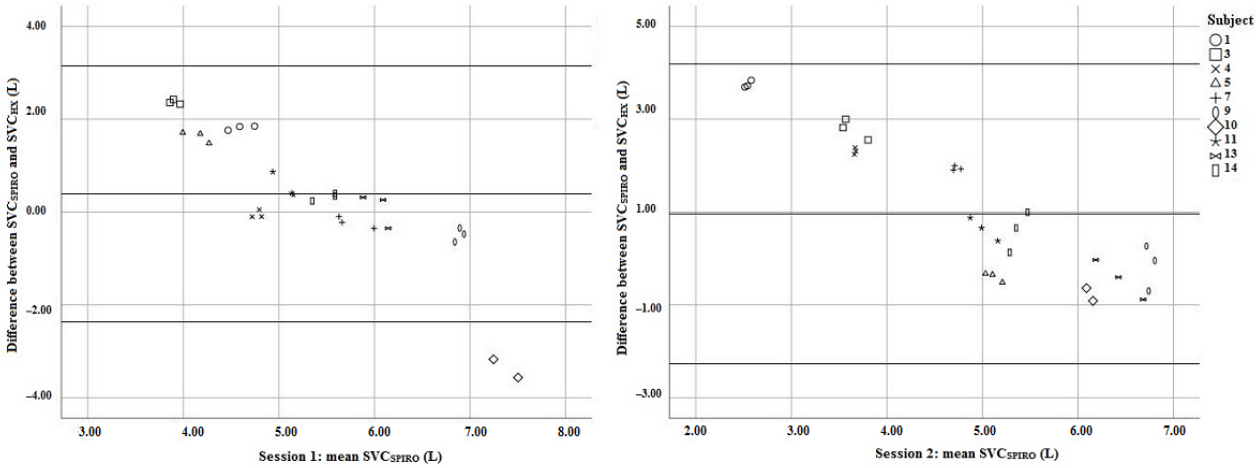


Table 3. Bland-Altman analysis was used to calculate the bias (limits of agreement [LOA]) between various lung parameters ($volume_{spirometer}$–$volume_{Hexoskin}$) in measurement B. Bias and LOA were based on 30 measurements (measurement of lung parameters in triplicate in 10 subjects).

Session	Lung parameter, bias (LOA)		
	Slow vital capacity, L	Forced vital capacity, L	Forced expiratory volume in the first second of expiration, L
Session 1	0.39 (2.76)	0.68 (2.23)	1.73 (2.20)
Session 2	0.96 (3.23)	1.26 (2.72)	1.85 (1.82)

Figure 5. Bland-Altman plots of slow vital capacity (SVC) (n=30). HX: Hexoskin; SPIRO: spirometer.



Questions on Experience

All subjects answered both questions. The mean rating for the fit of the shirt was 6 (SD 1). Comfort had a mean score of 8 (SD 2).

Discussion

Principal Findings

In this study, we determined the accuracy of the calibrated Hexoskin to measure TV in comparison to the spirometer in various tasks of daily living. In all tasks of session 1, accuracy of the Hexoskin was good, with biases within the criteria and acceptable LOA. Moreover, reapplicability of the calibration functions was tested in a repeated session. The high median bias in the second session showed that a recalibration was necessary after removal of the Hexoskin. However, LOA were comparable with the first session, which indicates that there was only an accuracy issue and not a precision problem. The Hexoskin is not yet a valid measurement tool to measure SVC, FVC, and FEV_1, based on bias and LOA. Lastly, the comfort level of subjects was high while wearing the Hexoskin.

Accuracy of TV Measurement During Activities of Daily Living

A priori, we determined that for good accuracy, the bias had to be between –5% and 5%, and the LOA had to be between –15% and 15%. Good accuracy of the Hexoskin in measuring TV was found in the first session. The median bias was found to be within ±0.6%, comparable to a bias of 3 mL on a TV of 0.500 L. Moreover, all LOA were within the ±15% that we determined a priori, with one exception for vacuuming (15.6%). The median biases were all negative, indicating that the Hexoskin overestimated TV slightly, on average.

The LOA found in the various tasks were slightly lower than the LOA in the previously mentioned studies [12,13], with the exception of the study by Clarenbach et al [11]. We hypothesize that the slight variations in outcome were the result of the performed analysis, in following three ways:

1. Calibration time. We used a calibration time of 5 minutes and Clarenbach et al [11] used 1 minute. It is hypothesized that the calibration is more accurate with longer calibration time and LOA are lower.
2. Bland-Altman analysis. We calculated a median bias over all subjects, while in all other studies the bias was calculated over all subjects. If the bias slightly varies among subjects, LOA will increase when one Bland-Altman analysis is performed.
3. Averaging window. We used an averaging window of 5 breaths; Clarenbach, for example, used 20 breaths. This could have increased the agreement, as a larger averaging window tends to decrease variations in RIP.

More demanding tasks (ie, bend-over sitting, vacuuming, walking with weights, and stair climbing) showed a slightly decreased agreement relative to the tasks at rest. The decrease in agreement was also reflected in the data we had to remove due to excessive artefacts; they were all part of a demanding task. Similarly, in the study by Brüllmann et al [12], agreement was decreased more highly in demanding tasks compared to agreement in a resting position. The increase in LOA from inactivity to more demanding tasks can be explained by the increase (1) in variation in breathing patterns [18,19] and (2) of movement artefacts.

Reapplicability of Calibration Factors

In session 2, we investigated the possibility of reusing calibration factors. Although the median biases were relatively small, in five out of seven tasks they did not meet the criteria. The LOA in the second session were comparable to the LOA found in session 1 in all tasks, with vacuuming as the exception, within 15%. This is in contrast with earlier studies. Heyde et al [20] found that the bias between RIP and the spirometer was similar in repeated measurements. Moreover, Brüllmann et al [12] compared calibration parameters between two measurements and found that the mean difference was zero.

Various reasons for poor accuracy in session 2 can be distinguished, based on poor calibration. The position and fit of the RIP belts on the subjects affect the calibration needed to get accurate results [12,14,15]. Various subjects were in between Hexoskin sizes; because of this, during measurements with these subjects, the Hexoskin tended to crawl up. Therefore, the chance is high that the Hexoskin and, thus, the thoracic and abdominal bands slightly changed positions (1) during measurements and

(2) in between measurements. A possible solution to this problem is presented in the study by Brüllmann et al [12]. The LifeShirt had a snug fit but was also affixed to the trousers of the subject with pins to prevent the upward movement of the shirt. Another explanation for not being able to reapply the calibration factors is magnetic coupling of the two belts in the Hexoskin. This appears to occur due to the mutual inductance between chest and abdominal bands modulating the output of the shirt. This is a known problem of RIP sensors [21]. The chance of poor reproducibility is high, as the influence of the bands on each other is not predictable when reapplying the Hexoskin.

Measurement of Functional Lung Parameters

A priori, we determined that all lung volumes should be measured with a bias close to zero and LOA equal to or smaller than 0.150 L. None of the lung volumes were measured within these criteria. To the knowledge of the authors, there have been no studies published investigating the feasibility of RIP technology to measure functional spirometry. Poor accuracy in this study can be explained by (1) movement of the shirt, (2) assumption of a linear relationship between circumference and lung volumes, and (3) the use of calibration factors determined from the "sitting" task. The first explanation has already been addressed in the previous section. The second and third argument are related. Based on earlier studies [22], the assumption was made that there is a linear relationship between circumference and lung volumes. However, this assumption is probably true for normal or resting breathing and not for forced maneuvers. We used the calibration of the "sitting" task. In this task, the volumes exhaled were, on average, 0.7 L, while in SVC and FVC maneuvers, volumes above 5 L, on average, were exhaled. This could explain the large deviation between volumes measured by RIP and the spirometer. The linear relationship between the mean and the difference in the Bland-Altman plots (Figure 5) can be explained by the difference in distribution of volumes (Figure 4) measured with both the Hexoskin and the spirometer.

Although we cannot determine the bias and LOA per subject, this would be interesting. Per subject, all SVC and FVC values measured with the Hexoskin seemed to be in very close range to each other (Figure 5). The mean coefficient of variation was comparable in both techniques. For SVC and FVC, the coefficient of variation was <10% for both techniques. This suggests that it is possible to measure functional lung volumes accurately; however, a proper calibration procedure should be defined containing a set with variations in lung volumes (eg, 200 mL to 2000 mL).

Limitations

In this study, we identified some limitations. In session 1, the same data were used for determination and validation of the calibration factors. We used this construction to minimize the measurement time for the subjects, and to make the calibration procedure realistic and applicable in a clinical setting for patients with pulmonary diseases. However, if the calibration varies over time within a task without removal of the equipment, we should have found larger LOA. Therefore, it is unlikely that the bias found in this study would change to a significant extent if extra validation data were to be included.

In measurement B, the use of the calibration from the "sitting" task could be the reason for poor accuracy of the measured lung volumes. Accuracy could have been decreased if calibration factors were derived from a sample of breaths larger than 70% of SVC or FVC.

Clinical Insights

We conclude that the Hexoskin accurately measures TV compared to spirometry values if the calibration procedure is used prior to measurement, as was the case in session 1. Based on our analysis, we conclude that 95% of the measured TV_{HX} values were accurately measured compared to TV_{SPIRO}. The other 5% of the measurements will not affect outcomes in home monitoring. During telemonitoring, it is common to measure over hours or days. In such a time span, the outliers will not have a significant impact on the average measurement of TV.

However, the implementation of the shirt as a telemonitoring device is not yet feasible. Currently, based on the results in session 2, calibrations have to be performed prior to each measurement. Due to the long calibration time (approximately 30 minutes), this would not be a practical procedure. Modifications to the shirt to deal with the problems of magnetic coupling will probably solve problems with the nonreproducibility of the calibration. Other solutions are to (1) investigate other calibration procedures that are simpler or shorter in nature or (2) look at uncalibrated data. We suggest, for example, a fixed-volume calibration [15], in which a subject breathes from a known volume for several minutes to obtain the calibration factors. In this case, spirometry measurements become unnecessary.

The shirt can be worn with a high level of comfort during measurements. This is a clinically relevant conclusion, as use of the shirt would not be feasible for patient monitoring if the comfort level was rated lower. Moreover, the manufacturer introduced a zipper into the shirt. This will make use of the shirt even more feasible for patient monitoring, as it will make putting on the shirt easier.

Conclusions

The spirometer-calibrated Hexoskin was able to accurately measure TV in healthy subjects during various tasks; however, calibration factors could not be used after removal and reapplication of the Hexoskin. The Hexoskin cannot be used to assess calibrated lung volumes reliably. In the near future, we will evaluate other calibration options to improve the reproducibility of the calibration. Moreover, we will determine the accuracy in patients and investigate the usability of the Hexoskin for home monitoring of patients with pulmonary disease.

Acknowledgments

Materials for this study were provided by MedWear BV, the Netherlands.

Conflicts of Interest

None declared.

References

1. Wedzicha JA, Brill SE, Allinson JP, Donaldson GC. Mechanisms and impact of the frequent exacerbator phenotype in chronic obstructive pulmonary disease. BMC Med 2013 Aug 14;11:181 [FREE Full text] [doi: 10.1186/1741-7015-11-181] [Medline: 23945277]
2. Cruz J, Brooks D, Marques A. Home telemonitoring effectiveness in COPD: A systematic review. Int J Clin Pract 2014 Mar;68(3):369-378. [doi: 10.1111/ijcp.12345] [Medline: 24472009]
3. Segrelles Calvo G, Gómez-Suárez C, Soriano J, Zamora E, Gónzalez-Gamarra A, González-Béjar M, et al. A home telehealth program for patients with severe COPD: The PROMETE study. Respir Med 2014 Mar;108(3):453-462 [FREE Full text] [doi: 10.1016/j.rmed.2013.12.003] [Medline: 24433744]
4. Lenferink A, Brusse-Keizer M, van der Valk P, Frith P, Zwerink M, Monninkhof E, et al. Self-management interventions including action plans for exacerbations versus usual care in patients with chronic obstructive pulmonary disease. Cochrane Database Syst Rev 2017 Aug 04;8:1-151 [FREE Full text] [doi: 10.1002/14651858.CD011682.pub2] [Medline: 28777450]
5. Lenferink A, Frith P, van der Valk P, Buckman J, Sladek R, Cafarella P, et al. A self-management approach using self-initiated action plans for symptoms with ongoing nurse support in patients with chronic obstructive pulmonary disease (COPD) and comorbidities: The COPE-III study protocol. Contemp Clin Trials 2013 Sep;36(1):81-89. [doi: 10.1016/j.cct.2013.06.003] [Medline: 23770110]
6. Jones P, Miravitlles M, van der Molen T, Kulich K. Beyond FEV1 in COPD: A review of patient-reported outcomes and their measurement. Int J Chron Obstruct Pulmon Dis 2012 Oct 16;2012(7):697-709. [doi: 10.2147/copd.s32675]
7. Lahaije AJMC, van Helvoort HAC, Dekhuijzen PNR, Vercoulen JH, Heijdra YF. Resting and ADL-induced dynamic hyperinflation explain physical inactivity in COPD better than FEV1. Respir Med 2013 Jun;107(6):834-840 [FREE Full text] [doi: 10.1016/j.rmed.2013.02.017] [Medline: 23514950]
8. Elliot CA, Hamlin MJ, Lizamore CA. Validity and reliability of the Hexoskin wearable biometric vest during maximal aerobic power testing in elite cyclists. J Strength Cond Res 2019 May;33(5):1437-1444. [doi: 10.1519/JSC.0000000000002005] [Medline: 28759538]
9. Banerjee T, Anantharam P, Romine W, Lawhorne L. Evaluating a potential commercial tool for healthcare application for people with dementia. In: Proceedings of the International Conference on Health Informatics and Medical Systems. 2015 Presented at: International Conference on Health Informatics and Medical Systems; July 27-30, 2015; Las Vegas, NV URL: https://corescholar.libraries.wright.edu/cgi/viewcontent.cgi?article=2442&context=knoesis
10. Villar R, Beltrame T, Hughson RL. Validation of the Hexoskin wearable vest during lying, sitting, standing, and walking activities. Appl Physiol Nutr Metab 2015 Oct;40(10):1019-1024. [doi: 10.1139/apnm-2015-0140] [Medline: 26360814]
11. Clarenbach CF, Senn O, Brack T, Kohler M, Bloch KE. Monitoring of ventilation during exercise by a portable respiratory inductive plethysmograph. Chest 2005 Sep;128(3):1282-1290. [doi: 10.1378/chest.128.3.1282] [Medline: 16162719]
12. Brüllmann G, Fritsch K, Thurnheer R, Bloch KE. Respiratory monitoring by inductive plethysmography in unrestrained subjects using position sensor-adjusted calibration. Respiration 2010;79(2):112-120 [FREE Full text] [doi: 10.1159/000212117] [Medline: 19365103]
13. Hollier CA, Harmer AR, Maxwell LJ, Menadue C, Willson GN, Black DA, et al. Validation of respiratory inductive plethysmography (LifeShirt) in obesity hypoventilation syndrome. Respir Physiol Neurobiol 2014 Apr 01;194:15-22. [doi: 10.1016/j.resp.2014.01.014] [Medline: 24468468]
14. Banzett RB, Mahan ST, Garner DM, Brughera A, Loring SH. A simple and reliable method to calibrate respiratory magnetometers and Respitrace. J Appl Physiol (1985) 1995 Dec;79(6):2169-2176. [doi: 10.1152/jappl.1995.79.6.2169] [Medline: 8847288]
15. Grossman P, Wilhelm FH, Brutsche M. Accuracy of ventilatory measurement employing ambulatory inductive plethysmography during tasks of everyday life. Biol Psychol 2010 Apr;84(1):121-128. [doi: 10.1016/j.biopsycho.2010.02.008] [Medline: 20176075]
16. Lahaije A, van Helvoort H, Dekhuijzen P, Heijdra Y. Physiologic limitations during daily life activities in COPD patients. Respir Med 2010 Aug;104(8):1152-1159 [FREE Full text] [doi: 10.1016/j.rmed.2010.02.011] [Medline: 20346641]
17. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, ATS/ERS Task Force. Standardisation of spirometry. Eur Respir J 2005 Aug;26(2):319-338 [FREE Full text] [doi: 10.1183/09031936.05.00034805] [Medline: 16055882]
18. Cabiddu R, Pantoni CBF, Mendes RG, Trimer R, Catai AM, Borghi-Silva A. Inductive plethysmography potential as a surrogate for ventilatory measurements during rest and moderate physical exercise. Braz J Phys Ther 2016 Mar 15;20(2):184-188 [FREE Full text] [doi: 10.1590/bjpt-rbf.2014.0147] [Medline: 26982454]

19. Pfaltz MC, Grossman P, Michael T, Margraf J, Wilhelm FH. Physical activity and respiratory behavior in daily life of patients with panic disorder and healthy controls. Int J Psychophysiol 2010 Oct;78(1):42-49. [doi: 10.1016/j.ijpsycho.2010.05.001] [Medline: 20472006]
20. Heyde C, Mahler H, Roecker K, Gollhofer A. A wearable respiratory monitoring device--The between-days variability of calibration. Int J Sports Med 2015 Jan;36(1):29-34. [doi: 10.1055/s-0034-1385864] [Medline: 25254900]
21. Cohen KP, Panescu D, Booske JH, Webster JG, Tompkins WJ. Design of an inductive plethysmograph for ventilation measurement. Physiol Meas 1994 May;15(2):217-229. [doi: 10.1088/0967-3334/15/2/009] [Medline: 8081197]
22. Konno K, Mead J. Measurement of the separate volume changes of rib cage and abdomen during breathing. J Appl Physiol 1967 Mar;22(3):407-422. [doi: 10.1152/jappl.1967.22.3.407] [Medline: 4225383]

Abbreviations

AE-COPD: acute exacerbation of chronic obstructive pulmonary disease
FEV_1: forced expiratory volume in the first second of expiration
FVC: forced vital capacity
LOA: limits of agreement
RIP: respiratory inductance plethysmography
SVC: slow vital capacity
TV: tidal volume
TV_{HX}: tidal volume measured with the Hexoskin
TV_{SPIRO}: tidal volume measured with spirometry

Edited by G Eysenbach; submitted 03.06.21; peer-reviewed by I Yang, G Hawthorne; comments to author 04.08.21; revised version received 16.08.21; accepted 17.08.21; published 18.10.21.

Please cite as:
Mannée D, de Jongh F, van Helvoort H
The Accuracy of Tidal Volume Measured With a Smart Shirt During Tasks of Daily Living in Healthy Subjects: Cross-sectional Study
JMIR Form Res 2021;5(10):e30916
URL: https://formative.jmir.org/2021/10/e30916
doi:10.2196/30916
PMID:34661546

Original Paper

Evaluation of a Digital Handheld Hydrogen Breath Monitor to Diagnose Lactose Malabsorption: Interventional Crossover Study

Simon C Mathews[1], MD; Sandy Templeton[2], BSc, PGCE; Stephanie K Taylor[3], MHA; Sten Harris[3], RMA; Margaret Stewart[3], RN, MS; Shruti M Raja[3], MD, MHS

[1]Johns Hopkins Medicine, Baltimore, MD, United States

[2]Electronics Program, Penn Foster College, Scottsdale, AZ, United States

[3]Duke Early Phase Clinical Research Unit, Durham, NC, United States

Corresponding Author:
Simon C Mathews, MD
Johns Hopkins Medicine
1800 Orleans St
Baltimore, MD, 21287
United States
Phone: 1 410 955 5000
Email: smathe14@jhmi.edu

Abstract

Background: Lactose malabsorption is a common condition that affects a broad segment of the population. Clinical diagnosis based on symptom recall can be unreliable and conventional testing can be inconvenient, requiring expensive laboratory-based equipment and conduction of the testing in a clinical setting.

Objective: The aim of this study is to assess the performance of a digital handheld hydrogen breath monitor (GIMate) in diagnosing lactose malabsorption compared to a US Food and Drug Administration (FDA)–cleared device (H2 Check) for the same indication.

Methods: An interventional crossover study was performed in adult participants with a prior confirmed diagnosis of lactose malabsorption or a suspected history of lactose intolerance.

Results: A total of 31 participants (mean age 33.9 years) were enrolled in the study. There was 100% positive percent agreement and 100% negative percent agreement between the GIMate monitor and the H2 Check. Correlation between gastrointestinal symptoms and hydrogen values was positive at 0.82 (P<.001).

Conclusions: The digital handheld GIMate breath monitor achieved equivalent diagnostic performance to that of an FDA-cleared device in the diagnosis of lactose malabsorption.

Trial Registration: ClinicalTrials.gov NCT04754724; https://clinicaltrials.gov/ct2/show/NCT04754724

(JMIR Form Res 2021;5(10):e33009) doi:10.2196/33009

KEYWORDS

digital health; lactose intolerance; digestive disease; evaluation; medical device; detection; diagnostic; digestion; testing; performance; gastrointestinal; diagnosis

Introduction

Lactose malabsorption is a common condition due to lactase deficiency; for many, it results in gastrointestinal symptoms, which is termed lactose intolerance [1]. Lactase is an enzyme occurring in the intestinal mucosa that hydrolyzes lactose into its constituent parts, galactose and glucose [2]. The enzyme is normally present in neonates; however, for a majority of individuals worldwide, there is an inherited and irreversible reduction in enzyme activity as individuals age [3]. Secondary lactose malabsorption can also occur when there is injury to the intestinal mucosa from a reversible condition, such as infection [4]. Not all individuals with lactose malabsorption will experience bothersome gastrointestinal symptoms (ie, lactose intolerance). Those with lactose intolerance are often diagnosed clinically—that is, their response is observed to lactose challenges that come through dietary exposure, followed by a trial of avoidance of lactose-containing products [5]. However, self-reported intolerance can often be unreliable [6,7], and formal diagnosis is still helpful in many cases.

Conventional testing is conducted through a hydrogen breath test in the ambulatory clinical setting. Individuals usually present to the testing site fasting, and a baseline breath reading is obtained. They are subsequently challenged with a lactose-containing solution, with follow-up hydrogen readings obtained on an hourly basis for 3 hours [8].

The mechanism of hydrogen detection is based on undigested lactose in the colon being fermented by bacteria and resulting in the production of hydrogen, which is then partially absorbed into the bloodstream and ultimately exhaled by the lungs via the pulmonary circulation and gas exchange. Direct lactase activity can also be measured on tissue obtained through jejunal biopsy via endoscopy. This approach, however, is more invasive, costly, and potentially less reliable given issues relating to sampling bias [5].

Current methods for hydrogen breath testing can be costly for physicians to purchase and inconvenient for patients because they must take time out of their day to go to the testing site for several hours. In addition, given that most conventional breath-testing equipment is reusable, infectious contamination can occur through aerosolized breath contents. As a result, the validation of an alternative breath test that is portable, handheld, and disposable presents an opportunity to improve the value, safety, and experience associated with hydrogen breath testing.

In this study, our primary aim was to compare the performance of the GIMate (Vivante Health), a novel digital handheld hydrogen breath monitor, to that of the H2 Check (Micro Medical Limited), a device cleared by the US Food and Drug Administration (FDA) for the diagnosis of lactose malabsorption.

Methods

Study Design and Procedure

The study was an interventional crossover design, with all participants receiving both diagnostic interventions. The order of which intervention was received first was alternately assigned at random (Multimedia Appendix 1). The study was conducted at Duke University and was reviewed and approved by the Duke University Institutional Review Board. Upon screening as eligible for the study, participants were provided with best practice pre–breath testing guidance [9], including dietary restrictions and fasting overnight. Baseline breath hydrogen measurements were performed the following day using the GIMate and H2 Check. This was followed by ingestion of a 25 g lactose solution and subsequent measurement of breath hydrogen on both devices at 1-hour, 2-hour, and 3-hour time points. During each of these measurements, participants completed a Likert-scale assessment of gastrointestinal symptom severity (0, none; 1, mild; 2, moderate; 3, moderately severe; 4, severe; 5, very severe).

Patient Population

Adults aged 18-55 years with a self-reported history of lactose malabsorption or lactose intolerance were recruited. Exclusion criteria included history of prior gastrointestinal surgery; self-reported history of any chronic gastrointestinal disease (eg, gastroesophageal reflux disease, celiac disease, Crohn disease, ulcerative colitis, pancreatitis); self-reported history of endocrine or metabolic disease that may impact gastrointestinal or colonic function (eg, hyper/hypothyroidism, diabetes); clinically significant cardiovascular, respiratory, renal, hepatic, hematologic, neurologic, or psychiatric disease for which chronic therapy (prescription or nonprescription) is required; self-reported history of allergic reaction to any drug or drug component; antibiotic use within 28 days of lactose malabsorption testing; use of nonantibiotic prescription or over-the-counter products (dietary or digestive supplements and laxatives) within 14 days of testing; self-reported use of nicotine-containing products or chronic secondhand smoke exposure within 14 days of testing; pregnancy; any other condition which in the Investigator's opinion may adversely affect the participant's ability to complete the study or its measures or which may pose significant risk to the participant based on medical history or physical examination; and consumption of food after midnight on the day of testing (within 12 hours) of testing or consumption of a nonwater beverage after midnight (or less than 8 hours) prior to testing.

Devices

The Micro H2 is a hydrogen monitor that has been cleared by the FDA [10] for the diagnosis of lactose malabsorption using an automatic sensor drift detection approach, which requires gas calibration [7]. The GIMate is a portable, handheld digital hydrogen monitor with a touch screen interface that detects hydrogen using a metal-oxide sensor, which does not require calibration (Figure 1). Results from breath tests are displayed to users on the digital touch screen and stored on the device.

Figure 1. The GIMate digital hydrogen monitor.

Outcome Measures

The primary outcome measure was positive percent agreement (PPA) in the diagnosis of lactose malabsorption of the GIMate compared to the H2 Check. Additional outcome measures included negative percent agreement (NPA) and correlation between GIMate hydrogen levels and patient self-reported gastrointestinal symptom severity. The protocol for measuring hydrogen levels was consistent with recent guidelines with a baseline measurement performed, followed by consumption of a 25 g lactose solution with subsequent hourly breath hydrogen measurements for a 3-hour period using both monitors. A positive diagnosis of lactose malabsorption was defined as a breath hydrogen level increase by 20 ppm or more from baseline at any point during the 3-hour measurement period based on guidelines [8]. A secondary outcome was the correlation between self-reported gastrointestinal symptoms and GIMate hydrogen readings.

Statistical Calculations and Analyses

The calculation for PPA was (number of individuals diagnosed with lactose malabsorption with GIMate) / (total number of individuals diagnosed with lactose malabsorption with H2 Check). The calculation for NPA was (number of individuals negative for lactose malabsorption with GIMate) / (total number of individuals negative for lactose malabsorption with H2 Check). Correlations between gastrointestinal symptoms and GIMate readings were calculated using the two-sided Spearman rank-based correlation measure of association. The *P* value was computed for testing for correlation estimate=0.

Results

Population Demographics

A total of 39 individuals were screened; of these, 31 were eligible to complete the lactose challenge. Demographic characteristics for the study participants are included in Table 1. The 8 participants who were excluded from the study did not meet the eligibility criteria, including no self-reported history of lactose malabsorption or intolerance (n=1); history of prior gastrointestinal surgery (n=1); history of chronic gastrointestinal disease (n=1); clinically significant condition requiring ongoing therapy (n=1); history of allergic drug reaction (n=2); and investigator determination of a condition that would pose unnecessary risk or would adversely affect participation in the study (n=2: 1 participant had a milk allergy, and 1 participant was unable to participate during normal work hours).

Table 1. Demographics of the study participants (N=31).

Characteristic	Value
Age (years), mean (SD)	33.9 (7.3)
Sex, n (%)	
Female	17 (55)
Male	14 (45)
Race, n (%)	
Asian	3 (10)
Black/African-American	14 (45)
White/Caucasian	14 (45)
Ethnicity, n (%)	
Hispanic or Latino	3 (10)
Non-Hispanic or non-Latino	28 (90)

Outcomes

The results for the primary outcomes regarding diagnostic performance are detailed in Table 2, demonstrating a PPA of 100% and an NPA of 100%.

The relationship between GIMate and H2 Check on a per-subject basis is depicted in Figure 2.

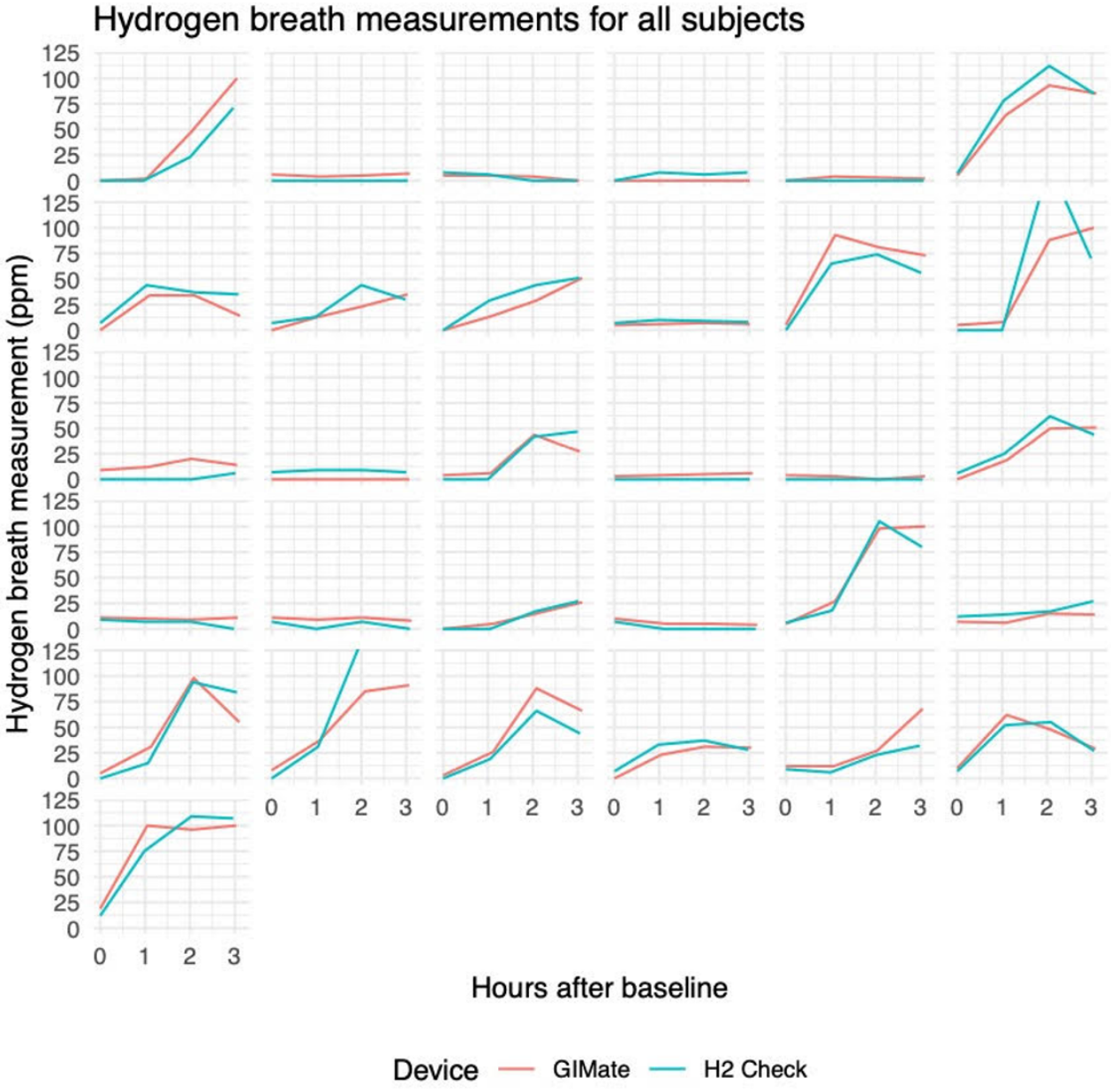
The relationships between GIMate and H2 Check readings for all participants across all time periods are demonstrated in Multimedia Appendix 2.

The secondary outcome of correlation between gastrointestinal symptoms and GIMate hydrogen is detailed in Multimedia Appendix 3, demonstrating an overall Spearman rank correlation of 0.82 across all time periods.

Table 2. Lactose malabsorption diagnosis by detection method (N=31).

	H2 Check test results	
GIMate test results	Positive	Negative
Positive	18	0
Negative	0	13
Total	18	13

Figure 2. GIMate and H2 Check hydrogen breath measurements by participant.



Discussion

Principal Results

The primary finding of this study is the 100% positive and negative percent agreement between GIMate and H2 Check for the diagnosis of lactose malabsorption, indicating equivalent diagnostic performance between both devices. This finding was also supported by the close relationship between individual hydrogen readings on both devices for each participant (Figure 1). The correlation estimate of 0.82 between GIMate hydrogen readings and gastrointestinal symptoms indicated a strong relationship between these two variables. This finding also suggests that the rise in hydrogen levels was likely associated with lactose intolerance in our study population, which is diagnosed based on gastrointestinal symptoms in the presence of lactose malabsorption. The highest correlation between symptoms and hydrogen levels was seen at the 2-hour interval, which suggests that symptom response was greatest prior to the end of formal testing at the 3-hour mark.

Prior studies examining portable hydrogen breath testing have been conducted, including a study focused on 29 adult and pediatric patients, which required a nasal prong and syringe to obtain samples and specifically included participants with comorbid gastrointestinal conditions (eg, irritable bowel syndrome, bacterial overgrowth) [11]. Another study included 12 patients with suspected lactose intolerance using a proprietary score calculated in “arbitrary units” and did not include diagnostic criteria for lactose malabsorption [12]. This study also required the use of a phone app to display data. The H2 Check was previously evaluated in 44 patients (77% female) and compared to a composite gold standard assessment that included breath, blood, and urine testing [7]. Prior studies did not conform to the 2017 consensus guidelines on hydrogen breath testing, which provided best practice recommendations

on key experimental elements such as dose of lactose, frequency of breath testing interval, and diagnostic criteria.

As a result, this study was the first to validate a portable, handheld device in diagnosing lactose malabsorption using the strongest and most current evidence-based approach. In addition, it demonstrated the first completely digital and portable measurement of breath hydrogen using clinically validated endpoints in a standalone, handheld device. This study also used the most stringent eligibility criteria compared to prior studies to minimize the risk of confounding due to the impact of comorbid gastrointestinal, surgical, and medical conditions on the production and detection of hydrogen. In addition, it is the first study to compare a novel breath hydrogen device to a previously FDA-cleared device in diagnostic performance for lactose malabsorption.

Lastly, while this was not explicitly evaluated, the compact, digital interface and substantially lower cost of the GIMate make it a potentially more convenient and safer alternative to conventional testing because it can be discarded after single-person use. These features also highlight its potential application in nonclinical settings, including home use. In this context, written instructions or integrated decision support could alert patients as to when to contact their physician based on the results. Although our study was the first to demonstrate the digital transformation of hydrogen breath testing, other studies have examined digital breath testing of other gases, including carbon dioxide [13], carbon monoxide [14], and hydrogen peroxide [15].

Limitations of this study included its relatively small sample size, although it was comparable to that in prior studies and included a more balanced and diverse representation of the population. In addition, our study was the first to examine a US-based population using a portable device. An additional limitation was the inability to directly compare our results with prior studies of portable breath hydrogen measurement, as prior studies used heterogeneous testing methods that did not conform to the most recent guidelines (two of the three studies were published prior to the establishment of these guidelines).

Conclusions

This study demonstrated that the GIMate has equivalent diagnostic performance to the H2 Check in the diagnosis of lactose malabsorption. It represents the first entirely digital approach to diagnosing lactose malabsorption with a portable, handheld device using validated clinical endpoints. These findings indicate that the GIMate is a potential viable alternative for portable, handheld detection of lactose malabsorption.

Acknowledgments

The authors acknowledge Vitaly Druker for his contributions to the statistical analysis. This study was funded by Vivante Health.

Authors' Contributions

SCM designed the intervention and wrote the manuscript. ST created training protocols for the use of equipment and provided critical review of the manuscript. SMR and MS provided critical review of the manuscript and oversaw the administration of the trial, including recruitment, allocation, scheduling, and collection and cleaning of data. SKT and SH coordinated and conducted all participant study visits. All authors reviewed and approved the manuscript.

Conflicts of Interest

SCM is an officer at Vivante Health with stock options in the company. ST has consulted for Vivante Health and has stock options in the company. SMR, SKT, SH, and MS have no conflicts to declare.

Multimedia Appendix 1
CONSORT (Consolidated Standards of Reporting Trials) diagram.
[PNG File , 127 KB - formative_v5i10e33009_app1.png]

Multimedia Appendix 2
Scatterplot of all GIMate and H2 Check readings.
[PNG File , 59 KB - formative_v5i10e33009_app2.png]

Multimedia Appendix 3
Spearman rank correlations between gastrointestinal symptoms and GIMate readings.
[PNG File , 44 KB - formative_v5i10e33009_app3.png]

References

1. Montalto M, Curigliano V, Santoro L, Vastola M, Cammarota G, Manna R, et al. Management and treatment of lactose malabsorption. World J Gastroenterol 2006 Jan 14;12(2):187-191 [FREE Full text] [doi: 10.3748/wjg.v12.i2.187] [Medline: 16482616]

2. Gilat T, Russo S, Gelman-Malachi E, Aldor TA. Lactase in man: a nonadaptable enzyme. Gastroenterology 1972 Jun;62(6):1125-1127. [Medline: 5068338]
3. Wang Y, Harvey CB, Hollox EJ, Phillips AD, Poulter M, Clay P, et al. The genetically programmed down-regulation of lactase in children. Gastroenterology 1998 Jun;114(6):1230-1236. [doi: 10.1016/s0016-5085(98)70429-9] [Medline: 9609760]
4. Labayen I, Forga L, González A, Lenoir-Wijnkoop I, Nutr R, Martínez JA. Relationship between lactose digestion, gastrointestinal transit time and symptoms in lactose malabsorbers after dairy consumption. Aliment Pharmacol Ther 2001 Apr;15(4):543-549 [FREE Full text] [doi: 10.1046/j.1365-2036.2001.00952.x] [Medline: 11284784]
5. Shaw AD, Davies GJ. Lactose intolerance: problems in diagnosis and treatment. J Clin Gastroenterol 1999 Apr;28(3):208-216. [doi: 10.1097/00004836-199904000-00005] [Medline: 10192605]
6. Jellema P, Schellevis FG, van der Windt DAWM, Kneepkens CMF, van der Horst HE. Lactose malabsorption and intolerance: a systematic review on the diagnostic value of gastrointestinal symptoms and self-reported milk intolerance. QJM 2010 Aug;103(8):555-572. [doi: 10.1093/qjmed/hcq082] [Medline: 20522486]
7. Peuhkuri K, Poussa T, Korpela R. Comparison of a portable breath hydrogen analyser (Micro H2) with a Quintron MicroLyzer in measuring lactose maldigestion, and the evaluation of a Micro H2 for diagnosing hypolactasia. Scand J Clin Lab Invest 1998 May;58(3):217-224. [doi: 10.1080/00365519850186607] [Medline: 9670345]
8. Rezaie A, Buresi M, Lembo A, Lin H, McCallum R, Rao S, et al. Hydrogen and methane-based breath testing in gastrointestinal disorders: the North American consensus. Am J Gastroenterol 2017 May;112(5):775-784 [FREE Full text] [doi: 10.1038/ajg.2017.46] [Medline: 28323273]
9. Instructions for patients. Johns Hopkins Medicine Gastroenterology and Hepatology Breath Testing Services. URL: https://www.hopkinsmedicine.org/gastroenterology_hepatology/_forms/patient_info/Breath_test_instructions_updated_6_5_18.pdf [accessed 2021-08-10]
10. 510(k) Premarket Notification. US Food and Drug Administration. URL: https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K963376 [accessed 2021-08-10]
11. Lee WS, Davidson GP, Moore DJ, Butler RN. Analysis of the breath hydrogen test for carbohydrate malabsorption: validation of a pocket-sized breath test analyser. J Paediatr Child Health 2000 Aug;36(4):340-342. [doi: 10.1046/j.1440-1754.2000.00511.x] [Medline: 10940167]
12. Shrestha A, Prodhan UK, Mitchell SM, Sharma P, Barnett MPG, Milan AM, et al. Validity of a portable breath analyser (AIRE) for the assessment of lactose malabsorption. Nutrients 2019 Jul 17;11(7):1636-1636 [FREE Full text] [doi: 10.3390/nu11071636] [Medline: 31319625]
13. Lorenz KA, Yeshurun S, Aziz R, Ortiz-Delatorre J, Bagley JR, Mor M, et al. A handheld metabolic device (lumen) to measure fuel utilization in healthy young adults: device validation study. Interact J Med Res 2021 May 17;10(2):e25371 [FREE Full text] [doi: 10.2196/25371] [Medline: 33870899]
14. Marler JD, Fujii CA, Wong KS, Galanko JA, Balbierz DJ, Utley DS. Assessment of a personal interactive carbon monoxide breath sensor in people who smoke cigarettes: single-arm cohort study. J Med Internet Res 2020 Oct 02;22(10):e22811 [FREE Full text] [doi: 10.2196/22811] [Medline: 32894829]
15. Neville DM, Fogg C, Brown TP, Jones TL, Lanning E, Bassett P, et al. Using the Inflammacheck device to measure the level of exhaled breath condensate hydrogen peroxide in patients with asthma and chronic obstructive pulmonary disease (the EXHALE pilot study): protocol for a cross-sectional feasibility study. JMIR Res Protoc 2018 Jan 30;7(1):e25 [FREE Full text] [doi: 10.2196/resprot.8768] [Medline: 29382628]

Abbreviations

FDA: US Food and Drug Administration
NPA: negative percent agreement
PPA: positive percent agreement

Edited by G Eysenbach; submitted 18.08.21; peer-reviewed by P Ray; comments to author 17.09.21; revised version received 17.09.21; accepted 20.09.21; published 18.10.21.

Please cite as:
Mathews SC, Templeton S, Taylor SK, Harris S, Stewart M, Raja SM
Evaluation of a Digital Handheld Hydrogen Breath Monitor to Diagnose Lactose Malabsorption: Interventional Crossover Study
JMIR Form Res 2021;5(10):e33009
URL: https://formative.jmir.org/2021/10/e33009
doi:10.2196/33009
PMID:34544034

Original Paper

A Mobile App to Identify Lifestyle Indicators Related to Undergraduate Mental Health (Smart Healthy Campus): Observational App-Based Ecological Momentary Assessment

Chris Brogly[1,2], MSc; J Kevin Shoemaker[3], PhD; Daniel J Lizotte[4,5], PhD; Jacqueline K Kueper[4,5], MSc; Michael Bauer[4], PhD

[1]Faculty of Information and Media Studies, Western University, London, ON, Canada

[2]Faculty of Health Sciences, Western University, London, ON, Canada

[3]School of Kinesiology, Western University, London, ON, Canada

[4]Department of Computer Science, Western University, London, ON, Canada

[5]Department of Epidemiology and Biostatistics, Western University, London, ON, Canada

Corresponding Author:
Chris Brogly, MSc
Faculty of Information and Media Studies
Western University
1151 Richmond Street
London, ON, N6A 3K7
Canada
Phone: 1 (519) 661 2111
Email: cbrogly@uwo.ca

Abstract

Background: Undergraduate studies are challenging, and mental health issues can frequently occur in undergraduate students, straining campus resources that are already in demand for somatic problems. Cost-effective measures with ubiquitous devices, such as smartphones, offer the potential to deliver targeted interventions to monitor and affect lifestyle, which may result in improvements to student mental health. However, the avenues by which this can be done are not particularly well understood, especially in the Canadian context.

Objective: The aim of this study is to deploy an initial version of the Smart Healthy Campus app at Western University, Canada, and to analyze corresponding data for associations between psychosocial factors (measured by a questionnaire) and behaviors associated with lifestyle (measured by smartphone sensors).

Methods: This preliminary study was conducted as an observational app-based ecological momentary assessment. Undergraduate students were recruited over email, and sampling using a custom 7-item questionnaire occurred on a weekly basis.

Results: First, the 7-item Smart Healthy Campus questionnaire, derived from fully validated questionnaires—such as the Brief Resilience Scale; General Anxiety Disorder-7; and Depression, Anxiety, and Stress Scale–21—was shown to significantly correlate with the mental health domains of these validated questionnaires, illustrating that it is a viable tool for a momentary assessment of an overview of undergraduate mental health. Second, data collected through the app were analyzed. There were 312 weekly responses and 813 sensor samples from 139 participants from March 2019 to March 2020; data collection concluded when COVID-19 was declared a pandemic. Demographic information was not collected in this preliminary study because of technical limitations. Approximately 69.8% (97/139) of participants only completed one survey, possibly because of the absence of any incentive. Given the limited amount of data, analysis was not conducted with respect to time, so all data were analyzed as a single collection. On the basis of mean rank, students showing more positive mental health through higher questionnaire scores tended to spend more time completing questionnaires, showed more signs of physical activity based on pedometers, and had their devices running less and plugged in charging less when sampled. In addition, based on mean rank, students on campus tended to report more positive mental health through higher questionnaire scores compared with those who were sampled off campus. Some data from students found in or near residences were also briefly examined.

Conclusions: Given these limited data, participants tended to report a more positive overview of mental health when on campus and when showing signs of higher levels of physical activity. These early findings suggest that device sensors related to physical

activity and location are useful for monitoring undergraduate students and designing interventions. However, much more sensor data are needed going forward, especially given the sweeping changes in undergraduate studies due to COVID-19.

(JMIR Form Res 2021;5(10):e29160) doi:10.2196/29160

KEYWORDS

smartphones; undergraduates; mental health; lifestyle; postsecondary institutions; mHealth; mobile application; ecological momentary assessment; mobile phone

Introduction

Undergraduate study is demanding, and students often feel overwhelmed by their obligations [1]. Depression and anxiety frequently occur in undergraduate students [2], and these issues lead to decreased student experience [1] and strain on existing health resources [3]. Somatic health problems on their own are a major burden for universities [3]. Therefore, exploring cost-effective, preventative options or ways to mitigate mental health issues using ubiquitous resources is a sensible response. Undergraduates are highly connected [4] and frequently use mobile technology. Previous research has identified that mobile technology can be used in mental health because of the abundance of useful sensors in modern devices [5]. For undergraduate students, mobile technology is uniquely positioned to combine these valuable sensor data with self-reports using apps for new insights into mental health, as approaches for translating sensor data into behavioral markers relevant to mental health have been previously outlined [5,6]. Studies in this area have been previously conducted [6] but not necessarily using every available sensor and, to the best of the author's knowledge, usually not at a Canadian university (although some do exist) [7]. Although Canadian universities can have similarities with universities in other countries (such as the United States), they are not identical, and there appears to be a general research gap for sensor-driven, smartphone-based mental health studies at Canadian universities. Similar studies conducted in other countries with comparable academic structures and settings can be relevant to Canadian undergraduates; however, they may not capture the nuanced experience of undergraduate education in Canada.

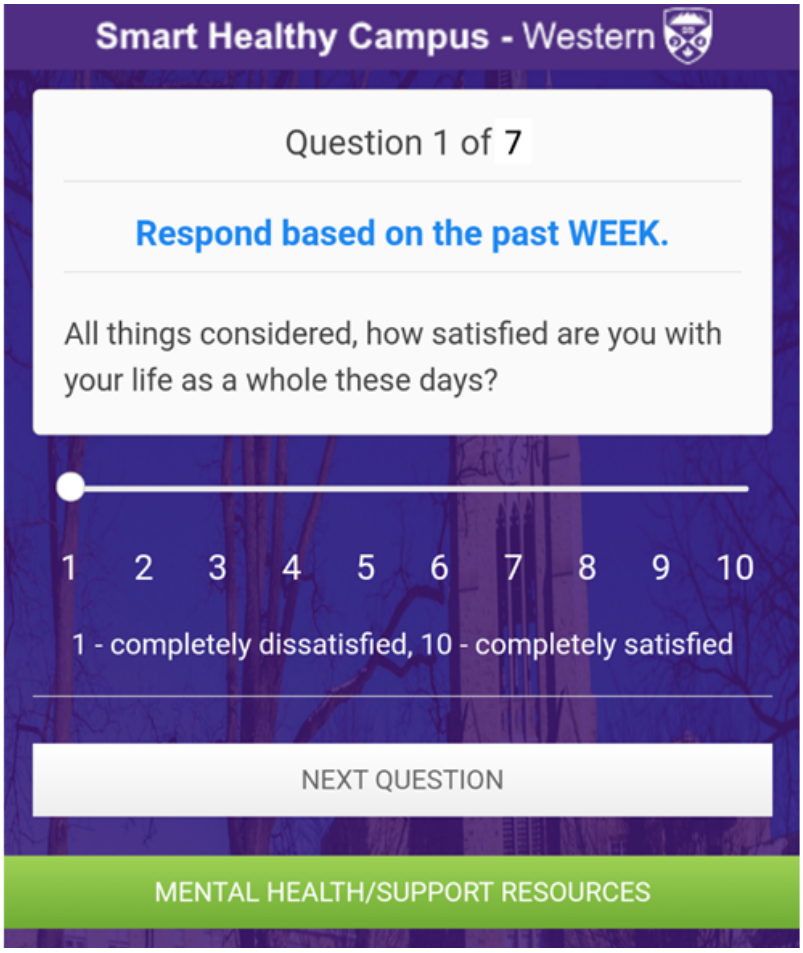
For the purposes of this research, it is important to clarify the term "mental health" as there are various definitions [8]. Here, the World Health Organization (WHO) definition is used, given the functional and practical requirements of an undergraduate study. Drawing upon the WHO definition for undergraduate study, it is essential that a student can "work productively" [9,10] and can "make a contribution" [9] to the university community. Students require the ability to "think" [9] for academic challenges and the ability to "emote or interact" [9], given that a university is ultimately an institution comprising people and not specific locations or buildings. Those who pursue undergraduate study essentially choose to do so as a step to "earn a living and enjoy life [9,11].

Given that Canadian universities [3] have recognized that undergraduate students are faced with stresses that challenge mental health, we developed the Smart Healthy Campus (SHC) app (Figure 1) to investigate the potential relationships between student lifestyle (measured by a short survey instrument) and relevant device sensors. The survey is a 7-item questionnaire that provides an overview of mental health and is included in Multimedia Appendix 1. The 7 questions were assembled from longer validated questionnaires [1]. The version of the app deployed at the time of writing contained a transaction-based sensor data collection system that captured readings relevant to aspects of lifestyle, which could be potential indicators of a student's mental health.

The SHC app was deployed at Western University (formally known as the University of Western Ontario [UWO]) in London, Ontario, Canada, from March 2019 to March 2020 to all full-time undergraduate students. London's population is about 383,000, and it is Canada's 11th largest city. Western University is a relatively large public university in Ontario, Canada, with approximately 25,000 full-time undergraduates out of approximately 40,000 students in total as of 2020. The main campus is technically urban, although it is situated in a highly residential neighborhood north of London's downtown core.

Participants were required to read the letter of information and consent at sign-up in the app. They then used their university email and self-identified as undergraduates and provided consent to this study, which was approved by the Health Sciences Research Ethics Board at Western University. Participants manually submitted questionnaire responses before any sensor readings were collected and sent to the study server.

Figure 1. Question displayed by the Smart Healthy Campus app.



The primary function of the SHC app is to facilitate an ecological momentary assessment (EMA). EMA is essentially used to observe and repeatedly sample study participants in their natural environments [12]. This preliminary iteration of the SHC study and app was sampled only on a weekly basis; however, the sample rate in subsequent work has been increased to a daily basis (see the Discussion section). The data collected comprised (1) the questionnaire results (providing a broad overview of mental health based on total score) and (2) sensor readings. The main goal of our work with this app is to determine if there were significant differences in the sensor readings between students who submitted a low-scoring response to the overall questionnaire compared with the sensor readings of those who submitted a high-scoring response. The reason for this interest is that, in the Canadian undergraduate context, sensors that might be the most useful with respect to understanding mental health have not been thoroughly investigated. If these are better understood, it would be considerably easier to build interventions aimed at evidence-based outcomes (including but not limited to notifications, suggestions, or app-based mentors or assistants) that might be included in new apps in this area. For instance, if there was evidence that Canadian students reporting a poor overview of mental health consistently displayed low readings from pedometers or minimal movement shown through GPS samples; interventions could attempt to alter lifestyle but then actually confirm a positive outcome by monitoring for increased activity in those sensors. We hope that this work will provide data to directly address this in the context of Canadian undergraduate students.

In addition, we investigated the relationship between distance to a university campus and sensor readings using GPS coordinates and geographic information system software. This was examined for 2 cases: (1) for samples that included questionnaire responses (providing a direct link to our focus on mental health) and (2) for samples that did not contain questionnaire responses (the app did not collect questionnaire responses for every consented transaction) as there were more samples in this case, and the data are still of interest.

Methods

Overview

This preliminary SHC study was conducted as a weekly EMA, which overall is a digitized, compacted version of most aspects of an in-person SHC pilot study [1]. The main advantage of moving to an app-based format was to make everything accessible to hundreds or potentially thousands of undergraduates rather than only a small classroom-sized group. However, although the in-person study focused on a mentorship intervention between upper- and lower-year students who hoped to improve mentors' physical activity, resilience, and mental health, this digital SHC study omitted mentorship and simply observed participants as they progressed normally through their undergraduate study while using the SHC app. The idea being

that SHC would still collect responses from participants regarding depression, anxiety, and resilience, but the extent to which these might occur would be connected to potential indicators coming from data from device sensors, such as pedometers or GPS, rather than from the impact of a mentorship intervention. In the pilot study, depression and anxiety were identified as having a major negative impact on undergraduates [1], and resilience was identified as the capacity to help cope with difficult life situations [1]. Having objective indicators of these items from device sensors could allow for new interventions such as notifications, suggestions, or app-based mentors or assistants targeting and monitoring those sensors. However, the questionnaires from the in-person study [1] for measuring these items would not fit well in an app format.

Development of the SHC Questionnaire

Although short, validated mental health–related questionnaires were already available, we wanted to attempt to capture the same items as the original in-person SHC pilot study, albeit with a much smaller number of questions. It was not possible to do this with any existing questionnaires, as they did not cover all the domains of the original in-person SHC pilot study. The in-person SHC pilot study used several long-form validated questionnaires to measure general well-being, depression, anxiety, and resilience. These were issued using traditional survey methods [1]. Some examples of these instruments include the Brief Resilience Scale; Depression, Anxiety, and Stress Scale; and Mental Health Inventory. The problem with these items is that they might be difficult to complete in an app because of their combined length. As a result, a new SHC 7-item questionnaire (Multimedia Appendix 1) was created for SHC 1.0. All 7 questions were taken from the validated questionnaires with some minor edits to clarify that they are asking about the participant's experience in the past week. These 7 questions would still provide similar measurements to the in-person SHC study, although at the expense of some accuracy, which was expected, given the significant reduction in the number of questions. Table 1 outlines the targeted areas of the questions and their scores. The total sum of the maximum scores for the 7 questions was 42. Table 2 outlines a tentative scoring system for the questionnaire, which was ultimately not used in this study. However, it is still related to how participant responses (and the corresponding sensor data) were grouped into the low- and high-scoring groups, as these were used for data analysis in the Results section.

In the Results section, we defined a low-scoring response as one with an SHC questionnaire total score ≤24 out of 42. We selected 24 as the low score cut-off, as it would include some marginal-to-acceptable total scores from 21 to 24 and all poor scores. We defined a high-scoring response as one with an SHC questionnaire total score >24, as it is difficult to obtain a full score of 42/42 in many cases. For instance, the last question asks about the days with significant exercise in the last week, and each day counts as 1 point, hence the seemingly marginal standard for a high-scoring response. This was done because of the limited amount of data for analysis.

Table 1. Coverage of the Smart Healthy Campus mental health overview questionnaire and scoring.

Question number[a]	Behavioral marker	Points	Question source
1	Life satisfaction	10	[13]
2	Psychological well-being	6	[14]
3	Resilience	5	[15]
4	Anxiety	4	[16]
5	Depression	4	[17]
6	Community connectedness	6	[18]
7	Physical activity	7	[19]

[a]Questionnaire score total: 42 points.

Table 2. Smart Healthy Campus 1.0 7-item questionnaire tentative scoring system.

Range	Tentative rating (disused in this work but relevant to the following column)	Assigned to a low-scoring or high-scoring group for data analysis in this work
0-6	Very poor	Low
7-13	Poor	Low
14-21	Marginal	Low
22-27	Acceptable	Low (score ≤24); high (score >24)
28-35	Good	High
36-42	Very good	High

Development of the SHC App

App Requirements

There were 3 key app requirements for SHC. The first key requirement of the app was to administer the 7-item SHC questionnaire and collect responses. Sliders and multiple-choice options were used depending on the question. The 7-item questionnaire incorporated into this study was asked on a weekly basis by the app.

The second key requirement was the ability to collect data from various device sensors. Although the focus is almost exclusively on smartphones for this work, we still use the term "device sensor" as tablets, iPads (Apple Inc), or iPods (Apple Inc) were also able to run the app (although no one used them to participate in this study). Data collection from device sensors was required to occur for any of 3 important events: when a participant requested to complete a weekly survey (a Request event), when a participant submitted responses to a survey (a Response event), or when a participant used a “Mental Health/Support Resources” panel (a HelpNow event).

The third key requirement was to make it easy for students to find information in a single place about the support services that were available to them. The app would contain a Mental Health Support Resources button, which opened a panel that provided participants with easy access to key crisis services if they felt that they needed them. This included phone numbers, website links, and general information about the types of support resources. This requirement was derived from the finding of one assessment that facilitating access to crisis support had evidence for suicide prevention [20]. If a participant pressed the Mental Health/Support Resources button, they were to be presented with a panel that asked them if they would like to consent to data collection about how they used the Mental Health/Support Resources panel. If consent was given, then data were to be sent to and recorded by the server on what parts of the panel were used, such as if they expressed interest in a service in the app or went further and selected a website link or a phone number. Combined with questionnaire and sensor data, these data are important, as they may provide markers that indicate that a student needs additional support. In addition, device sensor data were collected during this event.

Design and Architecture

We refer to the underlying client-server software for the SHC app as EMAX1 (EMA Extensions 1st edition software). The general idea behind EMAX1 is that it will eventually become a generic app for EMA-type studies instead of having a specific focus or name such as Smart Healthy Campus. Instead, participants will simply download a single app and select the survey they want to participate in; surveys will be able to adjust sensor data collection to their specific needs.

The EMAX1-based SHC software was designed using a standard client-server architecture. Participants used the EMAX1 client app to communicate with the EMAX1 server. The server software runs in the cloud and responds to any requests that are generated by clients over hypertext transfer protocol secure. In general, for SHC, which is the only EMAX1-based app, communication is minimized to improve responsiveness over potentially poor data connections. For instance, the survey questions were originally sent from the server to the app to allow for easy reconfiguration; however, eventually, as much as possible was moved into the client to maximize responsiveness at the expense of some flexibility.

Implementation

The EMAX1 client app was developed using Apache Cordova. Cordova apps are implemented as webpages and can be built into apps for Android and iOS. Cordova was used as Android and iOS share the same HTML, Cascading Style Sheets, or JavaScript codebase, which reduced development efforts. However, even with a shared codebase, some conditions needed to be in place to check system variables for detecting which operating system was running to correctly obtain hardware sensor data, because of differences between iOS and Android. The EMAX1 server software used for SHC was implemented in Python. The server is not described in detail in this paper, as the Python implementation relies on standard libraries and currently existing technology.

Recruitment and Consent

Recruitment was conducted at Western University primarily from March to December 2019. Web links to connect undergraduate students to the SHC app were distributed to certain undergraduate classes and sent to relevant student wellness stakeholders in undergraduate faculties. In-class recruitment sessions were held in high-enrollment classes (≥150 students) where professors allowed them. Mass email recruitment to all undergraduates was also conducted. Participants were able to join the SHC at any time by downloading the app and registering to participate using their university (UWO) email account. Students were required to read the letter of information and had the option of consenting to express their interest in connecting their records from student health services and student recreation services to the SHC study. Unfortunately, connecting to the records was ultimately not possible in this preliminary study because of logistic issues, although subsequent studies are expected to specifically connect to campus health records.

The SHC app solicited responses via push notifications (small popup messages on the participant’s device) to existing users, usually 2 times per week: once on Tuesday in the afternoon and then again between Friday and Sunday as a final reminder. At least one previous study focused on the timing and frequency of push notifications for use in an app and suggested that static notifications delivered at a recurring time are acceptable [21]. Notification emails were also occasionally sent to registered participants at the start of each weekly survey on Monday. Occasionally, if the app experienced technical issues, additional push notifications or emails were sent to participants to let them know the system was running again.

Analysis

All data were analyzed as a single collection, and there was no analysis with respect to time because of the limited amount of data. In addition, as Shapiro-Wilk tests showed that data were not normally distributed, we relied on nonparametric statistical tests that compare the mean ranks of the groups. The

Mann-Whitney U test was used to compare the mean ranks of the 2 groups, and the Kruskal-Wallis test with a post hoc Wilcoxon rank-sum test with Bonferroni correction was used to compare the mean ranks of 3 or more groups. No corrective action was required for the missing data. For example, during a sensor sample, if we did not obtain a GPS coordinate but did obtain the battery level, then we would only include any associated data related to the battery level for statistical tests, such as a questionnaire score. Data analysis was conducted using a combination of R 3.4.4 and SPSS 26 for Linux.

Results

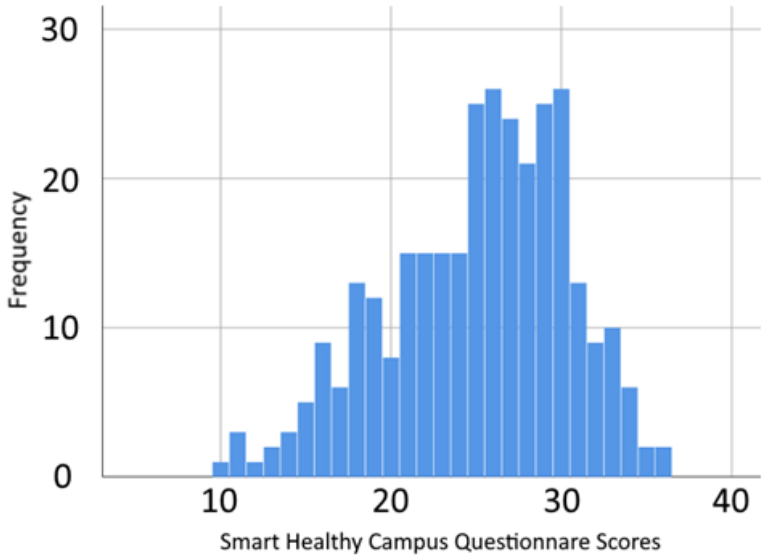
Overview

Data collection for this preliminary work occurred from March 2019 to March 2020 and ended around the time the WHO declared COVID-19 a pandemic. There was no specific sample size for this work; however, it was hoped that eventually, approximately 100-300 students would participate on a weekly basis (this would happen for our revised COVID-19 pandemic app called Student Pandemic Experience [SPE] for some time). For SHC 1.0, this did not turn out to be a feasible goal (possibly because of the lack of any incentive system), and enrollment and participation fell well short of this. Owing to technical limitations in both the study and the app, both being in the early stages of this work, demographic information was not collected. This omission limits the grouping variables for the analysis of the results. For instance, we did not examine any subgroups. This was later addressed after SHC 1.0 concluded to allow for demographic data collection in future studies (see the Discussion section). A summary of collected data is presented below in Table 3. Figure 2 is a histogram of all response scores.

Table 3. Summary of collected data.

Item and description		Values, n (%)
Participants (the total number of participants who enrolled in the study)		139 (100)
iOS users	How many participants used an Apple device?	121 (87.1)
Android users	How many participants used an Android device?	18 (12.9)
Participants responding for >1 week	The number of participants that completed more than one weekly survey.	42 (30.2)
Total number of samples (Response+Request+HelpNow event counts)		813 (100)
Response events	The number of questionnaire responses. Response events consist of a questionnaire response and data from all sensors at the time they occurred.	312 (38.4)
Request events	The number of times participants tried to obtain a questionnaire. Request events only contain data from all sensors at the time they occurred.	492 (60.5)
HelpNow events	The number of times participants used the mental health resources panel and consented to its data collection features.	9 (1.1)

Figure 2. Histogram plot of all participants' Smart Healthy Campus questionnaire scores. The maximum score is 42 (mean 25.08, SD 5.32; N=312).



In the first subsection, we show that the SHC questionnaire does have significant correlations with the selected mental health domains. Then, an analysis of the preliminary data collected is presented from different perspectives in the following 3 subsections. For the first 2 sections, focus is placed on how the participant responses to the app questionnaire (the self-reported mental health overview) related to the readings collected from device sensors (more objective indicators of lifestyle). These data came only from Response events (when participants submitted answers to a questionnaire). The third and final

analysis was based on any sensor samples containing GPS data, which came from both Request (only obtaining the weekly questionnaire) and Response (responding to the weekly questionnaire) events from participants. These were combined for analysis, as including the readings from Request events increased the value of N for statistical tests based on geographic information for on- or off-campus locations. No analysis of the HelpNow events could be presented here because of a lack of data.

The results of the analyses shown below provide some limited data as to what sensors would be most relevant for monitoring or designing interventions (ie, targeted suggestions, notifications, or app-based mentors or assistants) for undergraduate mental health.

Correlations Between the SHC Questionnaire and Mental Health Domains

Tests for correlations were conducted between the SHC app questionnaire and domains for resilience, anxiety, and depression. These domains comprised individual questions taken from fully validated questionnaires used in the original SHC course-based study [1]. Individual questions were selected from established questionnaires using a data-driven method (devised but not yet published by authors; Kueper and Lizotte, unpublished data, April 2019) applied to completed questionnaire data from a similar population, as well as undergraduate students at the same university who were participants in the original in-person SHC study. The method essentially ranks questions based on explained variability and was used to select a subset of the most highly ranked questions to cover each of the target measurement domains. When questions were similarly ranked, the one that would require minimal modification for quick administration using an app was selected. Where needed, questions included in the app were modified in terms of the period for participants to consider but not in terms of major word content. These domains are shown in Textbox 1.

Textbox 1. Domains and questions for correlation tests.

Resilience

- Entire Brief Resilience Scale questionnaire

Anxiety

- Entire General Anxiety Disorder-7 questionnaire
- Depression, Anxiety, and Stress Scale (DASS): 1, 3, 8, 10, 15, 19, 20
- Mental Health Inventory (MHI): 4, 10, 11, 18
- One Visual Analog Scale–type question

Depression

- DASS: 2, 5, 9,12, 14, 17, 21
- MHI: 2, 9, 12, 14

Other

- DASS: 4, 6, 7, 11, 13, 16, 18
- MHI: 1, 3, 7, 8, 13 15, 16, 17

For correlation tests, we compared the results of 1 week of in-class responses to the questions covering the 4 domains with those of responses to relevant parts of the app questionnaire reduction. The Spearman rank correlation test was used as the main test statistic (although Pearson correlation tests were run anyway and results were very similar; they are not included here). The results are shown in Table 4.

Table 4. Results of Spearman correlation between in-class responses and app questionnaire reduction responses.

Correlation	Domains	Sample, n	Spearman ρ (SE)	*P* value
Sum of questions 1-5	Resilience, anxiety, depression, and other	48	0.39 (0.13)	.005
Life satisfaction question	Resilience, depression, and anxiety	48	−0.23 (0.14)	.11
Resilience question	Resilience	55	−0.62[a] (0.10)	<.001
Depression question	Depression	52[b]	0.52 (0.10)	<.001
Anxiety question	Anxiety	50[b]	0.71 (0.07)	<.001

[a]Question coding was reversed during this test, which resulted in a negative correlation.

[b]n is different for some tests, which shows that some participants only completed certain full surveys in class.

On the basis of these results, the Spearman rank correlation between the selected items that were chosen for the app at the time of writing was generally moderate (Spearman ρ=0.39; *P*<.001) to strong (Spearman ρ=0.71; *P*<.001). An exception was app question 1 from the World Values Survey 2012, which turned out not to have a statistically significant correlation with the resilience, depression, and anxiety domains, although the focus was not quite the same.

Overall, the results showed that the 7-item SHC questionnaire implemented in the app had a useful correlation with the full mental health surveys, whereas the single question measures for resilience, depression, and anxiety had strong correlations with their respective domains from the full surveys. Some loss of precision was expected, given that the app questionnaire was very short. We argue that some loss of precision is acceptable for this type of observational research as the advantage is that the questionnaire reduction can be completed in far less time than the full surveys, making it appropriate for use in the app.

Difference Between On-Campus and Off-Campus Questionnaire Response Scores

The first analysis of interest was to determine, using GPS data, if there were any differences in the questionnaire scores between samples from students that were responding to weekly surveys on campus compared with samples from those that were responding to the surveys off campus. This was examined, as living off campus can help contribute to a decrease in daily physical activity during the undergraduate study [22]. To do this, all Response event records with identifiable GPS information (some records did not have GPS information because of transmission issues, or participants denied the request to access GPS) were used. On-campus students were within 1.5 km of coordinates at the center of the main campus of UWO. Everyone else was classified as an off-campus student. A Mann-Whitney U test was conducted between these 2 groups to determine if there were any significant differences (on campus, mean rank=121.41; off campus, mean rank=89.05; Mann-Whitney U=3914.5; *P*<.001).

The results of the test showed that there was a significant difference between the 2 groups (*P*<.001) and that the on-campus students actually had a higher mean rank than those off campus, which suggested, at least based on these samples, that students on campus actually reported a more positive overview of mental health than those who were off campus.

Difference Between Sensor Readings From Low-Scoring Responses Compared With High-Scoring Responses

The second analysis was conducted to determine if there were significant differences in the sensor readings from those submitting low responses to weekly questionnaires compared with those who submitted higher responses. Differences in sensor readings were tested from 2 perspectives: one without GPS data used as a grouping variable and one with GPS to create on- and off-campus groups.

We defined a low-scoring response as one with an SHC questionnaire total score ≤24 out of 42. We selected 24 as the low score cut-off, as it would include some marginal total scores from 21 to 24 and the poor (≤50%) scores below that. We defined a high-scoring response as one with an SHC questionnaire total score >24, as it is difficult to obtain a full score of 42/42 in many cases. For instance, the last question asks about the days with significant exercise in the last week, and each day counts as 1 point, hence the seemingly marginal standard for a high-scoring response.

Differences Between Low-Scoring and High-Scoring Responses—No Additional GPS Grouping

The first perspective was simply based on total questionnaire scores, where a low-scoring response group and a high-scoring response group were tested for any significant differences in sensor data using a Mann-Whitney U test. The results are shown in Table 5.

The results of the Mann-Whitney U tests in Table 5 yielded some significant findings with the sensors. The mean ranks of these items are listed in Table 6.

Participants with low-scoring responses, based on mean rank, tended to be plugged in (charging) more, had higher uptimes (possibly as a result of charging), showed less physical activity based on pedometer readings, used the app less, and took less time to complete the weekly responses.

Conversely, participants with high-scoring responses, based on mean rank, were plugged in less and had their devices turned on less. However, they spent more time using the app and responding to the weekly surveys. They also showed greater levels of physical activity based on the pedometer readings.

Table 5. Mann-Whitney U tests between low- and high-scoring samples for significant differences between sensor readings.

Sensor item	New use?	Low-scoring samples[a]	High-scoring samples[b]	Mann-Whitney U	*P* value[c]
Plugged-in flag	No	55	83	1477.5	<.001
Battery level	No	64	135	4074.5	.51
User CPU[d] time	Yes[e]	87	169	6916.0	.44
Idle CPU time	Yes	87	169	6818.0	.34
Total CPU time	Yes	87	169	6897.0	.42
User CPU percentage	Yes	80	153	5515.0	.22
User idle percentage	Yes	80	153	5537.0	.23
Available RAM	Yes	87	169	6386.0	.09
System uptime	No	121	189	9557.0	.006
Uptime with sleep	No	121	189	9524.0	.005
Pedometer step count	No	73	58	1564.0	.01
Pedometer distance	No	77	61	1738.0	.009
Pedometer floors up	No	77	61	1600.5	.001
Pedometer floors down	No	77	61	1617.5	.002
Pedometer 2 step count	No	13	10	51.0	.23
App use time	Yes	123	189	9082.0	.001
Survey time to complete	Yes	123	189	9678.0	.01
Resource panel time	Yes	123	189	11591.0	.76

[a]Total Smart Healthy Campus questionnaire score ≤ 24.

[b]Total Smart Healthy Campus questionnaire score >24.

[c]*P*<.01 is considered significant.

[d]CPU: central processing unit.

[e]Anything marked with "Yes" was, to the best of the author's knowledge in 2020, a new use of this sensor for this type of work.

Table 6. Mean ranks for sensor items with significant differences, low versus high questionnaire response scores.

Significant sensor item	Low-scoring questionnaire response scores, mean rank	High-scoring questionnaire response scores, mean rank
Plugged-in flag	84.14	59.80
System uptime	171.02	145.57
System uptime with sleep	171.29	145.39
Pedometer step count[a]	58.42[a]	75.53[a]
Pedometer distance	61.57	79.51
Pedometer floors up	59.79	81.76
Pedometer floors down	60.01	81.48
App use time	135.84	169.95
Survey time to complete[a]	140.68[a]	166.79[a]

[a]These items had *P* values very close to the level considered significant (*P*=.01), suggesting that tests with more data may report direct significance.

Differences Between Low-Scoring and High-Scoring Responses With On- or Off-Campus GPS Grouping

The second perspective was based on both total questionnaire scores and GPS data, resulting in 4 groups: low-scoring on-campus responses, high-scoring on-campus responses, low-scoring off-campus responses, and high-scoring off-campus responses. On-campus students were within 1.5 km of coordinates at the center of the main campus of UWO. Testing for any significant differences in sensor data among any of these 4 groups was completed using a Kruskal-Wallis test. These results are shown in Table 7, with the mean ranks from the Wilcoxon post hoc test in Table 8.

Table 7. Kruskal-Wallis tests for significant differences between sensor readings, including low- and high-scoring samples for both on- and off-campus students.

Sensor	New use?	Sample, n				Chi-square (*df*)	*P* value
		On campus		Off campus			
		Low scoring	High scoring	Low scoring	High scoring		
Plugged-in flag	No	14	53	28	14	32.3 (3)	<.001
Battery level	No	20	91	28	18	3.9 (3)	.27
User CPU[a] time	Yes[b]	34	103	44	48	0.9 (3)	.82
Idle CPU time	Yes	34	103	44	48	7.3 (3)	.06
Total CPU time	Yes	34	103	44	48	4.5 (3)	.22
User CPU percentage	Yes	34	94	37	42	18.1 (3)	<.001
User idle percentage	Yes	34	94	37	42	17.7 (3)	<.001
Available RAM	Yes	34	103	44	48	7.0 (3)	.07
System uptime	No	34	103	44	48	63.7 (3)	<.001
Uptime with sleep	No	34	103	44	48	68.9 (3)	<.001
Pedometer step count	No	19	20	28	34	1.9 (3)	.59
Pedometer distance	No	22	23	28	34	0.7 (3)	.88
Pedometer floors up	No	22	23	28	34	2.9 (3)	.41
Pedometer floors down	No	22	23	28	34	2.1 (3)	.55
Pedometer 2 step count	No	N/A[c]	N/A	13	N/A	2.2 (3)	.33
App use time	Yes	34	103	44	48	42.3 (3)	.001
Survey time to complete	Yes	34	103	44	48	30.4 (3)	.001
Resource panel time	Yes	34	103	44	48	1.2 (3)	.75

[a]CPU: central processing unit.

[b]Anything marked with "Yes" was, to the best of the author's knowledge in 2020, a new use of this sensor for this type of work.

[c]N/A: not applicable.

Table 8. Mean ranks for sensor items with significant differences, low-scoring versus high-scoring questionnaire response scores, on campus and off campus.

Significant sensor item	Mean rank			
	On campus		Off campus	
	Low-scoring	High-scoring	Low-scoring	High-scoring
Plugged in	48.71	41.79	77.36	66.57
User CPU[a] percentage	86.44	98.71	140.76	97.67
User idle percentage	121.74	109.13	67.68	110.17
Uptime	117.32	80.78	156.07	149.14
Uptime with sleep	116.54	79.46	156.95	151.71
App use time	112.79	143.72	94.66	73.58
Survey time to complete	95.91	138.65	113.98	78.71

[a]CPU: central processing unit.

Although the results shown in Tables 7 and 8 are limited, given the small values of N, they suggest that sensor values can be interpreted depending on various contexts, given our experimentation with GPS coordinates here. In this paper, context is essentially the grouping variable used. For instance, the questionnaire score was always intended to be used to contextualize the data for this initial study as we compare low scores with high scores here. However, GPS location also appears to provide some useful context, given that we found some significant differences among the 4 groups, shown initially in Table 7 through significant *P* values and then in Table 8 with mean ranks from the Wilcoxon post hoc test.

The results of the Kruskal-Wallis tests, using GPS as an on- or off-campus grouping variable, showed some significant findings. The mean ranks of these items are listed in Table 8.

Pairwise comparisons using the Wilcoxon rank-sum test with Bonferroni correction were conducted on the 4 groups shown in Table 5 to determine specific differences, the significance of which is shown by any mention of *P* values in the remainder of this section.

For the plugged-in sensor, there was a significant difference between the high-scoring on-campus samples and low-scoring off-campus samples ($P<.001$), and high-scoring off-campus samples ($P=.008$). This suggests that more participants with higher scores were not plugged in (charging or charged) on campus. This may suggest more movement on campus.

User central processing unit (CPU) time is consumed by the apps launched by the user. User CPU time percentage and user CPU idle time percentage, in general, have an inverse relationship. User CPU time, based on mean ranks from 3 pairwise comparisons, tended to be higher for low-scoring off-campus samples ($P<.001$; $P=.003$; $P=.004$) compared with any on-campus samples and high-scoring off-campus samples. This finding suggests that low-scoring off-campus participants put more computational load from user software on their devices, suggesting higher general use.

Low-scoring on-campus samples ($P=.004$), low-scoring off-campus samples ($P<.001$), and high-scoring off-campus samples ($P<.001$) tended to have significantly higher uptimes (in milliseconds) based on mean rank compared with high-scoring on-campus samples. Similar results were found for uptime with sleep (when the device is left on but is not used for a significant amount of time, such as overnight). This suggests that high-scoring on-campus students use devices less.

For app use time, high-scoring on-campus samples tended to have, based on their mean ranks, significantly higher SHC app use time than low-scoring off-campus and high-scoring off-campus samples but not low-scoring on-campus samples. This result somewhat contradicts the general test between low- and high-scoring questionnaire scores in Tables 7 and 8, where the low-scoring group had, based on mean rank, less app use time.

For survey time, there was only a significant difference between on-campus high-scoring samples and off-campus high-scoring samples ($P<.001$), where the on-campus high-scoring samples tended to report longer survey completion times.

Differences Among On-Campus, Limited In-Residence, and Off-Campus Sensor Data Using All Samples, No Questionnaire Scores

The last analysis was completed on the remaining preliminary data using additional Request events, where on- or off-campus GPS data were available but questionnaire response scores were not. The Request samples were combined with the Response sample sensor data to increase the value of N for the statistical tests. These results are suspected to be less reliable than those from the previous 2 subsections, given the amalgamation of the events.

First, Wilcoxon rank-sum tests were conducted between the Request and Response event groups for each sensor item to determine if there was a statistically significant difference between the 2 types of requests. If there was a significant difference between the Request and Response events for a sensor item, it was not analyzed in this section. For instance, app use time will always be significantly lower in a Request compared with a Response, as time will have passed in the Response event. The results for the acceptable sensor items are shown in Table 9.

Table 9. Mean ranks for sensor items with significant differences, on-campus versus residence versus off-campus locations.

Significant sensor item[a]	On-campus		Residence		Off-campus	
	Mean rank	Sample, n	Mean rank	Sample, n	Mean rank	Sample, n
Battery level	177.45	240	96.28	16	202.02	105
User time	211.38	166	323.51	35	213.66	241
Idle time	215.84	166	367.09	35	204.26	241
Total time	213.06	166	363.54	35	206.68	241
Uptime	202.88	280	364.36	47	379.41	260
Uptime with sleep	199.59	280	353.03	47	385.00	260

[a]$P<.01$.

Pairwise comparisons using the Wilcoxon rank-sum test with Bonferroni correction were conducted on the 3 groups shown in Table 6 to determine specific differences.

For battery level, there was a significant difference between the on-campus and residence samples ($P=.008$). On the basis of mean rank, on-campus samples tended to have higher battery levels than those from residence samples (possibly as participants returned to their residence with drained batteries).

There was also a significant difference between samples from residence compared with samples from off campus ($P<.001$), with off-campus samples reporting, based on their mean ranks, higher charge levels. This suggests that more participants may have been reporting in with access to a charger, likely at home.

For user CPU time, there was a significant difference between the on-campus and residence samples ($P=.008$) and residence and off-campus samples ($P<.001$), with higher user CPU times

tending to be from residence samples based on their mean ranks. The same results were found for idle CPU time ($P<.001$; $P<.001$) and total CPU time ($P<.001$; $P<.001$). These limited results may suggest that participants in residences are running their phones longer and actively using them more. However, this finding may be less reliable than previous ones in terms of CPU time because of the amalgamation of the Request and Response sensor data.

Discussion

Limitations

Overall, the results presented here are preliminary, given the limited amount of data. As this was exploratory observational research, the effects of confounding factors could not be controlled. In addition, we relied on self-identification of undergraduate status because of privacy issues related to the release of information from the university registrar. However, sign-ups largely coincided with recruitment from undergraduate classes and mass emails to undergraduates. In general, personal smartphones are commonly used by undergraduates at Western University; however, it is possible that some students may not have permanent access to one for various reasons, such as losing a device or perhaps for financial reasons, which may affect participation in any smartphone-based study from time to time and as such is always a possible limitation.

Other limitations were that the number of samples (n=813), total number of participants (n=139), and retention to a second weekly sampling (42 participants) were low, given that this study was accessible to a university campus over nearly a year. This stemmed primarily from incremental revisions to the study, resulting in limited recruitment, system downtime for maintenance or other issues, and possibly from limited interest as no incentives were available or offered to participants over this period. Owing to technical limitations with the study and the SHC app, it was not possible to collect information on why participation was low or if participants uninstalled the app. In addition, we did not take differences in the undergraduate year of study, or program, sex, or any other demographic items into account for this preliminary analysis. In part, this was because we expected to be able to obtain this from the university; we now ask for it directly in the new SHC 2.0 and SPE apps. Another issue was that there was no continuous or longitudinal analysis of individual responses over time; it was not feasible to perform this, given that retention rates were low. It is expected with version 2.0 of the app and intensified recruitment efforts that a longitudinal analysis may become possible.

It is also important to note that the results may not necessarily be representative of all institutions, as Western University is a relatively large (approximately 40,000 total enrollments) publicly funded university with an urban campus situated in a heavily residential area. In addition, these data were collected before the shift to web-based learning at Western University, resulting from COVID-19. It is important to note that going forward, numerous measures resulting from COVID-19 (even as vaccination rates increase in Canada) are likely to affect Western University in 2021 and in some time beyond. As a result, the limited data from this study will not be representative of conditions for some time, given the massive impact of COVID-19 on universities.

Impact

This work begins with the process of establishing an evidence base for future digital interventions for the Canadian undergraduate population. It is the first study in a line of research to identify associations between psychosocial factors (captured by our SHC questionnaire) and behaviors (measured by smartphone sensors) in undergraduates at a Canadian university. When the associations between psychosocial factors and behaviors are known, evidence-based, cost-effective digital interventions, such as notifications and chatbots, can be developed to intervene in the everyday experiences of students. These offer the potential to alter lifestyle patterns to align with behaviors related to positive mental health. As app-based interventions can be widely distributed, they are expected to be used to reduce the extent of mental health concerns and burdens on campus health resources. Overall, in this initial study, we found that on-campus location and higher levels of physical activity were associated with more positive mental health. Similar studies on undergraduates from the United States and China have found generally comparable results [23,24].

Next Steps

These initial results regarding sensor indicators are promising but limited, and it is apparent that increased data collection would yield more meaningful results. As a result, the underlying EMAX1 client or server software has been upgraded to EMAX2 (EMA Extensions 2nd edition software) and is supporting a newer SHC 2.0 app, which includes a number of general improvements, such as increasing the number of sensors used for data collection, background data sampling, daily sampling, and the addition of a points-based incentive system. The SHC 2.0 app also includes a more robust sign-up process that addresses differences in enrollment status and some demographic information, items that were not always feasible to obtain another way, such as from the registrar. The SPE app to study COVID-19 was also built on EMAX2 and is available on iOS and Android at the time of writing. It is expected that the SHC and SPE EMAX2–based research apps will eventually be replaced by a single unified EMAX3 (EMA Extensions 3rd edition software) app for iOS and Android, which will include a full configuration system to conduct additional EMA-based studies.

Conclusions

Although previous studies have been conducted in this area, data directly from Canadian undergraduates are limited, impeding the development of evidence-based mental health interventions using the capabilities of readily accessible smartphones. In this work, an initial attempt was made to address this in the context of Canadian undergraduates, where mental health remains an issue of concern for both students and administrators.

Acknowledgments

This work was supported in part by the University of Western Ontario's Interdisciplinary Development Initiative. CB thanks Ms Arlene Fleischhauer RN for administrative support of this work. CB thanks Dr Alan Salmoni, Mr Brad Matushewski, Dr Matthew Heath, and Dr Eva Pila for their support of this work in its early stages. CB also thanks Dr Lorie Donelle for suggestions to the work.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Smart Healthy Campus questionnaire.
[DOCX File , 95 KB - formative_v5i10e29160_app1.docx]

References

1. Fried R, Karmali S, Irwin J, Gable F, Salmoni A. Making the grade: mentors' perspectives of a course-based, smart, healthy campus pilot project for building mental health resiliency through mentorship and physical activity. Int J Evid Based Coach Mentor 2018 Jan;16(2):84-98. [doi: 10.24384/000566]
2. American College Health Association. American College Health Association-National College Health Assessment Spring 2008 Reference Group Data Report (abridged): the American College Health Association. J Am Coll Health 2009;57(5):477-488. [doi: 10.3200/JACH.57.5.477-488] [Medline: 19254888]
3. In it together: taking action on student mental health. Council of Ontario Universities, Toronto. 2017. URL: https://cou.ca/wp-content/uploads/2017/11/In-It-Together-PSE-Mental-Health-Action-Plan.pdf [accessed 2021-09-14]
4. Jacobsen WC, Forste R. The wired generation: academic and social outcomes of electronic media use among university students. Cyberpsychol Behav Soc Netw 2011 May;14(5):275-280. [doi: 10.1089/cyber.2010.0135] [Medline: 20961220]
5. Mohr DC, Zhang M, Schueller SM. Personal sensing: understanding mental health using ubiquitous sensors and machine learning. Annu Rev Clin Psychol 2017 May 08;13:23-47 [FREE Full text] [doi: 10.1146/annurev-clinpsy-032816-044949] [Medline: 28375728]
6. Wang R, Chen F, Chen Z, Li T, Harari G, Tignor S, et al. Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In: Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2014 Presented at: UbiComp '14: The 2014 ACM Conference on Ubiquitous Computing; Sep 13-17, 2014; Seattle Washington p. 3-14. [doi: 10.1145/2632048.2632054]
7. Grégoire S, Chénier C, Doucerain M, Lachance L, Shankland R. Ecological momentary assessment of stress, well-being, and psychological flexibility among college and university students during acceptance and commitment therapy. Can J Behav Sci / Revue canadienne des sciences du comportement 2020 Jul;52(3):231-243. [doi: 10.1037/cbs0000175]
8. Manwell LA, Barbic SP, Roberts K, Durisko Z, Lee C, Ware E, et al. What is mental health? Evidence towards a new definition from a mixed methods multidisciplinary international survey. BMJ Open 2015 Jun 02;5(6):e007079 [FREE Full text] [doi: 10.1136/bmjopen-2014-007079] [Medline: 26038353]
9. Chisholm D. Investing in mental health. Easter Mediterr Health J 2015 Jul 01;21(7):531-534 [FREE Full text] [doi: 10.26719/2015.21.7.531]
10. Hysenbegasi A, Hass S, Rowland C. The impact of depression on the academic productivity of university students. J Ment Health Policy Econ 2005 Sep;8(3):145-151. [Medline: 16278502]
11. Balloo K, Pauli R, Worrell M. Undergraduates' personal circumstances, expectations and reasons for attending university. Stud High Educ 2015 Oct 27;42(8):1373-1384. [doi: 10.1080/03075079.2015.1099623]
12. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
13. World Values Survey (Longitudinal Multiple-Wave Documentation). World Values Survey. URL: http://www.worldvaluessurvey.org/WVSDocumentationWVL.jsp [accessed 2021-09-14]
14. Berwick DM, Murphy JM, Goldman PA, Ware JE, Barsky AJ, Weinstein MC. Performance of a five-item mental health screening test. Med Care 1991 Feb;29(2):169-176. [doi: 10.1097/00005650-199102000-00008] [Medline: 1994148]
15. Smith BW, Dalen J, Wiggins K, Tooley E, Christopher P, Bernard J. The brief resilience scale: assessing the ability to bounce back. Int J Behav Med 2008;15(3):194-200. [doi: 10.1080/10705500802222972] [Medline: 18696313]
16. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
17. Henry JD, Crawford JR. The short-form version of the Depression Anxiety Stress Scales (DASS-21): construct validity and normative data in a large non-clinical sample. Br J Clin Psychol 2005 Jun;44(Pt 2):227-239. [doi: 10.1348/014466505X29657] [Medline: 16004657]

18. Mashek D, Cannaday LW, Tangney JP. Inclusion of community in self scale: a single-item pictorial measure of community connectedness. J Community Psychol 2007 Mar;35(2):257-275. [doi: 10.1002/jcop.20146]
19. Milton K, Bull FC, Bauman A. Reliability and validity testing of a single-item physical activity measure. Br J Sports Med 2011 Mar;45(3):203-208 [FREE Full text] [doi: 10.1136/bjsm.2009.068395] [Medline: 20484314]
20. Larsen ME, Nicholas J, Christensen H. A systematic assessment of smartphone tools for suicide prevention. PLoS One 2016 Apr 13;11(4):e0152285 [FREE Full text] [doi: 10.1371/journal.pone.0152285] [Medline: 27073900]
21. Morrison LG, Hargood C, Pejovic V, Geraghty AW, Lloyd S, Goodman N, et al. The effect of timing and frequency of push notifications on usage of a smartphone-based stress management intervention: an exploratory trial. PLoS ONE 2017 Jan 3;12(1):e0169162. [doi: 10.1371/journal.pone.0169162]
22. Small M, Bailey-Davis L, Morgan N, Maggs J. Changes in eating and physical activity behaviors across seven semesters of college: living on or off campus matters. Health Educ Behav 2013 Aug;40(4):435-441 [FREE Full text] [doi: 10.1177/1090198112467801] [Medline: 23232092]
23. Chow PI, Fua K, Huang Y, Bonelli W, Xiong H, Barnes LE, et al. Using mobile sensing to test clinical models of depression, social anxiety, state affect, and social isolation among college students. J Med Internet Res 2017 Mar 03;19(3):e62 [FREE Full text] [doi: 10.2196/jmir.6820] [Medline: 28258049]
24. Wu X, Tao S, Zhang Y, Zhang S, Tao F. Low physical activity and high screen time can increase the risks of mental health problems and poor sleep quality among Chinese college students. PLoS One 2015 Mar 18;10(3):e0119607 [FREE Full text] [doi: 10.1371/journal.pone.0119607] [Medline: 25786030]

Abbreviations

CPU: central processing unit
EMA: ecological momentary assessment
EMAX1: Ecological Momentary Assessment Extensions 1st edition software
EMAX2: Ecological Momentary Assessment Extensions 2nd edition software
EMAX3: Ecological Momentary Assessment Extensions 3rd edition software
SHC: Smart Healthy Campus
SPE: Student Pandemic Experience
UWO: University of Western Ontario (Western University, Canada)
WHO: World Health Organization

Edited by G Eysenbach; submitted 28.03.21; peer-reviewed by K Cohen, C Stunden, J Alvarez Pitti; comments to author 28.06.21; revised version received 08.07.21; accepted 01.08.21; published 19.10.21.

Please cite as:
Brogly C, Shoemaker JK, Lizotte DJ, Kueper JK, Bauer M
A Mobile App to Identify Lifestyle Indicators Related to Undergraduate Mental Health (Smart Healthy Campus): Observational App-Based Ecological Momentary Assessment
JMIR Form Res 2021;5(10):e29160
URL: https://formative.jmir.org/2021/10/e29160
doi:10.2196/29160
PMID:34665145

Original Paper

Comparing a Multimedia Digital Informed Consent Tool With Traditional Paper-Based Methods: Randomized Controlled Trial

Fuad Abujarad[1], MSc, PhD; Peter Peduzzi[2], PhD; Sophia Mun[3], MPH; Kristina Carlson[1], BSc; Chelsea Edwards[1], BSc; James Dziura[1], PhD; Cynthia Brandt[1,4], MPH, MD; Sandra Alfano[5], PharmD; Geoffrey Chupp[5], MD

[1]Department of Emergency Medicine, School of Medicine, Yale University, New Haven, CT, United States

[2]Department of Biostatistics, School of Public Health, Yale University, New Haven, CT, United States

[3]Department of Chronic Disease Epidemiology, School of Public Health, Yale University, New Haven, CT, United States

[4]VA Connecticut, New Haven, CT, United States

[5]Department of Internal Medicine, School of Medicine, Yale University, New Haven, CT, United States

Corresponding Author:
Fuad Abujarad, MSc, PhD
Department of Emergency Medicine
School of Medicine
Yale University
464 Congress Ave, Suite 264-J
New Haven, CT, 06519
United States
Phone: 1 (203)928 9259
Email: Fuad.Abujarad@yale.edu

Abstract

Background: The traditional informed consent (IC) process rarely emphasizes research participants' comprehension of medical information, leaving them vulnerable to unknown risks and consequences associated with procedures or studies.

Objective: This paper explores how we evaluated the feasibility of a digital health tool called *Virtual Multimedia Interactive Informed Consent* (VIC) for advancing the IC process and compared the results with traditional paper-based methods of IC.

Methods: Using digital health and web-based coaching, we developed the VIC tool that uses multimedia and other digital features to improve the current IC process. The tool was developed on the basis of the user-centered design process and Mayer's cognitive theory of multimedia learning. This study is a randomized controlled trial that compares the feasibility of VIC with standard paper consent to understand the impact of interactive digital consent. Participants were recruited from the Winchester Chest Clinic at Yale New Haven Hospital in New Haven, Connecticut, and healthy individuals were recruited from the community using fliers. In this coordinator-assisted trial, participants were randomized to complete the IC process using VIC on the iPad or with traditional paper consent. The study was conducted at the Winchester Chest Clinic, and the outcomes were self-assessed through coordinator-administered questionnaires.

Results: A total of 50 participants were recruited in the study (VIC, n=25; paper, n=25). The participants in both groups had high comprehension. VIC participants reported higher satisfaction, higher perceived ease of use, higher ability to complete the consent independently, and shorter perceived time to complete the consent process.

Conclusions: The use of dynamic, interactive audiovisual elements in VIC may improve participants' satisfaction and facilitate the IC process. We believe that using VIC in an ongoing, real-world study rather than a hypothetical study improved the reliability of our findings, which demonstrates VIC's potential to improve research participants' comprehension and the overall process of IC.

Trial Registration: ClinicalTrials.gov NCT02537886; https://clinicaltrials.gov/ct2/show/NCT02537886

(JMIR Form Res 2021;5(10):e20458) doi:10.2196/20458

KEYWORDS

digital consent; digital health; e-consent; informed consent; mobile phone

Introduction

Background

Informed consent (IC) is essential for upholding ethical conduct in research and medical treatment. The goal of the IC process is to provide research participants with sufficient information about the proposed research so that the participants can make an autonomous decision regarding their health and well-being [1-4].

Despite the implications and importance of the IC process, the Joint Commission for Transforming Healthcare has reported that an estimated 60% to 70% of individuals do not read or understand the information contained in the consent form, and 44% of the participants signing the IC documents do not understand the nature of the proposed procedure [5]. Although many providers have opted for electronic IC in an attempt to mitigate these issues, this method usually results in a mere electronic version of the standard paper-based form and does not address participant comprehension [6]. This lack of sufficient information and participant comprehension in the IC process negatively affects participant safety [7].

Related Works

Several studies have compared the efficacy of digital innovations in IC versus traditional paper consent. A systematic review conducted in 2019 to compare the 2 methods in published studies between 2012 and 2018 concluded that 67% of the included studies reported a positive effect on at least one of the studied outcomes. The efficacy of innovative interventions appeared high for interactive multimedia, with a positive effect on participants' comprehension and no negative effect on satisfaction or participation [8].

Another study that compared different methods of providing the same information in different formats to participants found that presenting plain text to participants with only audio narration proved to be the least effective when compared with other methods such as animated video or comics [9].

Several digital platforms have been developed to improve the consent process, and although their findings have been shown to be more effective than paper IC, either in a real study or a hypothetical study, some of these systems lacked 1 or more of the features that make Virtual Multimedia Interactive Informed Consent (VIC) unique [10-16]. These features include the use of avatars, supporting other languages, multimedia support, text-to-speech, quizzes and surveys, teach-back technique, accessibility among blind and deaf people, and electronic signature features [17].

Objective

Existing research suggests that the use of digital health interventions, such as virtual coaching and mobile apps, along with interactive audio and visual elements in a participant-centered IC, can increase the participant's interest and retention [3,18-21]. We applied these principles in the design, development, refinement, and testing of our web-based digital IC tool, *VIC* [17], and compared the results of our feasibility study with those of traditional paper consent methods.

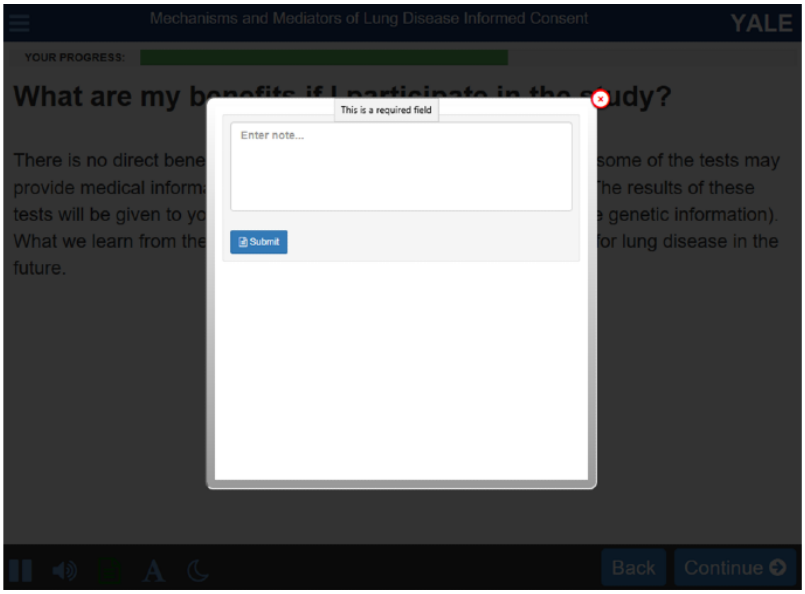
VIC Tool

The initial concept of the VIC tool was developed on the basis of previous work, literature findings in IC research, participant input, and subject matter expert interviews [17]. The theoretical framework of VIC is based on Mayer's cognitive theory of multimedia learning, and the principle that the use of multimedia in the presentation of the IC process will improve participant comprehension [22-25]. We adopted the user-centered design approach to design and develop a fully functional tool. Before using our tool, we conducted a usability study to evaluate user acceptability and satisfaction for the biorepository research study. Our tool uses virtual coaching with automated text-to-speech translation to conduct a brief and virtual interview with participants via tablet computers. VIC also features a comprehensive multimedia library (eg, video clips, animations, and presentations) to explain the risks, benefits, and alternatives of the proposed treatment or clinical study to enhance participants' awareness [17].

Our tool presents IC materials to participants with the option of going back and forth through each section as well as the ability to click on links within the sections, if desired, to drill down for more information. In addition, the VIC tool has an option to assess participant comprehension with automated quizzes, which can help emphasize the information presented. VIC provides many features and functions, including internet access to the consent, retrievable electronic records of IC, electronic signatures, and the potential for seamless integration with the electronic health record (Multimedia Appendix 1 and Figure 1). Moreover, VIC includes extensive security strategies that maintain the confidentiality and privacy of both participants and clinical information. It also provides access to the IC content via the internet before, during, and after the study or procedure, allowing the participants to benefit from supplemental resources as well [17].

Although more than half of the feasibility studies in the existing literature use hypothetical scenarios to test enhanced IC, we believe that the use of an actual research study improves the validity, accuracy, and reliability of the results [26,27]. This study is a randomized controlled trial to test VIC with participants involved in a real-world study and evaluated the tool's feasibility and utility compared with the standard, paper-based IC.

Figure 1. Virtual Multimedia Interactive Informed Consent features screenshot.



Methods

Trial Design

This study was a randomized controlled trial to test the feasibility of VIC in an ongoing, real-world biorepository research study titled "Yale Center for Asthma and Airway Disease Mechanisms and Mediators of Chronic Lung Disease Study" (GenEx 2.0). GenEx 2.0, the parent study for VIC, evaluated the pathophysiology and heterogeneity of airway disease in participants using several procedures, including a coordinator-administered, self-assessed questionnaire, lung function testing, blood draw, and hypertonic saline–induced sputum induction at the Winchester Chest Clinic (WCC). A detailed IC document was required so that the participants understood not only the risks and benefits of the procedures but also the ramifications of contributing biological and genetic samples to the GenEx 2.0 study.

Using the parent study's existing infrastructure and participants, the VIC trial recruited individuals and randomly assigned them to receive either the existing GenEx 2.0 paper consent document (control arm) or consent on the tablet through the VIC tool (intervention arm). Of the eligible participants, the allocation ratio for each group was 1:1. For both arms, the study coordinator responded to any questions from the participant regarding the IC process. After the consenting process, participants from both groups started the GenEx 2.0, which took an average of 2 hours to complete. The outcomes for the VIC trial were self-assessed using coordinator-administrated questionnaires at the WCC. Immediately following the GenEx 2.0, we surveyed the participants regarding their comprehension and satisfaction of the study to evaluate the effects of the VIC intervention.

Participants

The GenEx 2.0 study recruited participants with lung disease from the WCC at Yale New Haven Hospital in New Haven (CT) and healthy individuals from the community using fliers. We approached the participants who were considered for the parent study and asked if they were interested in participating in the VIC trial in addition to their participation in GenEx 2.0. For VIC participants, the trial included web-based components in addition to face-to-face components for the survey, or was conducted primarily face-to-face depending on the arm to which the participant was assigned.

The participants were eligible for both GenEx 2.0 and the VIC trial if they (1) spoke English, (2) were older than 21 years, (3) provided an email address, and (4) were willing to use an iPad. Computer literacy was not required for eligibility. The participants were excluded from the GenEx 2.0 study for (1) having a smoking history of more than 10 packs a year, (2) being active smokers within the past year, or (3) having other chronic lung disease or asthma variants. In addition, GenEx 2.0 participants were excluded from participation in the VIC trial if they were (1) not able to safely undergo the studies required for participation, (2) were unable to read or understand English, (3) refused to participate, or (4) had participated in the GenEx 2.0 trial in the past.

No participants withdrew during the study period from the VIC trial.

Randomization

Eligible and consenting participants were randomized to receive IC through standard paper consent or digital consent through the VIC tool. Because of the small sample size, we used the method of minimization [27] first to achieve balance on the following demographic characteristics—gender, race, education, employment type, marital status, household income, and technology confidence (Table 1). A computer algorithm belonging to the VIC back-end system maintained a record of all enrolled participants and automatically generated the randomization sequence after minimization. The study coordinator who handled enrollment would select a button in the VIC back-end system to prompt the randomization process, which would then display to which arm (control or intervention) the participant was randomized.

Table 1. Participant characteristics by method of informed consent administration (N=50).

Characteristic	VIC[a] (n=25)	Paper (n=25)
Age (years), mean (SD)	47.1 (15.3)	37.7 (14.7)
Gender, n (%)		
Male	9 (36)	10 (40)
Female	16 (64)	15 (60)
Race, n (%)		
White	17 (68)	15 (60)
Black or African American	6 (24)	6 (24)
Native American or American Indian	0 (0)	0 (0)
Asian or Pacific Islander	2 (8)	3 (12)
Other	0 (0)	1 (4)
Ethnicity, n (%)		
Hispanic or Latino	1 (4)	4 (16)
Non-Hispanic or Latino	24 (96)	21 (84)
Employment, n (%)		
Full-time	10 (40)	10 (40)
Part-time	3 (12)	5 (20)
Not employed	12 (48)	10 (40)
Education, n (%)		
High school graduate or GED[b]	7 (28)	6 (24)
Some college or associate's degree	6 (24)	5 (20)
College degree (bachelor's program)	5 (20)	7 (28)
Graduate or professional degree	7 (28)	6 (24)
Other	0 (0)	1 (4)
Household income before tax (US $), n (%)		
<50,000	15 (60)	19 (76)
50,000-99,999	4 (16)	4 (16)
≥100,000	6 (24)	2 (8)
Marital status, n (%)		
Single or widowed	14 (56)	17 (68)
Married or cohabitating	8 (32)	7 (28)
Divorced	3 (12)	1 (4)
Device use, n (%)		
Use a smartphone	20 (80)	22 (88)
Use a PC	20 (80)	18 (72)
Use a tablet	10 (62.5)	12 (48)
Confidence in using new technology, mean (SD)		
0: not confident and 10: very confident	7.8 (2.5)	7.7 (2.5)

[a]VIC: Virtual Multimedia Interactive Informed Consent.

[b]GED: General Educational Development.

Study Procedure

Individuals who were interested in and eligible for the VIC study consented to participate in the study and completed an initial demographic survey. The participants were then randomized and scheduled for their GenEx 2.0 study visit, which usually occurred within 1 to 2 weeks of enrollment in the VIC study. If assigned to paper consent for the GenEx 2.0 trial, the individual received a copy of the paper consent in the mail to review before the study visit. Because VIC is a web-based application that provides access to the consent before, during, and after the study visit, the individuals who were assigned to the VIC received an email with a link to the remote web-based version of the VIC consent session. This remote "review" link allowed VIC participants to preview what would be seen on the iPad during the study visit, but did not give access to the final signature needed to complete the consent process. Only VIC-assigned participants had access to the application itself during this time. If VIC participants had questions or concerns, they could leave comments in the "Notes" section of each slide, which would be addressed during the study visit. Remote sessions were separately tracked in the system to differentiate between the remote and study visit sessions. 10 of 25 participants accessed the remote session before their study visit.

During the GenEx 2.0 study visit, individuals completed the consent process according to the arm to which they were assigned. Individuals assigned to the control arm (traditional paper consent) completed the consent with a study coordinator who explained each section of the consent by reading it out loud and answering any questions during the process. The individual then signed the consent form if he or she wanted to participate in GenEx 2.0. Participants were not blinded, as it was not possible for the purposes of this study, and the study coordinator was aware of which group each participant belonged to as they oversaw both groups separately. The participants were also aware that the VIC was the comparator of interest for this trial.

Individuals assigned to the intervention arm (VIC tablet-based consent) completed the visit with a study coordinator who provided the individual with an iPad along with disposable headphones so that the individual could listen to the audio instructions comfortably. The individual would then go through the consent process alone and sign the consent on the iPad at the end if they were interested in participating in the GenEx 2.0 study. The format of the VIC process on the iPad allowed for the presentation of content to be displayed 1 section at a time with a "Continue" and "Back" button that participants could press to move forward or backward. For text-only sections, no more than approximately 90 words at a time were displayed, with an average of 76.6 words per slide overall. Some sections were transformed into interactive multimedia components, including animated videos that explained study procedures (videos specific to the GenEx 2.0 trial included demonstrations of blood draw and sputum collection) as well as videos specific to privacy and withdrawal information. In addition, these multimedia components also allowed interactivity with a simple menu that could pause, play, rewind, mute, and enable closed captioning if needed. Some sections were followed by interactive quizzes that emphasized key information to enhance participant comprehension, which would give automatic feedback to the participant on the answer and allow them to go back to the key section and revisit the material, or move forward, regardless of their answer. This method did not inhibit the participant, but rather encouraged the active retention of the material. The participants were able to either continue through each section (introduction, study procedures, risks, privacy, withdrawal, and so on) in order, or access a menu that allowed the participant to view any section of the consent in the order they wished. The VIC required that the participant views each section of the study's IC before allowing access to the signature portion of the tool. Once the IC was signed, the tool then ended and emailed a copy of the consent with the signature to the participant's enrolled email. Once the consents were collected, participants from both groups started the GenEx 2.0 study procedures.

After completing the GenEx 2.0 study session, the participants who enrolled in the VIC study from each arm were provided with a paper exit survey that asked questions about the feasibility of their respective IC process. After completing the survey, an incentive of US $60 was provided to thank them for their time in completing both the VIC trial and the GenEx 2.0 parent study. This incentive was the same for those who did not enroll in the VIC trial and completed only the GenEx 2.0 parent study.

Survey Assessment of Participant Comprehension and Satisfaction

The exit survey provided at the end of the GenEx 2.0 study procedures was designed to assess comprehension and satisfaction of the participants from each IC method (Multimedia Appendix 2). It included 13 comprehension questions that were structured as multiple-choice questions and were based on the validated Health Information Technology Usability Evaluation Scale and quality of IC surveys [26,28,29]. These comprehension questions measured six basic components of the IC—(1) why we asked individuals to participate in the GenEx 2.0 study, (2) the risks and benefits associated with the study, (3) their rights as participants, (4) whom to contact with questions or problems regarding the study, (5) study-specific procedures, and (6) coverage for potential study-related injuries.

The remaining 12 survey questions were administered using a 7-point Likert scale. Of these 12 questions, 4 questions assessed the participants' self-assessed understanding of the IC general concepts, with scores ranging from 1 (I did not understand this at all) to 7 (I understood this very well); 3 questions assessed satisfaction with the IC process, with scores ranging from 1 (Very dissatisfied) to 7 (Very satisfied); 1 question assessed the perceived length of the IC process, with scores ranging from 1 (Very long) to 7 (Very short); 1 question assessed the perceived difficulty in completing the IC process, with scores ranging from 1 (Very difficult) to 7 (Very easy); 2 questions assessed the likelihood of participating in future clinical trials, with scores ranging from 1 (Very unlikely) to 7 (Very likely); and 1 question assessed the importance of the IC process in the participant's decision to ultimately participate in the trial, with scores ranging from 1 (Not important at all) to 7 (Very important). Participants were also asked to provide an estimate of their perceived time (in minutes) to complete the IC process.

Study Outcomes

We measured six outcomes to assess the feasibility of VIC compared with the standard paper consent in a real-world clinical research study, which were recorded through a paper survey after completing the IC process. The study outcomes included each participant's (1) comprehension of the GenEx 2.0 study IC content, measured through a 13-question comprehension quiz; (2) satisfaction with the IC process, ranked on a 7-point Likert scale; (3) perceived time required to complete the IC process; (4) perceived ease of the IC process; (5) perceived likelihood of participating in future clinical trials; and (6) perception of the importance of IC in the decision to participate in clinical trials.

Data Analysis Plan

Data were analyzed according to intention-to-treat analysis (ie, all participants were analyzed as randomized). Baseline data are reported as mean and SD for continuous data and as counts and percentages of total for discrete data. The 13 comprehension questions were summarized as mean (SD) and proportions (SE). Differences in the mean values of correct answers between VIC and paper consent were presented with 95% CIs. The proportion of correct answers for each of the 13 individual questions was analyzed using risk ratios (VIC relative to paper) with 95% CIs. Likert scale data for each treatment group were summarized as mean (SD), as well as the perceived time to complete the IC process, and differences in the distribution between VIC and paper were tested using the Cochran–Mantel–Haenszel statistic. One participant in each treatment group reported 120 minutes as the perceived time, which were likely outliers and were removed from the mean calculations. Analyses were performed using SAS (version 9.4; SAS Institute).

Ethics and Security

The VIC trial protocol was approved by the Yale University Institutional Review Board. The research team included a chair of the institutional review board, who assisted with human rights perspectives. We anticipated potential risks with study participants consenting to a study via a digital medium and took multiple precautions to ensure comprehension of the digital IC for the GenEx 2.0 trial. The tool was tested with multiple users and subject matter experts before the study to determine its usability. The research coordinator had the ability to review the quiz results of the participant, as well as the time taken to read each section of the consent. The research coordinator evaluated each participant's quiz scores and the time spent going through the consent to determine if the participant was properly informed. If the research coordinator did not feel the participant understood the nature of the study, the coordinator had the ability to withdraw the consent.

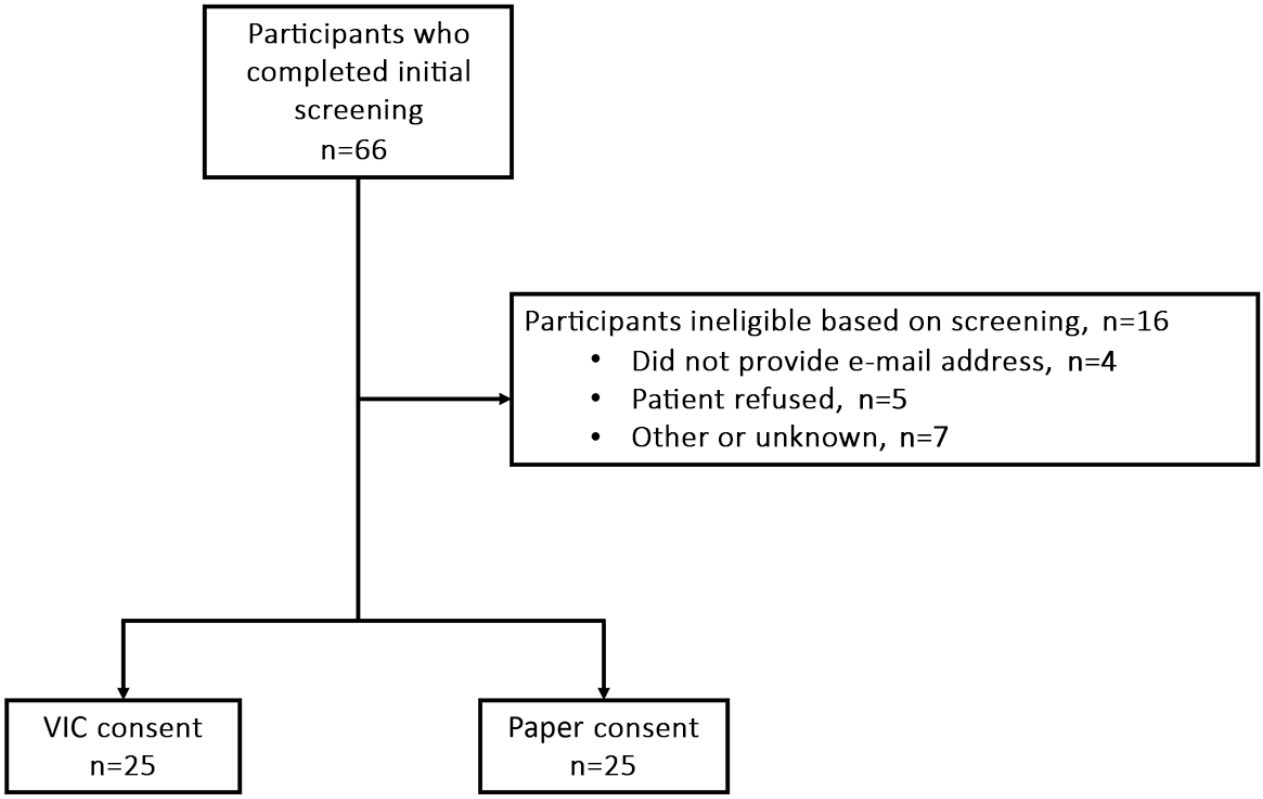
The VIC tool was hosted on a secure server located on the Yale University network and only accessible to those with explicit administrator access via the Yale University network through Yale's Central Authentication System. Databases storing participant data were also maintained on the Yale University network and hosted on separate servers from the tool itself to increase security measures. The databases were backed up and stored securely in a separate location on the Yale University network with access granted only to the system administrator and principal investigator. As with any digital application, we planned and anticipated common issues such as bugs, and although no bugs occurred during the trial itself, the developer was available to maintain the system and ensure stability of the application.

Results

Eligibility

A total of 91 individuals were approached for the study, of whom 25 were ineligible because they had already participated in GenEx 2.0. The remaining 66 individuals completed the initial screening questionnaire, and of them, 16 participants were deemed ineligible because they either did not provide an email address (n=4), refused to participate (n=5), or other unknown reasons (n=7). A total of 50 individuals were ultimately enrolled in the VIC trial, and 25 were randomized to the VIC intervention arm and 25 to the paper IC control arm (Figure 2).

Figure 2. Flow diagram of participant enrollment in the Virtual Multimedia Interactive Informed Consent trial CONSORT (Consolidated Standards of Reporting Trials) diagram. VIC: Virtual Multimedia Interactive Informed Consent.



Demographics

The mean age of the participants enrolled was higher in the VIC arm (47.0, SD 15.3 years) than in the paper arm (37.7, SD 14.7 years); with a mean difference 9.4 (95% CI 0.9-17.9). Overall, other demographic characteristics, such as race, ethnicity, household income, and relationship status, were comparable between both arms (Table 1). Regarding employment, 52% (13/25) of the VIC arm participants and 60% (15/25) of the paper consent group participants were employed at least part-time; education level was similar across both arms (Table 1). Most participants in both arms reported using or having access to a smartphone (20/25, 80% VIC arm, 22/25, 88% paper arm), less than half reported using or having access to tablets (10/25, 40% VIC arm vs 12/25, 48% paper arm), and most used or had access to PCs (20/25, 80% VIC arm vs 18/25, 72% paper arm). When asked about confidence in using new technology, participants in both arms reported similar levels of confidence on an 11-point Likert scale with 0 being “Not confident at all” and 10 being “Very confident” (mean 7.8 VIC vs 7.7 paper).

Survey Results

Overall, participant comprehension was high in both the IC process groups. For the comprehension outcome (Table 2), VIC participants scored a mean of 11 correct answers of 13 compared with the paper IC group mean of 10.6 correct answers; the mean difference in the number of correct answers between VIC and paper group was 0.4 (95% CI −0.5 to 1.3). For each of the 13 individual comprehension survey questions, the proportion of correct responses was generally comparable for VIC and paper group with risk ratios not different from one (Table 2). However, VIC participants appeared to have better knowledge about the use of their personal health information (PHI) and study withdrawal (risk ratios>1.20), which were sections in the tablet-based tool formatted uniquely as animated videos rather than plain text.

Table 2. Proportion of correct responses on the Participant Comprehension Survey by assigned study arm (N=50).

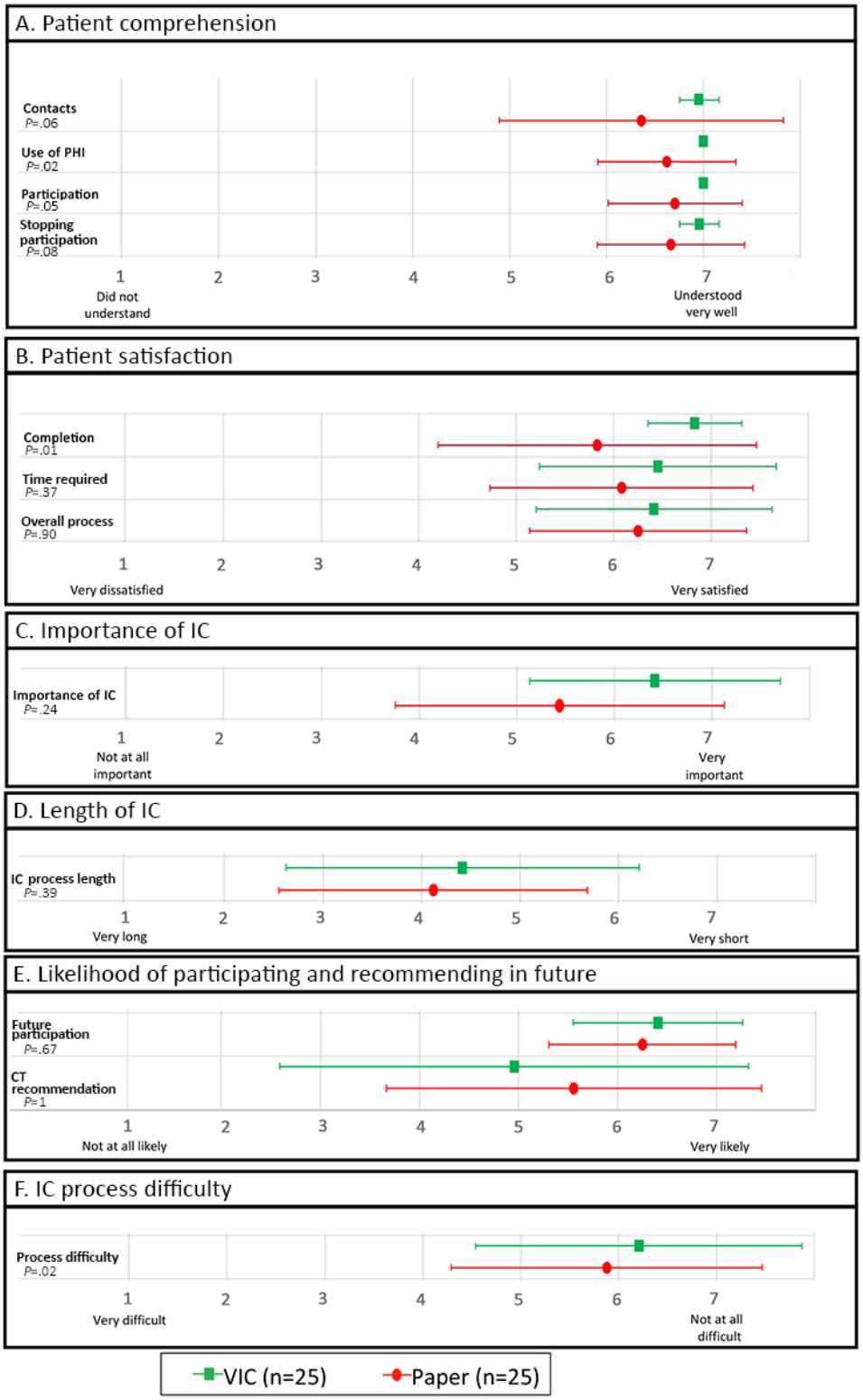
Number	Question	Participants with correct answers, n (%)		VIC[a] vs paper, risk ratio (95% CI)
		VIC arm (n=25)	Paper arm (n=25)	
1	Why did we ask you to participate in this study?	18 (72)	20 (80)	0.90 (0.66-1.23)
2	Which of the following are benefits of this study?	18 (72)	17 (68)	1.06 (0.74-1.52)
3	Which of the following procedures is part of the study?	25 (100)	24 (96)	1.04 (0.96-1.13)
4	Which of the following is true if you choose to participate?	23 (92)	24 (96)	0.96 (0.83-1.10)
5	Which of the following is an expected risk from participating in the study?	22 (88)	20 (80)	1.10 (0.86-1.40)
6	While you are in this research study, what will happen to your personal health information?	24 (96)	19 (76)	1.26 (1.00-1.60)
7	Which of the following statement is true about your participation in the study?	20 (80)	18 (72)	1.11 (0.81-1.52)
8	How can you withdraw from this study?	24 (96)	20 (80)	1.20 (0.97-1.48)
9	Who can you call if you have questions about your rights as a participant in this study?	18 (72)	16 (64)	1.13 (0.77-1.65)
10	Your driver's license number will be collected for purposes of this study	25 (100)	25 (100)	1.00 (1.00-1.00)
11	To participate in this study, you need to provide blood sample	24 (96)	25 (100)	0.96 (0.89-1.04)
12	If you are injured while participating in this study, you or your insurance carrier will be expected to pay the costs of this treatment	9 (36)	12 (48)	1.23 (0.76-1.99)
13	If you participate in the procedures for this study, you will be paid for your time	25 (100)	25 (100)	1.00 (1.00-1.00)

[a]VIC: Virtual Multimedia Interactive Informed Consent.

The participants were also asked questions to ascertain their level of satisfaction and perception with the 2 consent processes (Figure 3). Overall, mean levels of satisfaction and perception on a 7-point Likert scale were higher with the VIC tool for every category except for the "Clinical Trial Recommendation" category. When participants were asked how well they understood whom to contact with any questions or concerns related to the research study, VIC participants had a higher mean score (7.0) than the paper consent participants (6.4) (P=.05).

Furthermore, on average, VIC participants had a slightly better understanding of who in the study had access to their PHI and what would happen if they chose to participate in the study (mean scores of 7 for both), although paper consent participants scored a mean of 6.6 and 6.7, respectively (P=.02 and P=.045, respectively). When asked about their level of understanding on how to withdraw from the study, VIC participants had a mean score of 7.0 versus 6.7 for paper consent participants. Furthermore, VIC participants reported a lower mean perceived time to complete the consent process, 12.9 (SD 7.6) minutes for VIC participants versus 16.6 (SD 9.7) minutes for paper consent participants; mean difference of −3.7 (95% CI −9.0 to 1.5) minutes. In terms of satisfaction with completion time, VIC participants had a higher mean score than paper consent participants (6.8 vs 5.8; P=.01). Regarding overall process difficulty, VIC participants scored the process as less difficult, with a mean score of 6.3 compared with paper consent participants with a mean score of 5.9 (P=.02).

Figure 3. Perceived participant comprehension and satisfaction by the type of informed consent (IC) based on a 7-point Likert scale of levels of comprehension, satisfaction, perceived importance, perceived length, participation in clinical trials, and IC process difficulty. Mean (SD) values are plotted in the figure along with Cochran–Mantel–Haenszel test *P* values. CT: clinical trial; IC: informed consent; PHI: personal health information; VIC: Virtual Multimedia Interactive Informed Consent.



Regarding satisfaction with the overall consent process (Figure 3), VIC participants reported an average satisfaction level of 6.4 (7 being "Very satisfied"), whereas paper consent participants had an average satisfaction level of 6.3. More specifically, VIC participants reported an average satisfaction level of 6.8 for their ability to complete the consent process on their own without any help from research staff versus 5.8 for paper consent participants. VIC participants also scored a slightly higher average of 6.4 for their satisfaction with the time required to complete the IC process, whereas paper consent participants scored an average of 6.1. Finally, participants in the VIC group found the process of completing the IC process to be easier than participants in the paper group (6.4 VIC vs 5.4 paper).

The VIC participants were also more likely to recommend that research studies in the future use their method of IC (ie, electronic) than participants from the paper consent group (6.4 VIC vs 5.9 paper). Both IC groups reported that they were very likely to participate in the future clinical trials (6.4 VIC vs 6.3 paper) and that the IC process was relatively important in their decision to participate in the clinical research study (5.0 VIC vs 5.7 paper).

Discussion

Principal Findings

The study was conducted to test the feasibility of the VIC digital consent tool that enhances the IC process with interactive content on tablets. We built a VIC tool with a reusable infrastructure that allows for integration into the IC process for future research studies and may improve the clinical workflow through a more efficient IC process. The VIC tool can assess participant comprehension through automated quizzes and self-tests, emphasize topics using multimedia, and allow individuals to view demos and presentations. The participants can also listen to comments and explanations, get customized information, click on links to drill down for more information, ask questions, receive answers, and rewind and replay audio and visual components as needed. The VIC tool also provides a retrievable electronic record of the IC document, including the electronic signature, which can be integrated into modern electronic health records. Although not performed for the purposes of this study, VIC can also be performed remotely and securely with the electronic signature feature, and on a PC or smartphone if needed.

Significance

These innovative, dynamic features enhanced the overall participant experience in the ongoing research study, GenEx 2.0, compared with the standard paper IC process. In our study, the VIC trial participants reported significantly better understanding of the use of their PHI and participation in the study and greater satisfaction in completing the IC process because of the shorter duration. Previous systematic reviews have suggested that enhanced consent forms and extended discussions are most effective in improving participant understanding, and our findings confirm this claim [3,28-32]. This study provides additional evidence and reinforces the findings of previous studies implying that using digital health may enhance participant comprehension.

Systematic reviews have also suggested that the integration of audiovisual elements into the IC process can improve participant recall with no adverse impact on satisfaction or anxiety [33], and we similarly found that the use of an iPad for VIC did not negatively affect participants' willingness to participate or their satisfaction with the process. For the purposes of this study, the VIC tool used 5 videos (with closed caption options) in place of text. These videos were moderately dispersed throughout the tool, and the VIC ultimately contained less text than the paper version of the consent, which may have aided in reducing consent fatigue. It is important to note that VIC is not simply an electronic version of the IC process, but contains interactive multimedia components that visually demonstrate study procedures and certain sections of the consent document not normally seen in the traditional methods. Some studies referring to electronic consent processes may simply be a plain text version of the paper IC consent that offers text-to-speech as an audio feature and thus differ from VIC in this way.

A particular strength of our study was the option of having the participants provide consent to participate in an actual ongoing parent study in which they were enrolled prospectively, rather than asking them to imagine that they would take part in a hypothetical study. The real-world aspect of the study ensured that we were actively collecting data instead of relying on retrospective analyses, which have inherent limitations [26,27].

One comprehension question that provided indication of poor content delivery in both arms was the question of whether participants believed that they or their insurance carriers would be expected to pay for the costs of the treatment if they were injured while participating in the study. In the VIC arm, 64% (16/25) of the participants answered incorrectly compared with 52% (13/25) of those who answered incorrectly in the paper arm. In contrast, the other 12 questions indicated moderate to high comprehension of the content in both arms. This may be a reflection of the difficulty conveying language typical to the consent form regarding this specific section, because our results indicate it is atypical to the positive trend.

Limitations

There are some limitations with regard to the nature of this study, which may have affected certain findings. One limitation of this study was that we did not have masked research staff administering the survey to participants after the IC process. The study coordinator for GenEx 2.0 was the same person who administered the paper-based and VIC IC processes and collected study surveys, which limits our control over observer bias. Another limitation was that we did not directly and independently measure the time to complete the IC and instead relied on the participant-reported perceived time to complete the IC. We also limited participation in this study to individuals who spoke English, had an email address, and were willing to use the iPad. Although most participants were confident in using technology, we believe that these conditions could have potentially limited the generalizability of the findings because 6% (4/66) of the screened individuals did not provide an email address. Although the GenEx 2.0 study topic difficulty level could be described as moderate with very few study procedures involved, VIC may perform significantly better with more complicated topics with the use of multimedia and dynamic features to reinforce information, such as the "teach-back" quiz feature. In addition, our sample size of 50 participants could have affected the ability to detect differences between VIC consent participants and paper consent participants, which we plan to address in future research regarding the tool.

Potential future research topics that would be important to consider and explore with the VIC tool would be expanding future research using VIC to include participants of other languages other than English, as well as population with various hearing and vision impairments. The integrated text-to-speech and other audiovisual components of VIC are innovative language integration tools that would make switching to consent

in other languages a very feasible task. Moreover, the tablet-based methods of VIC also offer various accessibility tools that can be considered to reach a more inclusive target population, and with VIC's primary feature prioritizing a customized level of information, we feel that there is great potential for future topics to maximize its usability. Although cost analysis was not performed for this study, we believe that in future studies, the cost of creating a dynamic, digital IC may lessen as the tools become scalable and feasible to others.

Our team plans to disseminate the VIC tool in collaboration with other institutes to facilitate the adaptation of digital consent platforms with the goal of creating a scalable, dynamic, and effective IC process.

Conclusions

This study found that the VIC tool is feasible when integrated into a real-world research study, and the use of multimedia and other interactive features via a tablet-based IC process led to greater satisfaction in delivering important content compared with the standard paper process. VIC participants reported a lower perceived time to complete the IC process and higher comprehension, as well as higher overall satisfaction compared with the participants in the control group. Our preliminary findings suggest that compared with the standard paper consent process, dynamic digital IC processes can enhance comprehension and satisfaction and transform the consent process for human-based research studies.

Acknowledgments

The authors would like to acknowledge Sheinea Doram and Judy Kaya, the 2 patients with asthma who participated in the research team as co-researchers. This publication was made possible by the Yale Clinical and Translational Science Award grant UL1TR000142 from the National Center for Advancing Translational Science and the National Institutes of Health. The authors would also like to acknowledge Mr Thi Nguyen for coding the Virtual Multimedia Interactive Informed Consent system.

This study was funded by grant R21HS023987 from the Agency for Healthcare Research and Quality.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Virtual Multimedia Interactive Informed Consent demo video.
[MP4 File (MP4 Video), 32077 KB - formative_v5i10e20458_app1.mp4]

Multimedia Appendix 2
Virtual Multimedia Interactive Informed Consent trial survey.
[PDF File (Adobe PDF File), 105 KB - formative_v5i10e20458_app2.pdf]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 409 KB - formative_v5i10e20458_app3.pdf]

References

1. Jefford M, Moore R. Improvement of informed consent and the quality of consent documents. Lancet Oncol 2008 May;9(5):485-493. [doi: 10.1016/S1470-2045(08)70128-1] [Medline: 18452859]
2. Paterick TJ, Carson GV, Allen MC, Paterick TE. Medical informed consent: general considerations for physicians. Mayo Clin Proc 2008 Mar;83(3):313-319. [doi: 10.4065/83.3.313] [Medline: 18315998]
3. Miller FG, Emanuel EJ. Quality-improvement research and informed consent. N Engl J Med 2008 Feb 21;358(8):765-767. [doi: 10.1056/NEJMp0800136] [Medline: 18287598]
4. Hartgerink BJ, McMullen P, McDonough JP, McCarthy EJ. A guide to understanding informed consent. CRNA 1998 Nov;9(4):128-134. [Medline: 9866487]
5. 'What Did the Doctor Say?' Improving health literacy to protect patient safety. The Joint Commission. 2007. URL: https://www.jointcommission.org/assets/1/18/improving_health_literacy.pdf [accessed 2020-01-01]
6. Friedlander JA, Loeben GS, Finnegan PK, Puma AE, Zhang X, de Zoeten EF, et al. A novel method to enhance informed consent: a prospective and randomised trial of form-based versus electronic assisted informed consent in paediatric endoscopy. J Med Ethics 2011 Apr;37(4):194-200. [doi: 10.1136/jme.2010.037622] [Medline: 21245476]
7. Lorell BH, Mikita JS, Anderson A, Hallinan ZP, Forrest A. Informed consent in clinical research: Consensus recommendations for reform identified by an expert interview panel. Clin Trials 2015 Dec;12(6):692-695 [FREE Full text] [doi: 10.1177/1740774515594362] [Medline: 26178662]

8. Gesualdo F, Daverio M, Palazzani L, Dimitriou D, Diez-Domingo J, Fons-Martinez J, et al. Digital tools in the informed consent process: a systematic review. BMC Med Ethics 2021 Feb 27;22(1):18 [FREE Full text] [doi: 10.1186/s12910-021-00585-8] [Medline: 33639926]
9. Kraft SA, Constantine M, Magnus D, Porter KM, Lee SS, Green M, et al. A randomized study of multimedia informational aids for research on medical practices: Implications for informed consent. Clin Trials 2017 Feb;14(1):94-102 [FREE Full text] [doi: 10.1177/1740774516669352] [Medline: 27625314]
10. Bobb MR, Van Heukelom PG, Faine BA, Ahmed A, Messerly JT, Bell G, et al. Telemedicine provides noninferior research informed consent for remote study enrollment: a randomized controlled trial. Acad Emerg Med 2016 Jul;23(7):759-765 [FREE Full text] [doi: 10.1111/acem.12966] [Medline: 26990899]
11. Ham DY, Choi WS, Song SH, Ahn Y, Park HK, Kim HG, et al. Prospective randomized controlled study on the efficacy of multimedia informed consent for patients scheduled to undergo green-light high-performance system photoselective vaporization of the prostate. World J Mens Health 2016 Apr;34(1):47-55 [FREE Full text] [doi: 10.5534/wjmh.2016.34.1.47] [Medline: 27169129]
12. Harmell AL, Palmer BW, Jeste DV. Preliminary study of a web-based tool for enhancing the informed consent process in schizophrenia research. Schizophr Res 2012 Nov;141(2-3):247-250 [FREE Full text] [doi: 10.1016/j.schres.2012.08.001] [Medline: 22939457]
13. Kim H, Bell E, Kim J, Sitapati A, Ramsdell J, Farcas C, et al. iCONCUR: informed consent for clinical data and bio-sample use for research. J Am Med Inform Assoc 2017 Mar 01;24(2):380-387 [FREE Full text] [doi: 10.1093/jamia/ocw115] [Medline: 27589942]
14. Rothwell E, Wong B, Rose NC, Anderson R, Fedor B, Stark LA, et al. A randomized controlled trial of an electronic informed consent process. J Empir Res Hum Res Ethics 2014 Dec;9(5):1-7 [FREE Full text] [doi: 10.1177/1556264614552627] [Medline: 25747685]
15. Rowbotham MC, Astin J, Greene K, Cummings SR. Interactive informed consent: randomized comparison with paper consents. PLoS One 2013;8(3):e58603 [FREE Full text] [doi: 10.1371/journal.pone.0058603] [Medline: 23484041]
16. Tait AR, Voepel-Lewis T, McGonegal M, Levine R. Evaluation of a prototype interactive consent program for pediatric clinical trials: a pilot study. J Am Med Inform Assoc 2012 Jun;19(e1):43-45 [FREE Full text] [doi: 10.1136/amiajnl-2011-000253] [Medline: 21803924]
17. Abujarad F, Alfano S, Bright TJ, Kannoth S, Grant N, Gueble M, et al. Building an informed consent tool starting with the patient: the patient-centered virtual multimedia interactive informed consent (VIC). AMIA Annu Symp Proc 2017;2017:374-383 [FREE Full text] [Medline: 29854101]
18. Selby JV, Beal AC, Frank L. The Patient-Centered Outcomes Research Institute (PCORI) national priorities for research and initial research agenda. J Am Med Assoc 2012 Apr 18;307(15):1583-1584. [doi: 10.1001/jama.2012.500] [Medline: 22511682]
19. Tate EB, Spruijt-Metz D, O'Reilly G, Jordan-Marsh M, Gotsis M, Pentz MA, et al. mHealth approaches to child obesity prevention: successes, unique challenges, and next directions. Transl Behav Med 2013 Dec;3(4):406-415 [FREE Full text] [doi: 10.1007/s13142-013-0222-3] [Medline: 24294329]
20. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, et al. Mobile health technology evaluation: the mHealth evidence workshop. Am J Prev Med 2013 Aug;45(2):228-236 [FREE Full text] [doi: 10.1016/j.amepre.2013.03.017] [Medline: 23867031]
21. Sreenivasan G. Does informed consent to research require comprehension? Lancet 2003 Dec 13;362(9400):2016-2018. [doi: 10.1016/S0140-6736(03)15025-8] [Medline: 14683665]
22. Mayer R. Cognitive theory and the design of multimedia instruction: an example of the two-way street between cognition and instruction. New Direct Teach Learn 2002 Apr 16;2002(89):55-71. [doi: 10.1002/tl.47]
23. Cornoiu A, Beischer AD, Donnan L, Graves S, de Steiger R. Multimedia patient education to assist the informed consent process for knee arthroscopy. ANZ J Surg 2011 Mar;81(3):176-180. [doi: 10.1111/j.1445-2197.2010.05487.x] [Medline: 21342392]
24. Huber J, Ihrig A, Yass M, Bruckner T, Peters T, Huber CG, et al. Multimedia support for improving preoperative patient education: a randomized controlled trial using the example of radical prostatectomy. Ann Surg Oncol 2013 Jan 1;20(1):15-23. [doi: 10.1245/s10434-012-2536-7] [Medline: 22851045]
25. Sobel RM, Paasche-Orlow MK, Waite KR, Rittner SS, Wilson EA, Wolf MS. Asthma 1-2-3: a low literacy multimedia tool to educate African American adults about asthma. J Community Health 2009 Aug;34(4):321-327. [doi: 10.1007/s10900-009-9153-9] [Medline: 19353250]
26. Warriner AH, Foster PJ, Mudano A, Wright NC, Melton ME, Sattui SE, et al. A pragmatic randomized trial comparing tablet computer informed consent to traditional paper-based methods for an osteoporosis study. Contemp Clin Trials Commun 2016 Aug 15;3:32-38 [FREE Full text] [doi: 10.1016/j.conctc.2016.02.003] [Medline: 29736454]
27. Nishimura A, Carey J, Erwin PJ, Tilburt JC, Murad MH, McCormick JB. Improving understanding in the research informed consent process: a systematic review of 54 interventions tested in randomized control trials. BMC Med Ethics 2013;14:28 [FREE Full text] [doi: 10.1186/1472-6939-14-28] [Medline: 23879694]

28. Yen P, Bakken S. Review of health information technology usability study methodologies. J Am Med Inform Assoc 2012 Jun;19(3):413-422 [FREE Full text] [doi: 10.1136/amiajnl-2010-000020] [Medline: 21828224]
29. Joffe S, Cook EF, Cleary PD, Clark JW, Weeks JC. Quality of informed consent: a new measure of understanding among research subjects. J Natl Cancer Inst 2001 Jan 17;93(2):139-147. [doi: 10.1093/jnci/93.2.139] [Medline: 11208884]
30. Farrell EH, Whistance RN, Phillips K, Morgan B, Savage K, Lewis V, et al. Systematic review and meta-analysis of audio-visual information aids for informed consent for invasive healthcare procedures in clinical practice. Patient Educ Couns 2014 Jan;94(1):20-32. [doi: 10.1016/j.pec.2013.08.019] [Medline: 24041712]
31. Graham P. Type of consent does not influence patient recall of serious potential radiation toxicity of adjuvant breast radiotherapy. Australas Radiol 2003 Dec;47(4):416-421. [doi: 10.1046/j.1440-1673.2003.01212.x] [Medline: 14641195]
32. Krumholz HM. Informed consent to promote patient-centered care. J Am Med Assoc 2010 Mar 24;303(12):1190-1191. [doi: 10.1001/jama.2010.309] [Medline: 20332406]
33. Kalish LA, Begg CB. Treatment allocation methods in clinical trials: a review. Stat Med 1985;4(2):129-144. [doi: 10.1002/sim.4780040204] [Medline: 3895341]

Abbreviations

IC: informed consent
PHI: personal health information
VIC: Virtual Multimedia Interactive Informed Consent
WCC: Winchester Chest Clinic

Edited by R Kukafka; submitted 19.05.20; peer-reviewed by T Muto, L Genaro, M Doerr, G Wadley, MDG Pimentel, J Li, F Palmieri; comments to author 25.09.20; revised version received 14.10.20; accepted 01.08.21; published 19.10.21.

Please cite as:
Abujarad F, Peduzzi P, Mun S, Carlson K, Edwards C, Dziura J, Brandt C, Alfano S, Chupp G
Comparing a Multimedia Digital Informed Consent Tool With Traditional Paper-Based Methods: Randomized Controlled Trial
JMIR Form Res 2021;5(10):e20458
URL: https://formative.jmir.org/2021/10/e20458
doi:10.2196/20458
PMID:34665142

Original Paper

Impact of App-Delivered Mindfulness Meditation on Functional Connectivity, Mental Health, and Sleep Disturbances Among Physician Assistant Students: Randomized, Wait-list Controlled Pilot Study

Jeremy L Smith[1], MSDS, PhD; Jason W Allen[1,2], MD, PhD; Carla I Haack[3], MD; Kathryn L Wehrmeyer[4], BS; Kayley G Alden[4], MD; Maha B Lund[5], DHSc, PA-C; Jennifer S Mascaro[4], PhD

[1]Department of Radiology and Imaging Sciences, Emory University, Atlanta, GA, United States

[2]Department of Neurology, Emory University, Atlanta, GA, United States

[3]Department of Surgery, Emory University, Atlanta, GA, United States

[4]Department of Family and Preventative Medicine, Emory University, Atlanta, GA, United States

[5]Physician Assistant Program, Department of Family and Preventative Medicine, Emory University, Atlanta, GA, United States

Corresponding Author:
Jeremy L Smith, MSDS, PhD
Department of Radiology and Imaging Sciences
Emory University
1364 Clifton Road Northeast
Atlanta, GA, 30322
United States
Phone: 1 404 989 0524
Email: jsmi304@emory.edu

Abstract

Background: Health care provider and trainee burnout results in substantial national and institutional costs and profound social effects. Identifying effective solutions and interventions to cultivate resilience among health care trainees is critical. Although less is known about the mental health needs of physician assistants (PAs) or PA students, accumulating research indicates that they experience similarly alarming rates of burnout, depression, and emotional exhaustion. Mobile app–delivered mindfulness meditation may be an effective part of salubrious programming to bolster long-term resilience and health among PA students.

Objective: This study aims to examine the impact of app-delivered mindfulness meditation on self-reported mental health symptoms among PA students. A secondary aim is to investigate changes in brain connectivity to identify neurobiological changes related to changes in mental health symptoms.

Methods: We recruited PA students enrolled in their third semester of PA school and used a longitudinal, randomized, wait-list–controlled design. Participants randomized to the mindfulness group were provided 1-year subscriptions to the *10% Happier* app, a consumer-based meditation app, and asked to practice every day for 8 weeks. Before randomization and again after completion of the 8-week program, all participants completed resting-state functional magnetic resonance imaging as well as self-report assessments of burnout, depression, anxiety, and sleep impairment. App use was acquired as a measure of mindfulness practice time.

Results: PA students randomized to the mindfulness group reported improvements in sleep impairment compared with those randomized to the wait-list control group (η_p^2=0.42; *P*=.01). Sleep impairment decreased significantly in the mindfulness group (19% reduction; *P*=.006) but not in the control group (1% reduction; *P*=.71). There were no other significant changes in mental health for those randomized to app-delivered mindfulness. Across all students, changes in sleep impairment were associated with increased resting-state functional connectivity between the medial prefrontal cortex (a component of the default mode network) and the superior temporal gyrus, as well as between areas important for working memory. Changes in connectivity predicted categorical conversion from impaired to nonimpaired sleep in the mindfulness group.

Conclusions: This pilot study is the first to examine app-based mindfulness for PA students' mental health and investigate the impact of mindfulness on PA students' brain function. These findings suggest that app-delivered mindfulness may be an effective

tool to improve sleep dysfunction and that it may be an important part of the programming necessary to reduce the epidemic of suffering among health profession trainees.

(JMIR Form Res 2021;5(10):e24208) doi:10.2196/24208

KEYWORDS

mindfulness; meditation; resting state; fMRI; connectivity; mobile phone

Introduction

Background

Although the high prevalence of depression among medical students and residents is well characterized and increasingly appreciated [1,2], little is known about the mental health of physician assistant (PA) students [3,4]. The research that has been conducted indicates that practicing PAs report high levels of burnout [5] and depression [6], and this is despite the fact that PAs often report high levels of job satisfaction [6]. Although very few studies have evaluated burnout or depression among PA students, one recent study found that almost 80% of PA students report high levels of emotional exhaustion, with almost that many expressing interest in interventions to improve their well-being [7]. Overall, there is a critical need to examine the mental health needs of PA students and evaluate interventions to bolster PA student resilience.

A substantial and growing body of research indicates that mindfulness meditation enhances well-being [8]; reduces anxiety and depression [9]; and optimizes immune signaling, stress responsivity, and cognitive function [10-12]. Moreover, mindfulness meditation has shown great promise for improving sleep disruption and insomnia symptoms [13-15], which may be beneficial to health professional trainees who often experience high rates of sleep dysfunction [16]. With its apparent efficacy, several clinical trials have examined whether mindfulness training improves well-being among health care trainees, including medical and nursing students and medical residents [17]. Accumulating research indicates that mindfulness reduces anxiety and depression and enhances well-being among health profession trainees, and a recent meta-analysis highlighted the potential efficacy of mind-body wellness programs such as mindfulness meditation for improving trainee well-being [18,19]. However, most studies examining mindfulness among trainees have examined time-intensive interventions that are prohibitive for most trainees. Although a recent study found that incorporating mindfulness into PA coursework increased self-reported well-being [20], very few studies have examined the impact of mindfulness training on the mental health of PA students.

Even fewer studies have examined the impact of mindfulness on trainee brain function. Growing evidence indicates that the health-relevant effects of mindfulness are mediated by alterations to the default mode network (DMN), the salience network (SN), and the systems involved in executive control, often referred to as the central executive network [21,22]. In addition, at least 3 studies indicate that mindfulness training [23,24] or dispositional mindfulness [25] are related to increased functional connectivity between the prefrontal cortex and the amygdala, generally interpreted as augmented emotion regulation via top-down control of the amygdala. Although these studies indicate that benefits from mindfulness meditation are related to changes in functional connectivity within and among these brain regions, to date, no studies have examined changes in functional connectivity related to improvements in well-being among health profession trainees.

Objective

Here, we use a longitudinal, randomized wait-list–controlled design to examine the impact of app-delivered mindfulness meditation on self-reported mental health symptoms among PA students. A second aim is to examine whether changes in mental health are associated with changes in brain connectivity, indexed using a whole-brain connectome and multivariate pattern analysis approach to query within- and between-network connectivity across the entire brain before and after mindfulness training. On the basis of prior studies, we hypothesize that app-delivered mindfulness would reduce burnout, depression, anxiety, and sleep impairment; that changes in mental health outcomes would be associated with changes in DMN, SN, and central executive network connectivity; and that changes in mental health and connectivity would be positively associated with practice time, indexed as app use.

Methods

Recruitment

Students enrolled in their third semester of PA school were recruited for the study just before the commencement of their clinical rotations. Students were recruited via in-person presentations held after their classes, and 16 students were enrolled in the study. Upon providing informed consent in accordance with the university's institutional review board, participants were randomly assigned (in Microsoft Excel, using the *randbetween* function) to either a mindfulness meditation intervention using the *10% Happier* app (*practitioners* group) or to a wait-list (*control* group). Study personnel were blinded to group randomization, except for 1 researcher who corresponded with the participants and was not involved in data collection or analysis.

PA students randomized to the practitioner group were asked to practice app-guided mindfulness meditation for approximately 12 minutes per day for 8 weeks. Before randomization and again after completion of the 8-week program, all participants completed self-report assessments and functional magnetic resonance imaging (fMRI), both described in detail below. A total of 2 participants were removed from the analysis for excessive movement during fMRI, resulting in a cohort of 50% (7/14) practitioners (6/7, 86% women) and 50% (7/14) controls (5/7, 71% women; Table 1). Participants were compensated US $100 for completing both assessments. This study was part of

a larger preregistered clinical trial (NCT03452670) that included the planned enrollment of several trainee and employee populations. For the larger trial, self-reported perceived incivility was the primary outcome measure and burnout, depression, and anxiety were secondary outcome measures. Enrollment for the larger study did not meet the planned recruitment goals, and this was an exploratory analysis of a subset of the enrolled participants. In addition to the secondary outcome measures, we included a measure of sleep impairment given the prevalence of sleep dysfunction among health profession trainees and the relationship between sleep dysfunction and depression [26].

Table 1. Demographic information (N=14).

Characteristics	App, n (%; n=7)	Wait-list, n (%; n=7)	Chi-square (*df*)	*P* value
Gender				
Female	6 (86)	5 (71)	0.4 (1, 14)	.52
Male	1 (14)	2 (29)	0.4 (1, 14)	.52
Relationship status				
Single	1 (14)	3 (43)	5.7 (3, 14)	.23
Divorced	0 (0)	1 (14)	5.7 (3, 14)	.23
In a relationship	4 (57)	2 (29)	5.7 (3, 14)	.23
Married	2 (29)	1 (14)	5.7 (3, 14)	.23
Race				
White	6 (86)	5 (71)	3.0 (3, 14)	.21
African American or Black individual	1 (14)	0	3.0 (3, 14)	.21
Asian	0 (0)	2 (29)	3.0 (3, 14)	.21
Other	0 (0)	0 (0)	3.0 (3, 14)	.21

Mindfulness Meditation Intervention

Participants randomized to the meditation group were provided with a 1-year subscription to the app *10% Happier* [27,28]. We chose *10% Happier* among a variety of meditation apps currently in the market given its marketing toward *fidgety skeptics* and its orientation toward practical applications of mindfulness [27], which we thought would appeal to PA students. For example, the app notes, "Just in case you're worried, meditation does not require a lot of the things people think it might. For example, you don't have to sit in a particular position. (Unless you want to, of course). You also don't have to light incense, chant, or believe in anything in particular. There's nothing to join, no special outfits to wear" [29]. *10% Happier* provided subscriptions for the study participants, suggestions on recommended content, and anonymized app use data. App use was acquired as the elapsed time (in seconds) that the app meditation modules were used by each person.

Students randomized to the meditation group were given the following instructions:

> *We would like for you to try to practice every day for 8 weeks, even if it is only for one minute. Based on mindfulness research and on the suggestions of the app developers, we would like you to try the following programs: "The Basics" and "Emotional Agility." If you are pressed for time and cannot do a module from these programs, please do the "One minute counts."*

The *Basics* program included 16 modules with didactic instructions and mindfulness practice time varying between 4:20 to 13:22 minutes (average meditation length: 9:44 minutes, SD 3:17 minutes), and it serves as an introduction to mindfulness meditation. The practices include mindfulness of the sensations of the breath, mindfulness toward sensations and experiences of the body, and mindfulness toward the contents of the mind. In addition, the meditations encourage the practitioner to use the skill of *mental noting* to label their mental contents. The *Emotional Agility* program included 15 modules with didactic instructions and mindfulness toward mental content (focusing on emotions); meditation practice in the *Emotional Agility* program varied from 11:00 to 13:00 minutes (average meditation length: 12:18 minutes, SD 0:40 minutes). The meditation practices in these modules included mindfulness toward the sensations of the body and breath, coupled with other practices aimed at cultivating awareness and understanding of emotions as mental contents and the nonjudgmental stance toward emotions, with a goal of optimizing the response to one's emotions.

Self-reported Measures

We measured incivility using the Incivility in Nursing Education—Revised Survey [30], which contains 24 items that ask students about behaviors they exhibited or witnessed in the past 12 months (eg, "students made rude gestures or nonverbal behaviors towards others"). Participants indicated how often these behaviors occurred by selecting from 1=never, 2=rarely (once or twice), 3=sometimes (approximately once per month), or 4=often (more than once per month). Items were summed according to instances of low-level (15 items; eg, "Expressing disinterest, boredom, or apathy about course content or subject matter") and high-level incivility (9 items; eg, "Making condescending or rude remarks toward others"). Scores were

averaged, such that the range was 1-4, with higher scores indicating more incivility exhibited or witnessed.

We measured burnout using the School Burnout Inventory [31], a 9-item survey asking students about how much burnout they have felt in the past month (eg, "I feel overwhelmed by my schoolwork"). Respondents indicated the degree to which they agreed with each statement on a scale of 1-6, where 1=completely disagree and 6=completely agree. Total scores ranged from 9-54, with higher scores indicating more burnout.

The Depression Anxiety and Stress Scale [32] is a 42-item survey asking about feelings of depression, anxiety, and stress that the respondent has experienced in the past week. Participants indicated the degree to which they agreed with each statement on a scale of 0-3, where 0=does not apply to me at all and 3=applied to me very much or most of the time. Each of the three subscales included 14 items. The depression subscale assessed general dysphoria, anhedonia, self-contempt, and hopelessness. A score of 0-9 indicated no depression, 10-13 indicated mild depression, 14-20 indicated moderate depression, 21-27 indicated severe depression, and scores ≥28 indicated extremely severe depression. The anxiety subscale assessed symptoms and subjective feelings related to acute autonomic arousal. A score of 0-7 indicated normal levels of anxiety, 8-9 indicates mild anxiety, 10-14 indicated moderate anxiety, 15-19 indicated severe anxiety, and scores ≥20 indicated extremely severe anxiety.

Sleep dysfunction was measured using the 8-item PROMIS (Patient-Reported Outcomes Measurement Information System) sleep-related impairment (SI) short form 8b, which assessed the frequency with which participants experienced alertness, sleepiness, tiredness, and functional impairments associated with sleep problems during waking hours (eg, "I had difficulty falling asleep") in the previous 7 days [33]. Items were scored on a 5-point scale, such that higher scores indicated more sleep impairment. Raw scores for each of the eight items were totaled and converted to standardized scores using conversion tables published on the PROMIS website [34].

To assess safety, participants were asked to report their agreement with a number of statements reflecting positive (eg, "I enjoyed using the app") and negative experiences (eg, "I found it very difficult to do the meditations") with the app. In addition, they were asked to report anything else *good*, *bad*, *or neutral* that they wanted us to know about the app.

Resting fMRI Image Preprocessing

Baseline and postprogram resting-state fMRI (rsfMRI) data were acquired on a 3T Siemens Prisma FIT (Siemens Healthineers; 8-minute multiband acquisition with 2 seconds repetition time, 2.97×2.97×2.00 mm voxels, 70° flip angle, and echo train length 37). All preprocessing and bivariate correlation (connectome) analyses were performed in the CONN Toolbox (v19c) under MATLAB (vR2019a) [35]. Standard preprocessing methods were applied to the rsfMRI and anatomical volumes in CONN, which wraps SPM8 [36,37] and aCompCor [38] noise source removal functions. It comprised slice timing, field map, and motion correction; coregistration and normalization between rsfMRI images, anatomical images, and Montreal Neurological Institute, standard stereotactic space; smoothing at a 5 mm filter width at half-maximum, which limits intersubject differences and increases signal-to-noise ratio), linear detrending, bandpass filtering at 8-90 mHz; and regression of the 6 motion parameters and their first-order derivatives, along with cerebrospinal fluid and white-matter signals [39-41], by a general linear model (GLM). Scans (repetition time intervals [TRs]) that exhibited motion or global signal change beyond a SD 1.5 IQR limit were marked as *invalid scans* and nulled for the purposes of the GLM (see the Quality Assurance and Quality Control document, Multimedia Appendix 1). Three subjects exhibited 1-8 TRs, out of 240 total TRs, with global signal change beyond this tolerance. These TRs were tagged as *invalid* and were not included in further analyses. We did not add the mean whole-brain signal as a regressor, as there is some evidence that doing so may artificially introduce negative correlations and that the aCompCor method, in combination with bandpass filtering and orthogonalization of motion parameters, is preferable to global signal regression [42]. All structural and *denoized* functional data, gray matter, white matter, and CSF masks were manually inspected to confirm registration validity. In addition, a Fisher (inverse hyperbolic tangent) transformation was applied to bivariate correlation measures before between-subjects analysis to ensure that the connectivity data conformed to the normality assumptions of the GLM (see the Quality Assurance and Quality Control document in the Multimedia Appendix 1).

Identification of Regions of Interest and Computation of Connectivity Matrices

Regions of interest (ROIs) were computed from preprocessed rsfMRI data using multivariate pattern analysis (Norman et al [43]). The first step of the multivariate pattern analysis procedure, as implemented [31] in the CONN Toolbox, comprised dimensionality reduction into 64 components by singular value decomposition for each subject and condition (group and visit). Subject and condition-specific correlation maps—of size (*number of subjects* × *number of conditions* × *number of voxels in each data set*) and comprising standardized (Fisher transformed) bivariate correlation coefficients between each pair of voxels—were then generated from each subject's reduced-dimensionality data set. Next, a set of four principal components, capturing approximately 95% of the between-subjects or between-conditions variance, was obtained from the aggregate matrix (ie, with all subjects and conditions combined or *stacked*). Finally, the first two of these four principal components were selected for further analysis and subjected to a standard two-way analysis of variance (ANOVA; *group* × *visit*) to determine whether the components modulated with either condition. Application of a statistical threshold based on Gaussian random field theory [44-46], which estimates error fields within an fMRI statistical map after smoothing with the FWHM kernel (described above), yielded a map of 87 clusters that ostensibly represented any between-group or between-visit differences in voxel-to-voxel functional connectivity across the brain with a cluster growth threshold of $P \leq .05$ (uncorrected) and a topological false discovery rate (FDR [47]) of $P_{FDR} \leq .05$. A total of 16 of these clusters were excluded because of their localization in white matter. Another 15 ROIs representing

components of the default mode, sensorimotor, visual, salience, dorsal attention, frontoparietal, language, and cerebellar networks, predefined in the CONN Toolbox, were also included, for a total of 102 ROIs (Table 2 and Figure S1 in Multimedia Appendix 2).

To facilitate testing of our hypothesized relationships among meditation practice time, changes in mental health outcomes, and brain network connectivity, mean signals were extracted from the 71 a posteriori and 15 a priori ROIs for each subject and condition and aggregated into baseline (visit 1) and eighth-week (visit 2 follow-up) ROI-to-ROI connectivity matrices, each of size (*number of subjects* × *number of ROIs* × *number of ROIs*). As with the voxel-wise correlation maps, these ROI-ROI connectivity matrices comprised standardized (Fisher transformed) bivariate correlation coefficients between each pair of ROIs. To simplify the within-group analysis, a *delta matrix*, Δconn, was also computed as $\Delta conn = conn_{8wks} - conn_{baseline}$. The baseline, eighth-week, and delta ROI-ROI connectivity matrices (Figure S2 in Multimedia Appendix 2) were leveraged in lieu of voxel-to-voxel connectivity matrices for all further analyses.

Table 2. List of the 71 gray matter regions of interest derived from multivariate pattern analysis, with predefined atlas-based regions of interest. Coordinates for region of interest centers of mass and peak voxels (where applicable) are provided in MNI152 standard coordinate space.

Type and CLUSTER_ID	Voxels	Center of mass (MNI[a])			Peak (MNI)			ATLAS ROI[b]
		x	y	z	x	y	z	
Brain stem								
bs1_bstem	33	–1.2	–29.3	–45.6	0	–30	–46	Brainstem gray matter
Basal ganglia								
caud1_Rput	126	20.1	27.7	4.8	24	18	4	R putamen/globus pallidus
caud2_Rput1	108	29.3	7.1	1.8	34	4	6	R middle insula/putamen
caud3_Rgp	45	22.3	–9.4	–3.1	22	–14	–4	R globus pallidus/thalamus/motor thalamus
caud4_Rput	36	31.3	–18.7	1.9	32	–16	0	R putamen
Cerebellar								
cb1_Rlob9	364	4.2	–61.5	–45.7	12	–62	–54	R cerebellum (VIII) lobule IX
cb2_Llob6	269	–21.7	–49.8	–46.7	–28	–48	–38	L cerebellum (VI)/prob. WM
cb3_Rlob8b	99	16.4	–46.9	–50.3	18	–46	–50	R cerebellum (IX) lobule VIIIb
cb4_Rlob7b	82	32	–74.9	–54.1	28	–76	–54	R cerebellum (VII) lobule VIIb
cb5_Rcrus2	52	10.5	–72.2	–24.7	6	–72	–28	R cerebellar crus 2/vermis
cb6_Lcrus2	39	–48.7	–56.8	–42.6	–56	–56	–44	L cerebellar crus 2/lobule VIIa crus I
Cingulate								
cing1_Rpcc	426	11	–47.4	20.8	10	–48	18	R precuneus/posterior cingulate/BA23v
cing2_Rmicg	21	18.3	–33.3	49.8	18	–34	50	R middle cingulate cortex/ventromedial BA5
Frontal								
f1_Lba9	408	–3.1	48.3	45	–12	54	42	L superior frontal gyrus/mid BA9
f10_Lba6r	33	–50.9	4.3	16.5	–52	4	16	L precentral gyrus/BA44/rostral BA6/caudoventrolateral BA6
f11_Rba6ba8	28	18.9	31.7	60.1	20	32	62	R superior frontal gyrus/superior BA6-BA8 transitional area
f12_Lba6ba8	27	–31.1	4	59.8	–30	4	60	L middle frontal gyrus/inferior BA6-BA8 transitional area
f2_Rba10p	271	18.1	57.5	–13.3	24	62	2	R posterior BA10/superior frontal gyrus
f3_Rba8v	207	47.6	10.6	44.5	48	12	48	R ventral BA8A/caudal middle frontal gyrus/IFJ
f4_Rba8dl	163	28.7	18.5	49.2	28	20	54	R middle frontal gyrus/dorsolateral BA8A
f5_Lba8dl	133	–20	30.5	50.2	–20	30	48	L dorsolateral BA8A/superior frontal gyrus
f6_Lsma	124	–15.2	7.9	53.7	–10	12	50	L SMA/supplementary and cingulate eye area/medial frontal gyrus
f7_Rba24d	90	1.9	–24.6	46.1	2	–22	50	R dorsal BA24/SMA/superiomedial BA4
f8_Rba10d	88	–17.5	62.6	7.9	–22	60	10	L superior frontal gyrus/dorsal BA10/BA46/area Fp1
f9_Lba44v	81	–53.4	17.9	2.8	–54	18	4	L ventral BA44/inferior frontal gyrus pars triangularis
Medial temporal								
mtl1_Lphg1	137	–15.7	–46.5	–10.4	–16	–42	–10	L parahippocampal area 1/subiculum/lingual gyrus/area TH
mtl2_Rphg	89	27	–35.9	–13.4	28	–44	–6	R parahippocampal gyrus/rostral lingual gyrus/ventromedial visual area 1

Type and CLUSTER_ID	Voxels	Center of mass (MNI[a])			Peak (MNI)			ATLAS ROI[b]
		x	y	z	x	y	z	
mtl3_Renrc	83	21.2	–26.6	–28.5	22	–28	–26	R entorhinal cortex/presubiculum/caudal BA35/36
mtl4_Rprc	50	29.6	–1.1	–34.4	30	–2	–34	R perirhinal/ectorhinal cortex/rostral BA36/BA35/parahippocampal gyrus
mtl5_Lamg	34	–25.4	–4.1	–21	–24	–4	–16	L piriform cortex/amygdala/laterobasal amygdala
mtl6_Lenrc	33	–19.2	–9.4	–33.7	–22	–8	–34	L entorhinal cortex/BA28/BA34/parahippocampal gyrus
mtl7_Rhc	23	30.3	–16.2	–18	30	–16	–14	R hippocampus/CA3 fields/CA3 fields
Occipital/visual								
o1_Rba39rd	287	34.7	–69.8	41.6	36	–72	38	R intraparietal area1/rostro-dorsal BA39/middle occipital gyrus
o2_Llop	208	–16.2	–101.6	–2.1	–14	–106	–8	L lateral occipital pole
o3_Rv2v3d	29	23.5	–98.9	10.7	24	–98	10	R superior occipital gyrus/visual area V2/dorsal visual area V3
o4_Lv2	25	–24.9	–50.5	0	–24	–50	0	L precuneus/prostriate area/lingual gyrus/visual area V2
o5_Rba37mv	20	22.3	–42.3	–17.2	22	–44	–18	R ventral visual complex/area FG3/medioventral BA37
Opercular								
op1_Rro	60	53.9	2.4	2.9	54	4	6	R rolandic operculum/rostral BA6/BA43
Orbitofrontal								
orb1_Rba11	208	18.7	18.5	–22.8	20	30	–16	R superior orbital gyrus/lateral BA11/area Fo3
orb2_Lba10r14m	94	–5.3	41.3	–12.7	–6	44	–16	L rostral BA10/medial BA14/gyrus rectus
orb3_Rifgpo	89	32.2	22.9	–31.4	32	22	–26	R inferior frontal gyrus pars orbitalis/superior BA47/lateral BA11/area Fo3
orb4_L47l	76	–49.5	29.6	–5.9	–46	28	–6	L lateral BA47/inferior frontal gyrus pars orbitalis
orb5_Rba10v	49	4.2	59.6	–11.3	6	60	–10	R ventral BA10/medial BA11/middle orbital gyrus
orb6_Lpofc	36	–14	14.1	–14.4	–14	14	–14	L gyrus rectus/area Fo2/posterior orbitofrontal complex/subcallosal gyrus
orb7_Lba11l	34	–20.1	32.1	–18.8	–18	34	–22	L superior orbital gyrus/area Fo3/lateral BA11
orb8_Rofc	25	2.1	16.1	–21.7	2	16	–22	R orbitofrontal cortex/area Fo2
orb9_Rba14m	20	4.6	28.5	–13.2	4	30	–14	R middle orbital gyrus/area s32/area 25/medial BA14
Prefrontal								
pfc1_Lba9pv	50	–37.5	33.8	22	–40	28	26	L posterior ventral BA9/BA46
pfc2_Lba9av	40	–38.1	55.2	1.6	–38	54	0	L anteroventral BA9/46/rostral BA47
Parietal								
pl1_Lipl	280	–35	–67.6	25.1	–36	–68	24	L middle occipital gyrus/inferior parietal lobule/rostro-ventral BA39
pl2_Lba7m	233	–2.9	–53.5	58.1	–6	–52	66	L medial BA7A/medial BA5
pl3_Lba7m5ml	190	44.6	20.4	10.3	46	16	6	L medial BA7A/mediolateral BA5
pl4_Lpos2	114	–4.2	–65.9	42.6	–2	–70	44	L parieto-occipital sulcus area 2/medial BA7/precuneus

Type and CLUSTER_ID	Voxels	Center of mass (MNI[a])			Peak (MNI)			ATLAS ROI[b]
		x	y	z	x	y	z	
pl5_Rpos2	32	16.9	–59.5	28.2	20	–60	26	R precuneus/parieto-occipital sulcus area 2
Temporal								
t1_Lffc	95	–46.6	–56.9	–18.9	–44	–60	–16	L fusiform gyrus/fusiform face complex/lateroventral BA37/area FG2
t10_Ravsts	31	56.3	–7.2	–21.6	56	–6	–20	R anteroventral superior temporal sulcus/middle temporal gyrus
t11_Laud5	31	–61.5	–0.3	–8.7	–64	–4	–4	L superior temporal gyrus/area TE3/auditory 5 complex
t12_Rba38l	28	47.2	21.4	–26.3	48	22	–26	R temporopolar cortex/lateral BA38/dorsal area TG
t13_Lpiri	6	–41.7	6	–10.7	–40	4	–12	CLUSTER5/L piriform cortex/insula/posterior insula area 2
t14_Lba37vl	3	–48.7	–64	–3.3	–48	–64	–4	CLUSTER10/middle temporal gyrus/area PH/ventrolateral BA37
t2_Lba38l	68	–35.3	4	–23.2	–38	6	–20	L dorsal area TG/lateral BA38/temporopolar cortex
t3_Rba22	64	66.9	–17.3	0	66	–14	0	R auditory 4 complex/area TE3/caudal BA22
t4_Lba20il	62	–54.8	–11.9	–34.2	–56	–12	–32	L anterior area TE2/inferolateral BA20/inferior temporal gyrus
t5_Lba22c	60	–65.2	–46.7	11.8	–68	–48	12	L superior temporal visual area/caudal BA22
t6_Rins	45	34.2	–16.7	23.4	34	–14	22	R insula/area OP2-3/VS
t7_Lpins2	37	–35.4	–6.4	–6.5	–36	–8	–4	L posterior insular area 2/circular insula
t8_Lba38m	36	–45.8	24	–25.4	–44	26	–26	L temporopolar cortex/dorsal area TG/medial BA38
t9_Raud5	33	65.1	–38.6	5.5	66	–38	6	R middle temporal gyrus/auditory 5 complex/ventral superior temporal sulcus
Thalamic								
th1_Lpfthal	82	–5.7	–14.4	–3.5	–6	–14	–2	L thalamus/thalamic area IPF/prefrontal-directed thalamus
th2_Rpmthal	72	15.5	–18.6	4.9	16	–18	6	R thalamus/premotor-directed thalamus
Predefined								
medial PFC[c] (DMN[d])	—[e]	1	55	–3	N/A[f]	N/A	N/A	medial prefrontal cortex, default mode network
L lateral parietal (DMN)	—	–39	–77	33	N/A	N/A	N/A	L lateral parietal cortex/von Economo PG, default mode network
R lateral parietal (DMN)	—	47	–67	29	N/A	N/A	N/A	R lateral parietal cortex/von Economo PG, default mode network
posterior cingulate (DMN)	—	1	–61	38	N/A	N/A	N/A	posterior cingulate cortex, default mode network
L lateral sensorimotor	—	–55	–12	29	N/A	N/A	N/A	L lateral sensorimotor cortex
R lateral sensorimotor	—	56	–10	29	N/A	N/A	N/A	R lateral sensorimotor cortex
superior sensorimotor	—	0	–31	67	N/A	N/A	N/A	superior (mesial) sensorimotor cortex
medial visual	—	2	–79	12	N/A	N/A	N/A	medial visual cortex/Brodmann 18/calcarine gyrus/visual area V1
occipitopolar visual	—	0	–93	–4	N/A	N/A	N/A	occipitopolar visual cortex/Brodmann 17/calcarine gyrus

Type and CLUSTER_ID	Voxels	Center of mass (MNI[a])			Peak (MNI)			ATLAS ROI[b]
		x	y	z	x	y	z	
L lateral visual	—	–37	–79	10	N/A	N/A	N/A	L lateral visual cortex/visual area V4
R lateral visual	—	38	–72	13	N/A	N/A	N/A	R lateral visual cortex/visual area V4
anterior cingulate (SN[g])	—	0	22	35	N/A	N/A	N/A	anterior cingulate cortex, salience network
L anterior insula (SN)	—	–44	13	1	N/A	N/A	N/A	L anterior insula, salience network
R anterior insula (SN)	—	47	14	0	N/A	N/A	N/A	R anterior insula, salience network
L rostral PFC (SN)	—	–32	45	27	N/A	N/A	N/A	L rostral prefrontal cortex, salience network
R rostral PFC (SN)	—	32	46	27	N/A	N/A	N/A	R rostral prefrontal cortex, salience network
L supramarginal g. (SN)	—	–60	–39	31	N/A	N/A	N/A	L supramarginal gyrus/Brodmann 40, salience network
R supramarginal g. (SN)	—	62	–35	32	N/A	N/A	N/A	R supramarginal gyrus/Brodmann 40, salience network
L FEF (DAN[h])	—	–27	–9	64	N/A	N/A	N/A	L frontal eye fields, dorsal attention network
R FEF (DAN)	—	30	–6	64	N/A	N/A	N/A	R frontal eye fields, dorsal attention network
L inferior parietal s. (DAN)	—	–39	–43	52	N/A	N/A	N/A	L inferior parietal sulcus, dorsal attention network
R inferior parietal s. (DAN)	—	39	–42	54	N/A	N/A	N/A	R inferior parietal sulcus, dorsal attention network
L lateral PFC (FPN[i])	—	–43	33	28	N/A	N/A	N/A	L lateral prefrontal cortex/Brodmann 9/46, frontopolar network
L posterior parietal cortex (FPN)	—	–46	–58	49	N/A	N/A	N/A	L posterior parietal cortex/von Economo PFm, frontopolar network
R lateral PFC (FPN)	—	41	38	30	N/A	N/A	N/A	R lateral prefrontal cortex/Brodmann 9/46, frontopolar network
R posterior parietal cortex (FPN)	—	52	–52	45	N/A	N/A	N/A	R posterior parietal cortex/von Economo PFm, frontopolar network
L inferior frontal language area	—	–51	26	2	N/A	N/A	N/A	L inferior frontal gyrus
R inferior frontal language area	—	54	28	1	N/A	N/A	N/A	R inferior frontal gyrus
L posterior STG[j] language area	—	–57	–47	15	N/A	N/A	N/A	L posterior superior temporal gyrus
R posterior STG language area	—	59	–42	13	N/A	N/A	N/A	R posterior superior temporal gyrus
Anterior cerebellum	—	0	–63	–30	N/A	N/A	N/A	Anterior cerebellum
Posterior cerebellum	—	0	–79	–32	N/A	N/A	N/A	Posterior cerebellum

[a]MNI: Montreal Neurological Institute stereotactic coordinate system.

[b]ROI: regions of interest.

[c]PFC: prefrontal cortex.

[d]DMN: default mode network.

[e]Not available (for predefined ROIs, subject signals for correlation analyses were derived from the average signal over the entire region and cluster sizes are not available).

[f]N/A: not applicable (for predefined ROIs, subject signals for correlation analyses were derived from the average signal over the entire region and peak voxel locations are not available).

[g]SN: salience network.

[h]DAN: dorsal attention network.

[i]FPN: frontopolar network.

[j]STG: superior temporal gyrus.

Statistical Analysis

Descriptive statistics (means, SDs, and SEs) were used to characterize baseline demographics and responses to pre- and postintervention psychometric assessments. Meditation practice time was the sum of the 8 weeks of practice, which was provided by the app company. Missing items in the psychometric scales were estimated with expectation maximization [48] (missing items never accounted for >5% of the total data) using other items within the scale as predictor variables. Baseline differences between the app and wait-list groups were assessed using independent *t* tests for continuous variables and chi-square tests for categorical variables (Tables 1 and 3). Post- versus preintervention differences in burnout, depression, anxiety, or sleep impairment in the *mindfulness* group were assessed by repeated-measures ANOVA. Given the exploratory nature of our analyses, we first used an α level of .05. Tests of the hypotheses were also conducted using Bonferroni-adjusted α levels of .008 (.05/6 outcome variables). To evaluate whether statistically significant outcomes were related to mindfulness practice time, we conducted Spearman correlation analyses between practice time and changes in all relevant outcomes.

Table 3. Group means and differences with respect to the Patient-Reported Outcomes Measurement Information System sleep-related impairment short form 8b (sleep disturbance), School Burnout Inventory (burnout), and The Depression Anxiety and Stress Scale (depression and anxiety) scores at baseline and >8 weeks. Sleep disturbance indices were lower in the app group than the wait-list group at >8 weeks.

	Time 1			Time 2			*F* test (*df*)
	App group, mean (SD)	Wait-list group, mean (SD)	*t* test (*df*)	App group, mean (SD)	Wait-list group, mean (SD)	*t* test (*df*)	
Sleep disturbance	51.9 (9.99)	54.1 (2.99)	–0.57 (12)	42.2 (7.45)	53.4 (7.15)	–2.93[a] (12)	8.68[a] (1, 12)
Depression	3.00 (2.23)	4.86 (2.67)	–1.41 (12)	3.14 (2.41)	4.00 (2.83)	–0.61 (12)	0.48 (1, 12)
Anxiety	5.14 (3.08)	8.00 (3.96)	–1.51 (12)	3.43 (2.82)	4.57 (4.20)	–0.60 (12)	0.28 (1, 12)
Burnout	30.7 (6.21)	34.0 (5.48)	–1.05 (12)	29.9 (8.69)	34.9 (7.67)	–1.14 (12)	0.58 (1, 12)

[a]Statistically significant at $P<.05$.

Differences in Connectivity by Group, Visit, and Practice Time Bin

All within- and between-group statistical analyses were performed in MATLAB v2019a. Within-group differences in connectivity distributions (mindfulness at 8 weeks vs mindfulness at baseline and wait-list at 8 weeks vs wait-list at baseline) were assessed against each ROI pair in the baseline and eighth-week ROI-ROI connectivity matrix, as appropriate, using a 2-tailed *t* test assuming equal variances (MATLAB ttest2() function). Correction for FDR (the expected proportion of false discoveries between all ROI-ROI pairs with similar or larger effects) was performed at a critical value of α=.05 via the Benjamini-Hochberg procedure (MATLAB mafdr() function [49]). Similarly, between-group differences in connectivity (mindfulness at 8 weeks vs mindfulness at baseline and wait-list at 8 weeks vs wait-list at baseline) were assessed against each ROI pair in the Δconn matrix via one-way ANOVA omnibus test (anova1() function) using *group* as a predictor, with post hoc comparisons by 2-tailed *t* test. In addition, mindfulness subjects were divided into low (practice time <53 minutes), moderate (53 to <225 minutes), and high (≥225 minutes) practice time bins, and a 1-way ANOVA was conducted against the Δconn matrix using *practice time bin* as the predictor and post hoc comparisons by 2-tailed *t* tests. In each of the above cases, *P* values were FDR corrected using the Benjamini-Hochberg procedure at α=.05.

Differences in Connectivity by Practice Time, Mental Health Scores, and Mental Health Scores

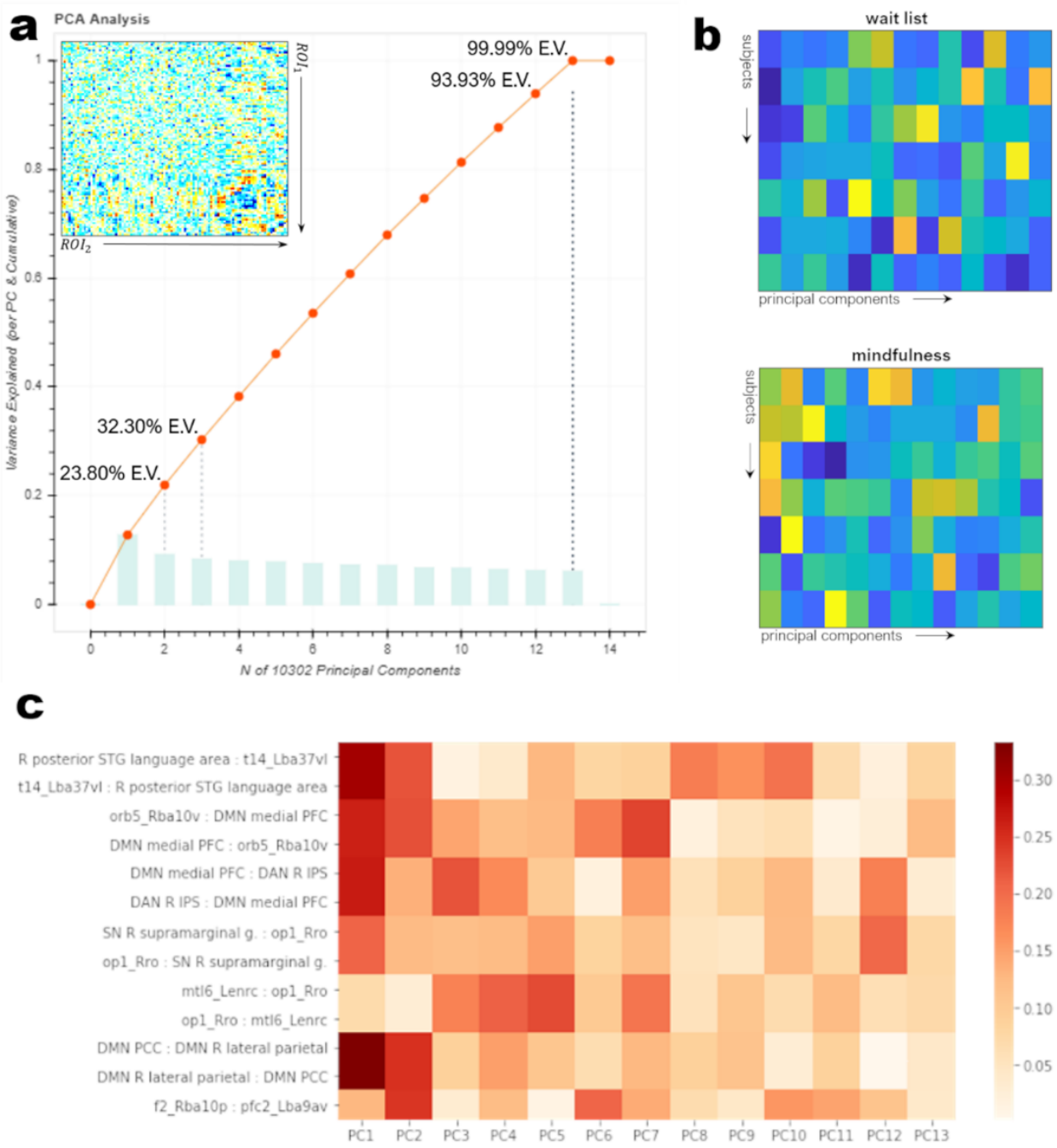
For any mental health outcomes that had a significant group-by-time interaction, we examined whether the change or changes were related to changes in functional connectivity. As our analyses indicated that the only significant impact of app-delivered mindfulness was on the PROMIS sleep impairment measure, we limited our analyses to this measure. Although connectivity data were normally distributed because of the application of the Fisher transform in preprocessing, the ordinal nature of PROMIS sleep impairment scores across all subjects, skewness of the practice time data within the mindfulness group, and the presence of outliers in SI and practice time data rendered parametric approaches inappropriate for analysis of these predictors relative to connectivity. Consequently, we used a rank-based method, Spearman rho, to investigate such relationships. Spearman rank-order correlations (ρ) are equivalent to the Pearson product-moment correlation coefficient but are applied to ranks rather than continuous values and are, therefore, less susceptible to strong outliers. Spearman correlations were calculated for practice time (mindfulness subjects only) and change in SI score ($\Delta SI=SI_{8wks}–SI_{baseline}$) versus Δconn using the MATLAB function corr(). FDR correction was performed as previously described, except that *P* value rankings were computed over all ROI-ROI pairs *within the mindfulness group* only for ρ (practice time and Δconn) analysis. All *P* values were FDR corrected, as previously described.

Machine Learning Classifiers for Sleep Impairment Based on Connectivity

Of the 7 subjects randomly assigned to the mindfulness group, 3 (43%) reported sleep impairment at baseline but no longer reported impairment at 8 weeks, whereas, of the 7 waitlisted subjects, 3 (43%) had sleep impairment at baseline and continued to exhibit impairment at 8 weeks (Multimedia Appendix 2, Figure S4). Therefore, we were interested in testing whether conversion between sleep impairment and nonimpairment might be predicted by a supervised learning algorithm based on the change in functional connectivity (Δconn) between baseline and >8 weeks. Accordingly, per-visit labels were assigned to each subject based on the PROMIS-derived classification of sleep impairment, such that a PROMIS SI score commensurate with sleep impairment was encoded as 1 and nonimpairment as 0. A ΔIMPAIRMENT score ($IMPAIRMENT_{8wks}$–$IMPAIRMENT_{baseline}$) was then computed and used as target classes (0=*no change*, 11 subjects total; 1=*improvement*, 3 subjects total; and −1=*decline*, 0 subjects). A corpus of six binary classifiers, comprising support vector, gradient boosting, random forest, Gaussian naïve Bayes, linear discriminant, and multilayer perceptron estimators were implemented in *scikit-learn*, v0.22.1, under Python 3.7.7 (see Multimedia Appendix 2 for training and testing details). After excluding undefined values (NaNs) corresponding to the conn diagonal from the matrix, conn samples and their associated labels {−1, 0, +1} were split into training (30%) and testing (70%) sets. The samples were reshuffled into train/test sets over 5 tuning rounds. In each round, the classifier hyperparameters were tuned by random search (RandomizedSearchCV() in *scikit-learn*) over 5000 iterations.

A second series of classification runs were performed to test whether further dimensionality reduction improved the classifier performance. This second series leveraged the same classifiers used previously but the *features* (ROI-ROI connections) used for classification comprised 13 principal components—accounting for 99.99% of the data variance—derived from the conn matrix using decomposition. PCA() in *scikit-learn*. The contribution of each feature (connection) to each of the 13 components was ascertained by taking the magnitude of the dot product of the conn matrix, renormalized to the explained variance, and the principal components, yielding a collinearity metric between each connection and component. The results of the PCA decomposition and the connections assigned to each feature are shown in Figure 1, which also shows the contribution of each feature (connection in the conn matrix) to the 13 principal components recovered by computing the dot product between the Δconn matrix and the derived components. Hyperparameters were computed separately for this second series of classifications, following the strategy previously described.

Figure 1. Results of principal components decomposition of Δconn matrix. (a) explained variance and cumulative explained variance for the first 14 principal axes. 13 components accounted for 99.99% of the explained variance in the delta matrix (Δconn), suggesting a large amount of multicollinearity across ROI-ROI connections. The recomposition of the Δconn matrix from the 13 selected components is shown in the inset. (b) within-group correlations between per-subject Δconn values and the 13 selected components, averaged across subjects within each group. (c) Primary contributors to each component based on the dot product of the original Δconn matrix and the 13-PC decomposition. Note the substantive contributions of posterior superior temporal, fusiform (t14), orbitofrontal, and default mode, dorsal attention, and salience-network components to components PC1 and PC2. These two components represented nearly one-third of the explained variance in the Δconn matrix. DAN: dorsal attention network; DMN: default mode network; PFC: prefrontal cortex; SN: salience network; STG: superior temporal gyrus.



Results

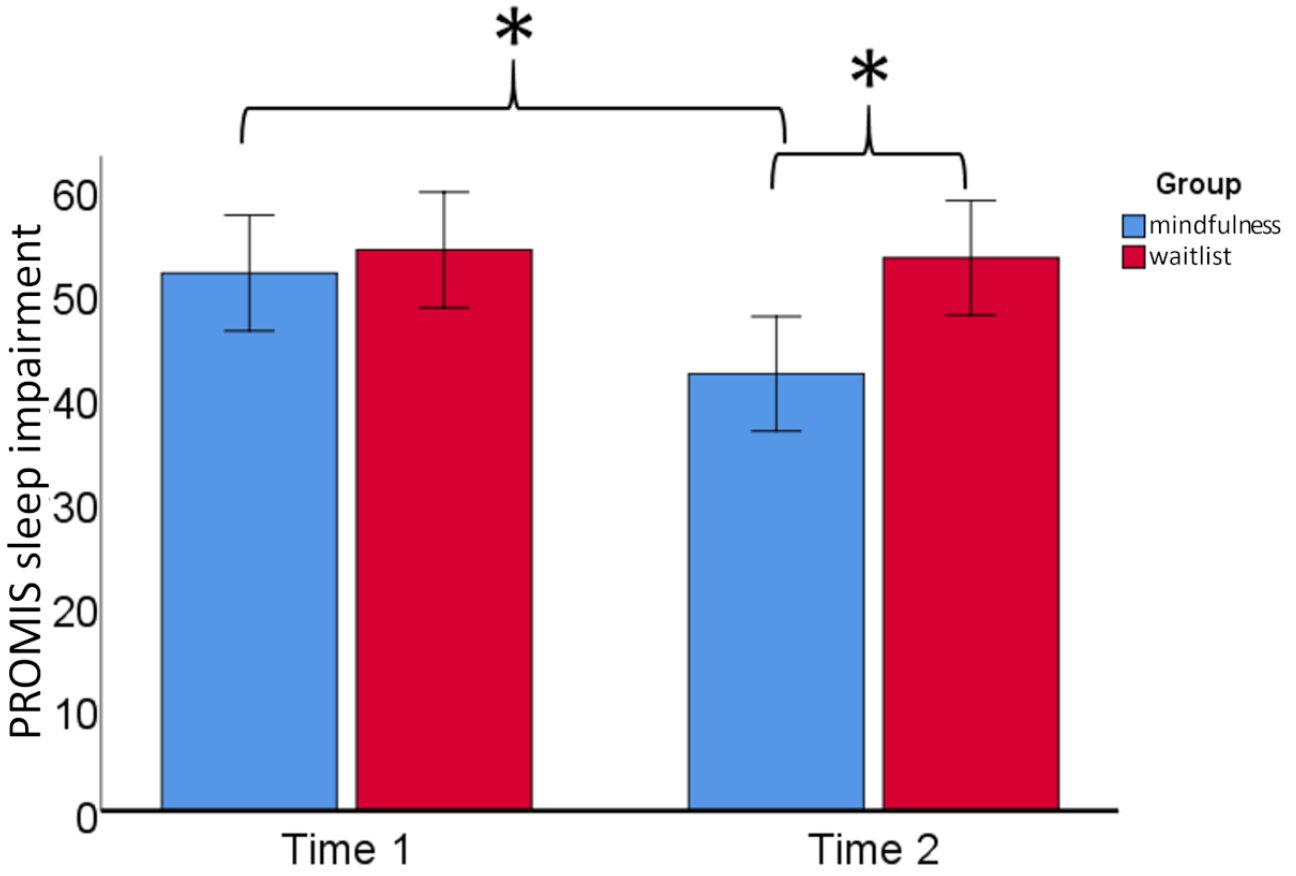
Overview

At time 1, before randomization, the groups showed no differences in any self-report variables (all *P* values>.16) or demographic variables (all *P* values>.21). Participants randomized to the mindfulness group used the app between 0 and 466.2 (mean 182.8, SD 182.8) minutes. A total of 2 trainees randomized to the practitioner group did not use the app at all; however, these PA students were included in all analyses in an intent-to-treat design. None of the participants reported adverse events or adverse experiences with the app.

There was a significant group-by-time interaction for sleep impairment, such that participants randomized to mindfulness group reported a reduction in impairment compared with those

randomized to the wait-list group (F_{12}=8.68; P=.01; η_p^2=0.42). No other self-reported outcomes were significant for the group-by-time interaction (low incivility: F_{12}=0.27; P=.61; high incivility: F_{12}=0.29; P=.60; burnout: F_{12}=0.56; P=.47; depression: F_{12}=0.53; P=.48; anxiety: F_{12}=1.23; P=.28). A paired sample t test indicated that participants in the mindfulness group reported significant reductions in sleep impairment (t_6=3.35; P=.02). Students randomized to the mindfulness group did not report any other significant changes, although there was a trend toward a reduction in burnout (t_6 2.20; P=.07). Paired sample testing indicated that the wait-list control group reported a reduction in anxiety (t_6=3.62; P=.01). Finally, independent sample t tests indicated that there was a significant difference between the mindfulness and wait-list groups in terms of sleep impairment at time 2 (t_{12}=–2.93; P=.02). There were no other significant differences between the groups at time 2. Finally, changes in sleep impairment were not significantly correlated with mindfulness practice time within the mindfulness group (Spearman r_s=–0.36; P=.43; Figure 2).

Figure 2. Differences in sleep impairment by group (mindfulness vs waitlist) and visit (time 1, baseline; Time 2, >8 weeks). An asterisk (*) indicates significant differences in sleep impairment at P≤.05. PROMIS: Patient-Reported Outcomes Measurement Information System.



Connectivity by Group and Visit

A 2-way ANOVA was conducted to assess whether *group* (mindfulness or control) and/or *visit* (baseline and >8 weeks) were significant predictors of between- and within-group variance in the baseline and eighth-week connectivity matrices or the *delta matrix* (Δconn=connectivity at >8 weeks − connectivity at baseline). ANOVAs were conducted on each ROI-ROI pair separately, and P values were corrected for FDR using the Benjamini-Hochberg procedure. We failed to find significant effects of the *group* × *visit* interaction for any connection, and subsequent inspection of the main effects indicated that although *group* and *visit* each accounted for a statistically significant amount of variance across the connectivity matrices, they did not affect the same connections. Consequently, main-effects ANOVAs and subsequent post hoc analysis by a 2-tailed t test were conducted for *group* and *visit* separately.

The ANOVA on *group* conducted against the baseline and eighth-week connectivity matrices indicated significant differences in control subjects' connectivity among the supplementary motor, middle temporal, inferior temporal, occipitopolar, and orbitofrontal cortices at baseline, and among the insular, cerebellar, lateral visual, and superior frontal cortices and thalamus at >8 weeks (Table 4). Post hoc t tests revealed that, at baseline, mindfulness participants exhibited stronger connectivity than controls with respect to left supplementary motor–left ventrolateral Brodmann area 37 (t_{12}=5.37; P_{FDR}=.02) and left inferolateral Brodmann area 20–right middle temporal gyrus (t_{12}=4.73; P_{FDR}=.05). The left BA37 ROI was localized to the ventrolateral aspect of the ipsilateral middle temporal gyrus corresponding to the area PH of Economo-Koskinas [50], and the left BA20 ROI was localized to the anterior aspect of visual area TE2. In contrast, controls exhibited stronger left lateral occipitopolar–ipsilateral superior temporal gyrus connectivity than the mindfulness subjects (t_{12}=−4.98; P=.03)

at baseline. The latter ROI was associated with area TE3 and/or auditory complex 5. At >8 weeks, mindfulness subjects exhibited stronger connectivity than control subjects with respect to the left dorsolateral Brodmann area 8–left posterior insula (t_{12}=4.84; P_{FDR}=.05), left cerebellar lobule VI–right lateral visual cortex (t_{12}=4.65; P_{FDR}=.05), and right premotor thalamus–right lateral visual cortex (t_{12}=4.32; P_{FDR}=.05).

Table 4. Results of *t* tests on group (mindfulness, or MF, vs wait-list, or CX) and visit (visit 1, v1, vs visit 2, v2). Connections surviving false discovery rate correction are presented as pairs of source and versus regions of interest.

Group and sources	Versus	*t* test (*df*)[a]	*P* FDR
MF v1>CX v1			
f6_Lsma	t14_Lba37vl	5.3665 (12)	.02
o2_Llop	t11_Laud5	–4.977 (12)	.03
t4_Lba20il	t9_Raud5	4.7331 (12)	.049
MF v2>CX v2			
f5_Lba8dl	t7_Lpins2	4.7415 (12)	.048
cb2_Llob6	R lateral visual	4.6543 (12)	.05
th2_Rpmthal	R lateral visual	4.3193 (12)	.05
MF v2>MF v1			
posterior cing. (DMN)	anterior cing. (SN)	5.7184 (7)	.009
orb2_Lba10r14m	t7_Lpins2	5.5659 (7)	.01
f11_Rba6ba8	anterior cingulate (SN)	4.9121 (7)	.02
R FEF (DAN)	R inf. frontal lang. area	4.8411 (7)	.04
posterior cing. (DMN)	L anterior insula (SN)	4.1621 (7)	.04
cing1_Rpcc	L anterior insula (SN)	4.1902 (7)	.04
orb5_Rba10v	t7_Lpins2	4.1199 (7)	.047
t7_Lpins2	medial PFC (DMN)	4.6925 (7)	.05
CX v2>CX v1			
o2_Llop	mtl5_Lamg	5.4641 (5)	.01

[a]Two-tailed *t* test.

The ANOVA on *visit* conducted against the delta matrix Δconn indicated that baseline-to-eighth-week connectivity differed between mindfulness and control subjects with respect to connections between the left lateral occipitopolar cortex and the ipsilateral amygdala and in mindfulness subjects' connectivity among the anterior and posterior cingulate, insular, orbitofrontal, and medial prefrontal cortices (Tables 4 and 5). Post hoc *t* tests subsequently revealed that control subjects exhibited higher connectivity between left lateral occipitopolar cortex and left amygdala at >8 weeks than at baseline (t=5.46; P_{FDR}=.01). In contrast, mindfulness participants exhibited higher connectivity at >8 weeks, relative to baseline, between anterior cingulate (an SN component) and posterior cingulate (a DMN component; t=5.72; P_{FDR}=.01), left rostral orbitofrontal Brodmann area 10 and ipsilateral posterior insula (t=5.57; P_{FDR}=.01), anterior cingulate and right superior frontal gyrus (t=4.91; P_{FDR}=.02), right frontal eye fields (a dorsal attention network component) and ipsilateral inferior frontal gyrus (t=4.84; P_{FDR}=.04), posterior cingulate and left anterior insula (an SN component; t=4.16; P_{FDR}=.04), the ventral aspect of the right orbitofrontal Brodmann area 10 and left posterior insula (t=4.12; P_{FDR}=.05), and left posterior insula and medial prefrontal cortex (t=4.69; P_{FDR}=.05). The superior frontal ROI was associated with an area denoted *i6-8* (Assem et al [51]) located in the superior aspect of the transition area between the premotor cortex (Brodmann area 6) and frontal eye fields (Brodmann area 8; Table 2).

Table 5. Results of one-way analysis of variance on delta matrix with group as the predictor. Post hoc analysis by 2-tailed *t* test. *P* values false discovery rate-corrected by Benjamini-Hochberg procedure. Note that, as the connectivity coefficients were normalized, they are directly comparable with a Cohen d statistic and thus indicate effect sizes. The differences listed here fall in the range of 0.15-0.35 and should be regarded as small-to-moderate effects.

ID1	ID2	ROI1	ROI2	*F* test (*df*)	MSE[a]	Model *P* value	Post hoc	Mean difference	*P* value
t1_Lffc	orb2_Lba10r14m	L fusiform face complex	L r. BA10/med. BA14	20.1465 (1, 12)	0.016	<.001	Control<mindfulness	–0.3031	<.001
pl2_Lba7m	orb1_Rba11	L med. BA7/med. BA5	R lat. BA11/area Fo3	15.3067 (1, 12)	0.0285	.002	Control<mindfulness	–0.3531	.002
Anterior cerebellum	Posterior cerebellum	Anterior cerebellum	Posterior cerebellum	8.8866 (1, 12)	0.0179	.01	Control<mindfulness	–0.2134	.01
mtl3_Renrc	cb4_Rlob7b	R entorhinal ctx./presubi	R cerebellar lobule VIIb	6.3201 (1, 12)	0.0162	.03	Control<mindfulness	–0.171	.03
pl4_Lpos2	caud2_Rput1	L parieto-occip. s./medial BA7	R mid. insula/putamen	5.9626 (1, 12)	0.0215	.03	Control<mindfulness	–0.1914	.03
pl3_Lba7m5ml	f4_Rba8d	L med. BA7/med. lat. BA5	R mid. frontal g./BA 8	5.9356 (1, 12)	0.051	.03	Control<mindfulness	–0.294	.03
cb2_Llob6	pl2_Lba7m	L cerebellar lobule VI	L med. BA7/med. BA5	4.9141 (1, 12)	0.018	.046	Control<mindfulness	–0.1588	.046
orb3_Rifgpo	mtl2_Rphg	R IFG orb./sup. BA47/lat. BA11	R parahippocampal g	4.889 (1, 12)	0.0167	.047	Control<mindfulness	–0.1527	.047
mtl1_Lphg1	f5_Lba8dl	L parahippocampal g./subi	L dors. lat. BA8	4.7757 (1, 12)	0.027	.049	Mindfulness<control	0.1919	.049

[a]MSE: mean square error.

Changes in Connectivity With the SI Score and Practice Time

Nearly all subjects exhibited a decrease in SI scores between baseline (control: 54.1 [SD 2.99]; mindfulness: 51.89 [SD 9.99]) and +8 weeks [control 53.36 [SD 7.15]; mindfulness: 42.19 [SD 7.45]). Mean ΔSI scores, defined as SI score at >8 weeks minus SI score at baseline, were −0.76 (SD 5.13) and −9.70 (SD 6.18) for the control and mindfulness groups, respectively. Spearman rank correlation, assessed against the delta matrix (Δconn, defined as connectivity among the 102 ROIs at >8 weeks minus connectivity at baseline) across all subjects, indicated that a greater decrease in SI score was associated with increased connectivity between the right superior frontal gyrus (BA6/8 transition area, *i6-8*) and ipsilateral inferior parietal sulcus (ρ=0.82, P_{FDR}=.03) and between the left superior temporal gyrus and medial prefrontal cortex (a component of the DMN; ρ=0.81; P_{FDR}=.05). The superior temporal ROI was localized to the caudal aspect of Brodmann area 22 (Wernicke area) and the superior temporal visual cortex. Greater decreases in SI score were also associated with reduced connectivity between the left supplementary motor area and the ventrolateral aspect of the ipsilateral middle temporal gyrus/Economo-Koskinas area PH (ρ=0.89; P_{FDR}<.01; Table 6).

Within the mindfulness group, Spearman correlations indicated an association between practice time and changes in the connectivity of four connections: between the right inferior parietal sulcus (a component of the dorsal attention network) and right lateral visual cortex, between the right inferior parietal sulcus and the occipitopolar visual cortex, between the right middle orbital gyrus and orbitofrontal Brodmann area 11, and between the right middle orbital gyrus and the left lateral sensorimotor cortex. The strengths of these connections increased linearly with practice time, with the exception of the left lateral sensorimotor–middle orbital gyrus connection, which decreased with practice time (Tables S1 and S2 and Figure S3 in Multimedia Appendix 2).

Table 6. Spearman correlations (rho) between the connectivity strength of selected connections and change in sleep impairment score (ΔSI, defined as sleep impairment at 8 weeks minus sleep impairment at baseline), taken across all subjects. Correlations are corrected for multiple comparisons using false discovery rate.

Source	Versus	rho	P_{FDR}
f6_Lsma	t14_Lba37vl	0.8896	.002
t5_Lba22c	DMN medial PFC	–0.8057	.05
f11_Rba6ba8	DAN R IPS	–0.8234	.03

Classifiers

Throughout 5 rounds of training or test group shuffling, each with 5000 iterations of hyperparameter tuning, the 6 estimators (support vector, gradient boosting, random forest, Gaussian naïve Bayes, linear discriminant, and multilayer perceptron) were equally proficient in classifying the change from sleep impairment at baseline to no sleep impairment at >8 weeks based on the full 102 × 102-ROI delta matrix (Δconn) for each subject. All estimators correctly classified true converters; conversely, they misclassified at least 1 nonconverter in each of the 5 rounds. Reducing the dimensionality of the Δconn matrix to 13 components (99.99% of the cumulative explained variance; Figure 1) via principal component analysis increased the accuracy of most classifiers by 3-5% by increasing estimator specificity (decreasing the false positive rate). Conversely, the estimators tended to misclassify true converters as nonconverters, increasing the false-negative rate, when predictions were based on the reduced-dimensionality delta matrix. Under this schema, Gaussian naïve Bayes and random forest classifiers slightly outperformed other estimators, with an accuracy rate of approximately 80%. We must, however, urge caution in overinterpreting these results, given the small number of true converters (n=3) to nonconverters (n=11) and the absence of subjects *converting* from nonsleep-impaired to sleep-impaired.

Discussion

Principal Findings

Provider burnout and depression have profound national and institutional economic costs, as well as deep societal and social effects. Burnout costs an estimated US $150 billion per year or almost 5% of the nation's health care expenditure [52]. Although these costs are generally estimated based on the effects of burnout and depression among physicians, PAs also report remarkably high levels of burnout [5] and depression [6]. Although little is known about the public health impact of PA burnout, physician burnout increases malpractice rates, exacerbates physician shortages, erodes both health care organization morale and patient experience [53], and reduces clinical effectiveness [54,55]. Given their overlapping roles and day-to-day activities, it is likely that burnout has a comparable effect on PAs.

Similar to the relative lack of data on burnout and well-being among practicing PAs, far less is known about the mental health needs of PA students than is known about medical residents and medical students. The PA profession is in a period of some flux, with a workforce population that is growing in overall number and that is increasingly young and female [4,56]. PA students may have unique needs, given the differences between PAs and physicians. PA students often report choosing their professional route based on concerns about debt load [57] and expectations of a healthier work-life balance [58]. Here, students reported low levels of depression (0% of students) and relatively low levels of anxiety. Approximately 21% (3/14) of students indicated moderate anxiety, and 7% (1/14) of students indicated severe anxiety at the outset of the study. More students (43%) reported sleep impairment; 29% (4/14) of students reported mild impairment, and 14% (2/14) of students reported moderate impairment before randomization. It is important to note that the students who enrolled in the study may not be representative of the entire PA student population.

Mindfulness meditation has shown great promise for improving sleep disruption and insomnia symptoms [13,14], and it may be beneficial to health professional trainees who often experience high rates of sleep dysfunction [16]. However, most studies examining mindfulness among health professionals and trainees have examined time-intensive interventions that are prohibitive for many. Previous studies have shown that app-delivered mindfulness may be effective in reducing anxiety among physicians [59]. Although we included a wait-list control group to control self-selection and the inevitable changes that occur during PA school but not attributable to the intervention, future studies should include an active control condition. Although this has been challenging in studies of app-delivered mindfulness to date, recent work has advanced in this area toward developing smartphone apps that can be used as active comparators (eg, Huberty et al [60]).

PA students randomized to mindfulness reported a significant reduction in sleep impairment compared with students randomized to the wait-list. Although 43% (3/7) of students randomized to mindfulness reported mild (t=55-59.9) or moderate (t=60-69.9) sleep impairment before randomization, none of the students randomized to mindfulness reported sleep impairment after training (t<55) [61]. These findings are consistent with previous research indicating that mindfulness-based interventions are effective in reducing sleep disturbance and for altering neurobiology related to the default mode in adults with sleep disturbances [62]. Our findings extend previous work to indicate that short, app-delivered mindfulness is beneficial for improving self-reported sleep impairment. A recent study found that most study participants using a popular mindfulness app (Calm) downloaded it to improve sleep impairment [63]; considering our data further highlights the potential importance of app-delivered mindfulness in the context of sleep and sleep dysfunction. This study also adds to what is known about the impact of mindfulness on medical trainees

[64] and may be part of a critical solution for sleep dysfunction, which is associated with an increased risk of depression, burnout, and medical errors [65,66]. Although changes in sleep impairment were not correlated with practice time, the relationship was in the expected direction, and it is likely that we were not powered to detect this relationship.

These data also indicate that improvements in sleep are associated with connectivity changes between the DMN and regions important for emotion, attention, and social cognition. Previous studies have shown that disordered sleep is related to altered brain function in the DMN as well as the SN [67-69]. Here, using principal components decomposition, we found that a small number of connections among DMN, SN, and dorsal attention network components, and with superior temporal, fusiform, and orbitofrontal areas, are closely associated with one another and with the explained variance of the delta-connectivity matrix. This area, explicitly characterized as i6-8, is functionally distinct from regions involved in simple eye movements and is considered part of a core complex involved in working memory, along with the inferior parietal sulcus [51]. The finding that changes in this network of regions are related to changes in sleep impairment further bolsters the existing evidence that the DMN and SNs are affected by or involved in sleep impairment.

PA well-being is a complex and multifactorial issue. Isolation, sleep deprivation and disturbance, and feeling overwhelmed by the amount of material they need to master are risk factors for depression among health care trainees [65,70]. Moreover, lack of time for self-care and stigma toward treatment-seeking are barriers to addressing mental health crises among trainees [2,71]. For these reasons, it is unlikely that a short-term, app-delivered mindfulness meditation program will be a stand-alone solution. Rather, addressing trainee mental health must be comprehensive and should include structural and organizational solutions alongside individualized resilience programming. Moreover, wellness programs for trainees must be sustainable and preventive in nature rather than reactive [64,72]. Medical training programs must make wellness feasible within the lives of trainees, and app-delivered programming may be a feasible and sustainable piece to foster a culture of resilience among PA students.

Limitations

This study had a small sample, and the findings may not be representative of all PA students. We were likely underpowered to detect small effects, and the changes in sleep impairment did not reach significance at alpha levels adjusted for multiple comparisons. Moreover, it is unclear whether improvements in sleep impairment reported by students randomized to mindfulness are enduring. There is some evidence that there are sex differences in how disordered sleep affects brain function [73]. The students in our sample were primarily women, and thus, the results may not be generalizable to male trainees. Despite these limitations, the methods used here are a novel approach to understanding sleep impairment and a mindfulness intervention that may improve it, and these data indicate that app-delivered mindfulness may be effective for PA students.

Acknowledgments

The authors would like to express their gratitude to the research assistants who helped collect the data reported here and the faculty in the Emory PA program for facilitating student recruitment. The authors are also grateful to Chris Larsen for his guidance and expertise in embarking on this research and *10% Happier* for providing app subscriptions. *10% Happier* provided input on recommended content during study design but did not have access to study data and was not involved in data analysis or interpretation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Quality control and assurance measures, including motion correction and signal change assessments, validity of subject-to-stereotactic space registrations, effects of denoizing on connectivity, and functional connectivity as a function of distance from the seed voxel.
[PDF File (Adobe PDF File), 283 KB - formative_v5i10e24208_app1.pdf]

Multimedia Appendix 2
Supplementary figures and tables, including results of multivoxel pattern analysis, connectivity matrices by group and visit, and additional regression and classification results.
[PDF File (Adobe PDF File), 1002 KB - formative_v5i10e24208_app2.pdf]

References

1. Mata DA, Ramos MA, Bansal N, Khan R, Guille C, Di Angelantonio E, et al. Prevalence of depression and depressive symptoms among resident physicians: a systematic review and meta-analysis. JAMA 2015 Dec 08;314(22):2373-2383 [FREE Full text] [doi: 10.1001/jama.2015.15845] [Medline: 26647259]

2. Rotenstein LS, Ramos MA, Torre M, Segal JB, Peluso MJ, Guille C, et al. Prevalence of depression, depressive symptoms, and suicidal ideation among medical students: a systematic review and meta-analysis. JAMA 2016 Dec 06;316(21):2214-2236 [FREE Full text] [doi: 10.1001/jama.2016.17324] [Medline: 27923088]
3. Dyrbye LN, Shanafelt TD, Sinsky CA, Cipriano PF, Bhatt J, Ommaya A, et al. Burnout among health care professionals: a call to explore and address this underrecognized threat to safe, high-quality care. National Academy of Medicine. 2017. URL: https://nam.edu/burnout-among-health-care-professionals-a-call-to-explore-and-address-this-underrecognized-threat-to-safe-high-quality-care/ [accessed 2021-09-10]
4. Essary AC, Bernard KS, Coplan B, Dehn R, Forister JG, Smith NE, et al. Burnout and job and career satisfaction in the physician assistant profession: a review of the literature. National Academy of Medicine. 2018. URL: https://nam.edu/burnout-and-job-and-career-satisfaction-in-the-physician-assistant-profession-a-review-of-the-literature/ [accessed 2021-09-10]
5. Tetzlaff ED, Hylton HM, DeMora L, Ruth K, Wong Y. National study of burnout and career satisfaction among physician assistants in oncology: implications for team-based care. J Oncol Pract 2018 Jan;14(1):e11-e22 [FREE Full text] [doi: 10.1200/JOP.2017.025544] [Medline: 29190209]
6. Varner D, Foutch B. Depression and burnout symptoms among Air Force family medicine providers. JAAPA 2014 May;27(5):42-46. [doi: 10.1097/01.JAA.0000446373.63790.c9] [Medline: 24758978]
7. Johnson AK, Blackstone SR, Simmons W, Skelly A. Assessing burnout and interest in wellness programs in physician assistant students. J Physician Assist Educ 2020 Jun;31(2):56-62. [doi: 10.1097/JPA.0000000000000303] [Medline: 32459443]
8. Demarzo MM, Montero-Marin J, Cuijpers P, Zabaleta-del-Olmo E, Mahtani KR, Vellinga A, et al. The efficacy of mindfulness-based interventions in primary care: a meta-analytic review. Ann Fam Med 2015 Nov;13(6):573-582 [FREE Full text] [doi: 10.1370/afm.1863] [Medline: 26553897]
9. Goyal M, Singh S, Sibinga EM, Gould NF, Rowland-Seymour A, Sharma R, et al. Meditation programs for psychological stress and well-being: a systematic review and meta-analysis. JAMA Intern Med 2014 Mar;174(3):357-368 [FREE Full text] [doi: 10.1001/jamainternmed.2013.13018] [Medline: 24395196]
10. Creswell JD, Lindsay EK. How does mindfulness training affect health? A mindfulness stress buffering account. Curr Dir Psychol Sci 2014 Dec 16;23(6):401-407. [doi: 10.1177/0963721414547415]
11. Black D, Slavich G. Mindfulness meditation and the immune system: a systematic review of randomized controlled trials. Ann N Y Acad Sci 2016 Jun;1373(1):13-24 [FREE Full text] [doi: 10.1111/nyas.12998] [Medline: 26799456]
12. Cásedas L, Pirruccio V, Vadillo MA, Lupiáñez J. Does mindfulness meditation training enhance executive control? A systematic review and meta-analysis of randomized controlled trials in adults. Mindfulness 2019 Dec 10;11(2):411-424. [doi: 10.1007/s12671-019-01279-4]
13. Ong JC, Moore C. What do we really know about mindfulness and sleep health? Curr Opin Psychol 2020 Aug;34:18-22. [doi: 10.1016/j.copsyc.2019.08.020] [Medline: 31539830]
14. Winbush NY, Gross CR, Kreitzer MJ. The effects of mindfulness-based stress reduction on sleep disturbance: a systematic review. Explore (NY) 2007;3(6):585-591. [doi: 10.1016/j.explore.2007.08.003] [Medline: 18005910]
15. Neuendorf R, Wahbeh H, Chamine I, Yu J, Hutchison K, Oken BS. The effects of mind-body interventions on sleep quality: a systematic review. Evid Based Complement Alternat Med 2015;2015:902708 [FREE Full text] [doi: 10.1155/2015/902708] [Medline: 26161128]
16. Sen S, Kranzler H, Didwania A, Schwartz A, Amarnath S, Kolars J, et al. Effects of the 2011 duty hour reforms on interns and their patients: a prospective longitudinal cohort study. JAMA Intern Med 2013 Apr 22;173(8):657-62; discussion 663 [FREE Full text] [doi: 10.1001/jamainternmed.2013.351] [Medline: 23529201]
17. Dobkin PL, Hutchinson TA. Teaching mindfulness in medical school: where are we now and where are we going? Med Educ 2013 Aug;47(8):768-779. [doi: 10.1111/medu.12200] [Medline: 23837423]
18. McConville J, McAleer R, Hahne A. Mindfulness training for health profession students-the effect of mindfulness training on psychological well-being, learning and clinical performance of health professional students: a systematic review of randomized and non-randomized controlled trials. Explore (NY) 2017;13(1):26-45. [doi: 10.1016/j.explore.2016.10.002] [Medline: 27889445]
19. Lomas T, Medina JC, Ivtzan I, Rupprecht S, Eiroa-Orosa FJ. A systematic review and meta-analysis of the impact of mindfulness-based interventions on the well-being of healthcare professionals. Mindfulness 2018 Nov 29;10(7):1193-1216. [doi: 10.1007/s12671-018-1062-5]
20. Hoover EB, Butaney B, Stoehr JD. Exploring the effectiveness of mindfulness and decentering training in a physician assistant curriculum. J Physician Assist Educ 2020 Mar;31(1):19-22. [doi: 10.1097/JPA.0000000000000288] [Medline: 31977963]
21. Creswell JD, Taren AA, Lindsay EK, Greco CM, Gianaros PJ, Fairgrieve A, et al. Alterations in resting-state functional connectivity link mindfulness meditation with reduced interleukin-6: a randomized controlled trial. Biol Psychiatry 2016 Jul 01;80(1):53-61. [doi: 10.1016/j.biopsych.2016.01.008] [Medline: 27021514]

22. Mooneyham B, Mrazek M, Mrazek A, Schooler J. Signal or noise: brain network interactions underlying the experience and training of mindfulness. Ann N Y Acad Sci 2016 Apr;1369(1):240-256. [doi: 10.1111/nyas.13044] [Medline: 27038003]
23. Hölzel BK, Hoge EA, Greve DN, Gard T, Creswell JD, Brown KW, et al. Neural mechanisms of symptom improvements in generalized anxiety disorder following mindfulness training. Neuroimage Clin 2013 Mar 25;2:448-458 [FREE Full text] [doi: 10.1016/j.nicl.2013.03.011] [Medline: 24179799]
24. Doll A, Hölzel BK, Mulej Bratec S, Boucard CC, Xie X, Wohlschläger AM, et al. Mindful attention to breath regulates emotions via increased amygdala-prefrontal cortex connectivity. Neuroimage 2016 Jul 01;134:305-313. [doi: 10.1016/j.neuroimage.2016.03.041] [Medline: 27033686]
25. Modinos G, Ormel J, Aleman A. Individual differences in dispositional mindfulness and brain activity involved in reappraisal of emotion. Soc Cogn Affect Neurosci 2010 Dec;5(4):369-377 [FREE Full text] [doi: 10.1093/scan/nsq006] [Medline: 20147457]
26. Fang H, Tu S, Sheng J, Shao A. Depression in sleep disturbance: a review on a bidirectional relationship, mechanisms and treatment. J Cell Mol Med 2019 Apr;23(4):2324-2332 [FREE Full text] [doi: 10.1111/jcmm.14170] [Medline: 30734486]
27. Harris D, Warren J, Adler C. Meditation for Fidgety Skeptics: a 10% Happier How-To Book. New York: Harmony; 2018.
28. Ten percent happier. tenpercent.com. URL: https://www.tenpercent.com [accessed 2021-09-10]
29. The basic of mindfulness meditation. Ten Percent Happier. URL: https://www.tenpercent.com/mindfulness-meditation-the-basics [accessed 2021-09-10]
30. Clark CM, Barbosa-Leiker C, Gill LM, Nguyen D. Revision and psychometric testing of the Incivility in Nursing Education (INE) survey: introducing the INE-R. J Nurs Educ 2015 Jun;54(6):306-315. [doi: 10.3928/01484834-20150515-01] [Medline: 26057424]
31. Salmela-Aro K, Kiuru N, Leskinen E, Nurmi J. School Burnout Inventory (SBI): reliability and validity. Eur J Psychol Assess 2009 Jan;25(1):48-57. [doi: 10.1027/1015-5759.25.1.48]
32. Lovibond P, Lovibond S. The structure of negative emotional states: comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. Behav Res Ther 1995 Mar;33(3):335-343. [doi: 10.1016/0005-7967(94)00075-u] [Medline: 7726811]
33. Hanish A, Lin-Dyken D, Han J. PROMIS sleep disturbance and sleep-related impairment in adolescents: examining psychometrics using self-report and actigraphy. Nurs Res 2017;66(3):246-251 [FREE Full text] [doi: 10.1097/NNR.0000000000000217] [Medline: 28448375]
34. Sleep-related impairment: a brief guide to the PROMIS® Sleep-Related Impairment instruments. PROMIS. 2020. URL: https://www.healthmeasures.net/images/PROMIS/manuals/PROMIS_Sleep-Related_Impairment_Scoring_Manual.pdf [accessed 2021-09-10]
35. Whitfield-Gabrieli S, Nieto-Castanon A. Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks. Brain Connect 2012;2(3):125-141. [doi: 10.1089/brain.2012.0073] [Medline: 22642651]
36. Ashburner J, Barnes G, Chen CC, Daunizeau J, Flandin G, Friston K, et al. SPM12 Manual. In: SPM12 Manual. London, UK: Wellcome Trust Centre for Neuroimaging; 2014.
37. Penny W, Friston K, Ashburner J, Kiebel S, Nichols T, editors. Statistical Parametric Mapping: the Analysis of Functional Brain Images. London, UK: Elsevier; 2011.
38. Behzadi Y, Restom K, Liau J, Liu TT. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. Neuroimage 2007 Aug 01;37(1):90-101 [FREE Full text] [doi: 10.1016/j.neuroimage.2007.04.042] [Medline: 17560126]
39. Fox MD, Snyder AZ, McAvoy MP, Barch DM, Raichle ME. The BOLD onset transient: identification of novel functional differences in schizophrenia. Neuroimage 2005 Apr 15;25(3):771-782. [doi: 10.1016/j.neuroimage.2004.12.025] [Medline: 15808978]
40. Fox MD, Zhang D, Snyder AZ, Raichle ME. The global signal and observed anticorrelated resting state brain networks. J Neurophysiol 2009 Jun;101(6):3270-3283 [FREE Full text] [doi: 10.1152/jn.90777.2008] [Medline: 19339462]
41. Fransson P. How default is the default mode of brain function? Further evidence from intrinsic BOLD signal fluctuations. Neuropsychologia 2006;44(14):2836-2845. [doi: 10.1016/j.neuropsychologia.2006.06.017] [Medline: 16879844]
42. Chai XJ, Castañón AN, Ongür D, Whitfield-Gabrieli S. Anticorrelations in resting state networks without global signal regression. Neuroimage 2012 Jan 16;59(2):1420-1428 [FREE Full text] [doi: 10.1016/j.neuroimage.2011.08.048] [Medline: 21889994]
43. Norman KA, Polyn SM, Detre GJ, Haxby JV. Beyond mind-reading: multi-voxel pattern analysis of fMRI data. Trends Cogn Sci 2006 Sep;10(9):424-430. [doi: 10.1016/j.tics.2006.07.005] [Medline: 16899397]
44. Worsley KJ, Marrett S, Neelin P, Vandal AC, Friston KJ, Evans AC. A unified statistical approach for determining significant signals in images of cerebral activation. Hum Brain Mapp 1996;4(1):58-73. [doi: 10.1002/(SICI)1097-0193(1996)4:1<58::AID-HBM4>3.0.CO;2-O] [Medline: 20408186]
45. Friston KJ, Holmes AP, Worsley KJ, Poline J, Frith CD, Frackowiak RS. Statistical parametric maps in functional imaging: a general linear approach. Hum Brain Mapp 1994;2(4):189-210. [doi: 10.1002/hbm.460020402]

46. Poline J, Worsley KJ, Holmes AP, Frackowiak RS, Friston KJ. Estimating smoothness in statistical parametric maps: variability of P values. J Comput Assist Tomogr 1995;19(5):788-796. [doi: 10.1097/00004728-199509000-00017] [Medline: 7560327]
47. Chumbley J, Worsley K, Flandin G, Friston K. Topological FDR for neuroimaging. Neuroimage 2010 Feb 15;49(4):3057-3064 [FREE Full text] [doi: 10.1016/j.neuroimage.2009.10.090] [Medline: 19944173]
48. Graham JW. Missing data analysis: making it work in the real world. Annu Rev Psychol 2009;60(1):549-576. [doi: 10.1146/annurev.psych.58.110405.085530] [Medline: 18652544]
49. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Series B Stat Methodol 1995;57(1):289-300. [doi: 10.1111/j.2517-6161.1995.tb02031.x]
50. von Economo EC, Freiherr C, Koskinas G, Triarhou LC. Atlas of Cytoarchitectonics of the Adult Human Cerebral Cortex. Basel: Karger; 2008.
51. Assem M, Glasser M, Van Essen DC, Duncan J. A domain-general cognitive core defined in multimodally parcellated human cortex. Cereb Cortex 2020 Jun 30;30(8):4361-4380 [FREE Full text] [doi: 10.1093/cercor/bhaa023] [Medline: 32244253]
52. The Hidden Epidemic: The Moral Imperative for Academic Health Centers to Address Health Professionals' Well-Being. Atlanta: Emory University Woodruff Health Sciences Center; Jun 2017.
53. Rathert C, Williams ES, Linhart H. Evidence for the quadruple aim: a systematic review of the literature on physician burnout and patient outcomes. Med Care 2018 Dec;56(12):976-984. [doi: 10.1097/MLR.0000000000000999] [Medline: 30339573]
54. Del Canale S, Louis DZ, Maio V, Wang X, Rossi G, Hojat M, et al. The relationship between physician empathy and disease complications: an empirical study of primary care physicians and their diabetic patients in Parma, Italy. Acad Med 2012 Sep;87(9):1243-1249. [doi: 10.1097/ACM.0b013e3182628fbf] [Medline: 22836852]
55. Rakel DP, Hoeft TJ, Barrett BP, Chewning B, Craig B, Niu M. Practitioner empathy and the duration of the common cold. Fam Med 2009;41(7):494-501 [FREE Full text] [Medline: 19582635]
56. Cockerham W, Hinote B. PAs in a changing society: a sociologic perspective. JAAPA 2015 Aug;28(8):18-20. [doi: 10.1097/01.JAA.0000469268.71310.bf] [Medline: 26164046]
57. Essary A, Coplan B, Cawley J, Schneller E, Ohsfeldt R. Women, family medicine, and career choice: an opportunity cost analysis. JAAPA 2016 Sep;29(9):44-48. [doi: 10.1097/01.JAA.0000490949.02814.bc] [Medline: 27575904]
58. Hooker R, Robie S, Coombs J, Cawley J. The changing physician assistant profession: a gender shift. JAAPA 2013 Sep;26(9):36-44. [doi: 10.1097/01.jaa.0000433914.54617.a0] [Medline: 24069670]
59. Roy A, Druker S, Hoge EA, Brewer JA. Physician anxiety and burnout: symptom correlates and a prospective pilot study of app-delivered mindfulness training. JMIR Mhealth Uhealth 2020 Apr 01;8(4):e15608 [FREE Full text] [doi: 10.2196/15608] [Medline: 32234708]
60. Huberty J, Eckert R, Puzia M, Laird B, Larkey L, Mesa R. A novel educational control group mobile app for meditation interventions: single-group feasibility trial. JMIR Form Res 2020 Jul 21;4(7):e19364 [FREE Full text] [doi: 10.2196/19364] [Medline: 32706719]
61. Online Assessment Measures: LEVEL 2—Sleep Disturbance—Adult. American Psychiatric Association. 2012. URL: https://www.psychiatry.org/psychiatrists/practice/dsm/educational-resources/assessment-measures [accessed 2021-09-10]
62. Kurth F, Luders E, Wu B, Black DS. Brain gray matter changes associated with mindfulness meditation in older adults: an exploratory pilot study using voxel-based morphometry. Neuro 2014;1(1):23-26 [FREE Full text] [doi: 10.17140/NOJ-1-106] [Medline: 25632405]
63. Huberty J, Vranceanu A, Carney C, Breus M, Gordon M, Puzia ME. Characteristics and usage patterns among 12,151 paid subscribers of the calm meditation app: cross-sectional survey. JMIR Mhealth Uhealth 2019 Nov 03;7(11):e15648 [FREE Full text] [doi: 10.2196/15648] [Medline: 31682582]
64. Wasson LT, Cusmano A, Meli L, Louh I, Falzon L, Hampsey M, et al. Association between learning environment interventions and medical student well-being: a systematic review. JAMA 2016 Dec 06;316(21):2237-2252 [FREE Full text] [doi: 10.1001/jama.2016.17573] [Medline: 27923091]
65. Kalmbach D, Arnedt J, Song P, Guille C, Sen S. Sleep disturbance and short sleep as risk factors for depression and perceived medical errors in first-year residents. Sleep 2017 Mar 01;40(3):zsw073 [FREE Full text] [doi: 10.1093/sleep/zsw073] [Medline: 28369654]
66. Rosen IM, Gimotty PA, Shea JA, Bellini LM. Evolution of sleep quantity, sleep deprivation, mood disturbances, empathy, and burnout among interns. Acad Med 2006 Jan;81(1):82-85. [doi: 10.1097/00001888-200601000-00020] [Medline: 16377826]
67. Nie X, Shao Y, Liu SY, Li HJ, Wan AL, Nie S, et al. Functional connectivity of paired default mode network subregions in primary insomnia. Neuropsychiatr Dis Treat 2015 Dec 16;11:3085-3093 [FREE Full text] [doi: 10.2147/NDT.S95224] [Medline: 26719693]
68. Khazaie H, Veronese M, Noori K, Emamian F, Zarei M, Ashkan K, et al. Functional reorganization in obstructive sleep apnoea and insomnia: a systematic review of the resting-state fMRI. Neurosci Biobehav Rev 2017 Jun;77:219-231 [FREE Full text] [doi: 10.1016/j.neubiorev.2017.03.013] [Medline: 28344075]

69. Pace-Schott EF, Zimmerman JP, Bottary RM, Lee EG, Milad MR, Camprodon JA. Resting state functional connectivity in primary insomnia, generalized anxiety disorder and controls. Psychiatry Res Neuroimaging 2017 Jul 30;265:26-34 [FREE Full text] [doi: 10.1016/j.pscychresns.2017.05.003] [Medline: 28500965]
70. Lapinski J, Yost M, Sexton P, LaBaere RJ. Factors modifying burnout in osteopathic medical students. Acad Psychiatry 2016 Feb 25;40(1):55-62. [doi: 10.1007/s40596-015-0375-0] [Medline: 26108394]
71. Williford ML, Scarlet S, Meyers MO, Luckett DJ, Fine JP, Goettler CE, et al. Multiple-institution comparison of resident and faculty perceptions of burnout and depression during surgical training. JAMA Surg 2018 Aug 01;153(8):705-711 [FREE Full text] [doi: 10.1001/jamasurg.2018.0974] [Medline: 29800976]
72. Slavin SJ. Medical student mental health: culture, environment, and the need for change. JAMA 2016 Dec 06;316(21):2195-2196. [doi: 10.1001/jama.2016.16396] [Medline: 27923076]
73. Dai X, Nie X, Liu X, Pei L, Jiang J, Peng D, et al. Gender differences in regional brain activity in patients with chronic primary insomnia: evidence from a resting-state fMRI study. J Clin Sleep Med 2016 Mar 15;12(3):363-374 [FREE Full text] [doi: 10.5664/jcsm.5586] [Medline: 26715399]

Abbreviations

ANOVA: analysis of variance
DMN: default mode network
FDR: false discovery rate
fMRI: functional magnetic resonance imaging
GLM: general linear model
PA: physician assistant
PROMIS: Patient-Reported Outcomes Measurement Information System
ROI: region of interest
rsfMRI: resting-state fMRI
SI: sleep-related impairment
SN: salience network
TR: repetition time interval

Edited by J Torous; submitted 09.09.20; peer-reviewed by J Huckins, H Cramer, K Sanada; comments to author 14.12.20; revised version received 23.04.21; accepted 23.07.21; published 19.10.21.

Please cite as:
Smith JL, Allen JW, Haack CI, Wehrmeyer KL, Alden KG, Lund MB, Mascaro JS
Impact of App-Delivered Mindfulness Meditation on Functional Connectivity, Mental Health, and Sleep Disturbances Among Physician Assistant Students: Randomized, Wait-list Controlled Pilot Study
JMIR Form Res 2021;5(10):e24208
URL: https://formative.jmir.org/2021/10/e24208
doi:10.2196/24208
PMID:34665153

Original Paper

Reach Outcomes and Costs of Different Physician Referral Strategies for a Weight Management Program Among Rural Primary Care Patients: Type 3 Hybrid Effectiveness-Implementation Trial

Gwenndolyn Porter[1], PhD; Tzeyu L Michaud[2], PhD; Robert J Schwab[3], MD; Jennie L Hill[4], PhD; Paul A Estabrooks[1], PhD

[1]Department of Health Promotion, University of Nebraska Medical Center, Omaha, NE, United States

[2]Center for Reducing Health Disparities, University of Nebraska Medical Center, Omaha, NE, United States

[3]University of Nebraska Medical Center, Omaha, NE, United States

[4]Department of Population Health Sciences, University of Utah, Salt Lake City, UT, United States

Corresponding Author:
Gwenndolyn Porter, PhD
Department of Health Promotion
University of Nebraska Medical Center
984365 Nebraska Medical Center
Omaha, NE, 68198
United States
Phone: 1 4025591082
Email: gwenndolyn.porter@unmc.edu

Abstract

Background: Rural residents are at high risk for obesity; however, little resources exist to address this disproportional burden of disease. Primary care may provide an opportunity to connect primary care patients with overweight and obesity to effective weight management programming.

Objective: The purpose of this study is to examine the utility of different physician referral and engagement processes for improving the reach of an evidence-based and technology-delivered weight management program with counseling support for rural primary care patients.

Methods: A total of 5 rural primary care physicians were randomly assigned a sequence of four referral strategies: point-of-care (POC) referral with active telephone follow-up (ATF); POC referral, no ATF; a population health registry–derived letter referral with ATF; and letter referral, no ATF. For registry-derived referrals, physicians screened a list of patients with BMI ≥25 and approved patients for participation to receive a personalized referral letter via mail.

Results: Out of a potential 991 referrals, 573 (57.8%) referrals were made over 16 weeks, and 98 (9.9%) patients were enrolled in the program (58/98, 59.2% female). Differences based on letter (485/991, 48.9%) versus POC (506/991, 51.1%) referrals were identified for completion (100% vs 7%; $P<.001$) and for proportion screened (36% vs 12%; $P<.001$) but not for proportion enrolled (12% vs 8%; $P=.10$). Patients receiving ATF were more likely to be screened (47% vs 7%; $P<.001$) and enrolled (15% vs 7%; $P<.001$) than those not receiving ATF. On the basis of the number of referrals made in each condition, we found variations in the proportion and number of enrollees (POC with ATF: 27/190, 50%; POC no ATF: 14/316, 41%; letter ATF: 30/199; 15.1%; letter no ATF: 27/286, 9.4%). Across all conditions, participants were representative of the racial and ethnic characteristics of the region (60% female, $P=.15$; 94% White individuals, $P=.60$; 94% non-Hispanic, $P=.19$). Recruitment costs totaled US $6192, and the overall recruitment cost per enrolled participant was US $63. Cost per enrolled participant ranged from POC with ATF (US $47), registry-derived letter without ATF (US $52), and POC without ATF (US $56) to registry-derived letter with ATF (US $91).

Conclusions: Letter referral with ATF appears to be the best option for enrolling a large number of patients in a digitally delivered weight management program; however, POC with ATF and letters without ATF yielded similar numbers at a lower cost. The best referral option is likely dependent on the best fit with clinical resources.

Trial Registration: ClinicalTrials.gov NCT03690557; http://clinicaltrials.gov/ct2/show/NCT03690557

(JMIR Form Res 2021;5(10):e28622) doi:10.2196/28622

KEYWORDS

weight management; rural; RE-AIM; hybrid effectiveness-implementation; primary care; obesity; physicians; digital health; health technology; mobile phone

Introduction

Background

Obesity is a pressing health concern nationwide, particularly in small rural communities. Rural residents are more obese, on average, than their urban counterparts [1] and often have no or limited access to obesity prevention and treatment programming [1,2]. Furthermore, individuals who use primary care are proportionally more obese than the general public [3], highlighting rural primary care patients as a high-need population regarding weight management. Primary care providers are often the only resource to support healthful eating, physical activity, and weight management in rural communities [4]. Primary care systems may offer a practical and sustainable venue for implementing evidence-based weight management interventions; however, little is known about how rural primary care physicians can pragmatically refer and enroll a large and representative group of individuals into an evidence-based weight management program.

A challenge for weight management interventions is ensuring that not only is an intervention effective but it also has the potential to reach populations at risk that could most benefit [5]. Factors beyond program effectiveness and total sample size, such as proportional yield from those recruited and sample representativeness, are important indicators of an intervention's impact [6]. Few studies provide a comprehensive report on the methods used to recruit participants, and even fewer report on the representativeness of the sample when compared with the target population [7,8]. In a systematic review of rural weight loss interventions, only 2 of 53 studies compared the demographic characteristics of the intervention sample with those of the target population [8]. In a 2016 systematic review of recruitment strategies for young adult weight gain prevention interventions, 23 of 25 studies were reported to have insufficiently described the recruitment process [7]. Documenting recruitment methods and the representativeness outcomes of those efforts can lead to a better overall understanding of how best to engage patients and maximize the reach of evidence-based behavioral interventions. This understanding is important for scale-up efforts, for preventing underrepresentation of populations experiencing disparities in research and in health care, and for physicians to improve their standard of care [9].

However, when drawing more broadly from the behavioral intervention literature, a number of examples of both active and passive recruitment strategies can be found that have been used to engage various target populations [7,10-13]. On the basis of the available literature, active recruitment strategies—those with direct interaction with potential participants, such as outreach telephone calls—appear to yield a lower absolute number of participants but a higher proportion of those exposed to recruitment strategies when compared with passive recruitment strategies (ie, those without direct interaction with potential participants, such as flyers or targeted mailings), which yield a higher number but a lower proportion of participants [5]. Active recruitment strategies may also yield a more representative sample than passive recruitment strategies [12,14,15]. The limited knowledge in this area warrants further study.

Objectives

The purpose of this study is to examine the utility and cost of different physician referral and engagement processes for improving reach (ie, number, proportion, and representativeness of participants) for a rural, evidence-based, and technology-delivered weight management program with counseling support. Our design is based on feedback gathered from prior qualitative work [16,17]; that is, an expressed clinical interest in testing the relative reach of engaging patients at the point of care (POC) during a well or chronic care visit or proactively reaching out to patients using a population health management approach facilitated by an electronic health record.

We use the RE-AIM (Reach, Effectiveness, Adoption, Implementation, and Maintenance) framework to guide our assessment of reach [18] and hypothesize that active recruitment strategies would yield a more representative sample than passive recruitment strategies. Reach refers to the absolute number, proportion, and representativeness of the participants compared with those who were exposed to recruitment efforts [18]. We also hypothesize that POC referrals, compared with the population health registry–derived letter referrals, would yield a higher proportion of enrolled participants from those who received a referral.

Methods

Intervention Selection

Before this pilot trial, we conducted focus groups with primary care staff employed at rural primary care clinics regarding the feasibility of implementing a weight loss program through primary care. Overwhelmingly, primary care staff agreed that a program to which physicians and nurses could refer eligible patients and track their progress throughout was more favorable than a program that would require physician- or nurse-led delivery [16,17]. We then assembled primary care physicians, staff, and obesity treatment experts from the Great Plains Practice-Based Research Network to engage in a participatory selection process [19] to identify and, if necessary, adapt an evidence-based intervention to pilot-test through a rural primary care clinic. The selection process is described elsewhere [17]. Ultimately, a digitally delivered intervention was selected by the group for local testing, and it was agreed that testing potential recruitment strategies was a priority over testing the relative effectiveness of the different weight loss programs.

Study Design

Overview

The study design used the hybrid methodologies described by Curran et al [20], which allow for blended design components targeting effectiveness and implementation but focus on systems-based approaches to improve dissemination at the participant level rather than implementing the evidence-based intervention. As such, we classified this trial as a hybrid type 3 effectiveness-implementation trial focused on dissemination at the participant level (ie, reach). This allowed us to test the utility of different dissemination strategies to increase program reach as a primary outcome while concurrently gathering information on intervention effectiveness [20]. We partnered with a rural primary care clinic and the 5 physicians serving their patients to implement an evidence-based and digitally delivered weight loss program. The 5 physicians were randomly assigned to a sequence of four referral strategies over a span of 16 weeks: POC referral with active telephone follow-up (ATF); POC referral, no ATF; a population health registry–derived letter referral with ATF; and letter referral, no ATF.

Dissemination Strategies: Referral Methods

Referral strategies varied by POC versus a population health electronic health record–derived letter referral and ATF versus no telephone follow-up. For POC referral, physicians were instructed to refer any adult patient to an evidence-based weight loss program with diet and physical activity counseling based on (1) BMI≥25 kg/m^2, (2) no contraindications to participation, and (3) visiting the clinic for a chronic care or well visit. In the population health approach, a clinic administrator pulled a list of patients from the electronic health record system with BMI≥25 who had visited the clinic in the previous 2 weeks. Each physician reviewed this list and removed any patients with contraindications to participation in the evidence-based weight loss program. The remaining patients were mailed a personalized invitation letter to participate in the weight loss intervention that was signed by their physician.

For ATF, the referred patient was informed that they would be contacted by a member of the research team to determine if they would like to participate in the program (an opt-out telephone number was also provided). In conditions without ATF, the referred patient was provided a telephone number to call if they were interested in discussing participation in the program. The study was reviewed and approved by the University of Nebraska Medical Center Institutional Review Board (#581-18-EP).

We designed this pilot study to recruit over a 16-week period, with the goal of enrolling approximately 100 participants and projected a 10% enrollment rate based on the overall denominator of potential referrals based on prior work [21,22]. Each physician (n=5) was randomly assigned a sequence of the four referral strategies, shifting strategies every 2 weeks (Table 1). Over the course of 16 weeks, physicians used each of the four referral strategies twice to eliminate any potential time effect on a single strategy. Each physician's referral strategy sequence was randomized to prevent any order effect on the yield of patients per referral strategy. A member of the research team visited the clinic every 2 weeks to remind physicians of their new referral strategy; a nurse at the clinic monitored the physicians daily to ensure that the referral strategies were implemented with fidelity.

Table 1. Randomization sequence of referral strategy by physician.

Weeks	Physician A	Physician B	Physician C	Physician D	Physician E
1-2	L[a]	POC+[b]	POC[c]	L+[d]	POC+
3-4	POC+	POC	L+	L	L+
5-6	L+	L+	L	POC	L
7-8	POC	L	POC+	POC+	POC
9-10	L+	POC+	L+	POC	L
11-12	L	POC	POC+	L	L+
13-14	POC+	L+	L	L+	POC
15-16	POC	L	POC	POC+	POC+

[a]L: letter referral without active telephone follow-up.

[b]POC+: point-of-care referral with active telephone follow-up.

[c]POC: point-of-care referral without active telephone follow-up.

[d]L+: letter referral with active telephone follow-up.

Enrollment

All referred and interested patients were required to undergo a brief telephone screening with a member of the research team and, if eligible, were given instructions over the phone and via email on how to complete web-based program enrollment. In addition, all interested and eligible patients were scheduled for an in-person enrollment visit with a member of the research team where they were given intervention program materials (ie, home scale) and a tutorial of the intervention's mobile app, provided with tips on how to stay engaged with the program, and completed the web-based program enrollment (if not already complete). Patient consent to participate in the weight loss program was obtained during the web-based enrollment process. Patients were given the opportunity to raise questions before consent and during the in-person enrollment visit.

Evidence-Based Weight Loss Program

All referred patients were offered a 12-month, digitally delivered, evidence-based weight management program free of charge. The program featured a social cognitive theory–based curriculum with counseling support delivered through daily emails and text messages. Program features also included daily meal plans and physical activity recommendations [23]. Estabrooks et al [23] provided a comprehensive overview of the program. In addition, modest financial incentives were offered with the intent to increase program reach and retention [24]. Incentives were offered to participants who lost a minimum of 5% of their initial body weight (US $15/quarter reward) graded up to a maximum of 30% body weight reduction (US $150/quarter reward). Program participants were provided a Bluetooth-enabled home scale (Smart Scale, incentaHEALTH, LLC), which connected to the program smartphone app that was installed on their smartphones during the enrollment visit. Participants were instructed to record their weight using this scale no less than once per quarter. The program also featured a website where participants could receive feedback on their weigh-ins, take health quizzes, and self-assess their progress with regard to healthy eating, physical activity, and weight loss. Upon completion of the program, participants could keep the home scale. This paper focuses on reach, which is the primary outcome of this study; data on weight loss are not presented here.

Data Analysis

Program reach was measured by the number and proportion of individuals who were (1) referred, (2) screened, and (3) enrolled and was compared across referral strategies relative to the number of eligible patients who visited the clinic during the recruitment period. Representativeness of the enrolled sample was assessed relative to the demographic characteristics of the region, as measured by the US Census [25]. We used chi-square tests to examine group differences in terms of screening and participation rates among referral strategies and to determine whether screening and participation rates differed according to (1) ATF versus no ATF and (2) POC versus letter. We applied a one-sample test of proportion (in the case of comparing proportions) to examine representativeness in terms of demographic characteristics of the enrolled participants compared with census data. We further conducted group comparisons using one-way analysis of variance tests for continuous variables and chi-square tests for categorical variables. Two-tailed *P* values <.05 were considered statistically significant for this study.

Costs of recruitment were prospectively and retrospectively estimated based on the costs of recruitment materials (eg, handouts or flyers), supplies, and recruitment activities, including telephone follow-up, telephone screening, and in-person sessions. Labor costs were calculated using the research assistant (RA) annual salary (US $25,000) and publicly available average salary estimates for primary care physicians (US $187,013) and clinic managers (US $65,356). RAs tracked the time spent on various recruitment activities, number of phone calls made, enrollment sessions completed, and number of flyers printed in a custom computer database. Cost results are presented for each referral strategy.

Results

Reach

Over a period of 16 weeks, 2534 eligible patients visited the clinic. The maximum number of referrals that could have been made during well or chronic care visits over this time were approximately 30% of patient visits (n=991; hereafter referred to as potential referrals). The actual number of referrals made by the 5 physicians was 573 (274, 47.8% women; average age 55.7, SD 16.8 years) and, out of the 573 referrals, 98 (17.1%) patients were enrolled, representing an overall enrollment rate of 10% of the potentially eligible patient population. Of the 485 potential letter referrals, 485 (100%; 46% women; average age 56.3 years) were completed, 229 (47.2%; 48% women; average age 56.1 years) patients were screened by telephone, and 57 (11.8%; 58% women; average age 55.8 years) were enrolled. Of the 506 potential POC referrals, 88 (17.4%; 56% women; average age 52.0 years) patients were referred, 60 (11.9%; 57% women; average age 51.6 years) were screened, and 41 (8.1%; 61% women; average age 47.6 years) were enrolled. Patients receiving ATF were more likely to be screened (49% vs 7%; $P<.001$) and enrolled (15% vs 7%; $P<.001$) than those without ATF. Chi-square test results revealed significant differences in terms of patient screening ($\chi^2_3[573]=238.6$; $P<.001$) and enrollment status ($\chi^2_3[573]=69.2$; $P<.001$) among referral strategies (Table 2). Specifically, based on the number of referrals completed, there were variations in the proportion and absolute number of enrollees among the four referrals strategies (POC with ATF: 27/190, 50%; POC no ATF: 14/316, 41%; letter ATF: 30/199; 15.1%; letter no ATF: 26/286, 9.1%). Table 2 outlines the number and proportion of potential referrals, referrals made, patients screened, and patients enrolled.

Table 2. Reach results by referral strategy (N=991).

Strategy	Potential referrals	Referrals made, n (%)	Patients screened, n (%)	Patients screened and eligible, n (%)	Patients enrolled, n (%)
POC[a]+ATF[b]	190	54 (28.4)	45 (23.7)[c,d,e]	27 (14.2)	27 (14.2)[c]
POC	316	34 (10.8)	14 (4.4)[d,e,f]	14 (4.4)	14 (4.4)[d,e,f]
Letter+ATF	199	199 (100)	147 (73.9)[c,e,f]	30 (15.1)	30 (15.1)[c]
Letter	286	286 (100)	30 (10.5)[c,d,f]	27 (9.4)	26 (9.1)[c]
Total	991	573 (57.8)	236 (23.8)	98 (9.9)	97 (9.8)

[a]POC: point of care.

[b]ATF: active telephone follow-up.

[c]Significantly different from POC at α=.05.

[d]Significantly different from letter+ at α=.05.

[e]Significantly different from letter at α=.05.

[f]Significantly different from POC+ at α=.05.

Tables 3 and 4 present the demographic characteristics of participants and the target population and include comparisons of those characteristics among the referral strategies. Among those who enrolled, 59% were women, had an average age of 52.3 (SD 14.3) years, and an average BMI of 35.5 (SD 7.5) kg/m^2. When compared with the target population (Butler County, Nebraska), participants did not significantly differ among demographic characteristics. Comparing by referral strategies, participants differed significantly in age.

Table 3. Representativeness of study sample (N=97).

Demographics	Target population: Butler County, NE[a]	Total enrolled participants
Age, median (years)	43.3	55.0[b]
Sex (female), n (%)	4026 (50.1)	58 (59)
Obese at baseline, %	(32)[c]	75 (77)[b]
White, n (%)	7846 (97.7)	91 (94)
Hispanic or Latino, n (%)	287 (3.6)	3 (3)

[a]NE: Nebraska; data as per census.gov 2019 estimates.

[b]Significant difference between target population and sample at α=.05.

[c]Only percentage estimate available.

Table 4. Representativeness comparisons among participants by referral strategies (N=97).

Demographics	Total sample	POC[a]+ATF[b] (n=27)	POC no ATF (n=14)	Letter+ATF (n=30)	Letter no ATF (n=26)	Chi-square (*df*)	ANOVA[c] (*df*)	*P* value
Age (years), mean (SD)	52.3 (14.3)	46.0 (14.7)[d]	51.9 (9.3)	58.7[e] (13.1)	53.4 (15.1)	N/A[f]	F_{93}=4.1 (3)	.009
Sex (female), n (%)	58.8	59	64	57	58	0.2 (3)	N/A	.97
Baseline BMI mean (SD), kg/m^2	35.5 (7.5)	37.1 (6.8)	38.1 (9.1)	38.1 (9.1)	33.2 (6.6)	N/A	F_{72}=1.5 (3)	.22
Age >65, n (%)	23 (23.7)	7.4[d]	7.1	43.3[e]	26.9	12.6 (3)	N/A	.07
White individuals, n (%)	94	93	93	100	92	2.3 (3)	N/A	.51
Hispanic or Latino, n (%)	3	3.7	0	0	7.7	3.1 (3)	N/A	.37

[a]POC: point of care.

[b]ATF: active telephone follow-up.

[c]ANOVA: analysis of variance.

[d]Significantly different from letter+ATF at α=.05.

[e]Significantly different from POC+ATF at α=.05.

[f]N/A: not applicable.

Costs

Table 5 provides an overview and categorization of costs across referral strategies. Costs were determined based on nonlabor costs (US $738) and labor costs (US $5380), which were summed to provide the overall costs of recruitment (US $6118). Table 6 provides costs by referral strategy. All reported costs are rounded to the nearest dollar. We estimated an average of 2 minutes per POC referral based on anecdotal reports from the physicians; therefore, costs related to labor for POC referrals were estimated at US $162 for POC with ATF and US $102 for POC without ATF. Labor costs related to RA time spent making recruitment calls were the highest single line item costs and varied greatly between ATF referrals (US $1788) and referrals without ATF (US $135). The varying labor and nonlabor costs yielded different costs per enrolled participant via each referral strategy: letters with ATF was the costliest, at US $86 per enrolled participant, whereas POC with ATF (US $50), POC without ATF (US $61), and letter without ATF (US $51) costs were comparable.

Table 5. Breakdown of recruitment costs, rounded to the nearest dollar.

Cost element	Cost (US $)
Nonlabor	
Printing letters and program descriptions, n=970	43
Postage	267
Envelopes, n=485	58
POC cards	170
IT[a] (phone line)	200
Labor	
Physician time spent making referrals	264
Clinic manager time spent pulling patient list, 2 hours	63
Research assistant time	
Training clinical staff, 2 hours	48
Letter preparation, 24 hours	577
POC[b] preparation, 8 hours	192
Recruitment calls, 80 hours	1923
Enrollment visit preparation, 4 hours	96
Enrollment visit, 49 hours	1178
Mileage reimbursement to enrollment visits	1039
Total recruitment costs	6118
Total recruitment costs per enrolled participant	63

[a]IT: information technology.

[b]POC: point of contact.

Table 6. Recruitment costs in terms of yield by referral strategy, rounded to the nearest dollar (N=98).

Cost element	POC[a]+ (US $; n=27)	POC (US $; n=14)	Letter+ (US $; n=30)	Letter (US $; n=27)	Total sample (US $)
Nonlabor					
Printing letters and program descriptions, n=970	N/A[b]	N/A	18	25	43
Postage	N/A	N/A	110	157	267
Envelopes, n=485	N/A	N/A	24	34	58
POC cards	99	71	N/A	N/A	170
IT[c,d] (phone line)	50	50	50	50	200
Labor					
Physician time spent making referrals	162	102	0	0	264
Clinic manager time spent pulling patient list, 2 hours	0	0	31.50	31.50	63
Research assistant time					
Training clinical staff, 2 hours[d]	12	12	12	12	48
Letter preparation, 24 hours	N/A	N/A	237	340	577
POC preparation, 8 hours	73	119	N/A	N/A	192
Recruitment calls, 80 hours	346	58	1442	77	1923
Enrollment visit prep, 4 hours	24	24	24	24	96
Enrollment visit, 49 hours	324	165	365	324	1178
Mileage reimbursement to enrollment visits[d]	259.75	259.75	259.75	259.75	1039
Total recruitment costs	1350	861	2573	1334	6118
Total recruitment costs per enrolled participant	50	61	86	51	63

[a]POC: point of contact.

[b]N/A: not applicable.

[c]IT: information technology.

[d]These items or activities were not specific to any one referral strategy; therefore, their costs were split evenly among each of the referral strategies.

Discussion

Principal Findings

Our results provided some support for the hypothesis that a physician letter referral with ATF is most effective for enrolling a large number of participants into a digitally delivered weight management program compared with POC referrals with and without telephone follow-up and physician letter referrals without follow-up. However, when considering the combined outcomes of penetration into the target population, overall and proportional yield, and cost of recruitment, letter referral without ATF may be the most attractive option for rural primary care clinics that wish to recruit patients into a weight management program. Importantly, all (100%) of the 485 potential letter referrals were made, as compared with the 88 (17%) of the 506 potential POC referrals. Letter with ATF (n=30), letter without ATF (n=27), and POC with ATF (n=27) yielded a similar number of participants. Finally, the costs associated with the letter without ATF referral strategy were relatively low (US $51 per enrolled participant).

Owing to the financial constraints of this pilot study, we limited the number of potential referrals to 991, knowing that this would be far less than the total number of eligible patients seen by the 5 physicians during the recruitment period. Awareness of the finite number of referrals may have influenced the physicians' decision to refer a patient to the program, perhaps causing them to pass on referring an eligible patient to *save* that referral for a patient in greater need. However, we do not believe this influence was significant because not all of those potential referrals were used; <60% (573) of the potential referrals were distributed to eligible patients. Although not assessed in this study, primary care physicians often cite a lack of time during patient visits as a barrier to health counseling [26,27], which may also be the reason why the referral distribution rate was low among POC referrals. In addition, it is possible that a physician attempted to refer a patient during a clinic visit but the patient immediately declined the referral to the program. We were unable to quantitatively track this, which may have caused our POC distribution rate to reflect an underrepresentation of reality.

We calculated the proportional yield per referral strategy using potential referrals and referrals made as the denominator. These two analyses revealed different stories. When considering enrollees compared with potential referrals, the letter with active follow-up (15%) and POC with active follow-up (14%)

strategies appear to be the best for enrolling a large number of participants into the weight loss program. However, when considering enrollees compared with the number of referrals made, POC with active follow-up (50%) emerges as the clear leader for proportional yield. When considering costs, POC with ATF again appeals as the least expensive recruitment strategy (US $50 per participant), although letter referrals without ATF (US $51) had similar costs. The time-intensive nature of recruitment strategies with ATF not only drives up costs but also places a higher burden on recruitment personnel compared with strategies without ATF. In kind, the cumulative time associated with POC referrals may be similar to that of clinic administrator time spent preparing the letter referrals. Primary care clinics that are considering referring patients to a weight management program should reflect on these approaches with regard to clinical flow and costs. Resource costs in the large volume of letters and telephone calls for letter referral with ATF may make POC referral strategies with active follow-up appealing to small rural clinics, or if system processes are in place to make patient identification and mailing easy, the letter referral without ATF may be the most attractive.

Similar to previous studies [5,11], our findings supported our hypothesis that referrals with ATF would yield a higher proportion of enrollees from those who received a referral. Conversely, our results did not match prior literature [5] and did not support our hypothesis that ATF would yield a lower absolute number of participants. The two conditions with ATF yielded the highest absolute number of participants (letter with ATF, 30 participants; POC with ATF, 27 participants), whereas the letter-only condition yield was comparable (26 participants). Our results supported our hypothesis that POC referrals would have a higher proportional yield of enrolled participants from those who received a referral compared with those who received a letter referral. The behavioral intervention literature has not definitively ruled on a dominant recruitment strategy [10,15]. Our findings provide valuable insight into the reach, representativeness, and cost of POC referrals and electronic health record registry–derived referral letters with and without ATF.

Our sample was older and had a higher proportion of females than the target population. This is easily explained by the following: (1) our intervention was limited to adults aged ≥19 years, whereas the male age of Butler County represents children and adults; (2) women traditionally have higher health care use than men [28]; and (3) excluding infants aged <1 year, older adults aged ≥65 have the highest physician rate by age group [29]. Our sample did not differ on any other demographic variable that was measured, which aligns with representativeness comparisons of active and passive recruitment strategies made by Lee et al [14] but contrasts with results reported by Linnan et al [12], who found that passive strategies yielded lower enrollment but a more diverse and higher-risk sample. Furthermore, among the four referral strategies, those with ATF had the highest (58.7 years, letter with ATF) and lowest (46.0 years, POC with ATF) mean age among participants, therefore providing no support for our hypothesis that ATF strategies would yield a more representative sample than referral strategies without ATF. Although not a significant difference, baseline BMI among enrollees from both POC referral strategies was higher than both letter referral conditions. This may point to an unintentional bias by physicians to refer patients with much higher BMIs than is required for eligibility. Interestingly, POC strategies were more successful at enrolling younger patients (<65 years) than the letter strategies. Although the literature is scant on this specific outcome and drawing any generalizable conclusions from this finding is unjustified, it nonetheless provides an intriguing area for further investigation.

Identifying the best method for maximizing the reach of a weight management program is important for future scale-up efforts. However, maximizing participant retention is a critical next step in the scale-up of any behavioral intervention. An interesting future point of investigation will be the intervention attrition gap of our intervention—the proportion of patients who initially enroll in the program and subsequently do not engage with any program features. This is an especially curious investigation given the conflicting literature regarding the effectiveness of active versus passive recruitment strategies for retaining participants in behavioral interventions [12,13].

There are limitations to our study that should be considered when interpreting our results. As the selected intervention and referral methods were tailored to fit the needs of our target population, the findings of our feasibility study may not be fully generalizable to other primary care clinics. However, the process by which we selected an evidence-based intervention is likely translatable to other researchers and practitioners attempting to select and implement an evidence-based intervention. As previously mentioned, we were not able to track the number of times a physician attempted to make a referral during a clinic visit but was immediately rejected by the patient being seen. Anecdotally, we can attest that this was a rare occurrence, as the physicians reported that the majority of patients were willing to receive or inquire for more information about the program. We did not collect demographic or qualitative data from the 5 physicians involved in referrals and therefore cannot comment on their national representativeness to physicians caring for rural patients or their perceived lack of time for distributing referrals. In addition, census data used to compare our sample with the target population were drawn from Butler County, Nebraska. However, the clinic from which we recruited participants may draw patients from outside Butler County, and comparisons may not directly align. However, when our sample is compared with the demographic characteristics of all adjacent counties (9 total), it remains a representative sample. Indeed, our study has several strengths. Our sample was 59% female, which is a higher proportion of male participants than typical community weight loss programs [30]. Our sample was representative of the racial and ethnic characteristics of the region. Although this feasibility study is small in scale, it addresses key areas of translational research as defined by RE-AIM [31-33] as well as areas of focus cited as integral to feasibility studies, such as adaptation, acceptability, practicality, and integration [34].

Conclusions

We conclude that primary care physicians serving a rural community are capable of referring patients to a digitally

delivered behavioral weight management intervention through in-person and population health record–generated letter referrals. When compared with POC methods, letter referrals were more effective at achieving high penetration in the target population and had a slightly higher yield of enrolled patients in this study. However, POC referrals yielded a considerably higher proportion of enrollees from those who received a referral. Participants who received ATF were more likely to be screened and enrolled in the program. These results suggest that either a letter or POC referral strategy can be effective with ATF; however, when resource costs are considered, POC with ATF may be the best method to engage rural primary care patients in a weight management program.

Acknowledgments

The authors would like to express their gratitude to the physicians, nurses, and administrative staff involved in recruitment and data collection.

Conflicts of Interest

None declared.

References

1. Befort C, Nazir N, Perri M. Prevalence of obesity among adults from rural and urban areas of the United States: findings from NHANES (2005-2008). J Rural Health 2012;28(4):392-397 [FREE Full text] [doi: 10.1111/j.1748-0361.2012.00411.x] [Medline: 23083085]
2. Phillips CD, McLeroy KR. Health in rural America: remembering the importance of place. Am J Public Health 2004 Oct;94(10):1661-1663. [doi: 10.2105/ajph.94.10.1661] [Medline: 15451725]
3. Befort CA, VanWormer JJ, DeSouza C, Ellerbeck EF, Kimminau KS, Greiner A, et al. Protocol for the Rural Engagement in Primary Care for Optimizing Weight Reduction (RE-POWER) trial: comparing three obesity treatment models in rural primary care. Contemp Clin Trials 2016 Mar;47:304-314. [doi: 10.1016/j.cct.2016.02.006] [Medline: 26898748]
4. Phillips SM, Glasgow RE, Bello G, Ory MG, Glenn BA, Sheinfeld-Gorin SN, MOHR Study Group. Frequency and prioritization of patient health risks from a structured health risk assessment. Ann Fam Med 2014;12(6):505-513 [FREE Full text] [doi: 10.1370/afm.1717] [Medline: 25384812]
5. Estabrooks P, You W, Hedrick V, Reinholt M, Dohm E, Zoellner J. A pragmatic examination of active and passive recruitment methods to improve the reach of community lifestyle programs: The Talking Health Trial. Int J Behav Nutr Phys Act 2017 Jan 19;14(1):7 [FREE Full text] [doi: 10.1186/s12966-017-0462-6] [Medline: 28103935]
6. Glasgow RE, Nelson CC, Strycker LA, King DK. Using RE-AIM metrics to evaluate diabetes self-management support interventions. Am J Prev Med 2006 Jan;30(1):67-73. [doi: 10.1016/j.amepre.2005.08.037] [Medline: 16414426]
7. Lam E, Partridge SR, Allman-Farinelli M. Strategies for successful recruitment of young adults to healthy lifestyle programmes for the prevention of weight gain: a systematic review. Obes Rev 2016 Feb;17(2):178-200. [doi: 10.1111/obr.12350] [Medline: 26663091]
8. Porter GC, Laumb K, Michaud T, Brito F, Petreca D, Schwieger G, et al. Understanding the impact of rural weight loss interventions: a systematic review and meta-analysis. Obes Rev 2019 May;20(5):713-724 [FREE Full text] [doi: 10.1111/obr.12825] [Medline: 30633845]
9. Estabrooks P, Harden S, Almeida F, Hill JL, Johnson SB, Porter GC, et al. Using integrated research-practice partnerships to move evidence-based principles into practice. Exerc Sport Sci Rev 2019 Jul;47(3):176-187 [FREE Full text] [doi: 10.1249/JES.0000000000000194] [Medline: 31008840]
10. Foster CE, Brennan G, Matthews A, McAdam C, Fitzsimons C, Mutrie N. Recruiting participants to walking intervention studies: a systematic review. Int J Behav Nutr Phys Act 2011 Dec 15;8(1):137 [FREE Full text] [doi: 10.1186/1479-5868-8-137] [Medline: 22171531]
11. Ives DG, Kuller LH, Schulz R, Traven ND, Lave JR. Comparison of recruitment strategies and associated disease prevalence for health promotion in rural elderly. Prev Med 1992 Sep;21(5):582-591. [doi: 10.1016/0091-7435(92)90066-q] [Medline: 1438108]
12. Linnan LA, Emmons KM, Klar N, Fava JL, LaForge RG, Abrams DB. Challenges to improving the impact of worksite cancer prevention programs: comparing reach, enrollment, and attrition using active versus passive recruitment strategies. Ann Behav Med 2002;24(2):157-166. [doi: 10.1207/S15324796ABM2402_13] [Medline: 12054321]
13. Raynor HA, Osterholt KM, Hart CN, Jelalian E, Vivier P, Wing RR. Evaluation of active and passive recruitment methods used in randomized controlled trials targeting pediatric obesity. Int J Pediatr Obes 2009;4(4):224-232 [FREE Full text] [doi: 10.3109/17477160802596189] [Medline: 19922036]
14. Lee RE, McGinnis KA, Sallis JF, Castro CM, Chen AH, Hickmann SA. Active vs. passive methods of recruiting ethnic minority women to a health promotion program. Ann Behav Med 1997;19(4):378-384. [doi: 10.1007/BF02895157] [Medline: 9706365]

15. UyBico SJ, Pavel S, Gross CP. Recruiting vulnerable populations into research: a systematic review of recruitment interventions. J Gen Intern Med 2007 Jun 21;22(6):852-863 [FREE Full text] [doi: 10.1007/s11606-007-0126-3] [Medline: 17375358]
16. Porter G. Identifying, adapting, and implementing a weight management program into rural primary care. Theses & Dissertations, University of Nebraska Medical Center. 2020. URL: https://digitalcommons.unmc.edu/etd/477 [accessed 2021-03-08]
17. Porter GC, Schwab R, Hill JL, Bartee T, Heelan KA, Michaud TL, et al. Examining the feasibility and characteristics of realistic weight management support for patients: focus groups with rural, micropolitan, and metropolitan primary care providers. Prev Med Rep 2021 Apr 30;23:101390 [FREE Full text] [doi: 10.1016/j.pmedr.2021.101390] [Medline: 34026468]
18. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health 1999 Sep;89(9):1322-1327. [doi: 10.2105/ajph.89.9.1322] [Medline: 10474547]
19. Estabrooks P, Smith-Ray R, Almeida F, Hill J, Gonzales M, Schreiner P, et al. Move More: translating an efficacious group dynamics physical activity intervention into effective clinical practice. Int J Sport Exerc Psychol 2011 Mar;9(1):4-18 [FREE Full text] [doi: 10.1080/1612197x.2011.563123]
20. Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care 2012 Mar;50(3):217-226 [FREE Full text] [doi: 10.1097/MLR.0b013e3182408812] [Medline: 22310560]
21. Almeida FA, Michaud TL, Wilson KE, Schwab RJ, Goessl C, Porter GC, et al. Preventing diabetes with digital health and coaching for translation and scalability (PREDICTS): a type 1 hybrid effectiveness-implementation trial protocol. Contemp Clin Trials 2020 Jan;88:105877. [doi: 10.1016/j.cct.2019.105877] [Medline: 31682941]
22. Almeida FA, Pardo KA, Seidel RW, Davy BM, You W, Wall SS, et al. Design and methods of "diaBEAT-it!": a hybrid preference/randomized control trial design using the RE-AIM framework. Contemp Clin Trials 2014 Jul;38(2):383-396 [FREE Full text] [doi: 10.1016/j.cct.2014.06.010] [Medline: 24956325]
23. Estabrooks PA, Wilson KE, McGuire TJ, Harden SM, Ramalingam N, Schoepke L, et al. A quasi-experiment to assess the impact of a scalable, community-based weight loss program: combining reach, effectiveness, and cost. J Gen Intern Med 2017 Apr;32(Suppl 1):24-31 [FREE Full text] [doi: 10.1007/s11606-016-3971-0] [Medline: 28271423]
24. Almeida FA, You W, Harden SM, Blackman KC, Davy BM, Glasgow RE, et al. Effectiveness of a worksite-based weight loss randomized controlled trial: the worksite study. Obesity (Silver Spring) 2015 Apr;23(4):737-745 [FREE Full text] [doi: 10.1002/oby.20899] [Medline: 25678325]
25. 2007-2011 American Community Survey 5-year estimates. United States Census Bureau. 2011. URL: https://data.census.gov/cedsci/ [accessed 2021-03-08]
26. Kolasa KM, Rickett K. Barriers to providing nutrition counseling cited by physicians: a survey of primary care practitioners. Nutr Clin Pract 2010 Oct;25(5):502-509. [doi: 10.1177/0884533610380057] [Medline: 20962310]
27. Shuval K, Leonard T, Drope J, Katz DL, Patel AV, Maitin-Shepard M, et al. Physical activity counseling in primary care: insights from public health and behavioral economics. CA Cancer J Clin 2017 May 06;67(3):233-244 [FREE Full text] [doi: 10.3322/caac.21394] [Medline: 28198998]
28. Bertakis KD, Azari R, Helms LJ, Callahan EJ, Robbins JA. Gender differences in the utilization of health care services. J Fam Pract 2000 Feb;49(2):147-152. [Medline: 10718692]
29. Ashman J, Rui P, Okeyode T. Characteristics of Office-based Physician Visits, 2016. National Center for Health Statistics - Data Brief No. 331. 2019. URL: https://www.cdc.gov/nchs/products/databriefs/db331.htm [accessed 2021-03-07]
30. Harden SM, Gaglio B, Shoup JA, Kinney KA, Johnson SB, Brito F, et al. Fidelity to and comparative results across behavioral interventions evaluated through the RE-AIM framework: a systematic review. Syst Rev 2015 Nov 08;4(1):155 [FREE Full text] [doi: 10.1186/s13643-015-0141-0] [Medline: 26547687]
31. Glasgow RE, Goldstein MG, Ockene JK, Pronk NP. Translating what we have learned into practice. Principles and hypotheses for interventions addressing multiple behaviors in primary care. Am J Prev Med 2004 Aug;27(2 Suppl):88-101. [doi: 10.1016/j.amepre.2004.04.019] [Medline: 15275677]
32. Glasgow RE, Lichtenstein E, Marcus AC. Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. Am J Public Health 2003 Aug;93(8):1261-1267. [doi: 10.2105/ajph.93.8.1261] [Medline: 12893608]
33. Kessler RS, Purcell EP, Glasgow RE, Klesges LM, Benkeser RM, Peek CJ. What does it mean to "employ" the RE-AIM model? Eval Health Prof 2013 Mar 21;36(1):44-66. [doi: 10.1177/0163278712446066] [Medline: 22615498]
34. Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med 2009 May;36(5):452-457 [FREE Full text] [doi: 10.1016/j.amepre.2009.02.002] [Medline: 19362699]

Abbreviations

ATF: active telephone follow-up
POC: point of care
RA: research assistant

RE-AIM: Reach, Effectiveness, Adoption, Implementation, and Maintenance

Edited by G Eysenbach; submitted 08.03.21; peer-reviewed by H Belli, D Downs; comments to author 10.05.21; revised version received 15.07.21; accepted 26.07.21; published 20.10.21.

Please cite as:
Porter G, Michaud TL, Schwab RJ, Hill JL, Estabrooks PA
Reach Outcomes and Costs of Different Physician Referral Strategies for a Weight Management Program Among Rural Primary Care Patients: Type 3 Hybrid Effectiveness-Implementation Trial
JMIR Form Res 2021;5(10):e28622
URL: https://formative.jmir.org/2021/10/e28622
doi:10.2196/28622
PMID:34668873

Original Paper

Medical Students' Perceptions of a Blockchain-Based Decentralized Work History and Credentials Portfolio: Qualitative Feasibility Study

Anton Hasselgren[1], MSc; Katina Kralevska[2], PhD; Danilo Gligoroski[2], PhD; Arild Faxvaag[1], MD, PhD

[1]Department of Neuromedicine and Movement Science, Norwegian University of Science and Technology, Trondheim, Norway

[2]Department of Information Security and Communication Technology, Norwegian University of Science and Technology, Trondheim, Norway

Corresponding Author:
Anton Hasselgren, MSc
Department of Neuromedicine and Movement Science
Norwegian University of Science and Technology
Mellomila 71
Trondheim, 7018
Norway
Phone: 47 46948498
Email: anton.hasselgren@ntnu.no

Abstract

Background: Increased digitization of health care might challenge some of the trust functions that are established in a traditional health care system. We have, with the concept of VerifyMed, developed a decentralized service for work history and competence verification, as a means to increase trust in the virtual interaction between a patient and a caregiver, mitigate administrative burden, and provide patient-reported outcomes seamlessly for health professionals.

Objective: This research aimed to validate the use case of a decentralized credentials service for health care professionals in Norway. We also aimed to evaluate the proof-of-concept of VerifyMed, a blockchain-based credential service for health care professionals.

Methods: A qualitative approach was applied with data collection through 9 semistructured interviews and 2 focus groups (one with 4 participants and the other with 5 participants). The System Usability Scale (SUS) was used as a part of the interviews. Data were analyzed through the principles of systematic text condensation. The recruitment of participants ended when it was concluded that the data had reached saturation.

Results: The following 5 themes were identified from the interviews and focus groups: (1) the need for aggregated storage of work- and study-related verification, (2) trust in a virtual health care environment, (3) the potential use of patient feedback, (4) trust in blockchain technology, and (5) improvements of the VerifyMed concept. The SUS questionnaire gave a score of 69.7.

Conclusions: This study has validated the need for a decentralized system where health care professionals can control their credentials and, potentially, their reputation. Future work should update the VerifyMed system according to this input. We concluded that a decentralized system for the storage of work-related verifiable credentials could increase trust in a virtualized health care system.

(JMIR Form Res 2021;5(10):e33113) doi:10.2196/33113

KEYWORDS

blockchain; eHealth; qualitative research; VerifyMed

Introduction

Background

The COVID-19 pandemic has accelerated the digital transformation of the health care sector. Social distancing and other measures to reduce transmission of SARS-CoV-2 have forced health systems to deliver health services using innovative methods [1]. Virtual health care consultations, which often are referred to as telemedicine, are an example of this transformation that has had a rapid increase during the pandemic. Telemedicine visits increased by 683% in New York City during the spring of 2020 [2], and general practitioners in Norway reported that 81% of them used video consultation during the pandemic (most

of them did not use it at all before the pandemic) [3]. Since the advantages of telemedicine include cost-effectiveness, increased access, and availability [4], we can assume that this increase will be permanent. In previous work, it was suggested that telemedicine might challenge some of the established structures for trust in a patient–health care professional relationship [5]. The ability to verify the competence of health care professionals will be of increasing importance in telemedicine in order to enhance trust [6,7].

The administrative burden placed on health care professionals has perhaps always been present [8]. However, the administrative burden related to work mobility seems to have increased recently [9], and this trend is also reflected in increased mobility among health care professionals [10]. As a result, the administrative burden of verifying credentials and experiences among this working group is increasing.

For the last decades, there has been a focus on putting patients in the center of evaluating clinical care, combined with biomarkers of health improvements [11]. As a mean for this, patient-reported outcome measures (PROMs) have been introduced to measure patient-reported outcomes (PROs). PROs are referred to as the patient's health, quality of life, or functional status associated with health care or treatment [12]. PROMs are the tools to measure PROs, which could, for example, be a measure of the quality of life. To complement PROMs, patient-reported experience measures (PREMs) have been introduced as a tool to measure patients' experiences with health care or health services, often with a satisfactory score [12]. PROs may have increasing importance as a means of learning and improving health care professionals, as well as a way for health care professionals to verify their work history [13].

We have identified a need for a new decentralized service for work history and competence verification as a means to increase trust in the virtual interaction between a patient and a caregiver, mitigate administrative burden, and provide PROs seamlessly for health professionals. This concept is described in the next subsection.

VerifyMed

The proposed concept of VerifyMed provides a solution for enhancing trust between a caregiver and a patient within a virtualized health care environment. The cornerstone of this architecture is an approach for capturing the trust relationships within the health care system by utilizing a blockchain. This trust mechanism can be used by patients to confirm the credentials and potentially enhance their trust in a caregiver during their interaction. Furthermore, the architecture includes a mechanism for evaluating these interactions publicly on the blockchain, using PROMs and PREMs. These evaluations serve as a portfolio of the caregiver's experience and could potentially be used as a mechanism for continued learning among health care professionals. The concept of VerifyMed is presented further in other reports [5,14,15]. To achieve the objectives of this research, a mock-up of the user interface of the platform was designed using user-centric design theory [16]. The mock-up can be accessed online [17]. Figure 1 and Figure 2 illustrate examples of the user interface that was explored in this research.

Figure 1. The data sharing page of the VerifyMed user interface.

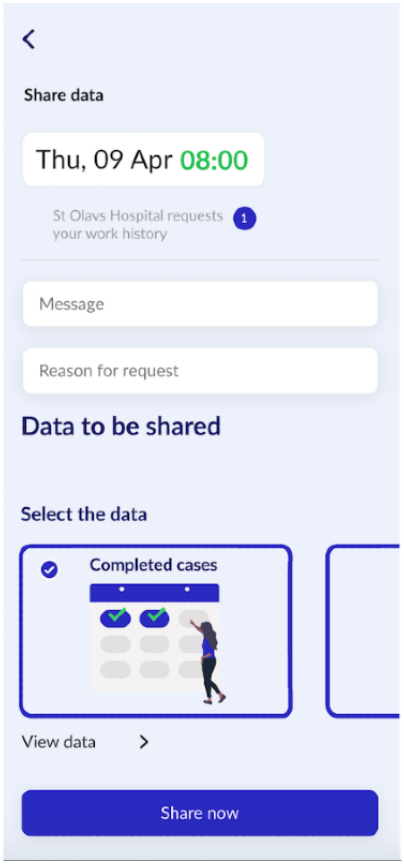
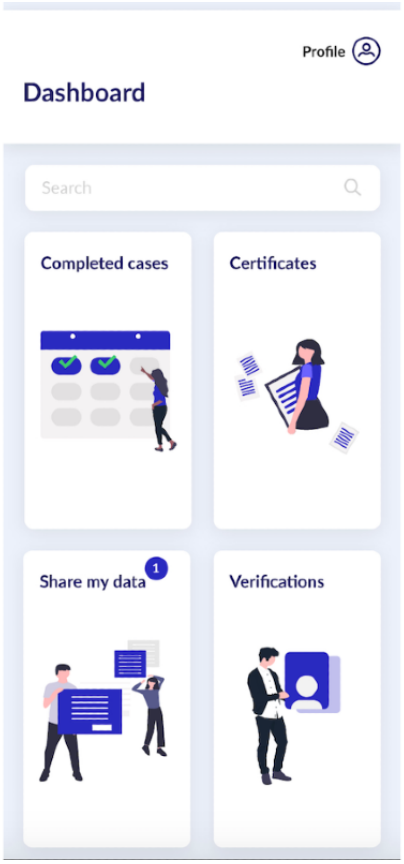


Figure 2. The dashboard page of the VerifyMed user interface.



Significance

It is important to gain user input early in technology development to further improve an application according to the needs of the users [18]. After identifying a problem in the health sector and designing a solution for that problem with a proof-of-concept, a feasibility study further validates the concept of VerifyMed and provides valuable user input for further development.

Aim and Objective

The objective of this work was 2-fold. First, the work aimed to validate a use case of a decentralized medical professional credentials service by mapping out the need for such a service in Norway. Second, we aimed to evaluate the proof-of-concept of VerifyMed, a blockchain-based credential service for health care professionals. We limited the scope of the work to medical students, as they might experience challenges with recording and managing their credentials and experience as they start and progress through their careers.

Research Questions

The research questions were as follows:

1. What are the potential scenarios of usage from the user group?
2. How will a decentralized system, such as VerifyMed, be accepted by future health care professionals?
3. Does the VerifyMed system meet the requirements of health care professionals who would be using the system?
4. What are the desired features of the users?
5. What are the opinions on a patient-feedback system?

Methods

Overview

To answer the research questions, a qualitative approach was applied. Data were collected through 9 semistructured interviews by using the System Usability Scale (SUS) [19] as a starting point for the interviews. In addition, 2 focus groups (one with 4 participants, and the other with 5 participants) were also conducted. The recruitment of participants ended when it was concluded that the data had reached saturation [20].

Data Collection

Medical students in Norway in study years 4 to 6 were recruited through student organizations. Two focus groups were performed prior to the individual interviews. The focus groups functioned as workshops where blockchain technology [21] and the concept of VerifyMed were presented by the moderator (AH) before any discussion. The moderator asked the participants to describe the current procedures they had experienced with skill verification, certificates, and trust in a virtual health care scenario. Finally, an open discussion on how the presented technology could be used to improve the current workflows was initiated. The focus groups were limited to 45 min. In addition to the moderator, a research assistant was present to take notes.

The duration of the individual interviews was limited to 30 min. The participants were invited to the mock-up of the VerifyMed user interface by an online link. They accessed the mock-up

through their web browsers on their laptops. After a short introduction, each participant was invited to perform several simple tasks in the prototype and was asked to explain his or her thoughts during this exploration phase, with minimal assistance from the moderator. The participant was then asked questions from the SUS questionnaire. The SUS is considered to be an easy, quick, and reliable test of usability that is technology agnostic [19]. Based on the answers, follow-up questions followed in a semistructured form. The focus groups and the individual interviews were conducted in an online format, using Zoom (Zoom Video Communications).

Data Management

The focus group sessions and the user-testing interviews were audio and video recorded. The recordings were analyzed after the sessions by the main researcher (AH). The 3 other researchers in the project (AF, KK, and DG) also reviewed the recordings when there were doubts by the first researcher. The recordings were stored locally on the first researcher's computer during the project. The recordings were transcribed and erased after transcription, and the notes from the focus group sessions were compared with the transcripts and then erased.

Data Analysis

Transcription of the collected data was performed according to the 6 steps of transcription proposed by Azevedo et al [22]. Since the data collection was conducted in Norwegian, an English translation of the used quotations was performed by the main researcher (AH). Data were analyzed according to the principles of systematic text condensation [23]. This procedure consists of the following 4 steps: (1) getting a total impression by reading all the text materials and identifying preliminary themes; (2) identifying meaningful units from both the technical aspects of the VerifyMed service and its use by medical students; (3) abstracting condensates from each group and subgroup; and (4) creating synthesized descriptions of the user's experiences and opinions about the use of a decentralized work history portfolio. The software NVivo (version 1.4.1; QSR International) was used for the analysis.

Ethical Considerations

All participants were asked to give written consent based on oral and written information about the study. Only those who gave their consent to participate in the study, according to the information in the consent form, were included (n=9). The study did not collect or otherwise handle patient- or health-related data. Therefore, ethical clearance from the Regional Ethical Committee (REK) was not obtained. The study was registered by NSD - Norwegian Center for Research Data and the Data Protection Officer at the Faculty of Medicine and Health Science (Norwegian University of Science and Technology) to be General Data Protection Regulation compliant.

Results

Participant Characteristics

A total of 9 participants were recruited in the study, and all 9 completed both participation in the focus group and the individual interview. The characteristics of the respondents are presented in Table 1.

Table 1. Characteristics of the informants.

Characteristic	Value (n=9), n (%)
Gender	
Male	4 (44)
Female	5 (56)
Age (years)	
23-24	3 (33)
25-26	2 (22)
27-28	3 (33)
>28	1 (11)
Study year (out of 6)	
4	1 (11)
5	5 (56)
6	3 (33)
University	
Norwegian University of Science and Technology	8 (89)
University of Oslo	1 (11)
Previous knowledge of blockchain	
No	6 (67)
Yes	3 (33)

Themes

The results from the SUS were mainly used as a starting point for the individual interviews. The quantitative results from the SUS were calculated using the standard formula for SUS [24]. The score was 69.7, with fairly equal responses from the respondents. A score above 70 is considered acceptable, according to validation studies [24]. In the data analysis, 5 themes were identified within the focus groups and individual interviews, and an overview is presented in Table 2. The results from both methods of data collection were intertwined. Several of the themes were discussed in both the focus groups and individual interviews, and they are therefore presented here jointly.

Table 2. Results overview.

Theme	Proportion[a] of data	Supporting quotes
The need for an aggregated storage of work- and study- related verifications	24.2%	*...large parts of the system is trust based. I don't know how to verify certificates, but as you say, paper-based certificates are an easy way to falsify knowledge and experience.*
Trust in a virtual health care environment	26.0%	*To showcase what you have done related to courses and such could contribute, it becomes the equivalent to have diplomas on the wall. It is not necessarily certain that the patient understands what it is, but it can improve the total impression.*
The potential use of patient feedback	14.5%	*The ones who write feedback are the patients how are either very pleased or they who are very displeased. ...the selection gets skewed.*
Trust in blockchain technology	7.3%	*I think I understand the value with that things could be verified and that falsification might be mitigated with time-stamping and such, that I see as positive...*
Improvements of the VerifyMed concept	6.5%	*I envision that in the future, when things get more digital and patients have a specific problem and want to get in contact with a doctor who has done research in that area or has any specific courses within the area...*

[a]The percentages do not add to 100 since other themes, not relevant to the research questions, also were discussed.

The Need for an Aggregated Storage of Work- and Study-Related Verifications

The first theme evolves around the need for a platform where medical students can collect and store verifications of their experiences. As the participants describe, as of now, there is no common digital system in use where they can store grades, certificates, references, and verifications of practical assignments. One participant expressed this as follows:

> *...it would be nice if it could be done digitally. Previously in the studies, we rotated to different departments of the hospital and were supposed to get one signature from each department. We were supposed to keep this piece of paper with over 20 signatures throughout the semester and try to not lose it. It would be an advantage if this could be done digitally.*

If this physical paper is lost by the students, they need to collect all the signatures again. This was expressed as a rather common problem and a lot of work. On this occasion, it seems like the supervisors were not always aware of what they approved by giving their signatures.

> *...if you needed to go back for the signature, it could happen that they were a bit uncertain but most of the time they signed anyway, or always.*

The risk of falsified documents with an analog trust-based system was further acknowledged in the discussion. As one respondent expressed:

> *...large parts of the system is trust-based. I don't know how to verify certificates, but as you say, paper-based certificates are an easy way to falsify knowledge and experience.*

Trust in a Virtual Health Care Environment

The second identified theme evolves around trust in the interaction between the medical doctor (student) and the patient. Since the respondents were in different stages of their education, they had experienced different exposures to patients. Their perceptions of trust in their encounters with patients also varied. Some respondents did experience a lack of trust towards them among patients.They expressed that this probably was a consequence of they being students and thus being considered less experienced and knowledgeable. However, most of the respondents experienced that trust could be established, and it was not considered a major disadvantage that they were students. Furthermore, trust in a virtual health care environment, mainly video consultations, was discussed. The respondents agreed that this way of providing health services will be an important part of their professional careers. They had so far been exposed to this medium in various degrees, mainly due to COVID-19, where restrictions enforced virtual meetings instead of physical meetings. Their perceptions of quality in virtual health services, compared to physical services, varied. Some experienced no difficulties in gaining the trust and confidence of patients. However, most seemed to agree that the lack of physical attributes and the lack of physical examinations may harm the trust-building mechanisms.

> *You get something for free in a hospital setting, you walk-in in a white coat, that looks professional. I believe most doctors perform virtual consultation from a setting that looks professional, otherwise, it can look suspicious.*

The individual interviews further explored the need for digital verifications, and the general opinions among the respondents were that this could have a purpose in a virtual environment.

> *To showcase what you have done related to courses and such could contribute, it becomes the equivalent to have diplomas on the wall. It is not necessarily certain that the patient understands what it is, but it can improve the total impression.*

However, participants were also hesitant about how this information would be interpreted by the patients, and if they would comprehend the meaning of such certificates and other proofs of competence.

> *...I'm a bit uncertain regarding this. What value would it bring if they could see this, it might be difficult for them to interpret. It's difficult to say what they would use this information for.*

The Potential Use of Patient Feedback

The third theme identified was the expectations and fears around a patient-feedback system, such as PROs. In this discussion, the Norwegian website Legelisten [25], a site where anyone can rate their general practitioner, was referred to several times. The respondents' general opinions around this service were negative, and the patient-feedback system was associated with the negative impressions of this service. For example, the risk of a biased selection of users of this service was expressed as follows:

> *The ones who write feedback are the patients who are either very pleased or they who are very displeased. ...the selection gets skewed.*

The participants also expressed a general fear of being publicly rated, similar to the rating system of Legelisten [25]:

> *...agree that it could be an individual asset but nothing that should be published publicly, how good you are in comparison with others because that will create competition rather than provide you with learning.*

This fear also extended to how data could be reported in a feedback system. Several respondents expressed the need for this kind of feature to be objective and systematic. Allowing patients to provide feedback without any systematic framework was expressed as being associated with a major risk of information overload and useless information from the patients.

> *...maybe you should not be able to write free text with no limit and maybe you should limit how the feedback is given, otherwise it will be a lot of irrelevant and unserious feedback, so it has to be a limitation for the patients' possibility to provide feedback.*

A feedback system as a means for health care professionals to learn was however expressed as something positive among the participants. At present, they have little or no opportunity to know the outcome of a given patient treatment, since they often rotate and may miss a revisit or the results when the patient gets referred to another department.

> *You often wonder how it went and what happened to the patient.*
>
> *It would be great to get a small correction and feedback on what you have done and how it went, and what conclusions were made further. That would be gold worth to know...*

Trust in Blockchain Technology

The fourth theme, trust in blockchain technology, was briefly discussed. As Table 1 indicates, a few of the respondents had knowledge about blockchain technology prior to participating in this research. Even though blockchain was introduced in a presentation by the moderator (AH) before the focus group discussions, several of the respondents reported that they did not understand the technology. However, none of them showed any negativity toward the technology and whether to trust the VerifyMed service.

> *I think I understand the value with that things could be verified and that falsification might be mitigated with time-stamping and such, that I see as positive. But I don't know enough about the technology to say if it gives any large advantages compared to other services. I think I understand it, but I'm not a technical person.*

As expressed by several respondents, the trust in the service was dependent on third-party validation and trust in the developers behind the service. One respondent commented as follows:

> *Yes, if the source is trustworthy and it helps if it is promoted by persons you trust. ...but if it is an unknown actor which I could not relate to I would be much more skeptical to provide any personal information.*

Improvements of the VerifyMed Concept

The last main theme that was discussed were general improvements and opinions regarding the VerifyMed user design and features experienced by all respondents. None of the respondents had any problems completing the 9-item task list given, and they all did so in a short amount of time (3-7 min). The general expression was that the solution could be useful and that they acknowledge the need for this kind of service. One respondent commented as follows:

> *I envision that in the future, when things get more digital and patients have a specific problem and want to get in contact with a doctor who has done research in that area or has any specific courses within the area, then it could be very useful for both the doctor to be able to show knowledge and interests in that particular area, then you might get more patients you can include in your research or that you find interesting.*

The informants expressed that the design and user flow were something that they were familiar and comfortable with. They had a few suggestions on improvements and additions of features, such as (1) make it clearer what data are being shared, for how long, and with whom, (2) make it possible to have direct communication with patients through a message system, and

(3) make it possible to showcase scientific publications or research projects as a part of the "portfolio."

Discussion

Principal Findings

This research aimed to validate a use case of a decentralized medical professional credentials service by mapping out the need for such a service in Norway and to evaluate the proof-of-concept of VerifyMed, a blockchain-based credentials service for health care professionals.

The informants expressed that the main area of use is a platform where they could store all the data they would need for a job application. This is perhaps an expected result since the respondents are already (or will soon be) in a job-seeking process. The general opinion was that they had no or little control over data, such as verifications of internships or practical assignments, at present. They were all positive about the idea of a system that could automate this and provide them with more control. Presently, it seems to be somewhat up to chance if they receive these paper-based verifications and how useful they are owing to a lack of systematization. This highlights the need for new services with features similar to those of VerifyMed.

Fear was generally expressed for a patient-feedback system among the participants, in case the data are used to evaluate them externally. This fear might be explained by the fact that young physicians (students) are already exposed to a lot of stress and have a fear of making mistakes [26]. The addition of another evaluation service could increase this stress. However, they were generally positive about receiving feedback for their own learning. They were also open to extend this and share the feedback with colleagues and take part in each other's feedback, for the objective of learning. Previous research has indicated that it might be difficult for health care professionals to learn from patient feedback [27]. The sample in this study (students) might explain this difference, as students are probably more inclined to learn and improve compared to more senior health care professionals. They did however see little or no use in sharing patient feedback with other patients, as they did not see the need for this. This is in line with previous research [28]. The existence of physician-rating websites, such as Legelisten [25], indicates that patients are interested in the feedback of other patients to evaluate physicians. This difference in perception between physicians and patients might again be explained by physicians' fears of being evaluated and potentially not having control over their reputation as health care professionals. Previous research has indicated that a physician's reputation on physician-rating websites is critical to attract patients [29], and there seems to be a lack of tools where physicians can take control over their online reputation [30]. This previous knowledge and our results clearly indicate the need for a service where physicians can control their online reputation. Considering this, future updates on the VerifyMed concept should include options to share or not to share patient feedback publicly. This control feature might enhance the acceptability of the service among health care professionals and enable reputation control in a virtualized health care environment.

The quantitative results from the SUS should be interpreted with the understanding that the small sample size prevents any strong conclusions from this quantitative result. However, it could serve as an indicator that the usability of the user design is acceptable [24] (the study showed a SUS score of 70). There were no indications that design changes need to be implemented in the platform based on the user testing.

The limited clinical experience of the informants may have influenced the results, and it is possible that another sample, with more experienced health care professionals, will have other opinions. However, the results from the current informant sample fulfill the objectives of this research. The individual interviews might have been influenced by the discussions in the focus groups and the presentation made by the main researcher (AH), which were both conducted before the individual interviews. The perception of the technology might have been influenced as a result.

Conclusion

This study validated the need for the concept of VerifyMed, and feedback from the users provided inputs that will further enhance the quality and fit-for-purpose aspect of the concept. Future work should update the system according to these inputs, enhance the data control of the user to provide reputation control, and move to the next step of system development. Furthermore, we concluded that a decentralized system for the storage of work-related verifiable credentials could increase trust in the health system, especially if there are less trusted institutions as a result of an increase in the number of health care providers in a digitally transformed health care system.

Acknowledgments

The authors acknowledge the contributions of Cathrin Brønbo Larsen, who assisted with validating the interview and focus group formats and took notes during the focus group sessions. Martha Skogen assisted with the validation of the mock-up design, and Torunn Hatlen Nøst assisted with the validation of the focus group format.

Conflicts of Interest

None declared.

References

1. Webster P. Virtual health care in the era of COVID-19. The Lancet 2020 Apr;395(10231):1180-1181. [doi: 10.1016/s0140-6736(20)30818-7]
2. Mann DM, Chen J, Chunara R, Testa PA, Nov O. COVID-19 transforms health care through telemedicine: Evidence from the field. J Am Med Inform Assoc 2020 Jul 01;27(7):1132-1135 [FREE Full text] [doi: 10.1093/jamia/ocaa072] [Medline: 32324855]
3. Johnsen TM, Norberg BL, Kristiansen E, Zanaboni P, Austad B, Krogh FH, et al. Suitability of Video Consultations During the COVID-19 Pandemic Lockdown: Cross-sectional Survey Among Norwegian General Practitioners. J Med Internet Res 2021 Feb 08;23(2):e26433 [FREE Full text] [doi: 10.2196/26433] [Medline: 33465037]
4. Kichloo A, Albosta M, Dettloff K, Wani F, El-Amir Z, Singh J, et al. Telemedicine, the current COVID-19 pandemic and the future: a narrative review and perspectives moving forward in the USA. Fam Med Community Health 2020 Aug 18;8(3):e000530 [FREE Full text] [doi: 10.1136/fmch-2020-000530] [Medline: 32816942]
5. Hasselgren A, Hanssen Rensaa J, Kralevska K, Gligoroski D, Faxvaag A. Blockchain for Increased Trust in Virtual Health Care: Proof-of-Concept Study. J Med Internet Res 2021 Jul 30;23(7):e28496 [FREE Full text] [doi: 10.2196/28496] [Medline: 34328437]
6. Pearson SD, Raeke LH. Patients' trust in physicians: many theories, few measures, and little data. J Gen Intern Med 2000 Jul;15(7):509-513 [FREE Full text] [doi: 10.1046/j.1525-1497.2000.11002.x] [Medline: 10940139]
7. O'Neill O. Intelligent Trust in a Digital World. New Perspectives Quarterly 2007 Nov 06;34(4):27-31. [doi: 10.1111/npqu.12105]
8. Woolhandler S, Himmelstein DU. Administrative Work Consumes One-Sixth of U.S. Physicians' Working Hours and Lowers their Career Satisfaction. Int J Health Serv 2014 Oct 01;44(4):635-642. [doi: 10.2190/hs.44.4.a]
9. Rao SK, Kimball AB, Lehrhoff SR, Hidrue MK, Colton DG, Ferris TG, et al. The Impact of Administrative Burden on Academic Physicians. Academic Medicine 2017;92(2):237-243. [doi: 10.1097/acm.0000000000001461]
10. Costigliola V. Mobility of medical doctors in cross-border healthcare. EPMA J 2011 Dec 12;2(4):333-339 [FREE Full text] [doi: 10.1007/s13167-011-0133-7] [Medline: 23199171]
11. Willke RJ, Burke LB, Erickson P. Measuring treatment impact: a review of patient-reported outcomes and other efficacy endpoints in approved product labels. Control Clin Trials 2004 Dec;25(6):535-552. [doi: 10.1016/j.cct.2004.09.003] [Medline: 15588741]
12. Weldring T, Smith SM. Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs). Health Serv Insights 2013;6:61-68 [FREE Full text] [doi: 10.4137/HSI.S11093] [Medline: 25114561]
13. Nelson EC, Eftimovska E, Lind C, Hager A, Wasson JH, Lindblad S. Patient reported outcome measures in practice. BMJ 2015 Feb 10;350:g7818. [doi: 10.1136/bmj.g7818] [Medline: 25670183]
14. Hanssen Rensaa JA, Gligoroski D, Kralevska K, Hasselgren A, Faxvaag A. VerifyMed-A blockchain platform for transparent trust in virtualized healthcare: Proof-of-concept. In: IECC 2020: Proceedings of the 2020 2nd International Electronics Communication Conference. 2020 Presented at: 2nd International Electronics Communication Conference; July 8-10, 2020; Singapore p. 73-80. [doi: 10.1145/3409934.3409946]
15. Hanssen Rensaa JA. VerifyMed - Application of blockchain technology to improve trust in virtualized healthcare services. Norwegian University of Science and Technology. 2020. URL: https://ntnuopen.ntnu.no/ntnu-xmlui/bitstream/handle/11250/2781112/no.ntnu%3Ainspera%3A54255071%3A20647165.pdf?sequence=1 [accessed 2021-10-02]
16. Still B, Crane K. Fundamentals of User-Centered Design: A Practical Approach. Boca Raton, FL: CRC Press; 2017.
17. Anton H. VerifyMed mock-up. Adobe. 2021 Mar 01. URL: https://xd.adobe.com/view/0f333da6-62a5-4928-bb32-b919452ff313-5a60/?fullscreenamp;hints=off [accessed 2021-06-01]
18. Choi YM. Utilizing end User Input in Early Product Development. Procedia Manufacturing 2015;3:2244-2250. [doi: 10.1016/j.promfg.2015.07.368]
19. Brooke J. SUS: A 'Quick and Dirty' Usability Scale. In: Jordan PW, Thomas B, McClelland IL, Weerdmeester B, editors. Usability Evaluation In Industry. Boca Raton, FL: CRC Press; 1996.
20. Fusch P, Ness L. Are We There Yet? Data Saturation in Qualitative Research. TQR 2015 Sep 8:1408-1416. [doi: 10.46743/2160-3715/2015.2281]
21. Hasselgren A, Kralevska K, Gligoroski D, Pedersen SA, Faxvaag A. Blockchain in healthcare and health sciences-A scoping review. Int J Med Inform 2020 Feb;134:104040 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.104040] [Medline: 31865055]
22. Azevedo V, Carvalho M, Costa F, Mesquita S, Soares J, Teixeira F, et al. Interview transcription: conceptual issues, practical guidelines, and challenges. Rev. Enf. Ref 2017 Sep 22;IV Série(Nº14):159-168. [doi: 10.12707/riv17018]
23. Malterud K. Systematic text condensation: a strategy for qualitative analysis. Scand J Public Health 2012 Dec 04;40(8):795-805. [doi: 10.1177/1403494812465030] [Medline: 23221918]
24. Bangor A, Kortum P, Miller J. Determining What Individual SUS Scores Mean: Adding an Adjective Rating Scale. Journal of Usability Studies 2009;4(3):114-123 [FREE Full text]
25. Legelisten. URL: https://www.legelisten.no/ [accessed 2021-09-27]
26. Dahlin M, Joneborg N, Runeson B. Stress and depression among medical students: a cross-sectional study. Med Educ 2005 Jun;39(6):594-604. [doi: 10.1111/j.1365-2929.2005.02176.x] [Medline: 15910436]

27. Sheard L, Marsh C, O'Hara J, Armitage G, Wright J, Lawton R. The Patient Feedback Response Framework - Understanding why UK hospital staff find it difficult to make improvements based on patient feedback: A qualitative study. Soc Sci Med 2017 Apr;178:19-27 [FREE Full text] [doi: 10.1016/j.socscimed.2017.02.005] [Medline: 28189820]
28. Patel S, Cain R, Neailey K, Hooberman L. General Practitioners' Concerns About Online Patient Feedback: Findings From a Descriptive Exploratory Qualitative Study in England. J Med Internet Res 2015 Dec 08;17(12):e276 [FREE Full text] [doi: 10.2196/jmir.4989] [Medline: 26681299]
29. Deng Z, Hong Z, Zhang W, Evans R, Chen Y. The Effect of Online Effort and Reputation of Physicians on Patients' Choice: 3-Wave Data Analysis of China's Good Doctor Website. J Med Internet Res 2019 Mar 08;21(3):e10170 [FREE Full text] [doi: 10.2196/10170] [Medline: 30848726]
30. Prabhu AV, Kim C, De Guzman E, Zhao E, Madill E, Cohen J, et al. Reputation Management and Content Control: An Analysis of Radiation Oncologists' Digital Identities. Int J Radiat Oncol Biol Phys 2017 Dec 01;99(5):1083-1091. [doi: 10.1016/j.ijrobp.2017.08.015] [Medline: 28939228]

Abbreviations

PREM: patient-reported experience measure
PRO: patient-reported outcome
PROM: patient-reported outcome measure
SUS: System Usability Scale

Edited by G Eysenbach; submitted 24.08.21; peer-reviewed by W Abramson, J Brooke; comments to author 01.09.21; revised version received 02.09.21; accepted 20.09.21; published 22.10.21.

Please cite as:
Hasselgren A, Kralevska K, Gligoroski D, Faxvaag A
Medical Students' Perceptions of a Blockchain-Based Decentralized Work History and Credentials Portfolio: Qualitative Feasibility Study
JMIR Form Res 2021;5(10):e33113
URL: https://formative.jmir.org/2021/10/e33113
doi:10.2196/33113
PMID:34677137

Original Paper

The Transition to an Entirely Digital Immunization Registry in Ha Noi Province and Son La Province, Vietnam: Readiness Assessment Study

Hong Duong[1*], MD, PhD; Sang Dao[2*], MPH; Huyen Dang[1], MD, PhD; Linh Nguyen[2], MPH; Tuan Ngo[2], MScIT; Trung Nguyen[1], MPH; Lan Anh Tran[1], MSc, MD; Doan Nguyen[1], MPH, MD; Maya Rivera[2], MPH; Nga Nguyen[2], MD, PhD

[1]National Institute of Hygiene and Epidemiology, Hanoi, Vietnam

[2]PATH Vietnam, Hanoi, Vietnam

[*]these authors contributed equally

Corresponding Author:
Nga Nguyen, MD, PhD
PATH Vietnam
#1101, 11th floor, Hanoi Towers, 49 Hai Ba Trung Street
Hoan Kiem District
Hanoi, 100000
Vietnam
Phone: 84 024 3936 2215 ext 130
Email: ntnguyen@path.org

Abstract

Background: Vietnam is one of the first low- to middle-income countries to develop and implement a national-scale electronic immunization registry. This system was finalized into the National Immunization Information System (NIIS) and scaled up to a national-level system in 2017. As a result, immunization coverage and the timeliness of vaccinations have drastically improved. The time spent on planning and reporting vaccinations has drastically reduced; as a result, vaccination planning and reporting has become more accurate and effective. However, to date, end users have been tasked with managing both the NIIS and paper-based systems in parallel until a formal assessment of the readiness to fully transition to the NIIS is conducted.

Objective: This study aims to evaluate the readiness to move to an entirely digital NIIS in 2 provinces of Vietnam—Ha Noi and Son La.

Methods: All health facilities were surveyed to assess their infrastructure, capacity, and need for human resources. NIIS end users were observed and interviewed to evaluate their NIIS knowledge and skill sets. Data from immunization cards and facility paper-based logbooks were compared with data from the NIIS, and vaccine stocks at selected facilities were tallied and compared with data from the NIIS.

Results: Of the 990 health facilities evaluated, most used the NIIS to enter and track immunizations (987/990, 99.7%) and vaccine stocks (889/990, 90.8%). Most had stable electricity (971/990, 98.1%), at least 1 computer (986/990, 99.6%), and ≥2 trained NIIS end users (825/990, 83.3%). End users reported that the NIIS supported them in managing and reporting immunization data and saving them time (725/767, 94.5%). Although many end users were able to perform basic skills, almost half struggled with performing more complex tasks. Immunization data were compiled from the NIIS and immunization cards (338/378, 89.4%) and paper-based logbooks (254/269, 94.4%). However, only 54.5% (206/378) of immunization IDs matched, 57% (13/23) of Bacillus Calmette-Guérin vaccination records were accurate, and 70% (21/30) of the facilities had consistent physical vaccine stock balances. The feedback received from NIIS end users suggests that more supportive supervision, frequent refresher training for strengthening their skill sets, and detailed standardized guides for improving data quality are needed.

Conclusions: The readiness to transition to a digital system is promising; however, additional resources are required to address the timeliness, completeness, and accuracy of the data.

(JMIR Form Res 2021;5(10):e28096) doi:10.2196/28096

KEYWORDS

immunizations; immunization registry; readiness assessment; electronic immunization records; Vietnam

Introduction

Background

Electronic immunization registries (EIRs) enable the collection and consolidation of immunization coverage and can therefore be an effective tool in the planning and delivery of immunization services [1,2]. EIRs are becoming more widespread in low- to middle-income countries because of more accessible and cheaper computer technologies [2]. Several studies have shown that the implementation of such systems improves data accuracy and timeliness and assists in identifying areas of low vaccine coverage and children who have missed vaccinations and as a result, saves health care workers a substantial amount of time [3,4]. However, the success and adoption of such systems are also contingent upon their ability to widely implement and successfully adopt these tools at a national scale [5-7]. The success of scale up of these systems is also highly contingent upon several factors such as adequate equipment, infrastructure, human resources and skills, reliable power supply, funding, and ongoing process evaluations [8].

Nonetheless, Vietnam is one of the first low- to middle-income countries to develop and implement a national-scale EIR. Beginning in 2012, PATH and the Vietnam National Expanded Program on Immunization (EPI) developed and piloted a digital immunization registry, ImmReg, in 1 district of the Ben Tre province. Between 2014 and 2015, ImmReg was integrated with VaxTrak (a vaccine tracking software) into a comprehensive software for the National EPI and implemented in the entire Ben Tre province. During 2016 and 2018, leveraging government-led efforts and private sector partnerships, this system was finalized into the National Immunization Information System (NIIS) and scaled up to a national-level system. Since its launch, the NIIS has been used in 99.8% of facilities across provinces and their districts nationwide, and >20 million children and women have been registered in the system. As a result, immunization coverage and timeliness of vaccinations have drastically improved [4]. Using the NIIS has also alleviated the burden on immunization officers across all levels. The time spent on planning and reporting vaccinations has drastically reduced, and as a result, it is more accurate and effective. A timely, complete, and transparent database has provided higher-level officers (district, provincial, and national) comprehensive information to better guide and strategically plan to improve immunization coverage, as well as respond promptly in cases of emergency.

However, to date, end users have been tasked with managing both NIIS and paper-based systems in parallel until a formal assessment on the readiness to fully transition to the NIIS is conducted. Even though this parallel process was burdensome for end users, it was an important and intentional part of the rollout process to troubleshoot any unanticipated issues with the NIIS and allowed end users to transition to the new digital registries.

Goal and Objectives

Thus, the main goal of this study is to evaluate and compare the readiness of 2 provinces in Vietnam in moving entirely from a paper-based to paperless system for immunization records, with an emphasis on accuracy, completeness, consistency, timeliness, and data quality.

The following objectives are used to determine readiness to transition to the NIIS: (1) assess the infrastructure, capacity, need for human resources, and indicators for data accuracy; (2) assess NIIS end users' perceptions and feedback; (3) evaluate the data quality of the NIIS; (4) compare the accuracy of stock data that were physically counted to stock data in the NIIS; and (5) evaluate the knowledge and skills of NIIS end users.

Methods

Study Area

This study was conducted in 2 provinces in Vietnam, Ha Noi City and Son La. These provinces were selected as they represented varying levels of readiness with respect to geography (urban, semiurban, rural, and mountainous areas), population density, facility type (fee-based, private, and public), and facility level (national, provincial, district, and commune). Ha Noi is the capital city and mostly urban with a large and growing population, high immigration rate, a good infrastructure system, and a higher number of private sector and fee-based facilities. In comparison, Son La is a primarily bordered mountainous province with limited resources and fee-based facilities.

Study Design and Data Collection

Table 1 shows the research methods used to address each specific objective as described below, including the sampling strategy, data collection method, and type of data collected. All data were collected via the Kobo toolbox software [9], which identified data errors and missing data via built-in programmed checks.

Table 1. Sampling strategy, data collection methods, and types of data collected for addressing the five study objectives.

Objective	Sampling strategy	Data collection methods	Type of data collected
1. Assess and compare infrastructure, capacity, and human resources needs in 2 provinces that represent different levels of readiness	• All immunization health facilities (N=1026) in Ha Noi and Son La Provinces	• Self-filled forms in Kobo toolbox sent via email	• Immunization services provided: EPI[a], non-EPI, and hepatitis B birth dose vaccine coverage • Infrastructure conditions: electricity, computers, type of internet connection, barcode printers, and readers • Current NIIS[b] use: NIIS log-in and username • End users implementing the NIIS: number of people trained in use of NIIS or need for additional trainings • Indicators for data accuracy: data entry for immunizations and stock balances
2. Assess NIIS end users' perceptions and feedback	• A total of 767 districts and communes were purposively selected	• Self-filled forms in Kobo toolbox sent via email	• Time spent planning and reporting before and after NIIS • Perceptions of NIIS system on managing data and quality of data • Feedback for system improvement
3A. Evaluate data quality between immunization cards and the NIIS	• Son La: All children born after July 1, 2017, purposefully selected from 1 village in each of the 6 selected communes[c] • Ha Noi: 7[d] children born after July 1, 2017, in each village or living quarter identified door-to-door, in each of the 12 selected communes[c]	• Review and compare data between immunization cards and the NIIS	• Immunization data on immunization cards of selected children • Immunization data in the NIIS of selected children
3B. Evaluate data quality between paper-based records and the NIIS	• 20 randomly selected children from each of 14[e] purposefully selected fee-based facilities and hospitals	• Review and compare data between paper-based logbooks at health facilities and the NIIS	• Immunization data written in the facility paper-based logbook of selected children • Immunization data in the NIIS of selected children
3C. Evaluate data quality between paper-based records and the NIIS for BCG[f] vaccination administered in previous month	• 18 randomly selected commune health centers and district health centers	• Review and compare count of BCG immunizations among children in facility paper-based logbooks and in NIIS for previous month	• Number of children vaccinated in previous month with the BCG vaccine per paper-based logbook • Number of children vaccinated in previous month with the BCG vaccine in the NIIS
4. Compare the accuracy of stock data that were physically counted to stock data in the NIIS	• Random selection of 3 vaccine lots in 30 health facilities with established NIIS system	• Review of actual stock of 3 randomly selected vaccine lots and in NIIS	• Physical stock balance of selected lots • Stock balance per NIIS of selected lots
5. Evaluate the knowledge and skills of NIIS end users	• Convenience sampling of 1 or 2 NIIS end users from each purposefully selected fee-based facility and hospital	• Observation and interviews with NIIS end users	• NIIS knowledge and skills of end users using the three modules: immunization registry, stock management, and reporting

[a]EPI: Expanded Program on Immunization.

[b]NIIS: National Immunization Information System.

[c]Overall, 6 out of 30 districts in Ha Noi and 3 out of 12 districts in Son La.

[d]Due to the large population size in Ha Noi province, the World Health Organization's 30-clusters methodology was applied.

[e]One hospital that did not use the NIIS at the time of survey was excluded.

[f]BCG: Bacillus Calmette-Guérin.

Objective 1

To describe and compare the characteristics of immunization facilities, human resources, and data accuracy indicators, all immunization health facilities in Ha Noi and Son La provinces were emailed a web-based survey about the immunization services provided; infrastructure conditions; current NIIS use; number of NIIS trained end users, including health care workers and managers at different levels, and at different types of facilities such as commune health centers, hospitals, and fee-based immunization facilities; implementing the NIIS; and NIIS acceptance and feedback. The number and frequency by province were evaluated and compared using the Wilcoxon Mann-Whitney test.

Objective 2

To evaluate perceptions and feedback, the NIIS end users were purposively selected from districts and communes as follows: (1) of the 30 districts in Ha Noi, 6 (20%) were selected to represent a mix of population densities, facility types (fee-based, private, and public), and geographical areas (urban, semiurban, and rural); (2) of the 12 districts in Son La, 3 (25%) were selected to represent urban, rural, and mountainous areas; (3) 2 communes within each district, 1 to represent the smooth operation of the NIIS and the other to represent a commune with challenges were selected; (4) 6 hospitals in Ha Noi to represent different levels and types (urban, rural, private, and public) were selected; if there was >1 hospital in each category, then the hospital with the highest number of newborns in 2018 served as a substitute for purposively selecting high-volume hospitals; in Son La, 3 hospitals representative of urban and rural hospitals were selected; and (5) 2 fee-based hospitals and 1 public facility were selected in Ha Noi, and 3 fee-based facilities were selected in Son La. The NIIS end users were asked a series of survey questions to estimate the time it took them to plan and report on immunization before and after NIIS implementation and to evaluate their acceptability of the NIIS and its data quality that were compiled and analyzed. In addition, users' feedback on how the NIIS system could be improved was elicited via open-ended questions; common themes were grouped and frequencies were noted to help the Ministry of Health (MOH) and software developers improve the NIIS.

Objective 3

To evaluate data quality between immunization cards and NIIS in objective 3A, all children born between July 1, 2017, and the date of data collection were purposefully selected from 1 village in each of the 6 representative selected communes in Son La. In Ha Noi, 7 children born between July 1, 2017, and the date of data collection were selected from each village or living quarter using the World Health Organization's 30-clusters methodology [10] in each of the 12 representative selected communes.

To evaluate the data quality between paper-based records and NIIS in objective 3B, 20 children from each of 14 purposefully selected fee-based facilities and hospitals in both provinces were randomly selected, and their records were compared. In addition, in purposively selected commune health centers and district health centers representative of facilities with and without NIIS implementation challenges, the number of children who received a Bacillus Calmette-Guérin (BCG) vaccination in the previous month per paper-based logbook at facilities was compared with the NIIS.

Objective 4

To evaluate vaccine stock accuracy, stock balances of 3 vaccine lots were randomly selected from 30 health facilities with available vaccine stocks across both provinces that had an established NIIS, and physical stock and NIIS counts were compared.

Objective 5

To address the knowledge and skills of NIIS end users, 1 or 2 NIIS end users from each purposefully selected fee-based facility and hospital were observed and interviewed on three modules: immunization registry, stock management, and reporting. Observations were used to evaluate end user skills, and interviews were used to evaluate end user knowledge.

Statistical Analyses

Descriptive analyses, including counts, proportions, and means, were estimated for each objective and stratified by province where applicable. Responses from open-ended questions were counted and tabulated. All statistical analyses were conducted in Stata 14 (StataCorp LLC).

Ethics

The study procedures were reviewed and received a nonresearch determination by PATH because the activity does not meet the definition of *research*, as defined in 45 Code of Federal Regulation 46.102(l). Interviewers and supervisors were trained on interview techniques, data collection procedures and tools, ethical issues (such as how to protect the identities of study participants where applicable and how to secure all the data collected), and quality control. Interviewees were informed of the study's objectives and their rights of participation. Verbal consent was obtained from all participants.

Ethical Approval

This research protocol was reviewed and determined as not human subjects research by the PATH research determination committee. Before the interviews, all participants were informed and received an explanation of the scope and purpose of the study, the right to participate, and confidentiality of their personal information. No personal information was provided.

Results

Objective 1

There were a total of 1026 immunization health facilities in Ha Noi and Son La provinces; 96.49% (990/1026) of health facilities completed the initial survey about infrastructure, capacity, and the need for human resources. Of the 990 facilities, 747 (75.5%) were in Ha Noi and 243 (24.5%) were in Son La. Table 2 shows the characteristics of these facilities by province.

Table 2. Characteristics of immunization facilities that completed the web-based survey, human resources, and data accuracy indicators by province (N=990).

Type of facilities	Total	Ha Noi (n=747)	Son La (n=243)	*P* value[a]
Commune health center, n (%)	788 (79.6)	584 (78.2)	204 (83.9)	.04
District health center, n (%)	42 (4.2)	30 (4)	12 (4.9)	N/A[b]
Fee-based facilities, n (%)	72 (7.3)	59 (7.9)	13 (5.4)	N/A
Government hospitals (center, provincial, and district), n (%)	61 (6.1)	49 (6.5)	12 (4.9)	N/A
Private hospitals, n (%)	23 (2.3)	22 (2.9)	1 (0.4)	N/A
Other, n (%)	4 (0.4)	3 (0.4)	1 (0.4)	N/A
Vaccination services provided, n (%)				
EPI[c,d] vaccines only	548 (55.4)	354 (47.4)	194 (79.8)	<.001
Non-EPI[a] vaccines only	92 (9.3)	79 (10.6)	13 (5.4)	N/A
Both EPI and non-EPI vaccines	224 (22.7)	222 (29.8)	2 (0.8)	N/A
Hepatitis B Birth dose	51 (5.2)	42 (5.6)	9 (3.7)	N/A
EPI vaccines and Hepatitis B Birth dose	26 (2.6)	10 (1.3)	16 (6.6)	N/A
Infrastructure				
Has stable electricity, n (%)	971 (98.1)	747 (100)	224 (92.2)	<.001
Has at least one computer, n (%)	986 (99.6)	747 (100)	239 (98.4)	<.001
Number of computers, mean (SD)[e]	2.9 (2.3)	3.0 (2.4)	2.5 (1.9)	<.001
Number of computers ready for NIIS[f], mean (SD)	1.9 (1.5)	2.0 (1.7)	1.7 (0.9)	.04
Has cable or Wi-Fi internet connection, n (%)	944 (95.3)	747 (100)	197 (81.1)	<.001
Has 3G or 3G internet connection, n (%)	46 (4.7)	0 (0)	47 (18.9)	N/A
Has a barcode printer, n (%)	89 (8.9)	38 (5.1)	51 (20.9)	<.001
Has a barcode reader, n (%)	85 (8.6)	74 (9.9)	11 (4.5)	.009
Current NIIS use, n (%)				
Has NIIS log-in username	990 (100)	747 (100)	243 (100)	N/A
Currently using the NIIS system	987 (99.7)	744 (99.6)	243 (100)	.32
Provided personal immunization ID	881 (89.3)	740 (99.5)	141 (58)	<.001
Have ≥2 end users use NIIS	766 (77.4)	610 (81.7)	156 (64.2)	<.001
NIIS trained end users				
Has only 1 health worker trained on NIIS, n (%)	147 (14.9)	105 (14.1)	42 (17.3)	.22
Have ≥2 end users trained on NIIS, n (%)	825 (83.3)	633 (84.7)	192 (79)	.04
Additional training on the NIIS not required, n (%)	165 (16.7)	140 (18.7)	26 (10.6)	.002
Number of end users trained on NIIS, mean (SD)	3.1 (2.4)	3.3 (2.7)	2.6 (1.5)	.004
Number of end users that need additional training on NIIS, mean (SD)	3.7 (2.4)	3.7 (2.4)	3.7 (2.2)	.27
Indicators of data accuracy[g], n (%)				
Data entered during the immunization session	697 (70.4)	644 (86.2)	53 (21.8)	<.001
Data entered after the immunization session	290 (29.3)	101 (13.5)	189 (77.7)	N/A
Data are not entered	3 (0.3)	2 (0.3)	1 (0.4)	N/A
Entered stock data into NIIS	899 (90.8)	677 (90.6)	222 (91.4)	.73
NIIS stock balance is accurate	884 (89.3)	679 (90.9)	205 (84.4)	.62
NIIS immunization data matched paper records	753 (76.1)	612 (81.9)	141 (58.0)	<.001
Correctly reported number of fully immunized children	762 (76.9)	612(81.9)	150 (61.7)	<.001

[a]Non–Expanded Program on Immunization vaccines are not in the list of Expanded Program on Immunization vaccines, and clients have to pay from their own pocket for vaccination.

[b]N/A: not applicable; too few estimates to determine significance.

[c]Expanded Program on Immunization vaccines are vaccines introduced and provided to people (children and women) free of charge.

[d]EPI: Expanded Program on Immunization.

[e]Wilcoxon Mann-Whitney test was applied.

[f]NIIS: National Immunization Information System.

[g]Per the Ministry of Health 4-step immunization process (among those who are using the NIIS in the last month).

The distribution of facilities differed in both provinces (*P*=.04); although 79.6% (788/990) were commune health centers, 9.5% (71/747) of facilities in Ha Noi were government and private hospitals compared with 5.3% (13/243) of facilities in Son La. The type of vaccination services provided was significantly different between the provinces (*P*<.001). EPI vaccines were the most common type of service provided: 79.8% (194/243) of facilities in Son La compared with 47.4% (354/747) of facilities in Ha Noi. Among the 990 health facilities, 971 (98.1%) had stable electricity, and 986 (99.6%) had at least one computer (Table 2). Compared with Son La, facilities in Ha Noi had a higher mean number of computers (3.0, SD 2.4 vs 2.5, SD 1.9; *P*<.001), higher mean number of computers ready for NIIS (2.0, SD 1.7 vs 1.7, SD 0.9; *P*=.04), more cable or Wi-Fi internet connections (747/747, 100% vs 197/243, 81.1%; *P*<.001), and more barcode readers (74/747, 9.9% vs 11/243, 4.5%; *P*=.009) but less barcode printers (38/747, 5.1% vs 51/243, 21%; *P*<.001). In addition, all 990 facilities had a NIIS log-in username, and 99.7% (987/990) reported having used the NIIS. A total of 3 hospitals in Ha Noi did not use NIIS. Of the 990 facilities, 881 (89%) provided personal immunization IDs to newborn babies and clients; this was higher in Ha Noi than in Son La (740/747, 99.5% vs 141/243, 58%; *P*<.001).

Of the 990 facilities, 825 (83.3%) had >2 end users trained on NIIS and 165 (16.7%) did not require additional training (Table 2). However, there was a significant difference (*P*=.004) in the mean number of trained NIIS end users in Ha Noi as compared with Son La (3.3, SD 2.7 vs 2.6, SD 1.5), even though the mean number of end users requiring additional training was the same in both provinces (mean 3.7, SD 2.4 in Hanoi and mean 3.7, SD 2.2 in Son La).

Data entry differed by province (Table 2); 86.2% (644/747) of facilities in Ha Noi entered the data during the immunization session, compared with 21.8% (53/243) of facilities in Son La (*P*<.001). For data accuracy, of the total 990 facilities, 899 (90.8%) entered stock data into the NIIS, and 884 (89.3%) reported that their stock balance accurately reflected their actual physical stock. Although most facilities reported NIIS immunization data matched paper records and correctly reported the number of children fully immunized, Ha Noi had a significantly higher proportion than those in Son La (612/747, 81.9% vs 141/243, 58%; *P*<.001).

Objective 2

A total of 767 NIIS end users, 580 (75.6%) from Ha Noi and 187 (24.4%) from Son La, provided estimates of monthly time spent planning and reporting on immunization before and after NIIS. Before NIIS was launched, it took end users an average of 10.9 hours to prepare immunization plans and 13.2 hours to report on the immunization data in their facility. After NIIS was launched, the average amount of time spent on planning and reporting was 3.1 hours and 4.5 hours, respectively. These time estimates were significantly different and resulted in 72% (7.8/10.9) of the time saved on planning and 66% (8.7/13.2) of the time saved on reporting. Similar results were found for each province (data not shown). In addition, among 767 end users, 725 (94.5%) reported that the NIIS supported them in managing and reporting the immunization data in their respective facilities; however, 9 (1.2%) reported that the system did not support them, and 33 (4.3%) reported that it increased their workload.

Feedback from end users for system improvement identified that there is still a need to develop detailed guidelines and standard operating procedures for all components of the NIIS, from registration to reporting; provide more direction and supervision to allow end users to strengthen their skill sets; offer annual refresher training when new functions are added to the system; upgrade software, technology, and internet connections; and improve functions to easily identify and resolve errors. End users also reported that simplifying the number of indicators tracked, reducing the reporting forms, and ultimately eliminating the need to maintain parallel systems would help improve data quality and encourage them to use the NIIS more frequently for planning and reporting.

Objective 3

Table 3 shows the indicators used to assess the data quality between the immunization cards and NIIS. A total of 378 immunization cards of children were evaluated; of the 378 cards, 229 (60.6%) were from Ha Noi, and 149 (39.4%) were from Son La. Overall, 62.4% (236/378) of immunization cards had an attached or written NIIS ID. Of the 378 immunization cards, 338 (89.4%) were registered in the NIIS, 206 (54.5%) had NIIS IDs recorded on both their immunization cards and NIIS records, 330 (87.3%) had demographic information that matched with NIIS, 372 (98.4%) of immunizations received since birth were recorded on the immunization cards, and 360 (95.2%) were entered into the NIIS but only 335 (88.6%) of immunizations recorded on both immunization cards and the NIIS matched. These indicators were all statistically different based on the province.

Table 3. Data Quality Indicators used to compare National Immunization Information System (NIIS) and immunization cards (N=378).

Immunization cards	Total	Ha Noi (n=229)	Son La (n=149)	*P* value
Indicators, n (%)				
Immunization card had an NIIS ID	236 (62.4)	202 (88.2)	34 (22.8)	<.001
Child was registered in the NIIS	338 (89.4)	217 (94.8)	121 (81.2)	<.001
Both immunization card and NIIS had an NIIS ID	206 (54.5)	179 (78.2)	27 (18.1)	<.001
Demographic information on immunization cards and NIIS matched	330 (87.3)	209 (91.3)	121 (81.2)	.01
Immunizations recorded on immunization card	372 (98.4)	226 (98.6)	146 (97.9)	.02
Immunizations recorded in the NIIS	360 (95.2)	224 (97.8)	136 (91.3)	<.001
Immunizations recorded on immunization cards and NIIS matched	335 (88.6)	217 (94.8)	117 (78.5)	<.001

The data quality indicators used to compare the facilities' paper-based logbooks and NIIS records for 269 children are shown in Table 4. Of these 269 records, 254 (94.4%) were registered in the NIIS, 251 (93.3%) had immunization data entered into the NIIS, and 242 (89.9%) had matching immunization data between the paper-based logbook and the NIIS. The proportions of the immunization indicators in Ha Noi were significantly higher than those in Son La (*P*<.001 and *P*=.03). There was no difference in immunization registration. When comparing the number of BCG vaccinations recorded in the logbook and NIIS in the previous month, only 57% (13/23) of facilities had accurate data.

Table 4. Data quality indicators used to compare National Immunization Information System (NIIS) and facility paper-based logbooks (N=269).

Facility paper-based logbook	Total	Ha Noi (n=196)	Son La (n=73)	*P* value
Indicators, n (%)				
Child was registered in the NIIS	254 (94.4)	186 (94.9)	68 (93)	.58
Immunizations recorded in the NIIS	251 (93.3)	188 (95.9)	63 (86)	<.001
Immunizations recorded on paper-based logbooks and NIIS system matched	242 (89.9)	181 (92.4)	61 (84)	.03

Objective 4

Among 30 health facilities, 19 (63%) in Ha Noi and 11 (37%) in Son La, 18 (60%) facilities in Ha Noi tracked vaccine and supplies of 3 randomly selected lots in NIIS, as compared with 8 (27%) facilities in Son La (Table 5). However, only 84% (16/19) and 45% (5/11) of facilities had physical stock balances that matched the NIIS in Ha Noi and Son La, respectively.

Table 5. Data quality indicators used to compare National Immunization Information System (NIIS) and physical stock balances by province (N=30).

Physical stock	Total	Ha Noi (n=19)	Son La (n=11)	*P* value
Indicators, (%)				
Vaccine and supplies of 3 lots tracked in the NIIS	26 (87)	18 (95)	8 (72)	.13
Physical vaccine stock balances of 3 lots matched NIIS	21 (70)	16 (84)	5 (46)	.04

Objective 5

The knowledge and skills of end users on the NIIS immunization registry, stock management, and reporting modules are shown in Table 6. Overall, a direct relationship between end users' knowledge and skills was observed. Most end users were able to perform basic skills such as registering a client, entering data following a 4-step procedure in the NIIS, and updating personal information for an existing client. However, almost half struggled with more complex tasks, such as creating a vaccination plan by type of vaccine, deactivating immunization reminders, and identifying and eliminating duplicates. Similar results were observed across both provinces.

Table 6. Overall knowledge and skills related to the National Immunization Information System (NIIS) among end users on the NIIS main modules.

Function	Knowledge, n (%)	Skills, n (%)	Correlation	*P* value
Immunization registry	51 (100)	51 (100)	N/A[a]	N/A
Register clients	48 (94)	49 (96)	0.8	<.001
Search a client	36 (71)	35 (69)	0.86	<.001
Search a newborn registered by a hospital	24 (47)	23 (45)	0.94	<.001
Update personal information for an existing client	40 (79)	40 (78)	1	<.001
Enter data following 4-step procedure in the system	45 (88)	44 (86)	0.88	<.001
Add a migrant child from a given location into the list of children due for vaccination[b]	22 (52)	23 (55)	0.95	<.001
Deactivate immunization schedule reminder of a child[b]	24 (57)	22 (52)	0.90	<.001
Create a vaccination plan for children by vaccines[b]	25 (60)	24 (57)	0.95	<.001
Duplication filter[b]	24 (57)	23 (55)	0.86	<.001
Duplication handling[b]	23 (55)	22 (52)	0.95	<.001
Stock management	51 (100)	51 (100)	N/A	N/A
Create a stock receipt from a vaccine distributor	19 (37)	18 (35)	0.96	<.001
Create a stock receipt from higher level[b]	23 (55)	23 (55)	1	<.001
Create a dispatch for low level[c]	7 (78)	7 (78)	1	<.001
Create a vaccine use voucher	30 (59)	31 (61)	0.96	<.001
Report only applied for CDC[d], DHC[e], and CHCs[f]	26 (100)	26 (100)	N/A	N/A
Generate and export monthly immunization report for children aged under 1 year	19 (73)	18 (69)	0.91	<.001
Generate and export monthly immunization report for children aged over 1 year	17 (65)	16 (62)	0.92	<.001

[a]N/A: not applicable.

[b]Applied for Center for Disease Control and Prevention, district health center, commune health center, and hospitals (overall: n=42; Ha Noi: n=32; Son La: n=10).

[c]Applied for Center for Disease Control and Prevention and district health center (overall: n=9; Ha Noi: n=6; Son La: n=3).

[d]CDC: Center for Disease Control and Prevention.

[e]DHC: district health center.

[f]CHC: commune health center.

Discussion

Principal Findings

The readiness to transition to an entirely digital immunization reporting system in 2 provinces, Ha Noi and Son La, in Vietnam, was promising. Overall, there was a high level of use of the NIIS for immunization registry (987/990, 99.7%) and for tracking vaccine stock data (889/990, 90.8%). A total of 3 hospitals in Ha Noi did not use NIIS because they did not provide any vaccinations to newborns, or they used their own system to record immunization data. Evidence suggests that ownership and acceptability are critical components that facilitate the use and scale up [11]. Given that the NIIS is an integrated system with established standards to support and guide transition and scale up, it is imperative that provincial health authorities encourage all facilities to use the NIIS, or at least share data with the NIIS, especially if they run a facility-owned information system.

Most facilities in both provinces had sufficient infrastructures, such as stable electricity, at least one computer, and cable or Wi-Fi internet connections. In Son La, fewer computers were ready for NIIS and 18.9% (47/243) of facilities were still using a 3G or 4G internet connection. In addition, there were few barcode readers and printers across all the facilities. This was primarily because of limited resources, where programs within the facility often must share the only computer available and the limited resources available to upgrade their internet connection. To encourage the uptake and sustainability of systems such as the NIIS, budgets should be revised so that related expenses, such as power supply, critical equipment, and internet connection needed to support NIIS, can be procured [12-14]. Moreover, planning and making provisions for possible offline data entry in the interim could be beneficial in areas with poor internet connections or unstable power supplies [13].

Sufficient training is essential for end users to adopt and use digital health solutions, such as the NIIS [15]. To ensure that NIIS is sustainable in that immunization data are entered and

updated in a timely manner, each facility should have at least two trained end users to use NIIS. However, only 77.4% (766/990) of facilities met these criteria, and facilities in both provinces requested additional training for an average of 3.7 staff members. This may be a result of varying skill levels and high turnover rates of end users. However, when comparing the knowledge and skills of a subset of NIIS end users, there was a clear correlation between the two. Hence, more and frequent refresher training is needed to ensure that end users can apply the knowledge and skills gained from training to their daily work. Findings from a recent study indicate that more time should be allocated for initial training, and training should be offered at various times to account for busy and demanding work schedules [2]. Trainings should also be revised over time, varied depending on the end users [2], and easy to replicate and roll out for end users across locations [15]. In addition, training using real-time data is often an overlooked factor that can empower users to track their own performance and, in turn, motivate peers [15]. In addition, facilities in remote areas or with limited resources can access remote e-learning via mobile apps or web-based platforms that have been shown to be less expensive but just as effective as traditional in-person training [16]. Furthermore, facilities can consider recognizing and rewarding high-performing staff in a transparent manner as a means to support, motivate, encourage, and hopefully retain end users [15,17,18].

With regard to system acceptance and feedback of end users for system improvement, 94.5% (725/767) of end users interviewed in both provinces reported that the NIIS supported them in managing and reporting immunization data and agreed that the quality of immunization data was better. These results are consistent with findings from Zambia, where 94% (83/89) of end users reported that data accuracy was good or excellent, and 28% (25/89) and 27% (24/89) reported an increase in their ability to identify areas with low vaccine coverage and children who have missed vaccines, respectively [3]. On the other hand, user satisfaction with an electronic medical record system use in 5 low-resource setting hospitals in Ethiopia was low with an overwhelming preference for paper-based records; respondents strongly disagreed that the system helped finish tasks faster or had a positive effect on the quality of care provided [13].

The finding that 4.3% (33/767) of end users reported that NIIS increased their workload is not surprising, as end users were purposefully asked to maintain paper-based records in addition to the NIIS primarily to identify bugs in the NIIS and update the system based on user feedback. A recent study on the analysis of EIR data in Tanzania showed that facilities that had transitioned to paperless reporting were more likely to use the EIR compared with facilities that were still responsible for reporting through parallel paper-based and paperless systems, when controlling for other factors. The authors hypothesized that this was mainly because of health care worker bandwidth and motivation [19]. Although the MOH had designated that on June 1, 2018, Vietnam would fully transition to only using the NIIS, there was no guidance on how this should be done. According to lessons learned from the Better Immunization Data initiative, overall data quality can suffer as a result of not setting expectations for both systems [20]. As expected, the mean time to create monthly immunization plans and reports after NIIS implementation on average was much lower than the mean time for these activities before NIIS implementation. These findings are consistent with results from Tanzania, where health care workers spent 41% less time registering and updating data, and as a result, saved 8 working days per year that could be reallocated to patient care [3]. Although the time spent on monthly planning and reporting was still higher than desired, this was mainly because of maintaining the NIIS along with the paper-based logbooks in parallel. However, as confidence in the NIIS increases and local authorities remove the paper-based system, end users will be able to fade out using the paper-based system, thus reducing the time spent on these activities.

Evaluation of the accuracy of data was encouraging, with 80.3% (795/990) of facilities reporting that the immunization data in the NIIS was accurate with the results from the paper system. This was validated upon further observation, where 88.6% (335/378) and 89.9% (242/269) of vaccination data matched between the NIIS and immunization cards and paper-based logbooks, respectively. In addition, vaccine stock counts were entered into the NIIS system at 90.8% (899/990) of the facilities. However, there is room for improvement; only 54.5% (206/378) of children had a matching ID on their immunization card and in the NIIS, 57% (13/23) of BCG vaccinations recorded in the NIIS were consistent with paper-based logbooks, and only 70% (21/30) of the facilities had consistent physical vaccine stock balances. These findings may be because of a delay in entering this information into the NIIS. Hence, the importance of the timeliness, completeness, and accuracy of the data should be reinforced regularly. Moreover, the feedback received from end users for system improvement suggests that there are still needs that should be addressed as soon as possible to ensure improvements in data quality and ultimately enable a transition to only using the NIIS.

It is important to highlight the importance of considering the feedback received from users on the need for more detailed guidelines and standard operating procedures across all facets of NIIS. User-based design has been the cornerstone of the successful expansion and use of digital health initiatives [15,21]. In Ghana, user-based design was critical in the expansion and rollout of telemedicine and led to the reduction of unnecessary hospital referrals [21].

Data storage, access, and confidentiality are additional essential components that can influence scale and transition [15]. End users can only use NIIS to access and enter the data into the system. The system was built to link tables and data elements or variables, and data dictionaries were aligned and formatted according to MOH national standards to allow for interoperability and the ability to connect with other systems. Per MOH regulations, built-in system user authentication allows users to tailor accessibility and permissions depending on their role and level. The NIIS database is hosted by Viettel, a state-owned company that hosts other government systems and is well known for its data security services. In addition, training includes components of user awareness and data protection and emphasizes that all users, regardless of role and level, are responsible for data accuracy, completeness, and privacy of data.

Limitations

This study has several limitations. First, as only 2 provinces were selected, these findings may not be generalizable to the country as a whole. Second, the selection of facilities, children, and end users may not be representative of the overall population in Vietnam or other low to middle-income resource settings. Third, responses from the facility assessment, including infrastructure, capacity, and the need for human resources and NIIS end users' perceptions and feedback, were self-reported by end users and, therefore, not objectively verified.

Conclusions

This study showed that Ha Noi and Son La provinces in Vietnam were almost ready to fully transition to the NIIS in terms of NIIS use, infrastructure, and end users' basic skills. However, there is still a need for additional support and resources to improve timeliness, completeness, and accuracy of the data and strengthen NIIS end users' skills. In the meantime, it is imperative that any software issues with the NIIS are fixed, barcode IDs are issued for all children, and e-cards are launched soon so as to encourage NIIS uptake, especially in fee-based facilities, and to improve the data quality of the NIIS.

The design of the readiness assessment could be replicated in other provinces in Vietnam, as well as in other countries considering scale up of similar EIRs. Moreover, the findings from this study can inform other provinces in Vietnam or other countries on the types of challenges that must be addressed before fully transitioning to paperless reporting. As others have recommended [20], countries that are planning to scale an EIR should develop a clear roadmap for the eventual transition away from paper-based reporting, which may include plans for a readiness assessment.

Acknowledgments

The authors thank the Northern Expanded Program on Immunization, Hanoi and Son La Centers for Disease Control and Prevention, and Hanoi Medical University Teams for their support and collaboration in conducting this assessment. The authors also thank Hung Pham, Thi Hoang, and Hai Le from the PATH team and Pham Quang Thai from the National Institute of Hygiene and Epidemiology for their contributions to this research. This study was funded by the Bill and Melinda Gates Foundation (grant OPP1198325). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The findings and conclusions of this study are those of the authors and do not necessarily reflect the positions or policies of the Bill and Melinda Gates Foundation. The authors would also like to thank their guarantor Dr Nga Nguyen for supporting this study.

Authors' Contributions

All authors contributed to, edited, and approved the final version of the manuscript.

Conflicts of Interest

None declared.

References

1. Seymour D, Werner L, Mwansa FD, Bulula N, Mwanyika H, Dube M, et al. Electronic immunization registries in Tanzania and Zambia: Shaping a minimum viable product for scaled solutions. Front Public Health 2019 Aug 07;7:218 [FREE Full text] [doi: 10.3389/fpubh.2019.00218] [Medline: 31440494]
2. Dolan SB, Alao ME, Mwansa FD, Lymo DC, Bulula N, Carnahan E, et al. Perceptions of factors influencing the introduction and adoption of electronic immunization registries in Tanzania and Zambia: A mixed methods study. Implement Sci Commun 2020 Mar 30;1:38 [FREE Full text] [doi: 10.1186/s43058-020-00022-8] [Medline: 32885195]
3. Bid initiative monitoring and evaluation reports (Tanzania and Zambia). Bid Learning Network. URL: http://bidinitiative.org/resource-library/posts/bid-initiative-monitoring-and-evaluation-reports-tanzania-and-zambia/ [accessed 2020-06-22]
4. Nguyen NT, Vu HM, Dao SD, Tran HT, Nguyen TXC. Digital immunization registry: Evidence for the impact of mHealth on enhancing the immunization system and improving immunization coverage for children under one year old in Vietnam. Mhealth 2017 Jul 19;3:26 [FREE Full text] [doi: 10.21037/mhealth.2017.06.03] [Medline: 28828373]
5. Wilson K, Gertz B, Arenth B, Salisbury N. The journey to scale: Moving together past digital health pilots. PATH. Seattle: PATH; 2014 Dec. URL: https://www.path.org/resources/the-journey-to-scale-moving-together-past-digital-health-pilots/ [accessed 2020-06-22]
6. Lemaire J. Scaling Up Mobile Health: Elements necessary for the successful scale up of mHealth in developing countries. Advanced Development for Africa, Geneva. 2011. URL: https://toolkits.knowledgesuccess.org/toolkits/mhealth-planning-guide/scaling-mobile-health-elements-necessary-successful-scale-mhealth-developing-countrie-0 [accessed 2020-06-22]
7. Labrique A, Vasudevan L, Weiss W, Wilson K. Establishing standards to evaluate the impact of integrating digital health into health systems. Glob Health Sci Pract 2018 Oct 10;6(Suppl 1):5-17 [FREE Full text] [doi: 10.9745/GHSP-D-18-00230] [Medline: 30305335]

8. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: Where is the evidence? PLoS Med 2013;10(2):e1001382 [FREE Full text] [doi: 10.1371/journal.pmed.1001382] [Medline: 23424286]
9. Kobo Toolbox. URL: https://www.kobotoolbox.org/ [accessed 2020-06-22]
10. Vaccination coverage cluster surveys: Reference manual. World Health Organization. URL: https://www.who.int/immunization/monitoring_surveillance/Vaccination_coverage_cluster_survey_with_annexes.pdf [accessed 2020-06-22]
11. Njoroge M, Zurovac D, Ogara EA, Chuma J, Kirigia D. Assessing the feasibility of eHealth and mHealth: A systematic review and analysis of initiatives implemented in Kenya. BMC Res Notes 2017 Feb 10;10(1):90 [FREE Full text] [doi: 10.1186/s13104-017-2416-0] [Medline: 28183341]
12. Syzdykova A, Malta A, Zolfo M, Diro E, Oliveira JL. Open-source electronic health record systems for low-resource settings: Systematic review. JMIR Med Inform 2017 Nov 13;5(4):e44 [FREE Full text] [doi: 10.2196/medinform.8131] [Medline: 29133283]
13. Tilahun B, Fritz F. Comprehensive evaluation of electronic medical record system use and user satisfaction at five low-resource setting hospitals in Ethiopia. JMIR Med Inform 2015 May 25;3(2):e22 [FREE Full text] [doi: 10.2196/medinform.4106] [Medline: 26007237]
14. Fraser HS, Blaya J. Implementing medical information systems in developing countries, what works and what doesn't. AMIA Annu Symp Proc 2010 Nov 13;2010:232-236 [FREE Full text] [Medline: 21346975]
15. Labrique AB, Wadhwani C, Williams KA, Lamptey P, Hesp C, Luk R, et al. Best practices in scaling digital health in low and middle income countries. Global Health 2018 Nov 03;14(1):103 [FREE Full text] [doi: 10.1186/s12992-018-0424-z] [Medline: 30390686]
16. Rowe AK, de Savigny D, Lanata CF, Victora CG. How can we achieve and maintain high-quality performance of health workers in low-resource settings? Lancet 2005 Sep 17;366(9490):1026-1035. [doi: 10.1016/S0140-6736(05)67028-6] [Medline: 16168785]
17. Franco LM, Bennett S, Kanfer R. Health sector reform and public sector health worker motivation: A conceptual framework. Soc Sci Med 2002 Apr;54(8):1255-1266. [doi: 10.1016/S0277-9536(01)00094-6] [Medline: 11989961]
18. Adzei FA, Atinga RA. Motivation and retention of health workers in ghana's district hospitals: Addressing the critical issues. J Health Organ Manag 2012;26(4-5):467-485. [doi: 10.1108/14777261211251535] [Medline: 23115900]
19. Carnahan E, Ferriss E, Beylerian E, Mwansa FD, Bulula N, Lyimo D, et al. Determinants of facility-level use of electronic immunization registries in Tanzania and Zambia: An observational analysis. Glob Health Sci Pract 2020 Sep 30;8(3):488-504 [FREE Full text] [doi: 10.9745/GHSP-D-20-00134] [Medline: 33008860]
20. Bid initiative lessons learned encyclopedia. Bid Learning Network. URL: http://bidinitiative.org/wp-content/uploads/FINAL_LessonsLearned_Encyclopedia_6July2018-Logo.pdf [accessed 2021-06-24]
21. Telemedicine project in Ghana healthcare. The Novartis Foundation. 2010. URL: https://www.novartisfoundation.org/past-programs/digital-health/ghana-telemedicine [accessed 2021-06-06]

Abbreviations

BCG: Bacillus Calmette-Guérin
EIR: electronic immunization registry
EPI: expanded program on immunization
MOH: Ministry of Health
NIIS: National Immunization Information System

Edited by G Eysenbach; submitted 21.02.21; peer-reviewed by K Harrington, R Lakhtakia, R Poluru, A Yazdanian; comments to author 14.05.21; revised version received 01.07.21; accepted 27.07.21; published 25.10.21.

Please cite as:
Duong H, Dao S, Dang H, Nguyen L, Ngo T, Nguyen T, Tran LA, Nguyen D, Rivera M, Nguyen N
The Transition to an Entirely Digital Immunization Registry in Ha Noi Province and Son La Province, Vietnam: Readiness Assessment Study
JMIR Form Res 2021;5(10):e28096
URL: https://formative.jmir.org/2021/10/e28096
doi:10.2196/28096
PMID:34694232

Original Paper

Evaluating the Clinical Feasibility of an Artificial Intelligence–Powered, Web-Based Clinical Decision Support System for the Treatment of Depression in Adults: Longitudinal Feasibility Study

Christina Popescu[1*], BSc, MSc; Grace Golden[2*]; David Benrimoh[1*], MSc, CM, MD; Myriam Tanguay-Sela[1], BASc; Dominique Slowey[3], BSc; Eryn Lundrigan[3], BSc; Jérôme Williams[3], BSc; Bennet Desormeau[3], BSc, MSc; Divyesh Kardani[1], BSCE; Tamara Perez[3], BSc, MSc; Colleen Rollins[4], BSc; Sonia Israel[1], BSc; Kelly Perlman[1,3], BSc; Caitrin Armstrong[1], BASc, MSc; Jacob Baxter[3], BSc; Kate Whitmore[3], BA; Marie-Jeanne Fradette[3], BSc; Kaelan Felcarek-Hope[3]; Ghassen Soufi[3], BSc; Robert Fratila[1], BSc; Joseph Mehltretter[3], MSc; Karl Looper[3], MD; Warren Steiner[3], MD; Soham Rej[3], MSc, MD; Jordan F Karp[5], MD; Katherine Heller[6], BSc, MSc, PhD; Sagar V Parikh[7], MD; Rebecca McGuire-Snieckus[8], BA, MSc, PhD; Manuela Ferrari[9], PhD; Howard Margolese[3], MSc, CM, MD; Gustavo Turecki[9], PhD, MD

[1]Aifred Health Inc., Montreal, QC, Canada
[2]University of Waterloo, Waterloo, ON, Canada
[3]McGill University, Montreal, QC, Canada
[4]University of Cambridge, London, United Kingdom
[5]University of Arizona, Tucson, AZ, United States
[6]Duke University, Durham, NC, United States
[7]University of Michigan, Ann Arbor, MI, United States
[8]Barts and the London School of Medicine, London, United Kingdom
[9]Douglas Mental Health University Institute, McGill University, Montreal, QC, Canada
[*]these authors contributed equally

Corresponding Author:
David Benrimoh, MSc, CM, MD
Aifred Health Inc.
1250 Rue Guy Suite #600
Montreal, QC, H3H 2T4
Canada
Phone: 1 5144637813
Email: david.benrimoh@mail.mcgill.com

Related Article:

This is a corrected version. See correction statement: https://formative.jmir.org/2024/1/e56570

Abstract

Background: Approximately two-thirds of patients with major depressive disorder do not achieve remission during their first treatment. There has been increasing interest in the use of digital, artificial intelligence–powered clinical decision support systems (CDSSs) to assist physicians in their treatment selection and management, improving the personalization and use of best practices such as measurement-based care. Previous literature shows that for digital mental health tools to be successful, the tool must be easy for patients and physicians to use and feasible within existing clinical workflows.

Objective: This study aims to examine the feasibility of an artificial intelligence–powered CDSS, which combines the operationalized 2016 Canadian Network for Mood and Anxiety Treatments guidelines with a neural network–based individualized treatment remission prediction.

Methods: Owing to the COVID-19 pandemic, the study was adapted to be completed entirely remotely. A total of 7 physicians recruited outpatients diagnosed with major depressive disorder according to the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* criteria. Patients completed a minimum of one visit without the CDSS (baseline) and 2 subsequent visits where the CDSS was used by the physician (visits 1 and 2). The primary outcome of interest was change in appointment length after the introduction of the CDSS as a proxy for feasibility. Feasibility and acceptability data were collected through self-report questionnaires and semistructured interviews.

Results: Data were collected between January and November 2020. A total of 17 patients were enrolled in the study; of the 17 patients, 14 (82%) completed the study. There was no significant difference in appointment length between visits (introduction of the tool did not increase appointment length; $F_{2,24}$=0.805; mean squared error 58.08; P=.46). In total, 92% (12/13) of patients and 71% (5/7) of physicians felt that the tool was easy to use; 62% (8/13) of patients and 71% (5/7) of physicians rated that they trusted the CDSS. Of the 13 patients, 6 (46%) felt that the patient-clinician relationship significantly or somewhat improved, whereas 7 (54%) felt that it did not change.

Conclusions: Our findings confirm that the integration of the tool does not significantly increase appointment length and suggest that the CDSS is easy to use and may have positive effects on the patient-physician relationship for some patients. The CDSS is feasible and ready for effectiveness studies.

Trial Registration: ClinicalTrials.gov NCT04061642; http://clinicaltrials.gov/ct2/show/NCT04061642

(JMIR Form Res 2021;5(10):e31862) doi:10.2196/31862

KEYWORDS

clinical decision support system; major depressive disorder; artificial intelligence; feasibility; usability; mobile phone

Introduction

Background

Clinical decision support systems (CDSSs) consolidate large quantities of clinical information to provide clinicians with the necessary data to support medical decision-making and assist with managing treatment protocols [1-3]. An emerging focus of medical informatics is the improvement of patient care through data-driven, patient-centered decision support systems. Artificial intelligence (AI) algorithms are increasingly being integrated into CDSSs, permitting predictive analytics to be used by clinicians as part of routine practice [3]. The overarching objective of these systems is to improve medical decision-making using a data-driven approach. However, although much has been written about machine learning techniques [4,5] that underpin the technical advancements that make these systems possible, comparatively less focus has been placed on the usability and feasibility of these kinds of systems in medicine in general and in mental health treatment in particular. In this paper, we discuss a feasibility study of a novel AI-powered CDSS aimed at improving the treatment of depression.

Feasibility and ease of use are major concerns as they roughly equate to the tolerability of drug treatment; with similar impact—much like a medication—a digital tool can only have a positive impact if patients (and, in this case, clinicians) use it and continue to use it. A recent meta-analysis of randomized controlled trials aimed to establish the dropout rates of studies on medical smartphone apps tracking depressive symptoms [6]. The analysis found that apps for depressive symptom tracking had a dropout rate of approximately 50% when accounting for bias. Despite this high dropout rate, there is some knowledge about how to reduce dropouts. For example, researchers found that the dropout rate was significantly lower—as low as 12%—in apps offering human feedback and in-app mood monitoring [6]. In addition, previous decision support systems have demonstrated the need for a tool that provides real-time utility [7] and the ability to personalize treatment choices and differentiate between medications [8] in a quantifiable manner [9] and incorporates clinical practice guidelines [10].

Only one-third of patients with depression who receive treatment will achieve remission during the first treatment course; most experience multiple treatment trials before entering remission [11]. Clinicians are faced with a wide range of treatment options, in combination with associated guidelines, to manage the selected treatments. However, there are no easily accessible point-of-care tools available to aid in the optimization of treatment success and minimize the time to remission. Furthermore, treatments are essentially equally effective at the population level; as such, to improve outcomes, treatment selection must address the individual's specific characteristics [10,12]. As such, there is a clear need for improved and personalized decision support for mental health care [13].

The Aifred CDSS

Aifred is a CDSS that uses AI to assist clinicians in selecting treatments for major depressive disorder (MDD). The tool incorporates a deep-learning model that was validated and trained on clinical and demographic baseline data to support treatment selection by providing individualized probabilities of remission for specific treatment options. Please see the study by Benrimoh et al [10] for a description of the tool, and the studies by Mehltretter et al [4,14] for a description of the machine learning model and model training and validation methodology. Clinicians accessing the app first see their patient's self-report questionnaire scores and can examine these through graphs showing trends over time or at the individual item level. They can then select the clinical algorithm, which is an operationalized version of the Canadian Network for Mood and Anxiety Treatments (CANMAT) 2016 guidelines for the treatment of depression [15]. These operationalized guidelines

function in an entirely rule-based manner and take in patient depression scores at baseline and at subsequent visits to determine if patients have achieved early response or remission (based on guideline-informed criteria and questionnaire remission thresholds, respectively) and provide guideline-appropriate information at each patient visit. At all times, clinicians remain in control of clinical decisions and may select treatments or modify treatments as they feel are clinically indicated; there were no automated clinical decisions in this study. The AI aspect of the CDSS is specifically meant to assist clinicians with treatment choices. It is directly integrated into the operationalized CANMAT guidelines, with personalized predicted remission probabilities for individual treatments presented within the guideline module when antidepressant treatments are being chosen. This occurs in the following manner: whenever first-line treatment options are presented in the clinical algorithm, the AI model will provide predictions in the form of remission probabilities for a number of antidepressant medications based on a mix of symptom and demographic information provided by patients at baseline. These remission probability predictions are generated by the AI model using questionnaires responded to early in treatment and are meant to help guide initial treatment selection or change; further treatment management support (eg, information about switching, dose adjustment, or augmentation options if patients do not show improvement with treatment) is provided longitudinally by the rule-based CANMAT algorithm. The AI is designed to support clinicians by considering complex interactions among multiple patient clinical, social, and demographic variables to help personalize treatment to improve upon a trial-and-error treatment approach and reduce the number of failed treatment trials [10]. To summarize, the app assists clinicians in providing measurement-based, treatment algorithm–guided, and AI-personalized care.

Providing patients with the ability to monitor their own mental health symptoms has attracted a great deal of interest across all ages [16]. Patients also have access to their own version of the Aifred app wherein they respond to questionnaires and can view their active and past treatments, as well as their symptoms graphed over time. This availability of data to both physicians and patients is intended to empower patients, enrich conversations, and facilitate shared decision-making [10].

Study Aims

Following a previous simulation center study [10] and ahead of larger clinical trials aimed at assessing safety and effectiveness, we decided to conduct a feasibility study aimed at exploring the feasibility of the CDSS in a real clinical setting and to assess its longitudinal impact on the patient-clinician relationship. This study has 4 aims:

1. To assess the feasibility of the CDSS for use in clinical practice
2. To assess physician and patient trust in the CDSS and its effect on the clinician-patient relationship
3. To assess the usability of the CDSS and study software and to ensure that major limitations are identified and rectified before clinical trials
4. To assess engagement with the app

One key metric brought up by clinicians interviewed during initial stakeholder conversations was appointment length; clinicians are increasingly required to interact with time-consuming digital systems, and the fear of yet another system adding time to assessments is a reasonable one [17]. We aim to measure appointment length as our primary outcome, as a key numerical proxy for real-world feasibility.

Methods

Overview

The study was approved by the research ethics board of the Douglas Mental Health University Institute (identifier: NCT04061642). All participants provided written informed consent to participate. The study was conducted according to the ethical principles stated in the Tri-Council Policy Statement on the Ethical Conduct for Research Involving Humans [18].

This was a single-arm, naturalistic follow-up study aimed at assessing software usability and acceptability conducted between January and November 2020. This study was not designed to assess the clinical effectiveness of the tool, which will be the focus of an upcoming clinical trial. It is important to note that physicians were provided access to the tool but were free to use the tool and its AI predictions or ignore it.

The study sample included 2 population groups: (1) physicians, including family physicians and psychiatrists and (2) patients of these physicians. The recruitment target was 10 physicians and 3-4 patients per physician (30-40 patients in total).

Physicians were recruited via recruitment email and direct contact by the study personnel. The sites consisted of university hospitals, primary care clinics, and psychiatric clinics in the Canadian province of Québec. Eligible physicians were family physicians or psychiatrists treating patients with depression on at least a monthly basis. Physicians who met the eligibility criteria were then invited to attend an introductory session with study personnel where the study and the AI model were described, and training on how to use the tool was administered.

Participating physicians informed their patients with MDD about the study and referred interested patients to the study personnel. Eligible patients were patients of enrolled physicians who were aged at least 18 years and diagnosed with MDD by the physician as per *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (*DSM-5*) criteria [19], able and willing to provide informed consent, and not diagnosed (or suspected) with bipolar affective disorder, as per DSM-5 criteria. Patients were required to be physically and mentally able to use a computer or smartphone (ie, to not be delirious or have sufficient cognitive function) but not necessarily to already be adept at it. Training was offered to patients who were not familiar with computer or mobile apps, and site-based computing resources were available to patients who did not have their own device; however, no patients required access to either these computing resources or training. Informed consent was obtained from patients, after which their account was created and linked to that of their physician. Patients accessed the web-based app via their own desktop or mobile devices; they were able to access the app in the setting of their choice, including at home or when

they met their clinician. Patients and clinicians were compensated for their time. Research assistants were available for clinicians and patients as needed to provide technical support and provided patients with a training session on the app after account creation.

Procedure

Upon account creation, patients were asked to complete the following questionnaires on the tool: Patient Health Questionnaire-9 (PHQ-9) to screen and track for depressive symptoms and their severity over time [20], General Anxiety Disorder-7 (GAD-7) to screen and track for anxiety symptoms and their severity over time [21], Alcohol Use Disorders Identification Test to screen for harmful alcohol use [22], Drug Abuse Screen Test to screen for the presence and severity of problematic drug use [23], and Standardized Assessment of Personality–Abbreviated Scale Self-Assessment to screen for personality disorders using a threshold of 3 points [24]. The results of the patient baseline questionnaire scores are summarized in Table 1. Patients identified other clinical comorbidities, such as migraines (2/14, 14% of patients) and anxiety disorders (7/14, 50% of patients), which are described in detail in Multimedia Appendix 1 [10,11,20,25-32], Table S2.

Patients were notified weekly by an automated email sent by the app to complete the PHQ-9; GAD-7; Patient Rated Inventory of Side Effects (to screen and track for specific antidepressant side effects and their severity); and Frequency, Intensity, Burden of Side Effects Rating (to assess the overall impact of antidepressant side effects) questionnaires [25].

Between obtaining informed consent and their next visit with their physician, patients met with study personnel to complete a demographic questionnaire (Table 2), Adverse Childhood Experiences questionnaire, and Life Events Checklist for DSM-5 (to screen for childhood or lifetime trauma [33,34]).

After enrollment into the study, patients had to complete a minimum of 3 visits: a baseline visit where their clinician did not use the tool, followed by at least 2 visits where the tool was used and appointment length was measured. These visits were intended to occur within a 4-month period per patient, with additional visits between or beyond visits 1 and 2 allowed and initially planned where the tool could be used; the visit frequency was intended to be at least monthly. However, as will be discussed, the impact of COVID-19 had a destabilizing effect on medical practices such that some patients experienced interappointment times that were longer than expected, leading to most patients completing only baseline, visit 1, and visit 2. However, this did not have an impact on the outcome, as only 2 postbaseline appointment lengths were intended to be measured. Patients were considered to have completed the study if they attended the baseline and visit 1 and 2 appointments at a minimum; completion of study-related tasks was not a criterion for study completion, given that this was a feasibility study focused on determining what patients would realistically complete. Research personnel recorded whether the baseline visit was an intake visit or a follow-up visit; this was relevant as initial intake visits are generally longer than follow-ups, and tracking this allowed for the adjustment of analyses such that initial intake visits would not artificially inflate the visit length at baseline. In the week preceding visit 1, patients completed the Quick Inventory of Depressive Symptomatology and met with study personnel to be administered the Inventory of Depressive Symptomatology, Clinician Rating [26,27]. These questionnaires were part of the set of questions used to generate the AI results.

Owing to COVID-19 and the public health recommendations released by the Québec government in March 2020, the study was adapted to be completed entirely from a distance. Originally, the protocol intended for appointment length to be recorded by research personnel on site, measured from the moment the patient entered the room to when they exited. However, because of the transition from in-person to telemedicine appointments (phone and video call), appointment length was measured as the length of the phone or video call during which the visit took place, as displayed on the physician or patient device and relayed verbally to research personnel. Further information about adaptation to COVID-19 can be found in the section *Telemedicine during COVID-19* in Multimedia Appendix 1.

After each visit where the tool was used, physicians were asked to complete a postappointment questionnaire describing device usability and any serious adverse events and to use the Udvalg for Kliniske Undersøgelser Side Effects Rating Scale [10,35] to record any side effects as perceived by the treating physician. Physician feedback was used to help identify and fix software errors (incorrect questionnaire score comparison in the CANMAT clinical algorithm noted at one visit and a broken link between pages in the CANMAT clinical algorithm noted at 2 back-to-back visits) and to refine the way the clinical algorithm presented the guidelines throughout the course of the study (by providing more context from the guideline paper for certain pieces of information, such as dose changes). Research personnel also administered the Brief Adherence Rating Scale to patients to estimate medication adherence since the prior visit [36].

Following visit 2, patients met with research personnel for end of study tasks, which consisted of completing the Quick Inventory of Depressive Symptomatology, the Scale to Assess Therapeutic Relationships in Community Mental Health Care-Patient (STAR-P) [37], a customized exit questionnaire designed specifically to capture elements of the experience of using this novel tool, as well as being administered the Inventory of Depressive Symptomatology–Clinician Rating and a custom semistructured interview. After all their patients completed the study, physicians were administered the Scale to Assess Therapeutic Relationships in Community Mental Health Care-Clinician (STAR-C), a customized exit questionnaire designed specifically to capture elements of the experience of using this novel tool, as well as a custom semistructured interview.

Table 1. Patient clinical baseline scores.

Questionnaire baseline score[a]	Patient, n (%)	Score, mean (SD)
SAPAS-SA[b] (n=15)		3.53 (2.23)
<3 points: negative screen for personality disorder	5 (33)	
≥3 points: positive screen for personality disorder. Further clinical evaluation is warranted.	10 (67)	
GAD-7[c] (n=14)		12.21 (5.81)
0-4: no or minimal anxiety	1 (7)	
5-9: mild anxiety	4 (29)	
10-14: moderate anxiety	3 (21)	
15-21: severe anxiety	6 (43)	
PHQ-9[d] (n=15)		14.80 (5.61)
0-4: minimal or no depression	1 (7)	
5-9: mild depression	1(7)	
10-14: moderate depression	4(27)	
15-19: moderately severe depression	7(47)	
20-27: severe depression	2(13)	
AUDIT[e] (n=15)		4.40 (3.58)
<8: negative screen for harmful alcohol use	12 (80)	
≥8: positive screen for harmful alcohol use	3 (20)	
DAST-10[f] (n=15)		2.40 (3.60)
0: no problems reported	8 (53)	
1-2: low level	3 (20)	
3-5: moderate level	0 (0)	
6-8: substantial level	3 (20)	
9-10: severe level	1 (7)	
WHODAS[g] (n=13)	N/A	54.62 (27.08)
QIDS[h] (n=12)		13.12 (6.04)
1-5: no depression	2 (17)	
6-10: mild depression	2 (17)	
11-15: moderate depression	3 (25)	
16-20: severe depression	4 (33)	
21-27: very severe depression	1 (8)	
IDS-C[i] (n=13)		30.00 (12.88)
0-11: no depression	0 (0)	
12-23: mild depression	4 (31)	
24-36: moderate depression	5 (38)	
37-46: severe depression	3 (23)	
47-84: very severe depression	1 (8)	
ACE[j] (n=13)		1.62 (1.45)
0	4 (31)	
1	3 (23)	
2	1 (8)	

Questionnaire baseline score[a]	Patient, n (%)	Score, mean (SD)
3	4 (31)	
>4	1 (8)	

[a]Life Events Checklist for Diagnostic and Statistical Manual of Mental Disorders-5 questionnaire results can be found in Multimedia Appendix 1.

[b]SAPAS-SA: Standardized Assessment of Personality–Abbreviated Scale Self-Assessment.

[c]GAD-7: General Anxiety Disorder-7.

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

[e]AUDIT: Alcohol Use Disorders Identification Test.

[f]DAST-10: Drug Abuse Screen Test-10.

[g]WHODAS: World Health Organization Disability Assessment Schedule.

[h]QIDS: Quick Inventory of Depressive Symptomatology.

[i]IDS-C: Inventory of Depressive Symptomatology, Clinician Rating.

[j]ACE: adverse childhood experiences.

Table 2. Patient demographics[a].

Patient characteristics	Values
Age (years; n=14), mean (SD)	36.43 (14.84)
Gender (n=14), n (%)	
Men	5 (36)
Women	9 (64)
Ethnicity (n=13), n (%)	
White	10 (77)
Caribbean	1 (8)
African or African American	1 (8)
Unanswered	1 (8)
Adoption status (n=13), n (%)	
Not adopted	12 (92)
Adopted	1 (8)
Residency status (n=13), n (%)	
Canadian citizen	11 (85)
Immigrant status (>5 years ago)	1 (8)
Immigrant status (<5 years ago)	1 (8)
Relationship status (n=13), n (%)	
Married	4 (31)
Divorced	1 (8)
Dating a single partner	2 (15)
Not in a relationship	6 (46)
Employment status (n=13), n (%)	
Full time	7 (54)
Part time	1 (8)
Disability (not working)	2 (15)
Unemployed and volunteer work	2 (15)
Unemployed	1 (8)
Highest level of education (n=12), n (%)	
Master's degree	3 (25)
Bachelor’s degree	1 (8)
University, no degree	3 (25)
Collège d’enseignement général et professionnel	3 (25)
High school or equivalent (GED[b])	2 (17)
Years of education (n=12), mean (SD)	15 (6.42)
Income (n=10, Can $; US $), mean (SD)	82,333.33 (70,099.93); 64,829.39 (55,196.80)

[a]14 patients completed the study; however, 1 patient did not complete the demographic questionnaire.

[b]GED: General Educational Development.

Analyses

Assessing the Feasibility of the CDSS for Use in Clinical Practice

Appointment lengths at baseline, visit 1, and visit 2 were extracted from the research assistant appointment log and analyzed using the SPSS (version 27) repeated measures analysis of variance (ANOVA; 3 factors, within-subject variables). Although complete appointment length data were available for 14 patients, one of the patients' baseline appointments could not be counted as the physician opened the tool to look at questionnaire answers and input current patient medications. Data from 13 patients were included in the final repeated measures ANOVA, with their means and SDs reported below. Descriptive data about the subjective view of appointment length were also extracted from the custom exit questionnaires.

To determine the validity of our sample with respect to the main outcome of appointment length, we conducted 2 sample size calculations using the G*Power package, version 3.1.9.7 [38]. The first was a sample size calculation for a repeated measures ANOVA with 1 group, three measures, a moderate effect size estimate (0.25), an α of .05, 80% power, and a correlation between responses of 0.5. This was intended to determine the power required to detect within-group differences over time. The calculated required sample size was 28, validating our target sample size of 30-40 patients. Our second sample size estimation was based on the need to detect a clinically significant increase in time, which was set as being 5 minutes based on physician comments in a previous study [10]. Here, we calculated the sample required to detect an increase of at least 5 minutes (ie, target sample mean of 25, SD 5 minutes) compared with a standard 20-minute interview (SD is an estimate based on a reasonably expected variation in appointment length). The α was set at .05 and power at 80%, resulting in a required sample size of 8 patients.

Assessing Physician and Patient Trust in the CDSS and Its Effect on the Clinician-Patient Relationship

Descriptive data about physician and patient trust in the CDSS were extracted from custom exit questionnaires. The physician-patient relationship was assessed by examining STAR-P and STAR-C scores, as well as patient ratings of the relationship extracted from the custom exit questionnaires.

Assessing the Usability of the CDSS

Descriptive data concerning physician and patient ratings of CDSS ease of use were extracted from custom exit questionnaires.

Assessing Engagement With the App

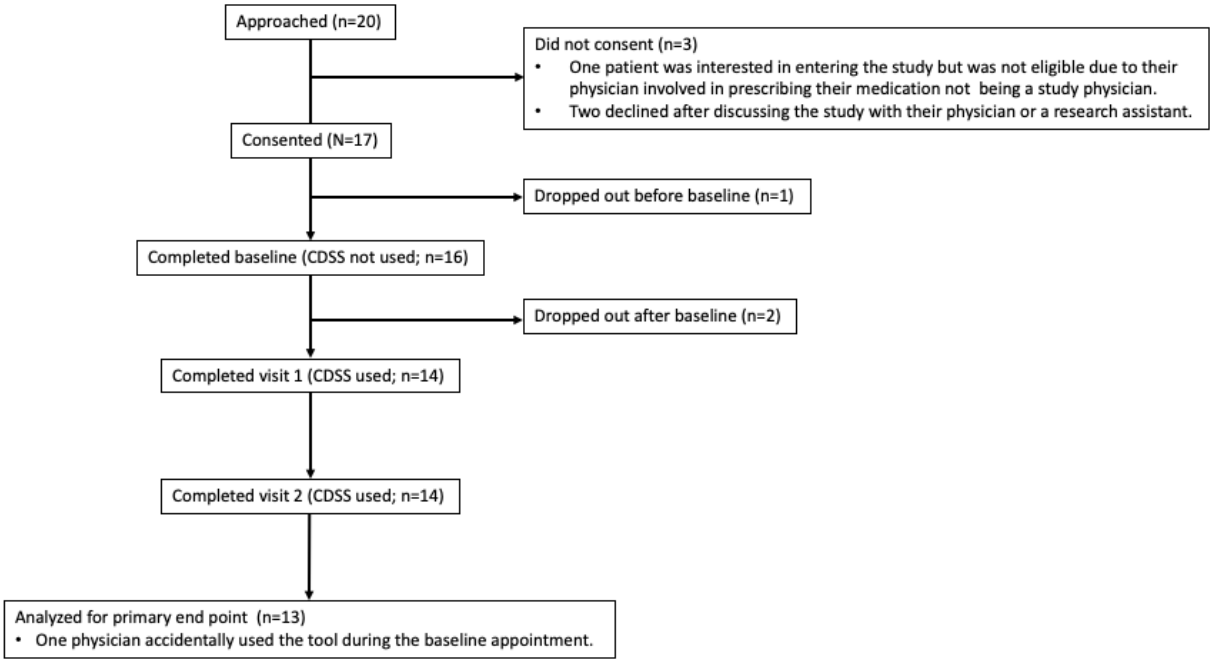
Physician engagement was assessed by determining the percentage of study visit days where the app was opened by the physician; this was determined by checking physician log-in dates and their correspondence to patient appointment dates. Patient engagement was assessed by measuring the percentage of questionnaires actually sent that were completed each week for PHQ-9 and GAD-7. Questionnaire completion was measured from the date of account creation (week 1 in the study) to week 12, a timeframe chosen to concord with the 12-week follow-up time planned for our upcoming clinical trial. For other timeframes, see Multimedia Appendix 1. The number of PHQ-9 questionnaires that were completed in the app by patients was calculated by subtracting those completed by physicians and taking the mean completion rate across all patients in each of the three time intervals. Note that only the PHQ-9 could be completed by physicians; all other questionnaires in the CDSS could be completed by patients only.

Results

Recruitment and Safety Data

A total of 10 physicians were initially recruited; however, of the 10 physicians, 3 (30%) psychiatrists were unable to recruit patients because of COVID-19 related interruptions in regular clinical practice and could not be included (of the 3 psychiatrists, 2 [67%] of the psychiatrists' day programs were closed, and 1 [33%] of the psychiatrist focused on providing consults rather than follow-up appointments during the pandemic). A total of 20 patients were approached by the 10 physicians recruited for the study (Figure 1) [39]. Of these 20 patients, 2 (10%) declined participation after discussing the study with their physician or a research assistant. One patient who was interested in entering the study was not eligible as another physician involved in prescribing their medication was not a study physician and as such would not be able to use the app to follow the patient. The recruiting physician was running a day hospital program, which the patient in question was attending. Of the 20 patients, 17 (85%) patients were recruited into the study. As such, 85% (17/20) of the patients approached were recruited. Of the 17 patients, 14 (82%; Table 2) completed the study (defined as attending baseline, visit 1, and visit 2 appointments). One patient withdrew before the baseline appointment, and 2 withdrew after the baseline appointment but before the CDSS was used at visit 1. Of the 14 patients, the sample of patients completing the study consisted of 9 (64%) women and 5 (36%) men with a mean age of 36.43 years (SD 14.84). See Table 2 for demographics and Table 1 for the baseline questionnaire scores. The pandemic was a reason for significantly reduced recruitment and for the withdrawal of several patients from the study.

Figure 1. Consolidated Standards of Reporting Trials study flow chart of participant recruitment and completion. CDSS: clinical decision support system.

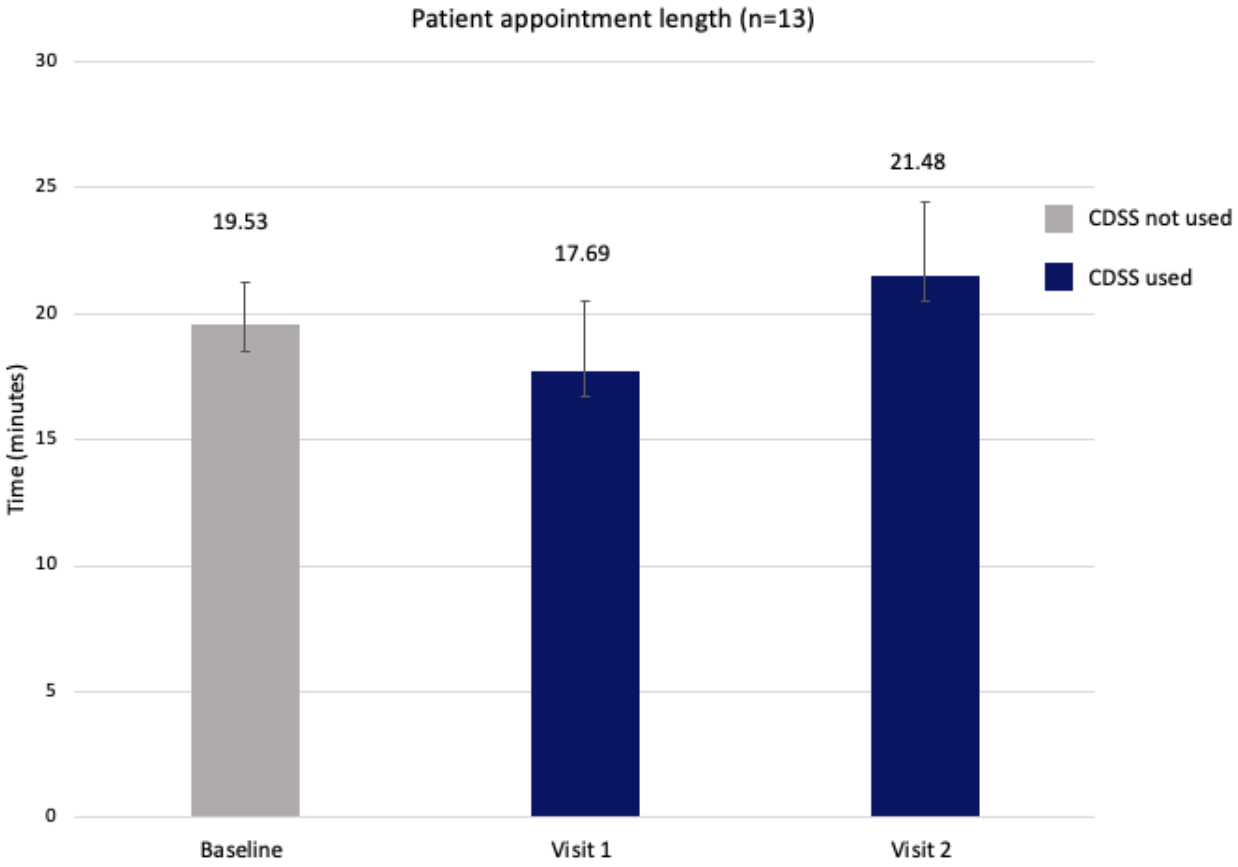


Of the 17 patients, 2 (12%) experienced side effects, as recorded on the Udvalg for Kliniske Undersøgelser Side Effects Rating Scale across the course of the study (see adverse events in Multimedia Appendix 1 for more details). Of the 17 patients, 1 (6%) discontinued treatment. We noted that discontinuation rates of psychiatric treatment could often be >40% [40]. There were no serious adverse events related to the tool; however, of 17 patients, 1 (6%) experienced 2 emergency room visits (a work-related injury and a rash that was thought to be viral by consultants in the emergency room; however, this may have been related to a new antidepressant prescription that was made by a physician without reference to the AI predictions) during the study.

Assessing the Feasibility of the CDSS for Use in Clinical Practice

Patients (n=14) were in the study for an average of 13.2 weeks or 92.4 days (SD 9.74 weeks or SD 68.18 days), excluding 2 patients who dropped out of the study within 1 week of creating their accounts. The mean time between the baseline appointment and visit 1 was 40.86 days (SD 29.40), and the mean time between visit 1 and visit 2 was 51.57 days (SD 62.58). For the 13 patients for whom the appointment length analysis was carried out, baseline visits lasted a mean of 19.53 minutes (SD 6.09). Visit 1 and visit 2 lasted a mean of 17.69 minutes (SD 10.12) and 21.48 minutes (SD 10.69), respectively (Figure 2). Our findings showed no significant difference between the baseline appointment time without the CDSS and subsequent visits using the CDSS ($F_{2,24}$=0.805; mean square error=58.08; P=.46).

Figure 2. Patient appointment times. A total of 14 patients completed the study; however, 1 patient's appointment time could not be counted at baseline because of their physician opening the clinical decision support system erroneously. CDSS: clinical decision support system.



With regard to the subjective physician view of appointment length, of the 7 physicians, 4 (57%) rated that using the tool "took about the same time as my usual practice," indicated by a rating of 3 on a 5-point Likert scale. In addition, of 13 patients, 8 (62%) felt that their appointment time did not change, whereas 1 (8%) patient felt that it decreased.

Assessing Physician and Patient Trust in the CDSS and Its Effect on the Clinician-Patient Relationship

With regard to the tool's trustworthiness, 62% (8/13) of patients and 71% (5/7) of physicians rated that they trusted the CDSS, indicated by a 4 or 5 on a 5-point Likert scale. The mean STAR-P and STAR-C scores were 42.69 (SD 5.57) and 40.29 (SD 5.65), comparable with 38.4 (SD 12.0) and 31.5 (SD 6.9) in the original Scale to Assess Therapeutic Relationships in Community Mental Health Care study [37], respectively, indicating no major negative effects of the CDSS on the clinician-patient relationship (a possible outcome, given we included a new piece of technology that was directly involved in the shared decision-making process). Further information about the Scale to Assess Therapeutic Relationships in Community Mental Health Care subscales is present in Multimedia Appendix 1. In addition, on their custom exit questionnaire, 46% (6/13) of patients felt that the patient-clinician relationship significantly or somewhat improved, whereas 54% (7/13) of patients felt that it did not change.

Assessing the Usability of the CDSS

Good overall usability of the CDSS, indicated by a 4 or 5 on a 5-point Likert scale, was rated by 92% (12/13) of patients and 71% (5/7) of physicians (Multimedia Appendix 1, Table S3).

Assessing Engagement With the App

At each patient's visit 1, all (7/7, 100%) physicians logged into the tool on the same day as the visit, and the clinical algorithm module (which contains the CANMAT guidelines and AI results, when available) portion of the tool was accessed on 93% (13/14) of appointment days. At the subsequent visit 2 appointments, once again, all (7/7, 100%) physicians logged into the tool, whereas the clinical algorithm component was again accessed at 93% (13/14) of appointments.

Figures 3-6 demonstrate the PHQ-9 and GAD-7 completion rates each week during the first 12 weeks of the study. The light bars in Figures 3 and 4 reflect the total number of questionnaires sent, given the number of patients that were active in the study during weeks 1 through 12. The total number of PHQ-9 and GAD-7 questionnaires completed by patients on the app for the first 12 weeks of the study were summed and are shown in the dark bars in Figures 3 and 4. In each of weeks 4, 5, 6, and 10, 1 patient completed their PHQ-9 questionnaire with a physician. For each of these weeks, one response was subtracted from the total number of PHQ-9 questionnaires completed to reflect only those done by patients (Figures 3 and 5).

Figure 3. Frequency of PHQ-9 completion by patients in the first 12 weeks of the study versus the total number sent in the clinical decision support systems (1 per week, per active patient). PHQ-9: Patient Health Questionnaire-9.

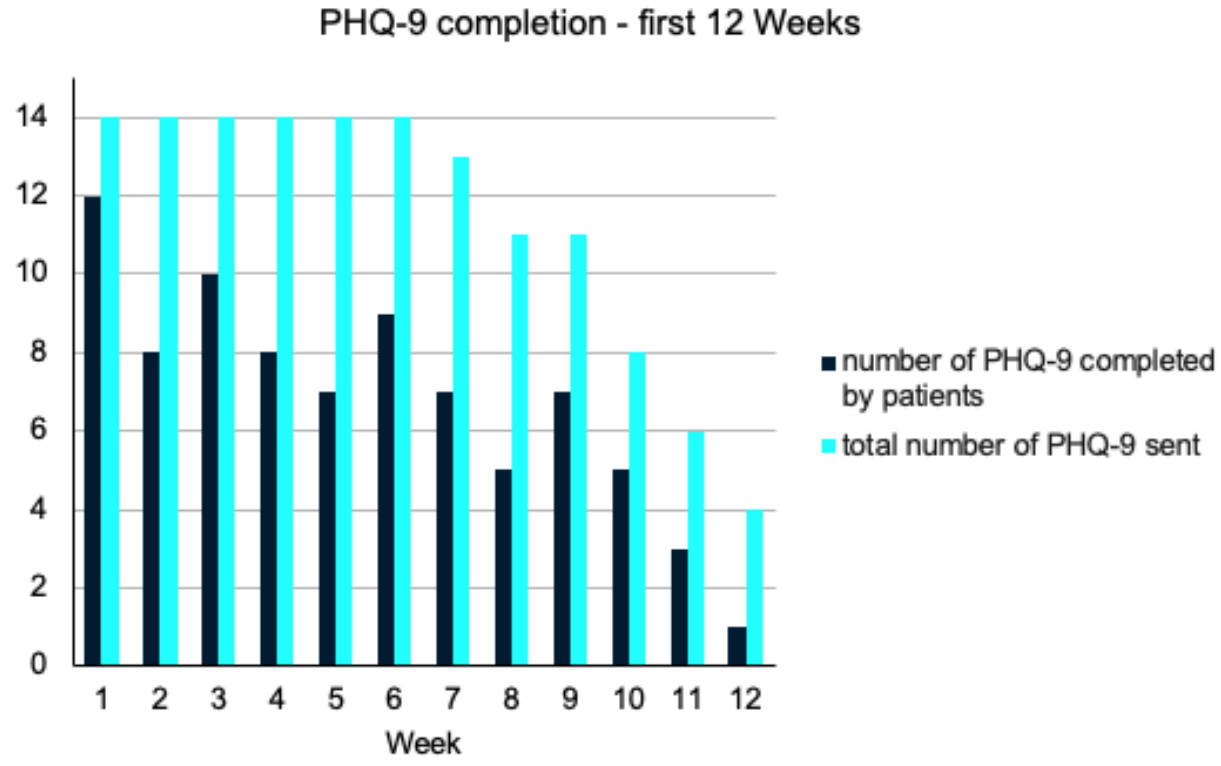


Figure 4. Frequency of GAD-7 questionnaires completed by patients in the first 12 weeks of the study versus the total number sent in the clinical decision support systems (1 per week, per active patient). GAD-7: General Anxiety Disorder-7.

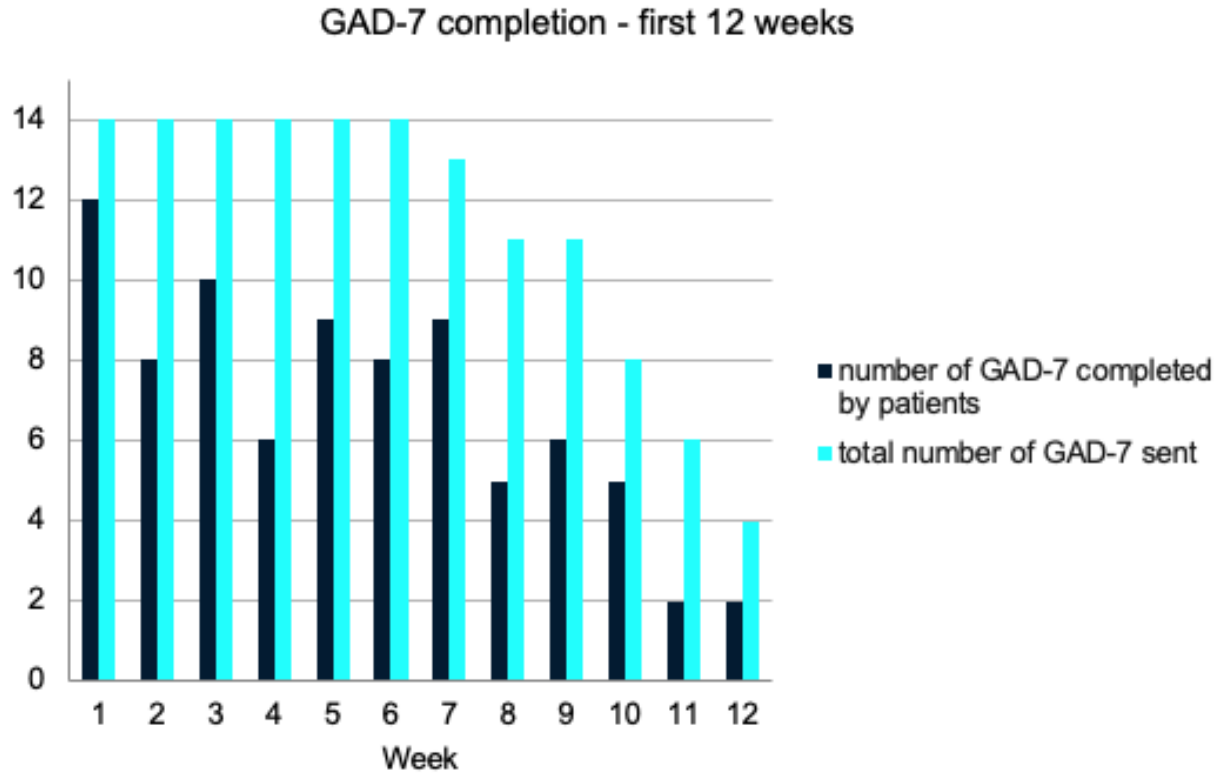


Figure 5. Percent of PHQ-9 completion by patients in the clinical decision support systems during the first 12 weeks of the study. PHQ-9: Patient Health Questionnaire-9.

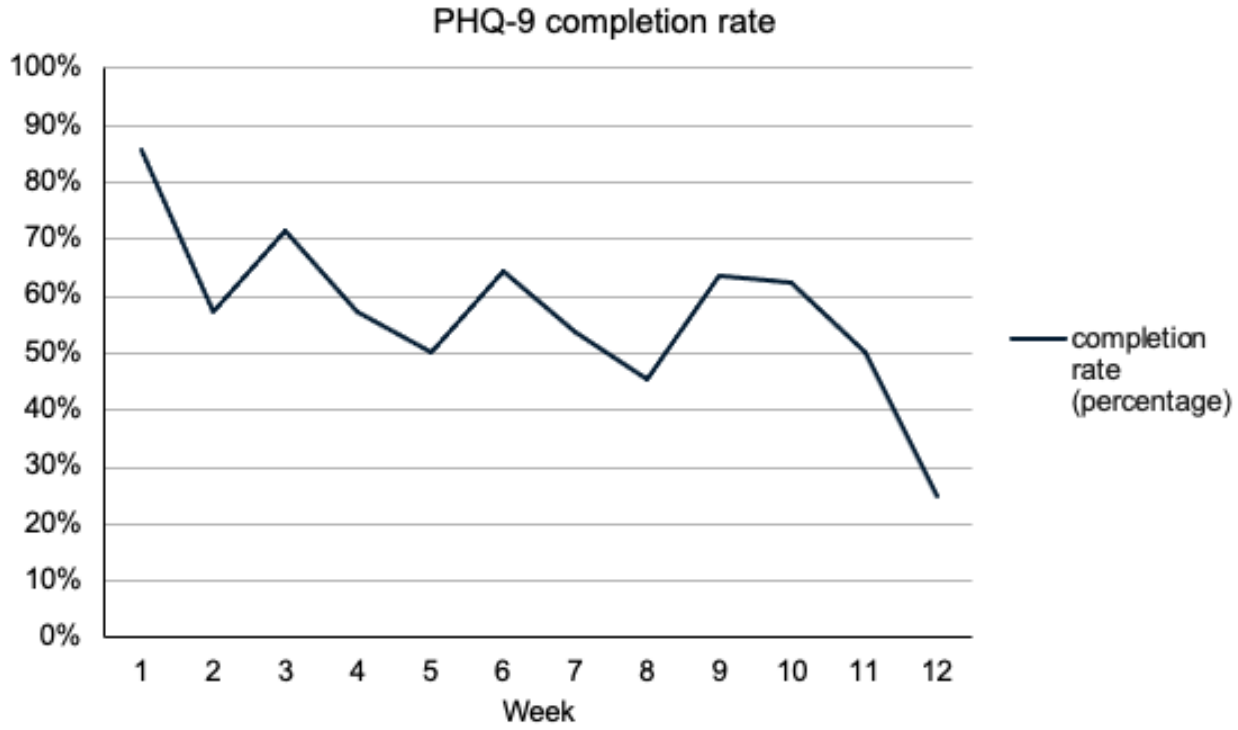
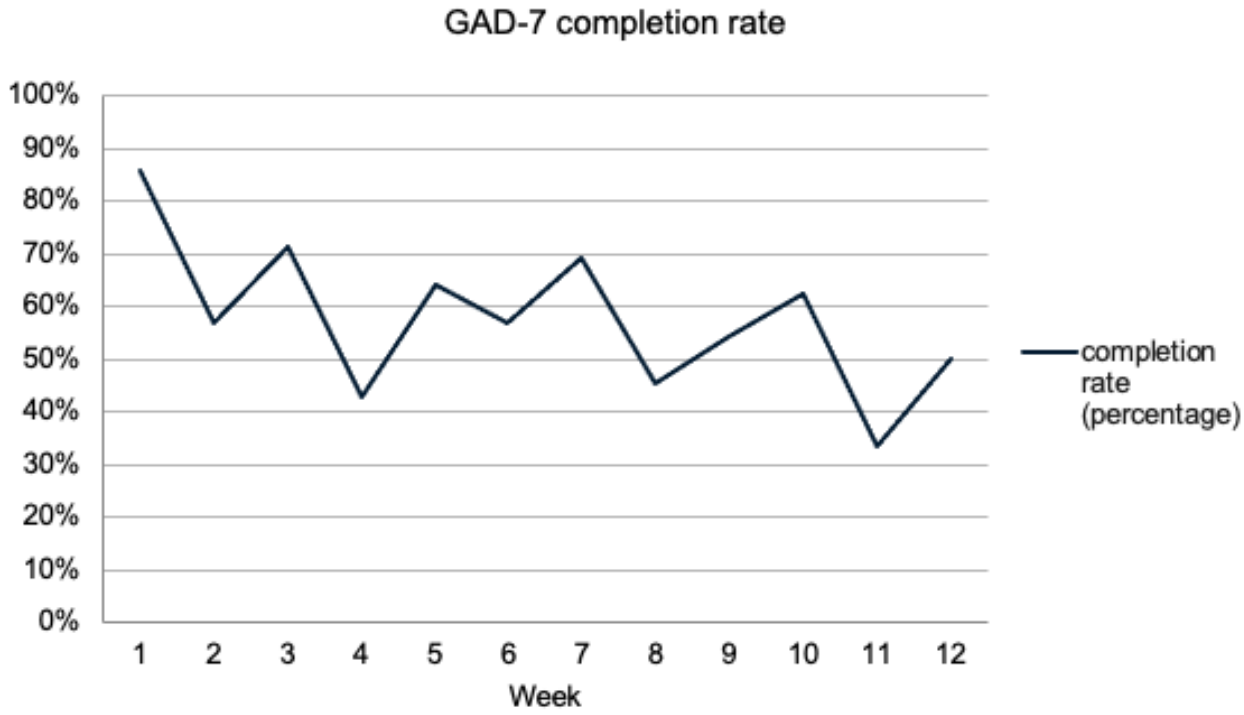


Figure 6. Percent of GAD-7 questionnaires completed by patients in the clinical decision support systems during the first 12 weeks of the study. GAD-7: General Anxiety Disorder-7.



The mean completion rate of all PHQ-9 questionnaires sent from account creation to week 12 of the study was 64% (SD 28%). The mean completion rate of the PHQ-9 by patients alone in this time frame was 59% (SD 28%). The GAD-7 had a mean completion rate of 60% (SD 28%) in the first 12 weeks of the study. Completion rates for other time frames can be found in Multimedia Appendix 1 but were similar.

Of the 14 participants, 10 (71%) had mean PHQ-9 and GAD-7 completion rates ≥50%. The lowest completion rate among the patients for both questionnaires sent in the first 12 weeks was 8%. Most patients regularly completed questionnaires, with 76% (11/14) completing ≥33% of app assessments and 71% (10/14) completing ≥50% of the PHQ-9 and GAD-7 assessments in the app over the first 12 weeks in the study.

Biweekly peaks in the completion rates were observed for the PHQ-9 and GAD-7 questionnaires (Figures 3 and 4). Although they were intended to be completed weekly as a part of the study, these questionnaires are often administered at ≥2-week intervals in practice, indicating the likely feasibility of a reduced questionnaire frequency.

In addition, an exploratory correlation analysis aimed at determining potential correlates of questionnaire completion rates revealed that patients who had appointments scheduled further apart were less likely to complete the PHQ-9 (r_{12}=–0.69; P=.006).

Results and a full discussion of the changes in depression and anxiety questionnaire scores can be found in Multimedia Appendix 1.

High rates of treatment adherence were noted as measured by the Brief Adherence Rating Scale and can be found in Multimedia Appendix 1.

Discussion

Principal Findings

The primary objectives of the study were to assess feasibility, usability, and ongoing engagement with the CDSS when integrated into clinical practice, as well as to measure physician and patient trust in the CDSS and its impact on the clinician-patient relationship. The primary feasibility outcome of interest was appointment length to determine whether the use of the CDSS required more time than the baseline appointment. We were able to confirm that appointment length did not significantly increase after the introduction of the tool. In addition, patients and clinicians provided high trust and usability ratings; the app resulted in improved patient-clinician relationship for some patients, and patients and clinicians both engaged with the app in a manner consistent that supports clinical feasibility. Digital application use for the purpose of promoting mental health is increasingly recommended by public health organizations. For example, in its Mental Health Action Plan 2013-2020, the World Health Organization proposed “the promotion of self-care, for instance, through the use of electronic and mobile health technologies” [41]. Meanwhile, the UK National Health Service’s website endorses a short list of web-based mental health resources, which includes smartphone-based apps [41]. Specifically, studies on app usability in the treatment of depression have demonstrated that telemedicine and internet-based approaches are feasible and as effective as in-person treatment [42].

However, these tools also face substantial barriers with regard to adherence. For example, a randomized clinical field trial conducted by Arean et al [42] compared 3 different mobile apps for depression to examine how individuals who download these tools typically use them. The authors’ findings show that most participants did not use their assigned intervention apps as instructed and experienced a significant drop off in use after 2 weeks [42]. In addition, a study that investigated the feasibility of using a smartphone app to assess schizophrenia prompted patients via SMS text messages to complete personalized questionnaires once per week. They found that participants (n=18) completed 65% of app assessments, “with 78% completing ≥33% app assessments and 72% completing ≥50% app assessments” [43], similar to the results observed in this study. In summary, response rates observed in our study (with 10/14, 71% of patients completing at least 50% of assessments) were reasonable in the context of previous reports, and engagement persisted fairly stably beyond 2 weeks, demonstrating that the app was able to retain patients at least as much or more consistently than applications in previous studies. In addition, physician engagement was high, with physicians opening the tool at each visit.

Physician engagement with mental health apps is related to their technological competency, their perception of patient access to technology, and organizational infrastructure that facilitates the adoption of the apps into their practice [44], which are all factors considered when designing the study; for example, physicians who were less technically oriented could rely on study staff to provide ongoing support for app use as needed. The results of our study demonstrate sustained patient and physician engagement beyond 2 weeks, potentially because the app was directly tied to clinical care and because high physician use of the app and the data patients inputted may have motivated patients to continue engaging. Indeed, higher rates of engagement are linked to the use of telephone and email reminders, as well as follow-up with a physician [45], a finding supported by our demonstration of lower PHQ-9 response rates as a result of longer interappointment lengths. In addition, the email reminders sent to patients to complete assessments likely had a positive impact on completion rates based on these previous findings.

More than half (4/7, 57%) of the physicians felt that using the tool in session took approximately the same time as their usual practice. A systematic literature review conducted by Kerst et al [46] found that 70.2% of physicians treating depression had never used applications in clinical practice before, suggesting that the integration of mental health tools remains quite novel. Therefore, it is possible that some physicians reported that their appointments felt longer than their usual practice simply because they were not yet familiar with the tool. Interestingly, most patients did not subjectively report that the tool increased their appointment time. Objectively, appointment length did not significantly increase when the tool was introduced, lending credence to the idea that the novelty of the tool use may have influenced the perception of time spent by physicians.

We found that 62% (8/13) of patients endorsed some degree of trust in the CDSS, somewhat lower than the percentage of clinicians (5/7, 71%) with some degree of trust. This may have in part been because of COVID-19: most clinicians followed up with their patients by phone, which meant that patients did not get to view the AI results on the physician’s screen as intended, which may have improved feelings of trust had it occurred more frequently; standardized patients noted that looking at the screen with their physicians was a positive experience in our previous simulation center study [10]. Nonetheless, patients’ mean score was 3.85 on a 5-point Likert scale, which indicates that patient trust trends in a positive direction.

Limitations

The main limitation of the study is the small sample size, which limits the strength of the conclusions that can be drawn from this study. All results should be considered preliminary and are only intended to demonstrate feasibility; similar metrics of feasibility and ease of use will be used during our upcoming clinical trial to confirm and expand upon these findings. Another weakness is the heterogeneity in the severity of the patients' depression, which limits our ability to comment on tool effectiveness and generalize these results to different strata of illness severity. Nevertheless, it also presents as a strength because it allowed us to demonstrate feasibility in a range of clinical situations.

Our study fell short of recruitment targets, which may have been in part because of several scenarios. One possibility is that there were sufficient patients available to recruit but that these patients were not interested in participating in the study. However, this is not supported by the high proportion of approached patients who agreed to participate. Furthermore, of the patients who did participate, dropout rates were comparable with other studies of mental health applications, as reviewed in the introduction. Another more likely scenario is that there were fewer patients available to recruit than expected; this seems to have been the case because of the COVID-19 pandemic. As discussed, 3 of our physicians were unable to recruit patients because their practices were significantly changed by the pandemic, and they either did not follow patients longitudinally or had very few new patients who were eligible to participate. In addition, physicians who successfully recruited patients frequently commented on the lack of patients presenting with MDD (although this was not formally recorded). As such, the low sample size in this study is not indicative of patient interest in participation but rather of the number of patients available to be recruited during the pandemic. We noted that in unpublished data from a quality improvement project, which included a control group using a version of the CDSS without the AI enabled (ie, providing measurement-based care and algorithm-guided treatment but no AI predictions), which was initiated later in the pandemic (once clinical practices had returned to some stability), recruitment speed was much faster, with 34 patients recruited in 7 months, indicating that recruitment should not be a barrier for our upcoming clinical trial.

However, the low sample size does mean that our study was underpowered to detect differences in appointment length within groups over time when considering an a priori sample size calculation for repeated measures as described above. Despite this, there are 2 reasons why this study provides useful information. First, the appointment lengths are interpretable data. Should appointment lengths have been different in the order of ≥5 minutes, which in previous work [10] has been identified by clinicians as the amount of time they would be willing to spend on the CDSS, and had this difference not been significant simply because of low power, then we would have to remain concerned that the tool might increase appointment lengths. However, the recorded appointment lengths differed by <5 minutes; even if this difference had been statistically significant, it would not have been clinically significant. Indeed, in the sample size calculation discussed earlier, where the objective was to detect an appointment length of at least five minutes more than a usual 20-minute length appointment, the required sample size was only 8, meaning that our study should have been sufficiently powered to detect this difference from a standard appointment length (which the baseline sessions matched closely). In addition, in the unpublished quality improvement project mentioned above, appointment lengths in the active group, which included 22 patients who used the tool, were also roughly 20 minutes long. Finally, as noted, most patients and physicians did not note a subjective increase in the appointment length. As such, although the study may be underpowered, there does not seem to be a signal in the available data to suggest a clinically significant increase in appointment length, which would reduce feasibility.

An additional and significant limitation of this study is that although its design allowed us to examine the impact of introducing the tool on the patient-clinician relationship and the clinician workflow, this at the same time prevented us from examining the effectiveness of the device in terms of improvement in depression scores. This is because the tool being introduced well into a patient's treatment course could not have its intended effect of assisting treatment selection or helping clinicians implement measurement-based care and algorithm-guided treatment across the entire length of the study. This was compounded by the delays between appointments and the reduced number of visits as a result of COVID-19. However, we note that the decision was made during study design not to focus on effectiveness and, because of the novelty of the device and the need to determine challenges to its introduction into clinical practice, to focus squarely on feasibility. As such, the modest improvements in depression and anxiety scores seen here are in line with expectations, given that the tool was not introduced in a manner where it could have its intended effect on patient care.

With regard to feasibility, ease of use, and the ability to correct any major limitations, it would be reasonable to be concerned about whether this study, with its small sample size, could on its own speak to the readiness of the tool for clinical trials. However, this study should be considered in the context of previously published evidence on the same tool, which demonstrated its ease of use in a simulation center environment [10]. In addition, several pilot projects and quality improvement projects have been undertaken with the non-AI version of this device. These pilots, although not undertaken as research projects, allowed for the testing of the user interface and the incorporation of user feedback and were conducted alongside intensive quality assurance in the development process. The purpose of this study was to test this tool with the AI enabled in the context of longitudinal follow-up. The present results in this context, combined with the adjustments and corrections that have been made through quality assurance, physician feedback in this study, and other pilot projects, allow us to conclude that the tool is ready for clinical trials.

Conclusions

In this paper, we have demonstrated that the Aifred CDSS is feasible and easy for clinicians and patients to use in a

longitudinal manner and that it does not require increased time to use in clinic. In addition, the tool had an interesting impact on the clinician-patient relationship. For roughly half of the patients, it did not negatively or positively affect the relationship, helping to allay concerns about technological solutions worsening relationships between clinicians and patients. For the other half of the patients, the relationship was actually rated as having improved, indicating that for some patients, the CDSS may have beneficial effects on the clinician-patient relationship. This latter point will be further elaborated in a future study and should be investigated in future work. Planned clinical trials will serve as an opportunity to confirm these feasibility results and to determine if the CDSS is effective in improving depression outcomes.

Acknowledgments

The authors thank the patients, participating doctors, and local staff for their involvement in this feasibility study. Funding was obtained between initial institutional review board approval (July 29, 2019) and the start of the study but was disclosed to research ethics boards as amendments were approved before the start of the study. This research was funded by the Aifred Health Inc; Innovation Research Assistance Program, National Research Council, Canada; ERA-Permed Vision 2020 supporting IMADAPT; Government of Québec Nova Science; and the MEDTEQ COVID-Relief Grant. Inquiries regarding access to study data should be directed to the corresponding author.

Authors' Contributions

CP, GG, and DB contributed equally to this study. DB conceptualized the study, contributed to the protocol, participated in the creation of the AI model used, helped coordinate the study, participated in data analysis, and contributed to manuscript writing and review. CP helped coordinate the study, participated in data collection and analysis, and contributed to manuscript writing and review. MTS helped coordinate the study, participated in data collection, and contributed to manuscript writing and review. DK participated in the data analysis and in building the digital platform used in the study. DS, EL, JW, BD, TP, CR, JB, KFH, and GS participated in data collection and analysis and contributed to manuscript writing and review. SI and KP contributed to the study conceptualization and manuscript review. CA, RF, and JM participated in the creation of the AI model used for this study and in the manuscript review. GG, KW, and MJF participated in the data analysis, manuscript writing, and review. KL, WS, SR, JFK, KH, SVP, and MF provided guidance on the study implementation, manuscript development, and manuscript review. HM and GT provided supervision, helped conceptualize the study, and participated in the manuscript review.

Conflicts of Interest

DB, CA, RF, SI, KP, and MC are shareholders and employees, directors, or founders of Aifred Health. DK and JM are employed by Aifred Health. MTS is employed by Aifred Health and is an options holder. GG, CP, KFH, GS, EL, JB, JW, TP, DS, BD, KW, and CR have been or are employed or financially compensated by Aifred Health. SVP, KH, and JFK are members of Aifred Health’s scientific advisory board and have received payments or options. WS, KL, and SR are members of the data safety monitoring board. HM has received honoraria, sponsorship, or grants for participation in speaker bureaus, consultation, advisory board meetings, and clinical research from Acadia, Amgen, HLS Therapeutics, Janssen-Ortho, Mylan, Otsuka-Lundbeck, Perdue, Pfizer, Shire, and SyneuRx International. After the study but during manuscript preparation, SR became a shareholder of Aifred Health; this was disclosed to their site and to other members of the DSMB; SR reports owning shares in Aifred Health.

Multimedia Appendix 1
Peer-reviewed supplementary material.
[DOCX File , 46 KB - formative_v5i10e31862_app1.docx]

References

1. Sim I, Gorman P, Greenes RA, Haynes RB, Kaplan B, Lehmann H, et al. Clinical decision support systems for the practice of evidence-based medicine. J Am Med Inform Assoc 2001;8(6):527-534. [doi: 10.1136/jamia.2001.0080527] [Medline: 11687560]
2. Zikos D, DeLellis N. CDSS-RM: a clinical decision support system reference model. BMC Med Res Methodol 2018 Nov 16;18(1):137 [FREE Full text] [doi: 10.1186/s12874-018-0587-6] [Medline: 30445910]
3. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020 Feb 06;3:17 [FREE Full text] [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
4. Mehltretter J, Rollins C, Benrimoh D, Fratila R, Perlman K, Israel S, et al. Analysis of features selected by a deep learning model for differential treatment selection in depression. Front Artif Intell 2020 Jan 21;2:31 [FREE Full text] [doi: 10.3389/frai.2019.00031] [Medline: 33733120]
5. Squarcina L, Villa FM, Nobile M, Grisan E, Brambilla P. Deep learning for the prediction of treatment response in depression. J Affect Disord 2021 Feb 15;281:618-622. [doi: 10.1016/j.jad.2020.11.104] [Medline: 33248809]

6. Torous J, Lipschitz J, Ng M, Firth J. Dropout rates in clinical trials of smartphone apps for depressive symptoms: a systematic review and meta-analysis. J Affect Disord 2020 Feb 15;263:413-419. [doi: 10.1016/j.jad.2019.11.167] [Medline: 31969272]
7. Rollman BL, Hanusa BH, Lowe HJ, Gilbert T, Kapoor WN, Schulberg HC. A randomized trial using computerized decision support to improve treatment of major depression in primary care. J Gen Intern Med 2002 Jul;17(7):493-503 [FREE Full text] [doi: 10.1046/j.1525-1497.2002.10421.x] [Medline: 12133139]
8. Harrison P, Carr E, Goldsmith K, Young AH, Ashworth M, Fennema D, et al. Study protocol for the antidepressant advisor (ADeSS): a decision support system for antidepressant treatment for depression in UK primary care: a feasibility study. BMJ Open 2020 May 24;10(5):e035905 [FREE Full text] [doi: 10.1136/bmjopen-2019-035905] [Medline: 32448796]
9. Adli M, Wiethoff K, Baghai TC, Fisher R, Seemüller F, Laakmann G, et al. How effective is algorithm-guided treatment for depressed inpatients? Results from the randomized controlled multicenter german algorithm project 3 trial. Int J Neuropsychopharmacol 2017 Sep 01;20(9):721-730 [FREE Full text] [doi: 10.1093/ijnp/pyx043] [Medline: 28645191]
10. Benrimoh D, Tanguay-Sela M, Perlman K, Israel S, Mehltretter J, Armstrong C, et al. Using a simulation centre to evaluate preliminary acceptability and impact of an artificial intelligence-powered clinical decision support system for depression treatment on the physician-patient interaction. BJPsych Open 2021 Jan 06;7(1):e22 [FREE Full text] [doi: 10.1192/bjo.2020.127] [Medline: 33403948]
11. Warden D, Rush AJ, Trivedi MH, Fava M, Wisniewski SR. The STAR*D project results: a comprehensive review of findings. Curr Psychiatry Rep 2007 Dec 07;9(6):449-459. [doi: 10.1007/s11920-007-0061-3] [Medline: 18221624]
12. Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, 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. Focus (Am Psychiatr Publ) 2018 Oct;16(4):420-429 [FREE Full text] [doi: 10.1176/appi.focus.16407] [Medline: 32021580]
13. Benrimoh D, Fratila R, Israel S, Perlman K, Mirchi N, Desai S, et al. Aifred health, a deep learning powered clinical decision support system for mental health. In: The NIPS '17 Competition: Building Intelligent Systems. Cham, Switzerland: Springer; Sep 28, 2018:251-287.
14. Mehltretter J, Fratila R, Benrimoh D, Kapelner A, Perlman K, Snook E, et al. Differential treatment benefit prediction for treatment selection in depression: a deep learning analysis of STAR* D and CO-MED data. Comput Psychiatr 2020 Dec:61-75. [doi: 10.1101/679779]
15. Kennedy SH, Lam RW, McIntyre RS, Tourjman SV, Bhat V, Blier P, CANMAT Depression Work Group. Canadian network for mood and anxiety treatments (CANMAT) 2016 clinical guidelines for the management of adults with major depressive disorder: section 3. Pharmacological treatments. Can J Psychiatry 2016;61(9):540-560 [FREE Full text] [doi: 10.1177/0706743716659417] [Medline: 27486148]
16. Torous J, Friedman R, Keshavan M. Smartphone ownership and interest in mobile applications to monitor symptoms of mental health conditions. JMIR Mhealth Uhealth 2014 Jul 07;2(1):e2 [FREE Full text] [doi: 10.2196/mhealth.2994] [Medline: 25098314]
17. Ventola CL. Mobile devices and apps for health care professionals: uses and benefits. P T 2014 May;39(5):356-364 [FREE Full text] [Medline: 24883008]
18. Tri-council policy statement: ethical conduct for research involving humans. Medical Research Council of Canada. 2005. URL: http://www.bcmhsus.ca/Documents/tri-council-policy-statement-ethical-conduct-for-research-involving-humans.pdf [accessed 2021-09-06]
19. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5). United States: American Psychiatric Pub; 2013.
20. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
21. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
22. Bohn MJ, Babor TF, Kranzler HR. The alcohol use disorders identification test (AUDIT): validation of a screening instrument for use in medical settings. J Stud Alcohol 1995 Jul;56(4):423-432. [doi: 10.15288/jsa.1995.56.423] [Medline: 7674678]
23. Skinner HA. The drug abuse screening test. Addict Behav 1982;7(4):363-371. [doi: 10.1016/0306-4603(82)90005-3] [Medline: 7183189]
24. Merlhiot G, Mondillon L, Vermeulen N, Basu A, Mermillod M. Adaptation and validation of the standardized assessment of personality – abbreviated scale as a self-administered screening test (SA-SAPAS). J Psychol Psychother 2014;4(6):1000164. [doi: 10.4172/2161-0487.1000164]
25. Wisniewski SR, Rush AJ, Balasubramani GK, Trivedi MH, Nierenberg AA, STAR*D Investigators. Self-rated global measure of the frequency, intensity, and burden of side effects. J Psychiatr Pract 2006 Mar;12(2):71-79. [doi: 10.1097/00131746-200603000-00002] [Medline: 16728903]
26. Yeung A, Feldman G, Pedrelli P, Hails K, Fava M, Reyes T, et al. The quick inventory of depressive symptomatology, clinician rated and self-report: a psychometric assessment in chinese americans with major depressive disorder. J Nerv Ment Dis 2012 Aug;200(8):712-715 [FREE Full text] [doi: 10.1097/NMD.0b013e318261413d] [Medline: 22850307]

27. Rush AJ, Carmody T, Reimitz P. The inventory of depressive symptomatology (IDS): clinician (IDS-C) and self-report (IDS-SR) ratings of depressive symptoms. Int J Method Psychiat Res 2006 Jun;9(2):45-59. [doi: 10.1002/mpr.79]
28. U.S. public sees multiple threats from the coronavirus – and concerns are growing. Pew Research Center. 2020. URL: https://www.pewresearch.org/politics/2020/03/18/u-s-public-sees-multiple-threats-from-the-coronavirus-and-concerns-are-growing/ [accessed 2021-10-01]
29. Williams N. The GAD-7 questionnaire. Occup Med 2014 Apr 03;64(3):224. [doi: 10.1093/occmed/kqt161]
30. Kroenke K. Enhancing the clinical utility of depression screening. CMAJ 2012 Feb 21;184(3):281-282 [FREE Full text] [doi: 10.1503/cmaj.112004] [Medline: 22231681]
31. 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 Mar 15;265:395-401. [doi: 10.1016/j.jad.2020.01.032] [Medline: 32090765]
32. Fortney JC, Unützer J, Wrenn G, Pyne JM, Smith GR, Schoenbaum M, et al. A tipping point for measurement-based care. Focus (Am Psychiatr Publ) 2018 Jul;16(3):341-350 [FREE Full text] [doi: 10.1176/appi.focus.16303] [Medline: 32015714]
33. Felitti VJ, Anda RF, Nordenberg D, Williamson DF, Spitz AM, Edwards V, et al. Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults: the adverse childhood experiences (ACE) study. Am J Prev Med 1998 May;14(4):245-258. [doi: 10.1016/s0749-3797(98)00017-8] [Medline: 9635069]
34. Weathers FW, Blake DD, Schnurr PP, Kaloupek DG, Marx BP, Keane TM. The life events checklist for DSM-5 (LEC-5). National Center for PTSD. URL: https://www.ptsd.va.gov/professional/assessment/documents/LEC5_Standard_Self-report.PDF [accessed 2021-09-06]
35. Lingjaerde O, Ahlfors UG, Bech P, Dencker SJ, Elgen K. The UKU side effect rating scale: a new comprehensive rating scale for psychotropic drugs and a cross-sectional study of side effects in neuroleptic-treated patients. Acta Psychiatr Scand Suppl 1987;334:1-100. [doi: 10.1111/j.1600-0447.1987.tb10566.x] [Medline: 2887090]
36. 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 Mar;100(1-3):60-69. [doi: 10.1016/j.schres.2007.12.470] [Medline: 18255269]
37. McGuire-Snieckus R, McCabe R, Catty J, Hansson L, Priebe S. A new scale to assess the therapeutic relationship in community mental health care: STAR. Psychol Med 2007 Jan;37(1):85-95. [doi: 10.1017/S0033291706009299] [Medline: 17094819]
38. Faul F, Erdfelder E, Lang A, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods 2007 May;39(2):175-191. [doi: 10.3758/bf03193146] [Medline: 17695343]
39. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
40. Maund E, Stuart B, Moore M, Dowrick C, Geraghty AW, Dawson S, et al. Managing antidepressant discontinuation: a systematic review. Ann Fam Med 2019 Jan;17(1):52-60 [FREE Full text] [doi: 10.1370/afm.2336] [Medline: 30670397]
41. Anthes E. Mental health: there's an app for that. Nature 2016 Apr 07;532(7597):20-23. [doi: 10.1038/532020a] [Medline: 27078548]
42. Arean PA, Hallgren KA, Jordan JT, Gazzaley A, Atkins DC, Heagerty PJ, et al. The use and effectiveness of mobile apps for depression: results from a fully remote clinical trial. J Med Internet Res 2016 Dec 20;18(12):e330 [FREE Full text] [doi: 10.2196/jmir.6482] [Medline: 27998876]
43. Eisner E, Bucci S, Berry N, Emsley R, Barrowclough C, Drake RJ. Feasibility of using a smartphone app to assess early signs, basic symptoms and psychotic symptoms over six months: a preliminary report. Schizophr Res 2019 Jun;208:105-113 [FREE Full text] [doi: 10.1016/j.schres.2019.04.003] [Medline: 30979665]
44. Chiauzzi E, Newell A. Mental health apps in psychiatric treatment: a patient perspective on real world technology usage. JMIR Ment Health 2019 Apr 22;6(4):e12292 [FREE Full text] [doi: 10.2196/12292] [Medline: 31008711]
45. Christensen H, Griffiths KM, Farrer L. Adherence in internet interventions for anxiety and depression. J Med Internet Res 2009 Apr 24;11(2):e13 [FREE Full text] [doi: 10.2196/jmir.1194] [Medline: 19403466]
46. Kerst A, Zielasek J, Gaebel W. Smartphone applications for depression: a systematic literature review and a survey of health care professionals' attitudes towards their use in clinical practice. Eur Arch Psychiatry Clin Neurosci 2020 Mar;270(2):139-152. [doi: 10.1007/s00406-018-0974-3] [Medline: 30607530]

Abbreviations

AI: artificial intelligence
ANOVA: analysis of variance
CANMAT: Canadian Network for Mood and Anxiety Treatments
CDSS: clinical decision support system
***DSM-5*:** *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition*
GAD-7: General Anxiety Disorder-7

MDD: major depressive disorder
PHQ-9: Patient Health Questionnaire-9
STAR-C: Scale to Assess Therapeutic Relationships in Community Mental Health Care-Clinician
STAR-P: Scale to Assess Therapeutic Relationships in Community Mental Health Care-Patient

Edited by J Torous; submitted 07.07.21; peer-reviewed by A Thunström, R Castilla-Puentes; comments to author 26.07.21; revised version received 23.08.21; accepted 23.08.21; published 25.10.21.

Please cite as:
Popescu C, Golden G, Benrimoh D, Tanguay-Sela M, Slowey D, Lundrigan E, Williams J, Desormeau B, Kardani D, Perez T, Rollins C, Israel S, Perlman K, Armstrong C, Baxter J, Whitmore K, Fradette MJ, Felcarek-Hope K, Soufi G, Fratila R, Mehltretter J, Looper K, Steiner W, Rej S, Karp JF, Heller K, Parikh SV, McGuire-Snieckus R, Ferrari M, Margolese H, Turecki G
Evaluating the Clinical Feasibility of an Artificial Intelligence–Powered, Web-Based Clinical Decision Support System for the Treatment of Depression in Adults: Longitudinal Feasibility Study
JMIR Form Res 2021;5(10):e31862
URL: https://formative.jmir.org/2021/10/e31862
doi:10.2196/31862
PMID:34694234

Original Paper

Usability of a Technology-Based Bystander Bullying Intervention for Middle School Students in Rural, Low-Income Communities: Mixed Methods Study

Diana M Doumas[1,2], DPhil; Aida Midgett[1,2], EdD; Valerie Myers[3], DPhil; Mary Klein Buller[3], MA

[1]Institute for the Study of Behavioral Health and Addiction, Boise State Universtiy, Boise, ID, United States

[2]Department of Counselor Education, Boise State University, Boise, ID, United States

[3]Klein Buendel, Golden, CO, United States

Corresponding Author:
Diana M Doumas, DPhil
Institute for the Study of Behavioral Health and Addiction
Boise State Universtiy
1910 University Drive
Boise, ID, 83702
United States
Phone: 1 2084261219
Email: dianadoumas@boisestate.edu

Abstract

Background: Students who are targets of bullying and who witness bullying are at high risk for negative mental health outcomes. Bystander training is essential to reduce bullying and the negative associated consequences for targets and bystanders. Resources necessary for program delivery, however, pose significant barriers for schools, particularly those in rural, low-income communities. Technology-based programs can reduce health disparities for students in these communities through cost-effective, easy-to-disseminate programming.

Objective: The aim of this study is to conduct usability testing of a bystander bullying web app prototype, STAC-T (technology-based STAC, which stands for the 4 bystander strategies Stealing the Show, Turning it Over, Accompanying Others, and Coaching Compassion) as an initial step in the development of a full-scale STAC-T intervention. Objectives include assessing usability and acceptability of the STAC-T prototype, understanding school needs and barriers to program implementation, and assessing differences in usability between school personnel and students.

Methods: A sample of 16 participants, including school personnel and students recruited from 3 middle schools in rural, low-income communities, completed usability testing followed by a qualitative interview. Descriptive statistics, 2-tailed independent sample *t* tests, and consensual qualitative research were used to assess usability and program satisfaction and to extract themes related to acceptability, feasibility, needs, barriers, and feedback for intervention development.

Results: Usability testing indicated that the app was easy to use, acceptable, and feasible. Both school personnel (mean rating 89.6, SD 5.1) and students (mean rating 91.8, SD 7.0) rated the app well above the standard cutoff score for above-average usability (ie, 68), and both school personnel (mean rating 5.83, SD 0.41) and students (mean rating 6.10, SD 0.57) gave the app high user-friendliness ratings (0-7 scale, with 7 as high user-friendliness). The overall ratings also suggested that school personnel and students were satisfied with the program. Of the 6 school personnel who said they would recommend the program, 1 (17%), 4 (66%), and 1 (17%) rated the program as 3, 4, and 5 stars, respectively; 80% (8/10) of students said they would recommend the program; and 60% (6/10) and 40% (4/10) rated the program as 4 stars and 5 stars, respectively. Qualitative data revealed that school personnel and students found the STAC-T app to be useful, user-friendly, and relevant, while providing feedback related to the importance of digital learning activities that engage the user. Data from school personnel also indicated positive perceptions regarding program feasibility and probability of program adoption, with the most significant barrier being cost, suggesting the importance of considering the financial resources available to schools in rural, low-income communities when setting the price point for the full-scale STAC-T intervention.

Conclusions: This study provides support for the full-scale development of the STAC-T app and provides key information for revision to enhance used engagement.

Trial Registration: ClinicalTrials.gov NCT04681495; https://clinicaltrials.gov/ct2/show/NCT04681495

(JMIR Form Res 2021;5(10):e32382) doi:10.2196/32382

KEYWORDS

technology-based bullying intervention; STAC-T; usability testing; middle school; rural; low-income; mobile phone

Introduction

Background

Bullying is a national public health issue in the United States, with 20.2% of students aged 12 to 18 years reporting being bullied at school and 15.3% reporting being cyberbullied in the previous year [1]. Both bullying and cyberbullying peak in middle school, with 28% of students reporting being a target of school bullying, and 33% reporting being cyberbullied [2]. Among middle school students, bullying victimization is associated with a variety of mental health problems, including somatic symptoms [3-7], anxiety [6-8], social anxiety [9-11], depression [5-9], suicidal ideation, and suicide attempts [8]. Similarly, being a target of cyberbullying is associated with internalizing symptoms and suicidal ideation [12-14]. Thus, the development of effective interventions for middle school students is important for reducing bullying and its negative consequences.

Youth in Rural and Low-Income Communities

Youth in rural [15-17] and low-income [18-20] communities are particularly vulnerable to school bullying and cyberbullying. US national data indicate a higher prevalence of school bullying victimization among students in rural areas (23.8%) compared with students in urban areas (19.9%) [1]. Furthermore, students in rural areas are 3% to 5% more likely to report bullying their peers [17]. Students at the lowest income levels report the highest rates of bullying (21%-26.6%) compared with students at higher income levels (16.6%-19.8%) [1]. Students from low-income households also report the highest rates of physical bullying, bullying-related injury, and cyberbullying, yet they are the least likely to report bullying to an adult [1]. In addition, low-income students report the highest rates of consequences associated with bullying victimization, including negative effects on school work, relationships, feelings about oneself, and physical health [1]. Among middle school students attending schools in rural, low-income communities, bullying victimization is associated with poor school relationships, negative school experiences [19], and depression and anxiety [19,21]. These data reveal significant mental health disparities for youth in rural and low-income communities.

Bullying Bystanders

Negative consequences are not only limited to targets of bullying but also extend to students who witness bullying [7,22] and cyberbullying [23-26] as bystanders. Students who witness school bullying are at increased risk for mental health problems, including somatic symptoms [7], sadness [27], helplessness [7,27], isolation, guilt [28], depression, anxiety [7,22], and suicidal ideation [7]. Similarly, witnessing cyberbullying is associated with anxiety, depression [23,25,26], and somatic symptoms [25], even when controlling for the effects of witnessing school bullying [23,25]. In addition, bystanders who intervene in bullying situations experience higher rates of anxiety and depressive symptoms than students who remain passive [29,30], possibly owing to using maladaptive behaviors to defend targets [30]. Research indicates 80% of students report witnessing bullying [7] and more than 50% witnessing cyberbullying [31]. Therefore, developing interventions to teach student bystanders how to appropriately intervene is important, as most students are bullying bystanders.

School-Based Bullying Interventions

Results from a recent meta-analysis indicate that training student bystanders to intervene as *defenders* is an important component of comprehensive school-based bullying programs [32]. Although up to 80% of students report witnessing bullying [7], only 20% to 30% intervene [33]; for cyberbullying, as few as 10% may intervene [34]. Researchers have demonstrated that when bystanders are trained to act, it not only reduces school bullying but also leads to improved mental health in bystanders [35-39]. Bystander disapproval of cyberbullying acts can also effectively limit its prevalence [40]. Furthermore, self-efficacy for defending is positively related to intervening in both bullying [41] and cyberbullying [42]; however, few comprehensive, school-based programs incorporate bystander training.

In addition, there are several barriers to implementing school-based bullying prevention. Available interventions require substantial resources, including demands on teachers, limited access to training, lack of funding, and few school mental health professionals [43], reducing access and posing significant barriers to implementation. Rural, low-income communities face economic disparities, creating further obstacles to implementing these programs [44]. These barriers include a lower tax base for funding programs, training costs inflated by transportation needs related to bringing in expert trainers, frequent staff turnover with limited resources to reestablish expertise, school closures, staff overload and burnout, and lack of program advocates and local expertise in bullying prevention [16]. Therefore, brief programs that focus on bystander training and reduce barriers for implementation are needed to reduce bullying and its negative consequences among middle school students in rural and low-income communities.

Technology-Based Interventions

National data in the United States indicate that the number 1 barrier to implementing educational technology is that school districts do not have the funding [45]. Technology-based interventions may be one way to increase access and reduce implementation challenges specific to rural and low-income communities [16]. Computer access and internet connectivity in middle school is high; 86% of students have computers in school [46], and only 6% of schools do not meet federal connectivity marks for broadband capacity [47]. Notably, low-income students have similar access to school computers compared with middle to high income peers [46]. Some rural areas have higher rates of poor internet connectivity; however, to assist in reducing this disparity, eligible rural and low-income

schools receive discounts ranging from 20% to 90% for telecommunications services, including internet access and broadband services through the Telecommunications Act of 1996 [48]. Grants are also available through the US Department of Agriculture to build broadband infrastructure in rural areas [49], and more recently, the US Department of Education for discounted internet service for low-income households [50]. Thus, most students in rural, low-income communities have access to the infrastructure needed for technology-based programs.

The STAC Intervention

STAC, which stands for the 4 bystander strategies Stealing the Show, Turning it Over, Accompanying Others, and Coaching Compassion, [51] is a brief, stand-alone bystander intervention developed to reduce barriers to program implementation for middle school students. STAC was originally designed as a 90-minute in-person intervention, including didactic and experiential role-play components, as well as 2 booster sessions. The didactic training includes education about bullying and cyberbullying, consequences of bullying, bystander roles, and a description of the 4 STAC strategies: (1) *stealing the show*—using humor or distraction to interrupt the bullying situation, removing the attention away from the target; (2) *turning it over*—informing an adult about the bullying and asking for help; (3) *accompanying others*—befriending or providing supporting the targeted student; and (4) *coaching compassion*—gently confronting the perpetrator to increase empathy for target. The experiential component includes role-plays in which students practice using the STAC strategies in hypothetical bullying situations. Students also participate in 2 biweekly 15-minute booster sessions to reinforce their skill acquisition. Research indicates that the STAC intervention is effective in reducing bullying [52,53] and negative mental health outcomes for bystanders [35-39]. Researchers have also adapted the STAC intervention to be culturally appropriate for middle school students in rural, low-income communities [54-56] with similar positive effects on bullying reduction [54,57] and improved mental health [54,58].

Need for a Technology-Based STAC Intervention

Most bullying programs are comprehensive school-wide interventions that require significant resources, creating implementation barriers for schools. Although the STAC intervention reduces some of these barriers, in-person interventions pose implementation challenges, particularly for schools in rural, low-income communities. To increase access for these schools, we propose to translate the STAC intervention into a technology-based format. As a first step, we conducted a needs assessment with the goal of understanding needs and perceived program implementation challenges for schools in rural, low-income communities to provide information on how to best serve students in these schools [59]. Findings from interviews and focus groups with key school personnel (ie, administrators, teachers, and school counselors) from 3 middle schools in rural, low-income communities indicated a strong interest in a technology-based bullying intervention. Participants also described positive conditions for implementation, including support from the administration and technology readiness of their schools. Participants identified implementation challenges, such as time and financial resources. In addition, participants provided feedback related to translating the intervention into a technology-based format, including the importance of activities that require user input to increase user engagement. Overall, the findings supported the need for the proposed technology-based STAC (STAC-T) intervention and provided feedback on challenges that need to be addressed for successful adoption and sustained implementation in middle schools in rural, low-income communities.

STAC-T Intervention

The STAC-T web-based app is intended to shift program delivery from an in-person intervention to a technology-delivered intervention, thereby increasing accessibility and eliminating implementation barriers. STAC-T will also allow large groups of students to be trained simultaneously, accessed from a computer, tablet, or smartphone. The STAC-T app is designed to be a modular program that can be customized to meet the needs of individual schools. The app will allow students to customize their experience by selecting avatars and bullying scenarios. Assessment and personalized feedback components are also infused through the program to individualize the user experience to promote behavior change [60,61]. The innovative, user-centered design of the STAC-T app will be inherently sensitive to the cultural needs of students and identify personally appropriate strategies. The STAC-T app addresses both bullying and negative mental health outcomes for targets and bystanders through an evidence-based approach that is adapted for a broader audience and uses technology to effectively implement bullying prevention.

The success of the STAC-T design can be evaluated through usability testing. Usability testing is an important step in the development of technology-based programs, providing information from end users on what works and gaps in how the program functions [62], as well as the acceptability and relevance of program content. Usability ultimately affects the likelihood that the program will be adopted [63] and thus is a critical component of technology-based intervention development [64]. Assessing usability with school personnel and students is also important, as school personnel make decisions about program adoption [65], and students, as end-users, need to understand and respond positively to the program to benefit from the content [62,66].

Study Objectives

This study aims to evaluate the usability and acceptability of the STAC-T prototype to inform full-scale development of the STAC-T intervention app. To achieve this aim, we implemented usability testing with key stakeholders (ie, school personnel and students) at 3 middle schools in rural, low-income communities (N=16). The study had the following objectives: (1) to assess the usability and acceptability of the STAC-T prototype, (2) to understand school needs for and barriers to program implementation, and (3) to assess the differences in usability between school personnel and students. We used a mixed-methods design to assess the usability and acceptability

of the app prototype, as well as ways to improve the app, likelihood of program use, and potential implementation barriers.

Methods

Participants

Participants were key school personnel (ie, administrators, teachers, and school counselors) and students were recruited from 3 middle schools in rural, low-income communities in the Northwest region of the United States. The schools were selected based on previous and ongoing research partnerships. The 3 schools were Title 1 schools, with 52.9% (339/641), 68.98% (725/1051), and 98.8% (663/671) of the student population at the 3 schools being below the poverty line. A total of 6 school personnel and 10 students were recruited. Among the school personnel, the age ranged from 28 to 59 years (mean 42.0 years, SD 10.8 years) and the majority were women (5/6, 83%) and White (6/6, 100%). For students, age ranged from 11 to 14 years (mean 12.2 years, SD 0.9 years), with 40% (4/10) in grade 6, 30% (3/10) in grade 7, and 30% (3/10) in grade 8, and the majority were girls (8/10, 80%) and White (7/10, 70%).

Development of the STAC Prototype App

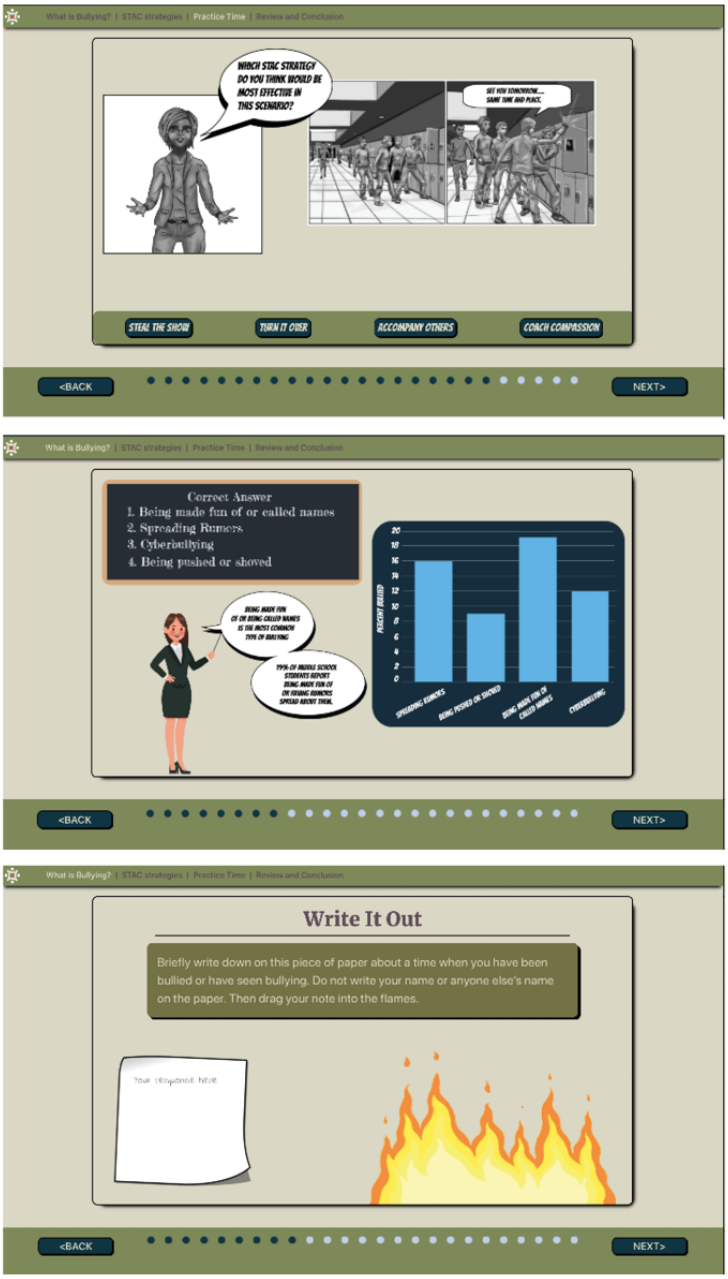
We used a multitheoretical framework to guide prototype production. We translated the STAC in-person intervention into a technology framework guided by Persuasive System Design, a theoretical guide for translating clinical aims to health-related technology frameworks [67-69]. Instructional objectives and design of STAC-T were guided by the existing STAC intervention for rural and low-income students, and input from an expert advisory board, school personnel, and students. The STAC-T prototype was developed using AGILE programming, a collaborative and incremental programming methodology [67-69]. The prototype was functional on all web browsers that support HTML5 and was built on a full stack web application using HTML or JavaScript as the main interface. React.js was used as the front-end framework. The system was accessible on desktop computers, iOS, Android tablets, and smartphones. Design ideas were created in written form, combined with scripts, flowcharts, and storyboards, before creating actual images and authoring elements. Programmers produced the STAC-T prototype, alpha- and beta tested it in-house for stability and code errors, tested it for usability, and revised it following an iterative, agile production process.

The initial prototype was created with clickable wireframes that showed the progression of content with still graphics (Figure 1). In addition to Persuasive System Design, User-centered Design [70,71], and the ADDIE (Analyze, Design, Develop, Implement, and Evaluate) model [72-74], we followed a user interface and instructional design approach to develop the STAC-T prototype. User input was solicited through iterative cycles, with adjustments made based on feedback [75], resulting in enhanced user experience and more reliable and effective results [76,77]. Design elements such as space (colors and visual space), components (characters and objects), and mechanics (actions) were determined for program features (ie, activities and games). Consistent with research on mobile health strategies for adolescents [78-80], the program development approach emphasized gaming as a teaching strategy, and interactive components were the centerpiece of the STAC-T intervention. STAC strategy practice required students to select between 2 avatars, view a bullying event, select actions to operationalize the STAC strategy, view the avatar enacting the selected action, and receive feedback on its effectiveness. To reward learning and bolster adherence, *badges* (visual reward icons eg, *Stealing theShow badge*) were included for intermittent awards to encourage user engagement.

The STAC prototype content comprised 3 overarching modules: (1) What is Bullying? Users were presented with background information on bullying, including bullying definitions (ie, physical, verbal, relationship, and cyberbullying), bullying facts and statistics, characteristics of students who bullied, and negative consequences of bullying; (2) What are Bystanders? Users were taught what a bystander is, and how bystanders affect bullying outcomes. This module explained the 4 bystander roles: (1) *Assistants*: those who intentionally help the bully; (2) *Reinforcers*: those who are not directly involved in hurting another student, but encourage the bully by standing around, laughing, or watching quietly; (3) *Outsiders*: those who do not take sides while witnessing bullying; and (4) *Defenders*: those who do something to stop the bullying situation or help the target in some way, and (3) STAC strategies: Users were introduced to 4 STAC strategies: (1) stealing the show, (2) turning it over, (3) accompanying others, and (4) coaching compassion. The module also included the STAC strategy practice using avatars selected by the user.

Iterative focus groups (3 rounds) with middle school students (N=21; 14 girls and 7 boys) attending schools in rural, low-income communities informed prototype development. Groups ascertained whether the app concept was easily understood and engaging and identified essential features for a successful prototype. The first round (n=10) was conducted with students who had been trained in the in-person STAC intervention. The second (n=6) and third (n=5) rounds were conducted with students who were naive to the STAC program. Group 1 provided feedback on the translation of the in-person STAC program content to a technology-based format for prototype 1. Group 1 reported that the content was realistic, expressed interest in a web translation, and made suggestions for modifications for web translation that were incorporated into prototype 1. For group 2, prototype 1 was very well received; participants thought the content was useful, important, and understandable and were enthusiastic about using the program. Group 2 participants suggested more interactivity (eg, short clips, videos, thought bubbles, graphs, and cartoons) to increase engagement. Group 3 evaluated prototype 2, reporting that they liked the program content and ease of navigation. Participants liked the sketches and discussed changes to characters (eg, wear hoodies instead of clothes with buttons) and surroundings (eg, school bus looked too clean) to more closely match the middle school environment. Input from the focus groups informed the development of the final prototype used in this study.

Figure 1. Samples from the STAC-T (technology-based Stealing the Show, Turning it Over, Accompanying Others, Coaching Compassion) prototype.



Procedures

Participant recruitment and usability testing occurred during the spring of 2020. All research procedures were approved by the University Institutional Review Board and by the School District or Administration. The researchers provided the school counselor from each school with an email script describing the purpose and procedures of the study. School counselors were also provided with rubrics developed by the research team to identify key school personnel and students who demonstrated the following characteristics assessed by the rubric: school personnel: (1) caring for students, (2) desire to be a positive influence on school climate, (3) approachable to students, (4) caring about addressing the problem of bullying, and (5) leadership qualities; and students: (1) leadership, (2) maturity, (3) responsibility, (4) caring toward others, (5) influence, and (6) a desire to be a positive influence on peers. For each item, school personnel and students were assessed on a 3-point scale, which included the ratings of *yes* and *somewhat* to *no* for each item described above. School personnel and students who scored *yes* or *somewhat* on all inclusion criteria were eligible to participate. The school counselor used the rubric to identify and contact key school personnel and students and then used the script to invite them to participate in the study. Usability testing and interviews were conducted remotely. Researchers obtained informed consent for school personnel and parental consent and student assent for students and collected demographic information from participants immediately before usability testing.

Participants interacted with the STAC prototype app by selecting an avatar and environment to tailor content and talked aloud while completing the tasks and identifying problems and solutions attempted. Researchers and users were on videoconference and shared screens. Researchers could see what

participants were doing, and they were able to communicate with each other in real time. The researchers observed the users as they worked through the tasks and asked questions to gather more data. Participants were asked to complete a brief usability survey followed by a semistructured interview protocol. All participants were asked to provide information about their perceptions of (1) program utility, (2) relevance and appropriateness of program content, and (3) ways they would improve the program. School personnel were also asked about (1) their thoughts on implementation feasibility, (2) likelihood of school program adoption, and (3) barriers to program use. All individual interviews lasted for 1 hour and were audio-recorded. School personnel received a web-based US $50 Amazon gift card as an incentive for participation in the usability testing and individual interview. There were no incentives for student participants.

Measures

Demographics

Participants self-reported their age, ethnicity, race, and gender. Students also reported their grade level.

Usability

Usability was assessed using the System Usability Scale (SUS) [81]. The SUS is a widely used 10-item validated tool used to measure the usability and acceptability of technology-based programs. Responses were measured on a 5-point Likert scale ranging from 0 (*strongly disagree*) to 4 (*strongly agree*). Items were summed, and the total was multiplied by 2.5, creating an overall SUS score ranging from 0 to 100. An SUS score of ≥68 was considered above average [82].

User-friendliness

One item was used to assess the user-friendliness of the program. Participants were asked to rate the user-friendliness with the question: *Overall, I would rate the user-friendliness of this program as...* with a 7-point scale ranging from 0 (*worst imaginable*) to 7 (*best imaginable*).

Program Satisfaction

Two items were used to assess the program satisfaction. Participants were asked the question, "Would you tell your friends/colleagues to use the program?" with response choices *yes*, *no*, and *don't know*. Participants were also asked how many stars they would give the program (1 star being the lowest and 5 stars being the highest).

Interview Questions

Following the usability testing sessions, participants were asked a series of open-ended questions about the utility and relevance of the app prototype, as well as ways to improve the app, likelihood of program use, and potential implementation barriers. School personnel and students were asked the following: (1) Please talk about your perception of how useful this program could be to helping to address the problem of bullying at school, (2) Please share your thoughts on whether you think the content of this program is relevant and appropriate for students at your school and your community, and (3) Can you talk about ways that you would improve the program? School personnel were asked the following: (4) What are your thoughts on how practical or workable you think it would be to use this program at your school? (5) What do you believe is the likelihood that your school would use this intervention? and (6) What, if anything, would keep you from using this program?

Data Analyses

Quantitative

Quantitative data from the questionnaires were analyzed using the SPSS version 25.0. Before conducting statistical analyses, the data were examined for outliers and normality, and all variables were within the normal range for skewness and kurtosis. Descriptive statistics were used and presented separately for the school personnel and students. We examined the differences between school personnel and students using 2-tailed independent sample *t* tests for continuous variables and chi-square analyses for categorical variables. All analyses were considered significant at $P<.05$.

Qualitative

Qualitative data from open-ended questions were analyzed separately for the school personnel and students. A team member who participated in conducting the usability tests transcribed the data verbatim. We used thematic analysis [83,84] to identify, analyze, organize, describe, and report themes found within the qualitative data. NVivo was used to track quotes and organize themes. Two trained master's students with previous experience in qualitative data analysis and a faculty member with expertise in qualitative methodologies analyzed the data. Before analyzing the data, the analysts discussed their assumptions and expectations regarding possible findings. Themes for each question were determined based on consensus. The process began with each team member individually developing initial themes for each question for school personnel and students. Next, the team met 2 times in person and conducted email communications over a 4-week period to arrive at a consensus on themes and frequency categories supported by participant quotations. During the first meeting, the lead analyst trained the team on thematic coding procedures, and each analyst coded the data separately. Next, the team met again for a second time to share each of their themes, followed by team members commenting and voicing agreement or disagreement. The analysts relied on participants' quotes to resolve disagreements. Once the team reached a consensus, an external auditor reviewed the interview transcripts and themes. Overall, the auditor agreed with the team's findings but provided an alternative way to organize one of the findings. The analysts used email correspondence to discuss and incorporate the auditor's feedback and obtain a final consensus for themes. The external auditor agreed with the team's final themes. Interview data were deidentified to ensure anonymity, and quotes were identified by participant type (ie, school personnel or students).

Results

Quantitative Analysis

Usability

The usability scores of the SUS are presented in Table 1. Overall, the scores for both school personnel and students suggested a very high level of usability, functionality, and acceptability. As presented in Table 1, there were no differences in any of the individual items or the SUS total score between school personnel and students, with both participant groups scoring the STAC-T app at a very high level of usability.

Table 1. Means and SDs for the System Usability Scale (SUS) by school personnel and students (N=16)[a].

	School personnel (n=6), mean (SD)	Students (n=10), mean (SD)	*t* test (*df*)	*P* value
I think that I would like to use the program frequently	3.33 (0.82)	3.40 (0.52)	−0.20 (14)	.84
I found the program to be more complex than it needed to be	0.67 (0.52)	0.40 (0.70)	0.81 (14)	.43
I thought the program was easy to use	3.50 (0.55)	3.80 (0.42)	−1.24 (14)	.24
I think that I would need the support of a technical person to be able to use this program	0.17 (0.41)	0.20 (0.42)	−0.16 (14)	.88
I found the various functions in the program were well put together with each other	3.33 (0.52)	3.50 (0.53)	−0.62 (14)	.55
I thought there was too much inconsistency in this program	0.17 (0.41)	0.40 (0.52)	−0.94 (14)	.36
I imagine that most people would learn to use this program very quickly	3.50 (0.55)	3.60 (0.70)	−0.30 (14)	.77
I found the program very awkward to use	0.17 (0.41)	0.10 (0.32)	0.37 (14)	.72
I felt very sure that I could use the program correctly	3.33 (0.82)	3.60 (0.52)	−0.81 (14)	.43
I needed to learn a lot of things before I could get going with this program	0.00 (0.00)	0.10 (0.32)	−0.76 (14)	.46
SUS total score	89.58 (5.10)	91.75 (6.98)	−0.66 (14)	.52

[a]Responses were scored on a 5-point Likert scale ranging from 0 (strongly disagree) to 4 (strongly agree).

User-friendliness

School personnel and students rated the program high regarding user-friendliness. For school personnel, scores on user-friendliness ranged from 5 to 6 (mean 5.83, SD 0.41). For students, scores on user-friendliness ranged from 5 to 7 (mean 6.10, SD 0.57). There were no differences in scores between school personnel and students in terms of user-friendliness (t_{14}=−1.00; *P*=.33).

Program Satisfaction

The program satisfaction ratings are listed in Table 2. Overall ratings suggested that school personnel and students were satisfied with the program. There were no differences in scores between school personnel and students on program recommendation (χ^2_1=1.3; *P*=.24, or star ratings, χ^2_1=2.3; *P*=.31).

Table 2. Program satisfaction by school personnel and students (N=16).

Variables	School personnel (n=6), n (%)	Students (n=10), n (%)
Recommend program		
Yes	6 (100)	8 (80)
No	0 (0)	0 (0)
Unsure	0 (0)	2 (20)
Star rating		
1 star	0 (0)	0 (0)
2 stars	0 (0)	0 (0)
3 stars	1 (17)	0 (0)
4 stars	4 (66)	6 (60)
5 stars	1 (17)	4 (40)

Qualitative Analysis

Overview

Qualitative feedback for the STAC-T prototype supported the quantitative findings and was very positive overall, with participants sharing the perception that the STAC app is useful, relevant, and appropriate, as well as ways to improve the program. In addition, school personnel shared positive thoughts about program feasibility, a high likelihood of program adoption, and implementation barriers. The results are presented below, organized by the following themes: (1) utility and user-friendliness, (2) relevance, (3) program feedback, (4) feasibility, and (5) program adoption.

Utility and User-friendliness

Regarding participants' perceptions on the program's utility, school personnel and students indicated that STAC-T is useful and delivers helpful educational content to intervene in bullying. For example, 1 teacher stated, "this program is definitely one that can give strategies to students that they can apply and they also have a variety of strategies that they can choose from."

A student shared:

> *I thought it was pretty good. It gave lots of information; gave different ways to deal with the problem or kind of bullying.*

Furthermore, students emphasized that the program allowed them to learn about bullying. One student reflected, "I think it'd be really really good cause there's some things I didn't know and I learned about it."

School personnel also indicated that the program is user-friendly and straightforward, and that developing the program on the web is a strength. A school counselor indicated:

> *I love that it's gonna be digital because that's right now what we do. We compete with the Minecraft's [videogames] of the world.*

Relevance

Participants were also asked to share their thoughts regarding the relevance and appropriateness of the program content for students at their school and their community. School personnel indicated that the program is relevant and can help increase students' understanding of relationships. For example, a school counselor stated, "I definitively think that it's relevant, just like we talked about, especially the online portion."

Another school counselor indicated that the program helps "...kids to continue to understand the differences..." (between people). Students also indicated that the program was age-appropriate. One student said:

> *I think it is very appropriate and relevant. I think it's a great idea, and it has great examples so kids can see [bullying] if they haven't seen it before or see different situations it [bullying] could be in.*

Another student shared:

> *Yeah I think it's really good cause I've had some programs that use really big words, and so you can't understand, or it explains every word like you're a little child. Your program was perfect, like right at my level.*

Program Feedback

When asked for feedback regarding how to improve the program, school personnel suggested that it is important to follow through while making the program interactive. One teacher stated:

> *I'd have to kind of see how all of the avatars work together, how the animation works together, how relevant that would work with keeping students' attention. Because right now currently I'd have to read everything and know students aren't going to read everything.*

A school counselor said, "I definitely think having, like you already talked about, the piece of it being interactive, definitely needs to have that."

Students suggested adding colors, including realistic pictures. One student stated:

> *I would definitely add color to the pictures, that's one thing. I bet that would catch somebody's eye. I know people who have more creative minds tend to pay attention to color more, cause I'm one of those people.*

Feasibility

School personnel were also asked to speak about the feasibility of the program implementation. Participants provided positive responses indicating that the program was user-friendly and would align with existing programs that address social and emotional development. One school counselor indicated:

> *...it's very user-friendly, even for me who had struggles getting going. It really gives you very clear, concise steps to be able to move forward and directions of what you need to do.*

A teacher stated, "So, for me I can easily put it into the current curriculum or the current places in which we are but I think that would be up to individual teachers."

Program Adoption

Finally, when asked about the likelihood of program adoption, the school personnel indicated that their school would be open to using this program. One teacher stated: "I know I would really enjoy it personally because I think it's the strategies that are important for kids to understand and be able to use and apply."

A school counselor indicated, "I think our school has been pretty open to learning new or picking up new tools that would enrich out student population."

The school personnel indicated that resource constraints were the major barriers to adopting the program. A school counselor shared:

> *Money, money, and cost. But honestly, as far as implementation, it's simple because it's online. It*

really does come down to the time and how much does it cost.

Discussion

Principal Findings

This study aims to examine the usability of a technology-based bystander bullying intervention designed specifically for middle schools in rural, low-income communities. We were particularly interested in perspectives from both middle school personnel and students, as these participant groups represent key stakeholders who are in the position of making decisions regarding bullying programming and are the end users of the program. We aimed to test the usability of the STAC-T prototype; assess program utility, user-friendliness, and relevance; and gain an in-depth understanding of the needs and challenges related to program adoption and feedback regarding program content and delivery. Quantitative analysis of the survey data demonstrated a very positive response to the program, which was supported by qualitative data from individual interviews. Overall, the results indicate that participants perceived the STAC-T app to be useful, user-friendly, and appropriate for students at their schools and reported high levels of satisfaction with the program. Findings from this study indicate that the STAC-T app is relevant and feasible for implementation in middle schools in rural, low-income communities.

The findings of this study support the usability of the STAC-T app. For both school personnel (mean 89.58, SD 5.10) and students (mean 91.75, SD 6.98), scores on the SUS indicated a very high level of usability, well exceeding the standard cutoff score of 68 [82]. The user-friendliness of the program was also rated very high by both school personnel, with all participants rating the STAC app at ≥5 on a 0 to 7 point scale. We found no differences between school personnel and students in terms of SUS scores or user-friendliness ratings, suggesting that all participants found the program to be highly usable. Qualitative data supported these results, with both school personnel and students indicating that they perceived the program to be user-friendly, as well as age-appropriate and relevant for middle school students in rural, low-income communities. Furthermore, both school personnel and students reported high levels of satisfaction with the program, with most participants indicating that they would recommend the program to others. These findings are particularly important, as usability and acceptability are associated with both program adoption and implementation [63].

Regarding feasibility of implementation, school personnel indicated that they believed that their school would use the STAC-T app if the intervention was cost-effective. This finding is consistent with research indicating that administrators in rural middle schools would support a technology-based bullying program [59] and parallels research on bullying prevention in rural communities, identifying cost as a barrier to program implementation [16,59]. These findings also echo research suggesting the number 1 barrier to implementing educational technology is the lack of school district funding [45] and that financial resources and program effectiveness are necessary conditions for program implementation and sustainability of delivery of school-based programs [65].

Both school personnel and students discussed STAC-T program strengths and considerations for increasing user engagement. Qualitative findings suggested that participants perceived the STAC-T app to be useful and helpful in addressing the problem of bullying. They also indicated that the content of the program was appropriate for their schools, confirming the need for bullying programming that teaches students skills to use to intervene in bullying situations. This finding is consistent with previous research in which school personnel in rural, low-income communities indicated a need for bystander training that teaches students strategies to intervene on behalf of targets of bullying and includes having students actively practice these strategies across different scenarios [59]. Regarding user engagement, participants suggested adding more color and more realistic pictures, as well as increasing interactivity. This feedback is consistent with research suggesting that the motivating elements of technology-based interventions, including program content, length, and interactivity, are important in promoting behavior change [85].

Limitations

This study supports the usability, relevance, and feasibility of the STAC-T prototype, providing valuable information for the development of a full-scale STAC-T app. However, certain limitations of this study must be noted. Participants were recruited from 3 schools in rural, low-income areas from 1 state in the Northwestern region of the United States. Although participants were recruited from 3 different counties to increase generalizability, school personnel and students from different regions of the country may have a different perspective. Furthermore, most participants in this study were female, further limiting the generalizability of the study. Although variability was low in the quantitative data and we did not identify divergent responses among participant interview responses, interpretation of the results for males should be made cautiously as females comprised approximately 81% (13/16) of the sample. Further formative research conducted during the development of the full-scale STAC app should include purposeful sampling to ensure appropriate representation of male participants. Furthermore, although most participants were White (13/16, 81%), this parallels the ethnic and racial composition of rural communities in the United States [86]. It is also possible that social desirability influenced participants as they were aware that the goal of the study was to translate the in-person STAC intervention into a technology-based format.

Implications

This study has important implications for the development of a technology-based bullying intervention that addresses the needs and challenges specific to middle schools in rural, low-income communities. First, participants provided very high usability ratings for the STAC-T prototype, with qualitative data supporting the usability, utility, and relevance of the program. In addition, participants indicated that program implementation is feasible as long as the program is cost-effective. These findings support the development of the full-scale STAC-T app, while keeping the price point as an

important consideration. Program cost considerations are particularly important in rural schools, as they face significant financial challenges due to a lower tax base for funding programs [16]. Furthermore, to enhance program implementation and sustainability, sufficient resources must be available [65]. Thus, the STAC-T app needs to be designed to be low in cost to be successfully adopted by schools in rural, low-income communities.

The findings from this study provide a strong scientific premise for moving forward with a full-scale production of the STAC-T app and have implications for program development. The full-scale STAC-T app will include 3 core modules, including strategy training, in which students select customized avatars as they move through role-plays in which they practice the STAC strategies. On the basis of participant feedback on the STAC-T prototype, all core modules will be built to include gaming and features requiring user input (eg, drag and drop, hover, click and reveal, and video). Real instances of game dynamics and mechanics will include several options, such as badges, leaderboards, levels, points, achievements, avatars, content unlocking, quests, social recognition, teams, and tokens. These items target the social cognitive theory components of modeling, outcome expectancies, self-efficacy, self-regulation, identification, and reciprocity. Consistent with research on mobile health strategies for adolescents [78-80], this design emphasizes gaming as a teaching strategy. The outcomes of each game will be used to ensure information uptake and demonstrate comprehension (eg, users must achieve a targeted score to receive a reward), thus providing feedback to participants early and often. To reinforce learning and bolster adherence, *badges* (ie, visual icons), which are highly effective gaming tools used to encourage user engagement and build game *loyalty* [87,88], will be awarded intermittently. These components will be the centerpiece of the program and are designed to increase engagement and learning.

Conclusions

Bullying is a significant public health issue for middle schools in rural, low-income communities. Schools in these communities face multiple and competing demands on their time and have limited access to training, funding, and mental health professionals to implement bullying programming. The findings of this study demonstrate the usability, relevance, and feasibility of a brief, technology-based bystander bullying intervention. This study provides support for the development of the full-scale development of the STAC-T app and provides information that can be used to enhance program usability while addressing the unique needs of schools in rural, low-income communities.

Acknowledgments

This study was registered with ClinicalTrials.gov (identifier NCT04681495). This research was supported by the National Institute on Minorities and Health Disparities of the National Institutes of Health under award number R41MD014943. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

MKB is the owner of Klein Buendel, Inc, the organization that received the National Institutes of Health grant award for this project. As it is a small business technology transfer, a concerted effort to commercialize a final product that may result from this and subsequent National Institutes of Health–funded research will be made, and the organization may receive royalties or revenue from the commercialization. DMD and AM may also receive royalties or revenue from the commercialization.

References

1. Student reports of bullying: results from the 2017 school crime supplement to the national crime victimization survey. U.S. Department of Education - NCES 2019-054. 2019. URL: https://nces.ed.gov/pubs2019/2019054.pdf [accessed 2021-09-20]
2. Injury prevention and control: #StopBullying. Centers for Disease Control and Prevention, National Center for Injury Prevention and Control. URL: https://www.cdc.gov/injury/features/stop-bullying/index.html [accessed 2021-09-21]
3. Ching CB, Lee H, Mason MD, Clayton DB, Thomas JC, Pope JC, et al. Bullying and lower urinary tract symptoms: why the pediatric urologist should care about school bullying. J Urol 2015 Feb;193(2):650-654. [doi: 10.1016/j.juro.2014.08.103] [Medline: 25196655]
4. van Geel M, Goemans A, Vedder PH. The relation between peer victimization and sleeping problems: a meta-analysis. Sleep Med Rev 2016 Jun;27:89-95. [doi: 10.1016/j.smrv.2015.05.004] [Medline: 26140869]
5. Fekkes M, Pijpers FI, Verloove-Vanhorick S. Bullying behavior and associations with psychosomatic complaints and depression in victims. J Pediatr 2004 Jan;144(1):17-22. [doi: 10.1016/j.jpeds.2003.09.025] [Medline: 14722513]
6. Hase C, Goldberg S, Smith D, Stuck A, Campain J. Impacts of traditional bullying and cyberbullying on the mental health of middle school and high school students. Psychol Schs 2015 May 22;52(6):607-617 [FREE Full text] [doi: 10.1002/pits.21841]
7. Rivers I, Poteat V, Noret N, Ashurst N. Observing bullying at school: the mental health implications of witness status. Sch Psychol Q 2009 Dec;24(4):211-223 [FREE Full text] [doi: 10.1037/a0018164]
8. Jack AH, Egan V. Childhood bullying, paranoid thinking and the misappraisal of social threat: trouble at school. School Ment Health 2018 Nov 22;10(1):26-34 [FREE Full text] [doi: 10.1007/s12310-017-9238-z] [Medline: 29503671]

9. Juvonen J, Graham S, Schuster MA. Bullying among young adolescents: the strong, the weak, and the troubled. Pediatrics 2003 Dec 03;112(6 Pt 1):1231-1237. [doi: 10.1542/peds.112.6.1231] [Medline: 14654590]
10. Siegel RS, La Greca AM, Harrison HM. Peer victimization and social anxiety in adolescents: prospective and reciprocal relationships. J Youth Adolesc 2009 Sep 22;38(8):1096-1109. [doi: 10.1007/s10964-009-9392-1] [Medline: 19636774]
11. Storch EA, Masia-Warner C. The relationship of peer victimization to social anxiety and loneliness in adolescent females. J Adolesc 2004 Jun;27(3):351-362. [doi: 10.1016/j.adolescence.2004.03.003] [Medline: 15159093]
12. Fisher BW, Gardella JH, Teurbe-Tolon AR. Peer cybervictimization among adolescents and the associated internalizing and externalizing problems: a meta-analysis. J Youth Adolesc 2016 Sep 22;45(9):1727-1743. [doi: 10.1007/s10964-016-0541-z] [Medline: 27447707]
13. Kowalski RM, Giumetti GW, Schroeder AN, Lattanner MR. Bullying in the digital age: a critical review and meta-analysis of cyberbullying research among youth. Psychol Bull 2014 Jul;140(4):1073-1137. [doi: 10.1037/a0035618] [Medline: 24512111]
14. Sticca F, Perren S. Is cyberbullying worse than traditional bullying? Examining the differential roles of medium, publicity, and anonymity for the perceived severity of bullying. J Youth Adolesc 2013 May 27;42(5):739-750. [doi: 10.1007/s10964-012-9867-3] [Medline: 23184483]
15. Dulmus CN, Theriot MT, Sowers KM, Blackburn JA. Student reports of peer bullying victimization in a rural school. Stress Trauma Crisis 2004 Jan;7(1):1-16. [doi: 10.1080/15434610490281093]
16. Leadbeater B, Sukhawathanakul P, Smith A, Thompson R, Gladstone E, Sklar N. Bullying and victimization in rural schools: risks, reasons, and responses. J Rural Community Dev 2013;8(1):31-47 [FREE Full text]
17. Nansel TR, Overpeck M, Pilla RS, Ruan WJ, Simons-Morton B, Scheidt P. Bullying behaviors among US youth: prevalence and association with psychosocial adjustment. J Am Med Assoc 2001 Apr 25;285(16):2094-2100 [FREE Full text] [doi: 10.1001/jama.285.16.2094] [Medline: 11311098]
18. Azeredo C, Rinaldi A, de Moraes C, Levy R, Menezes P. School bullying: a systematic review of contextual-level risk factors in observational studies. Aggr Violent Behav 2015 May;22:65-76 [FREE Full text] [doi: 10.1016/j.avb.2015.04.006]
19. Evans C, Smokowski P, Cotter K. Cumulative bullying victimization: an investigation of the dose–response relationship between victimization and the associated mental health outcomes, social supports, and school experiences of rural adolescents. Child Youth Serv Rev 2014 Sep;44:256-264 [FREE Full text] [doi: 10.1016/j.childyouth.2014.06.021]
20. Tippett N, Wolke D. Socioeconomic status and bullying: a meta-analysis. Am J Public Health 2014 Jun;104(6):e48-e59. [doi: 10.2105/AJPH.2014.301960] [Medline: 24825231]
21. Smokowski PR, Cotter KL, Robertson C, Guo S. Demographic, psychological, and school environment correlates of bullying victimization and school hassles in rural youth. J Criminol 2013 Jun 11;2013:1-13. [doi: 10.1155/2013/137583]
22. Midgett A, Doumas DM. Witnessing bullying at school: the association between being a bystander and anxiety and depressive symptoms. School Ment Health 2019 Jan 25;11(3):454-463. [doi: 10.1007/s12310-019-09312-6]
23. Doumas DM, Midgett A. Witnessing cyberbullying and internalizing symptoms among middle school students. Eur J Investig Health Psychol Educ 2020 Oct 04;10(4):957-966 [FREE Full text] [doi: 10.3390/ejihpe10040068] [Medline: 34542429]
24. Doumas D, Midgett A. The association between witnessing cyberbullying and depressive symptoms and social anxiety among elementary school students. Psychol Schs 2020 Dec 15;58(3):622-637 [FREE Full text] [doi: 10.1002/pits.22467]
25. Doumas DM, Midgett A. The relationship between witnessing cyberbullying and depressive symptoms and social anxiety among middle school students: is witnessing school bullying a moderator? J Child Adolesc Couns 2021 Aug 18:1-12. [doi: 10.1080/23727810.2021.1934369]
26. Wright MF, Wachs S, Harper BD. The moderation of empathy in the longitudinal association between witnessing cyberbullying, depression, and anxiety. Cyber Psychol 2018 Dec 29;12(4):11419. [doi: 10.5817/cp2018-4-6]
27. Janson G, Carney J, Hazler R, Oh I. Bystanders' reactions to witnessing repetitive abuse experiences. J Couns Dev 2009;87(3):319-326. [doi: 10.1002/j.1556-6678.2009.tb00113.x]
28. Hutchinson M. Exploring the impact of bullying on young bystanders. Educ Psychol Pract 2012 Nov 08;28(4):425-442 [FREE Full text] [doi: 10.1080/02667363.2012.727785]
29. Lambe LJ, Hudson CC, Craig WM, Pepler DJ. Does defending come with a cost? Examining the psychosocial correlates of defending behaviour among bystanders of bullying in a Canadian sample. Child Abuse Negl 2017 Mar;65:112-123. [doi: 10.1016/j.chiabu.2017.01.012] [Medline: 28131946]
30. Wu W, Luu S, Luh D. Defending behaviors, bullying roles, and their associations with mental health in junior high school students: a population-based study. BMC Public Health 2016 Oct 10;16(1):1066 [FREE Full text] [doi: 10.1186/s12889-016-3721-6] [Medline: 27724881]
31. DeSmet A, Bastiaensens S, Van Cleemput K, Poels K, Vandebosch H, Cardon G, et al. Deciding whether to look after them, to like it, or leave it: a multidimensional analysis of predictors of positive and negative bystander behavior in cyberbullying among adolescents. Comput Hum Behav 2016 Apr;57:398-415 [FREE Full text] [doi: 10.1016/j.chb.2015.12.051]
32. Gaffney H, Ttofi M, Farrington D. Evaluating the effectiveness of school-bullying prevention programs: an updated meta-analytical review. Aggr Violent Behav 2019 Mar;45:111-133 [FREE Full text] [doi: 10.1016/j.avb.2018.07.001]

33. Salmivalli C, Voeten M. Connections between attitudes, group norms, and behaviour in bullying situations. Int J Behav Dev 2016 Jun 30;28(3):246-258. [doi: 10.1080/01650250344000488]
34. Lenhart A, Madden M, Smit A, Purcell K, Zickuhr K. Teens, kindness and cruelty on social network sites. Pew Research Center. 2011. URL: https://www.pewresearch.org/internet/2011/11/09/teens-kindness-and-cruelty-on-social-network-sites/ [accessed 2021-09-21]
35. Doumas DM, Midgett A, Watts AD. The impact of a brief, bullying bystander intervention on internalizing symptoms: is gender a moderator of intervention effects? Sch Psychol Int 2019 Feb 14;40(3):275-293. [doi: 10.1177/0143034319830149]
36. Midgett A, Doumas DM. The impact of a brief bullying bystander intervention on depressive symptoms. J Couns Dev 2019 Jun 13;97(3):270-280. [doi: 10.1002/jcad.12267]
37. Midgett A, Doumas D, Trull R. Evaluation of a brief, school-based bullying bystander intervention for elementary school students. Prof Sch Couns 2018 Feb 15;20(1):172-183. [doi: 10.5330/1096-2409-20.1.172]
38. Midgett A, Doumas D, Trull R, Johnston A. A randomized controlled study evaluating a brief, bystander bullying intervention with junior high school students. J Sch Couns 2017;15(9):1-34 [FREE Full text]
39. Watts AD, Doumas DM, Midgett A. Efficacy of a brief, school‐based bystander bullying intervention on high school students' alcohol use. J Addict Offender Couns 2019 Oct 07;40(2):66-83. [doi: 10.1002/jaoc.12066]
40. Macháčková H, Dedkova L, Sevcikova A, Cerna A. Bystanders' support of cyberbullied schoolmates. J Community Appl Soc Psychol 2012 Dec 27;23(1):25-36. [doi: 10.1002/casp.2135]
41. Sjögren B, Thornberg R, Wänström L, Gini G. Bystander behaviour in peer victimisation: moral disengagement, defender self-efficacy and student-teacher relationship quality. Res Pap Educ 2020 Feb 12;36(5):588-610. [doi: 10.1080/02671522.2020.1723679]
42. Bussey K, Fitzpatrick S, Raman A. The role of moral disengagement and self-efficacy in cyberbullying. J Sch Violence 2014 Sep 30;14(1):30-46 [FREE Full text] [doi: 10.1080/15388220.2014.954045]
43. Reinke WM, Stormont M, Herman KC, Puri R, Goel N. Supporting children's mental health in schools: teacher perceptions of needs, roles, and barriers. Sch Psychol Q 2011 Mar;26(1):1-13. [doi: 10.1037/a0022714]
44. NetSupport DNA - IT asset management for schools. NetSupport DNA. URL: http://www.netsupportdna.com/us/education/ [accessed 2021-09-21]
45. Duffin E. E-learning and digital education - Statistics and facts. Statista. 2020. URL: https://www.statista.com/topics/3115/e-learning-and-digital-education/ [accessed 2021-09-21]
46. Survey questionnaires results: students' computer access and use. National Assessment of Educational Progress (NAEP). 2015. URL: https://www.nationsreportcard.gov/sq_computer/ [accessed 2021-09-21]
47. Riddell R. Is E-Rate doing enough to effectively expand broadband in rural schools? K-12 Dive. 2018. URL: https://www.k12dive.com/news/is-e-rate-doing-enough-to-effectively-expand-broadband-in-rural-schools/519788/ [accessed 2021-09-21]
48. Rachfal C, Gilroy A. Broadband internet access and the digital divide: federal assistance programs. Congressional Research Service. 2019. URL: https://sgp.fas.org/crs/misc/RL30719.pdf [accessed 2021-09-21]
49. USDA launches new program to create high-speed internet e-connectivity in rural America. U.S. Department of Agriculture. 2018. URL: https://www.usda.gov/media/press-releases/2018/12/13/usda-launches-new-program-create-high-speed-internet-e-connectivity [accessed 2021-09-21]
50. Department of education releases resource to help schools improve ventilation systems to prevent COVID-19. U.S. Department of Education. 2021. URL: https://www.ed.gov/news/press-releases/department-education-releases-resource-help-schools-improve-ventilation-systems-prevent-covid-19 [accessed 2021-07-23]
51. Midgett A, Doumas D, Sears D, Lundquist A, Hausheer R. A bystander bullying psychoeducation program with middle school students: a preliminary report. Prof Couns 2015 Dec;5(4):486-500. [doi: 10.15241/am.5.4.486]
52. Midgett A, Doumas DM, Johnston AD. Establishing school counselors as leaders in bullying curriculum delivery. Prof Sch Couns 2018 May 28;21(1):1-9. [doi: 10.1177/2156759x18778781]
53. Midgett A, Doumas DM, Trull R, Johnson J. Training students who occasionally bully to be peer advocates: is a bystander intervention effective in reducing bullying behavior? J Child Adolesc Couns 2017 Jan 19;3(1):1-13. [doi: 10.1080/23727810.2016.1277116]
54. Midgett A, Doumas DM. Acceptability and short-term outcomes of a brief, bystander intervention program to decrease bullying in an ethnically blended school in low-income community. Contemp School Psychol 2020 Sep 18;24(4):508-517. [doi: 10.1007/s40688-020-00321-w]
55. Midgett A, Doumas D, Moran M, Gallo L. Cultural adaptation of a school-based bullying intervention. J Educ Psycholo Res 2020:2-60 [FREE Full text]
56. Moran M, Midgett A, Doumas DM, Moody S, Porchia S. A mixed method evaluation of a culturally adapted, brief, bullying bystander intervention for middle school students. J Child Adolesc Couns 2019;5(3):221-238. [doi: 10.1080/23727810.2019.1669372] [Medline: 33015352]
57. Moran M, Midgett A, Doumas DM. Evaluation of a brief, bystander bullying intervention (STAC) for ethnically blended middle schools in low-income communities. Prof Sch Couns 2020 Jul 20;23(1):2156759X2094064. [doi: 10.1177/2156759x20940641]

58. Midgett A, Doumas DM, Peralta C, Bond L, Flay B. Impact of a brief, bystander bullying prevention program on depressive symptoms and passive suicidal ideation: a program evaluation model for school personnel. J Prev Health Promot 2020 Jul 26;1(1):80-103 [FREE Full text] [doi: 10.1177/2632077020942959] [Medline: 33738442]
59. Midgett A, Doumas D, Myers V, Moody S, Doud A. Technology-based bullying intervention for rural schools: perspectives on needs, challenges, and design. Rural Ment Health 2021 Jan;45(1):14-30 [FREE Full text] [doi: 10.1037/rmh0000151] [Medline: 33777286]
60. DiClemente CC, Marinilli AS, Singh M, Bellino LE. The role of feedback in the process of health behavior change. Am J Health Behav 2001;25(3):217-227. [Medline: 11322620]
61. Miller DT, Prentice DA. Changing norms to change behavior. Annu Rev Psychol 2016;67:339-361. [doi: 10.1146/annurev-psych-010814-015013] [Medline: 26253542]
62. Wichansky AM. Usability testing in 2000 and beyond. Ergonomics 2000 Jul 10;43(7):998-1006. [doi: 10.1080/001401300409170] [Medline: 10929833]
63. Lyon AR, Bruns EJ. User-centered redesign of evidence-based psychosocial interventions to enhance implementation-hospitable soil or better seeds? JAMA Psychiatry 2019 Jan 01;76(1):3-4. [doi: 10.1001/jamapsychiatry.2018.3060] [Medline: 30427985]
64. 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 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
65. Han SS, Weiss B. Sustainability of teacher implementation of school-based mental health programs. J Abnorm Child Psychol 2005 Dec;33(6):665-679. [doi: 10.1007/s10802-005-7646-2] [Medline: 16328743]
66. Ritterband LM, Gonder-Frederick LA, Cox DJ, Clifton AD, West RW, Borowitz SM. Internet interventions: in review, in use, and into the future. Prof Psychol Res Pr 2003;34(5):527-534. [doi: 10.1037/0735-7028.34.5.527]
67. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
68. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009;24(1):28. [doi: 10.17705/1cais.02428]
69. Oinas-Kukkonen H, Harjumaa M. A systematic framework for designing and evaluating persuasive systems. In: Oinas-Kukkonen H, Hasle H, Harjumaa M, Segerståhl K, Øhrstrøm P, editors. Persuasive Technology. Berlin, Heidelberg: Springer; 2008:164-176.
70. Abras C, Maloney-Krichmar D, Preece J. User-centered design. In: Bainbridge W, editor. Encyclopedia of Human-Computer Interaction. Vol 37. Thousand Oaks, CA: SAGE Publications; 2004:445-456.
71. Norman D, Draper S. User Centered System Design: New Perspectives on Human-Computer Interaction. Hillsdale, NJ: Lawrence Erlbaum Associates Inc; 1986.
72. Gustafson K, Branch R. What is instructional design? In: Reiser RA, Dempsey JV, editors. Trends and Issues in Instructional Design and Technology. Columbus, OH: Merril/Prentice Hall; 2002.
73. Molenda M. In search of the elusive ADDIE Model. Perf Improv 2015 Feb 19;54(2):40-42. [doi: 10.1002/pfi.21461]
74. Peterson C. Bringing ADDIE to life: instructional design at its best. J Educ Multimedia Hypermedia 2003;12(3):227-241. [doi: 10.1007/978-0-387-09506-6_3]
75. Mazzara ML. Application of theory: minimalism and user centered design. In: Society for Technical Communication. Hanover, Germany: TIB University Library; 2001:327-331.
76. Nicholson S. A user-centered theoretical framework for meaningful gamification. In: Proceedings of the Games+Learning+Society Conference 8.0. 2012 Presented at: Games+Learning+Society Conference 8.0; June 13-15, 2012; Madison, WI URL: https://scottnicholson.com/pubs/meaningfulframework.pdf
77. Pagulayan R, Keeker K, Fuller T, Wixon D, Romero R, Gunn DV. User-centered design in games. In: Sears A, Jacko JA, editors. Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications. Mahwah, NJ: Lawrence Erlbaum Associates, Inc; 2002:883-906.
78. Cafazzo JA, Casselman M, Hamming N, Katzman DK, Palmert MR. Design of an mHealth app for the self-management of adolescent type 1 diabetes: a pilot study. J Med Internet Res 2012;14(3):e70 [FREE Full text] [doi: 10.2196/jmir.2058] [Medline: 22564332]
79. Graafland M, Dankbaar M, Mert A, Lagro J, De WL, Schuit S, et al. How to systematically assess serious games applied to health care. JMIR Serious Games 2014;2(2):e11 [FREE Full text] [doi: 10.2196/games.3825] [Medline: 25654163]
80. Hieftje K. A tablet game for risk reduction and HIV prevention in adolescents. J Mob Technol Med 2012 Dec;1(4S):8. [doi: 10.7309/jmtm.32]
81. Brooke J. SUS: A 'Quick and Dirty' usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland AL, editors. Usability Evaluation in Industry. London: Taylor and Francis; 1996.
82. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Studies 2009;4(3):114-123 [FREE Full text]

83. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
84. Nowell LS, Norris JM, White DE, Moules NJ. Thematic analysis: striving to meet the trustworthiness criteria. Int J Qual Methods 2017 Oct 02;16(1):1-13. [doi: 10.1177/1609406917733847]
85. Lehto T, Oinas-Kukkonen H. Persuasive features in web-based alcohol and smoking interventions: a systematic review of the literature. J Med Internet Res 2011 Jul 22;13(3):e46 [FREE Full text] [doi: 10.2196/jmir.1559] [Medline: 21795238]
86. Rural America at a glance. U.S. Department of Agriculture. 2020. URL: https://www.ers.usda.gov/data-products/chart-gallery/gallery/chart-detail/?chartId=99538 [accessed 2021-09-21]
87. Casilli C, Knight E. 7 things you should know about badges. EDUCAUSE. 2012. URL: https://net.educause.edu/ir/library/pdf/ELI7085.pdf [accessed 2021-09-21]
88. Miller AS, Cafazzo JA, Seto E. A game plan: gamification design principles in mHealth applications for chronic disease management. Health Informatics J 2014 Jul 1:184-193. [doi: 10.1177/1460458214537511] [Medline: 24986104]

Abbreviations

ADDIE: Analyze, Design, Develop, Implement, and Evaluate
STAC: Stealing the Show, Turning it Over, Accompanying Others, Coaching Compassion
STAC-T: technology-based STAC
SUS: System Usability Scale

Edited by G Eysenbach; submitted 13.08.21; peer-reviewed by M Nitsch; comments to author 17.09.21; revised version received 22.09.21; accepted 26.09.21; published 26.10.21.

Please cite as:
Doumas DM, Midgett A, Myers V, Buller MK
Usability of a Technology-Based Bystander Bullying Intervention for Middle School Students in Rural, Low-Income Communities: Mixed Methods Study
JMIR Form Res 2021;5(10):e32382
URL: https://formative.jmir.org/2021/10/e32382
doi:10.2196/32382
PMID:34582355

Original Paper

Use of a Mobile App for the Process Evaluation of an Intervention in Health Care: Development and Usability Study

Winnie Szu Yun Chin[1,2,3], ScD, MS; Alicia Kurowski[1,4], ScD; Rebecca Gore[1,4], PhD; Guanling Chen[5], PhD; Laura Punnett[1,4], ScD, MS; SHIFT Research Team[1,4]

[1]Center for the Promotion of Health in the New England Workplace (CPH-NEW), Lowell, MA, United States

[2]Department of Public Health, University of Massachusetts Lowell, Lowell, MA, United States

[3]Division of Population Sciences, Dana-Farber Cancer Institute, Boston, MA, United States

[4]Department of Biomedical Engineering, University of Massachusetts Lowell, Lowell, MA, United States

[5]Department of Computer Science, University of Massachusetts Lowell, Lowell, MA, United States

Corresponding Author:
Winnie Szu Yun Chin, ScD, MS
Division of Population Sciences
Dana-Farber Cancer Institute
450 Brookline Ave, Room LW711
Boston, MA, 02215
United States
Phone: 1 617 632 5602
Email: winnies_chin@dfci.harvard.edu

Abstract

Background: Process evaluation measures the context in which an outcome was or was not achieved through the ongoing monitoring of operations. Mobile apps are a potentially less burdensome tool for collecting these metrics in real time from participants. Research-driven apps are not always developed while paying attention to their usability for target users. Usability testing uncovers gaps in researchers', developers', and users' mental models of what an efficient, effective, and satisfying product looks like and facilitates design improvement. Models may vary by user demographics.

Objective: This study describes the development of a mobile app for collecting process evaluation metrics in an intervention study with health care workers that uses feedback at multiple stages to refine the app design, quantify usage based on workers' overall adoption of the app and the app's specific function, and compare the demographic and job characteristics of end users.

Methods: An app was developed to evaluate the Center for Promotion of Health in the New England Workplace Healthy Workplace Participatory Program, which trains teams to develop solutions for workforce health obstacles. Labor-management health and safety committee members, program champions, and managers were invited to use the app. An accompanying website was available for team facilitators. The app's 4 functions were meeting creation, postmeeting surveys, project time logs, and chat messages. Google Analytics recorded screen time. Two stages of pilot tests assessed functionality and usability across different device software, hardware, and platforms. In stage 1, student testers assessed the first functional prototype by performing task scenarios expected from end users. Feedback was used to fix issues and inform further development. In stage 2, the app was offered to all study participants; volunteers completed task scenarios and provided feedback at deployment. End user data for 18 months after deployment were summarized and compared by user characteristics.

Results: In stage 1, functionality problems were documented and fixed. The System Usability Scale scores from 7 student testers corresponded to *good* usability (mobile app=72.9; website=72.5), whereas 15 end users rated usability as *ok* (mobile app=64.7; website=62.5). Predominant usability themes from student testers were *flexibility and efficiency* and *visibility of system status*; end users prioritized *flexibility andefficiency* and *recognition rather than recall*. Both student testers and end users suggested useful features that would have resulted in the large-scale restructuring of the back end; these were considered for their benefits versus cost. In stage 2, the median total use time over 18 months was 10.9 minutes (IQR 23.8) and 14.5 visits (IQR 12.5). There were no observable patterns in use by demographic characteristics.

Conclusions: Occupational health researchers developing a mobile app should budget for early and iterative testing to find and fix problems or usability issues, which can increase eventual product use and prevent potential gaps in data.

(JMIR Form Res 2021;5(10):e20739) doi:10.2196/20739

KEYWORDS

mobile apps; usability testing; user experience design; mobile phone; mhealth; iterative testing; participatory research; user demographics; worker participation

Introduction

Background

In intervention research, process evaluation has become increasingly important to reliably assess the reasons for the effectiveness of an intervention or a lack thereof. Process evaluation is the ongoing monitoring of operations to measure the implementation process, and it provides a detailed context for subsequent outcomes evaluation [1]. When evaluated in occupational health interventions, this is often measured specifically as context, reach, dose delivered, dose received, fidelity, implementation, and recruitment [2]. Measuring these items requires data collection on the diversity of participants or organizations, recruitment or retention of members, their role in teams or activities, number and type of events attended, amount of time spent in and outside of the teams' activities, benefits and challenges of participation, satisfaction with the work or process, and balance of power and leadership [1]. These research process activities can be hard to track during a participant's workday or shift, and data collection might be more efficient if delivered through one medium.

The widespread use of smartphones has made mobile apps popular for mobile health (mHealth) studies, defined as "the use of mobile and wireless devices to improve health outcomes, health care services, and health research" [3]. However, there have been a limited number of tools developed to support and evaluate workplace improvement studies [4,5]. With the use of mobile apps, participant information can be collected in real time, increasing the convenience and, thus, ideally, participation level and data quality. Surveys can be created and sent at set intervals or during times where data collection is time sensitive and may be easier to incorporate into daily life [6]. Event logs can be entered at any time by participants to document where and when actions were taken by members [1]. With the flexibility of mobile apps, intervention activities and participant engagement and satisfaction can be tracked, and all of these are the primary components of process evaluation. Mobile apps provide a unique medium of data collection that might overcome the organizational and logistical barriers to data collection, which are common in occupational health studies.

Most mHealth apps are developed using a consumer-driven approach and are motivated by the participants' perceived need to monitor their goals or manage their health condition. In contrast, apps created primarily for a research goal parallel a product-driven (or driving-markets) approach, which involves developing a unique product first and then influencing the structure or behavior of the existing market to gain a competitive edge [7]. These *research-driven* apps seek to fulfill a data collection need of the investigators [8] but may be at a disadvantage compared with consumer-driven apps, with regard to participants' intrinsic motivation to use the app.

The usability of mobile apps strongly influences their actual use. User-centered design principles are recommended for mobile app development and include 4 principles: specify the context of use, specify app requirements, create design solutions in stages, and evaluate designs iteratively [9]. Ongoing evaluation through end user usability testing and quality assurance protocols is intended to enhance user satisfaction and uncover obstacles to effective and efficient product use. Unfortunately, the target users are infrequently involved in designing the features [10-12].

User-centered design is surprisingly difficult, and little empirical evidence has guided app development [4,12,13]. Empathy and appreciation for how users think and work are critical [13]. App development teams must not assume that users will approach the app in the same way that they would [13]. Participatory methods, such as card sorting, engage end users during the early development stages to design the information architecture to resemble users' mental models [14,15]. This informs the development of prototypes that software developers often test on emulators. Real device testing is then needed to accommodate the various combinations of phone dimensions, screen resolutions, software versions, changing environmental contexts, and unreliable wireless networks that characterize mobile device interaction [16]. Testing with target users throughout the lifecycle of the app also helps with uncovering problems and discovering opportunities to improve the product while ensuring that the design is still flexible [17]. However, in the work context, end users may not have adequate time to devote to iterative and participatory design [15,18,19]; therefore, testing basic functionality in a nonrepresentative sample may be necessary to identify bugs before introducing the app to the user population. When usability testing is implemented only in later stages, fixes or feature requests are likely to be more costly and time consuming, as much of the structure has already been set [20]. Furthermore, research participants encountering early difficulties may become permanently discouraged from using the app throughout the study.

Some guidelines have been created for the iterative usability testing of mobile apps, but the form and extent of testing vary among studies, and validated instruments are not always used [21]. Laboratory-based testing is often tedious and expensive and has been criticized for not reflecting real use cases [22]. Others have proposed toolkits that can be embedded into the code of the mobile app to track user interface events from users [22]. A recent review suggested that combination approaches would be most useful [21]. Regardless of the method, testing with just 5 users helps identify 80% of the usability problems [23]. Information on usability testing should be documented as it has an impact on the adoption and use of the app [24].

Understanding relevant user demographics is important to assist in designing for a wide variety of target users [11]. The influence of user demographics on app use is not clear, particularly in

mHealth studies. In one study focusing on a diabetes mHealth system, younger users performed app tasks faster and had fewer errors [11]. Another study on a cardiovascular disease risk management app found that younger populations downloaded the app more often, but older populations demonstrated greater sustained engagement [25]. Younger age groups may find mobile technologies commonplace and readily acceptable, but older adults are also interested in technology and are capable of acquiring complex computer skills [26]. Relationships between age and use of computer software have been examined in the literature, with some authors finding inverse relationships, whereas other authors did not find inverse relationships [27].

A US survey revealed that those who were younger, had more education, reported excellent health, and had higher income were the main users of health apps [28]. Another study reported that the odds of users downloading health apps were higher in college or graduate school than in high school and decreased with increasing age [29]. The existing literature provides some information regarding the type of users that use mHealth apps to improve their own health, but these differences may or may not apply to research-driven applications, including program evaluation.

Objectives

The aforementioned gaps in the literature suggest that our mobile app is one of the very few apps developed to collect and conduct process evaluation in a participatory workplace change study. The aim of this study is to describe a user-centered development approach for a mobile app that tracks the process of a participatory intervention. In particular, this study seeks to (1) describe the iterative development of a mobile app to track a workplace change process, (2) identify functions most often used within the app by target users, and (3) examine demographic and job-related characteristics of app users.

Methods

Study Design

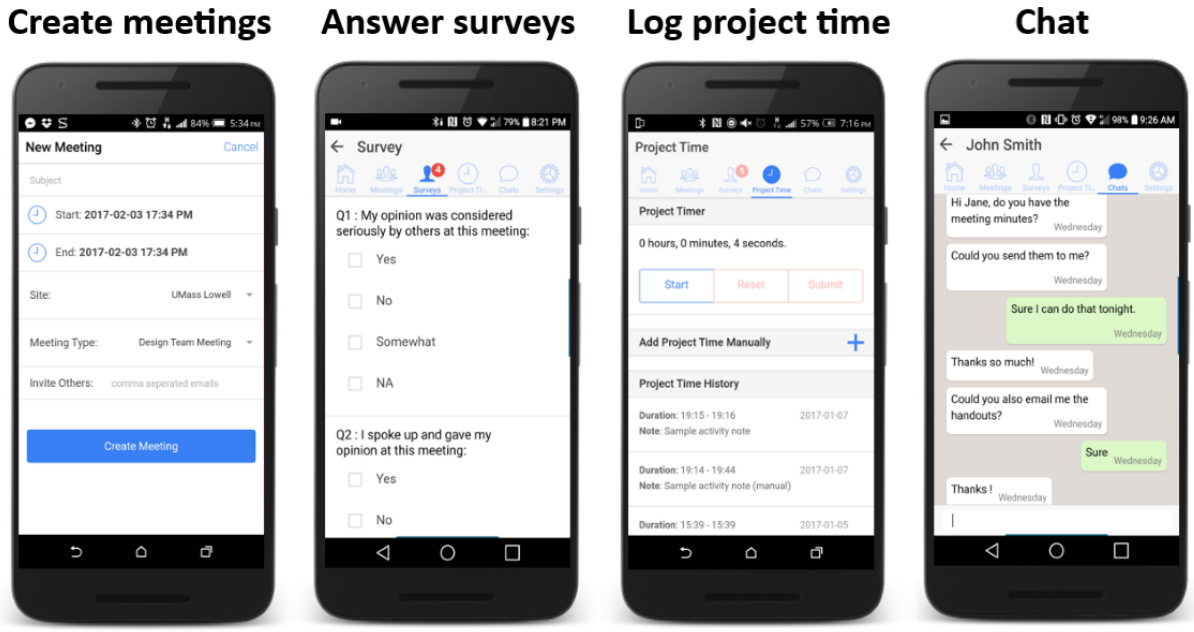
This descriptive study involved 5 health care facilities in the northeast United States participating in the Safety and Health through Integrated, Facilitated Teams (SHIFT) intervention study **(**Clinicaltrials.gov NCT04251429) [30]. SHIFT uses the Center for Promotion of Health in the New England Workplace Healthy Workplace Participatory Program (HWPP), a process for increasing the effectiveness of occupational health and safety committees through root cause analysis, identifying health and safety needs, and proposing solutions to leadership for implementation [31]. The joint labor-management Design Teams (DTs) are groups of 8 to 12 frontline workers from various departments, with 2 cofacilitators who chair the meetings and facilitate the HWPP process. The Steering Committee (SC) included upper-level managers responsible for budget and resource allocation. This usability study was conducted to guide the development of a data collection tool used in the trial; the number of subjects did not correspond to the anticipated enrollment for the trial itself.

System Development

Mobile App

The HWPP Assistant app was developed for iOS (version 8.0 or higher) and Android (version 4.1 or higher) platforms using an agile approach. Detailed specifications were developed by the researchers in consultation with a computer scientist. The app had 4 main functions that allowed users to (1) create meetings, (2) answer surveys, (3) log time spent on project-related tasks, and (4) converse privately or with the entire group regarding any questions or concerns (Figure 1).

Figure 1. Safety and Health through Integrated, Facilitated Teams project mobile app: HWPP Assistant functions. HWPP: Healthy Workplace Participatory Program.



The meeting function provided the ability to create and view upcoming meetings. The 3 meeting types were DT, SC, and small group meetings. Cofacilitators could set up DT meetings, which automatically invited all DT members and cofacilitators at the specified site. SC members could create SC meetings. Anyone could call small group meetings by entering the email addresses of the desired meeting attendees. Once meetings were created, the invited users were able to see a list of upcoming meetings and download agendas attached to them. Users were reminded of meetings 24 hours and 2 hours before the meeting date or time via push notifications.

Surveys were sent at the end of meetings to all members, with specific questions based on their role in the study, intervention phase, and intervention status (control or intervention). Survey question templates were created by the research team and uploaded onto the website, where the majority of the administrative tasks were performed.

Time spent on project-related tasks was reported in either of the 2 ways: by using a start or stop timer, before or after executing the task, or by selecting a predetermined time interval of 30, 60, 90, or 120 minutes. The second method was designed as a backup in case the user forgot to use the timer or was not able to use their phone during the task. Both methods require an activity note to be submitted, describing the task executed.

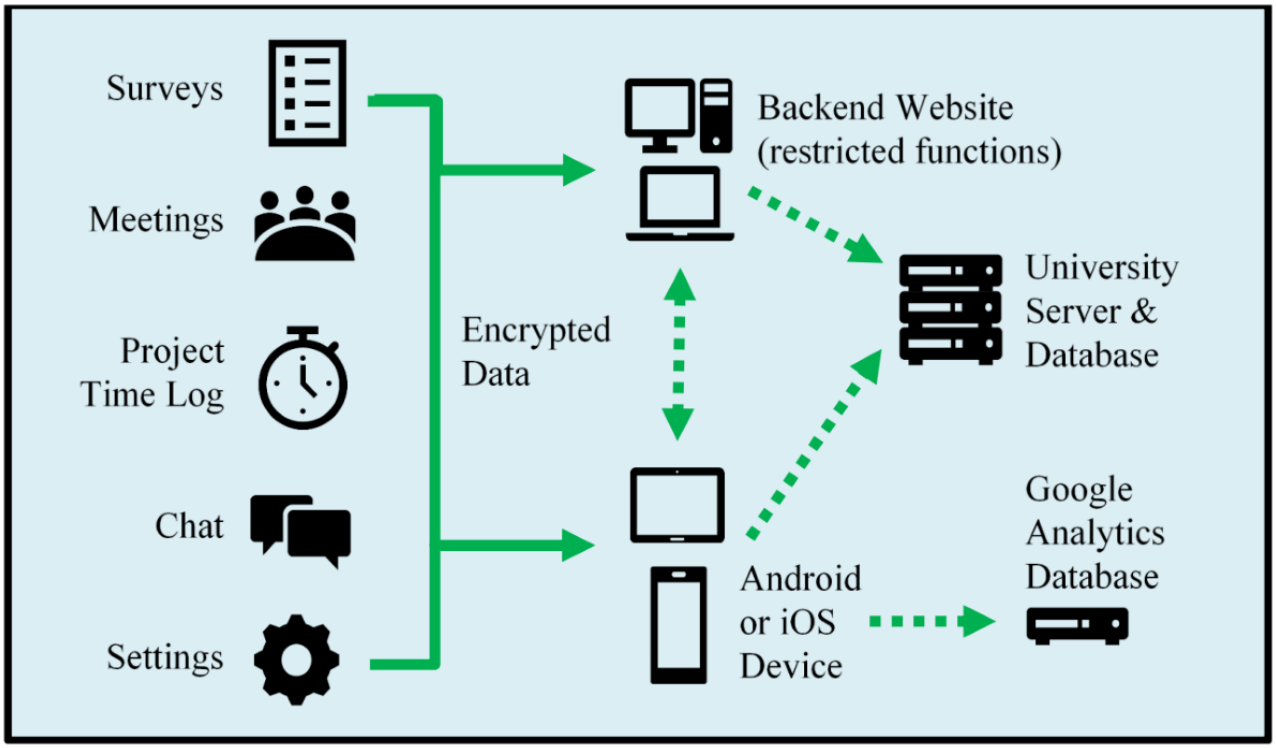
Finally, the chat function allowed the users to chat privately or broadcast a message to users within their team. Private and broadcast chats included a timestamp, and push notifications were sent when a message was received.

An integrated page timer in the background recorded the time that users spent on each page function when the app was open and reported the information to Google Analytics. The app was included in the study protocol approved by the University of Massachusetts Institutional Review Board (approval number: #16-131-PUN-XPD).

Website Application

Researchers entered users into the system through the accompanying website and monitored the incoming data in real time. Cofacilitators could set meeting times and upload meeting minutes and agendas using the website as well. All data were encrypted and sent to the back-end password-protected server hosted on the University of Massachusetts Lowell Department of Computer Science server (Figure 2) [32].

Figure 2. System architecture of the Safety and Health through Integrated, Facilitated Teams project mobile app HWPP Assistant. HWPP: Healthy Workplace Participatory Program.



Usability Testing

Stage 1: Student Testing

Initial testing of the app and accompanying website was done by 7 students because our formative research indicated that our target participants were overburdened with work demands. These undergraduate and graduate students were employed in the SHIFT project and were compensated for their time. Each student was assigned 1 to 6 roles to test. These roles were DT member, facilitator, SC member, champion, researcher, and administrator. Students were asked to provide informed consent and were instructed to think aloud while completing tasks, while the researcher observed the tester and took notes. A user guide was made available during the test. Once the student completed all the app tasks for one role, the researcher changed their role. This process was repeated for each student until all tasks were performed for the assigned roles. Depending on the students' availability, some students had the option of testing when they were not observed, but they were asked to provide detailed

descriptions and schedule a follow-up meeting if their feedback needed clarification.

The task scenarios were created based on the guidelines by Dumas and Redish [33] to mimic the functionality expected from the app across roles. Task scenarios were selected and developed based on (1) tasks that users would do with the product, (2) tasks that probed potential usability problems, and (3) tasks suggested from concerns and experience from initial testing by research team members [33]. Paper task scenarios were provided at the app pilot, where users were asked to perform the essential functions within the app and provide qualitative feedback. Further questions for each task included whether users encountered errors, whether they were able to complete the task, and whether they could see a more effective method to complete the task. Some roles had website tasks associated with them; this paper focuses on the app tasks.

At the end of the entire testing period, students answered the System Usability Scale (SUS) separately for the mobile app and the website [34]. The themes and issues were identified and reported back to the developer for the next iteration of the app.

Stage 2: End User Testing

The mobile app was deployed at all 5 sites among 94 participants who were engaged in the DTs and SCs at their respective sites. A walk-through tutorial was presented in person to all participants, along with hard-copy task scenarios and SUS forms for real-time evaluation. If necessary, users were allowed to complete task scenarios and SUS at home and mail them or return their evaluations at the next meeting.

Demographics Survey and Team Roster

As part of the larger SHIFT project, surveys were distributed to all employees at the 5 sites. Demographic information was collected by these surveys and added to a team roster with other observationally collected information from these meetings. Information from these sources provided demographic and occupational information on the subgroups of participants in this study. This information was combined with end user usability responses using individuals' randomized ID.

Data Management and Statistical Analysis

All task scenarios and SUS surveys were entered into the project database. In stage 2, the app data were exported via the website and Google Analytics. Surveys distributed, surveys answered, meeting dates or time, chat sessions, project time logs, and app screen time were compiled for each end user for 18 months after deployment to SHIFT study end users.

Data were stored on an encrypted, password-protected drive in the Computer Science department at the University of Massachusetts Lowell. Backups from the back-end server and Google Analytics were run in parallel on the SHIFT project's shared drive at 1-month intervals, which was the frequency expected for meetings and their associated surveys.

Usability scores were computed using the SUS scoring system [34]. SUS responses were scored from 0 to 100 and compared with a threshold of 68 and an adjective scale [35,36]. The NVivo 12 program (QSR International) was used to analyze the themes of qualitative feedback on the types of errors reported and fixed. One research assistant analyzed the feedback content by sorting through responses by the app's functions and interpreting whether the feedback provided was focused on usability or functionality. Unique usability feedback was categorized using the usability heuristics developed by Nielsen [37]. In the cases where suggestions from testers and users could not be implemented, suggestions were documented and (where possible) alternative solutions were proposed.

SAS version 9.4 (SAS Institute) was used to analyze use and demographic information. The Kruskal-Wallis test and Wilcoxon rank sum tests were used to compare median screen time between groups based on demographic and job characteristics.

Results

Overview

The first prototype was created in November 2016. Pilot testing was carried out on versions 0.1.0 to 0.3.2 from December 2017 to February 2018. Three student testers used an iPhone with iOS operating system of 8.1.3-11.1.2, whereas the other 4 used Samsung smartphones with Android operating systems 4.4.2-7.0. Screen sizes and resolutions ranged from 4 to 5.7 inches and 540×960 to 2560×1440 pixels, respectively. Testing time ranged from 15 minutes to 5 hours per person, depending on the number of roles that each student tested and whether the test was moderated or unmoderated.

The app was deployed in June 2018. Approximately one-fourth (23/95, 24%) of the invited end users downloaded the app for use during the SHIFT project. Most of the use was by participants while they were in the intervention period. Most of the users were female (15/23, 65%); were White (20/23, 87%); were not Latino or Hispanic (19/22, 86%); reported their health as "very good" (8/14, 57%); were members of a union (16/22, 73%); worked the day shift (21/23, 91%); reported an income of at least US $75,000 (9/14, 64%); and had at least a college or professional degree (10/14, 71%; Table 1). The median total use time over 18 months was 10.9 minutes (IQR 23.8). The median total number of page visits was 14.5 visits (IQR 12.5). There were no significant differences in the median total use time and page visits between the demographic groups. When compared with those who did not use the app, app users were more likely to have college or professional education and to earn US $75,000 or more.

The tested app versions ranged from 1.0.1 to 1.0.5. Most (18/23, 78%) of the end users had Apple devices with iOS versions from 9.3.2 to 12.1.2, whereas 22% (5/23) were Android users with operating systems ranging from 7.0 to 9.0. Screen sizes and resolutions ranged from 5 to 6.4 inches and 1280×720 to 2880×1440 pixels, respectively.

Table 1. Demographic and job characteristics of Safety and Health through Integrated, Facilitated Teams participants who downloaded the HWPP[a] Assistant app (n=23).

Characteristics	Participant[b], n (%)
Age group (years)	
25-39	10 (43)
40-54	8 (35)
≥55	5 (22)
Sex	
Male	8 (35)
Female	15 (65)
Race	
White	20 (87)
Unknown	3 (13)
Ethnicity	
Latino or Hispanic	3 (14)
Not Latino or Hispanic	19 (86)
BMI	
Normal	2 (15)
Overweight	6 (46)
Obese	5 (38)
Self-reported health	
Excellent	1 (7)
Very good	8 (57)
Good	4 (29)
Fair	1 (7)
Union status	
Member	16 (73)
Nonmember	6 (27)
Shift	
Day	21 (91)
Evening	2 (9)
Job title	
Administration	9 (39)
Clinical	9 (39)
Other	5 (22)
Income (US $)	
25,000-49,999	2 (14)
50,000-74,999	3 (21)
≥75,000	9 (64)
Education	
College or professional	10 (71)
Postgraduate	4 (29)

[a]HWPP: Healthy Workplace Participatory Program.

[b]Missing information is excluded.

Stage 1: Student Testing

Overview

The average SUS scores for the 7 student testers were similar for both the interfaces: 72.9 (SD 19.2) for the mobile app and 72.5 (SD 20.7) for the website, equating to *good* usability. The usability issues represented 6 different themes, with the 2 most common being *flexibility and efficiency* and *visibility of system status* (Table 2). Most problems were found in meetings and survey functions. The 2 functions were associated with each other, which meant that if an issue occurred in one, then the other was affected.

Table 2. HWPP[a] Assistant app usability issues reported by student testers (n=7).

App function and usability feedback	Remedied	Usability theme
Log-in		
"Was not automatically logged on website after logging into the app"	Yes	Flexibility and efficiency of use
"Quicker than expected"	No	Flexibility and efficiency of use
"Make sure the user is able to retrieve a lost email"	No	Recognition rather than recall
Meetings		
"App does not refresh to meeting tab after I hit 'create meeting'"	Yes	Visibility of system status
"Meetings are displayed but I cannot edit them"	No	User control and freedom
"Change times to a 12-hour [clock]"	Yes	Consistency and standards
Survey		
"Not sure if it posted or not even though it says it was submitted"	Yes	Visibility of system status
"Surveys I took still say 'ready' and do not say 'taken'."	Yes	Visibility of system status
"Slow to load"	Yes	Flexibility and efficiency of use
"Make sure questions are in order"	Yes	Match between system and the real world
"Could not go back to the survey and make edits to it"	No	Flexibility and efficiency of use
Time logs		
"Time is not displayed on the app so unclear if it posted"	Yes	Visibility of system status
"Quicker than expected"	No	Flexibility and efficiency of use
"Custom time option may be helpful"	No	Flexibility and efficiency of use
"Confirmation pop up was shown"	No	Visibility of system status
Chats		
"No confirmation...besides my sent message...that [my text] was received, read, or replied to."	No	Visibility of system status
"Space to add new contacts...add new chat members"	No	Flexibility and efficiency of use
"Download new messages faster"	Yes	Flexibility and efficiency of use

[a]HWPP: Healthy Workplace Participatory Program.

Download and Log-in

In stage 1, the email addresses of site users were pre-entered, and an initial generic password was set for them. For most tests, the download, log-in, and password change were successful and proceeded more quickly than expected. The functionality issues reported were all fixed by the developer. Testers requested automatic log-in to the accompanying website after logging into the app, but researchers decided that this would compromise confidentiality in a real-world use case where users have to log in using a shared work computer. One tester requested the ability to retrieve a lost email address, but researchers thought it was unlikely that end users would forget their address, and if necessary, they could contact the SHIFT team.

Meeting Creation and View

Designing the meetings function for both the computer and the smartphone simultaneously was challenging, as any changes had to be in sync with each other while also ensuring that some features were only on one medium (such as agenda upload on the website). These changes sometimes introduced functionality bugs that testers experienced, which were all fixed by the developer before reaching end users. Several testers requested the ability to edit meetings if they made a mistake, but this change would have required many engineering hours. As there was an existing feature to cancel individual meetings, researchers instead implemented an email feature to inform all study participants (including nonapp users) of changes in meeting date or time.

Meeting Survey Submission

Testers' responses for the survey function centered around data quality concerns such as whether questions were received, in the right order, and were provided for the right meeting type or role. This led to the implementation of a *subject* for meetings and their associated surveys to reduce potential confusion in end users. Other reported issues such as question order, inconsistent push notifications, and slow loading times were also fixed by the developer.

Time Log Reporting

All submitted time logs were received in the back end, but sometimes they did not appear under the project time history; this was fixed. Another tester requested more customization of time reported, outside the regular intervals offered by the app, but this was deemed unnecessary because of time and budget constraints.

Chat Communications

For the chat function, testers noted that sometimes communication between devices and push notifications were inconsistent and noted that there was no feature to indicate that a text was read or received. One student requested the ability to add new contacts to the chat, outside the project participants, but this was deemed unnecessary as the app is intended only for SHIFT study participants, with user entry by researchers.

Stage 2: End User Testing

Overview

After fixing the issues reported by the students, the app was deployed to end users. The average SUS scores for the 15 end users were similar for the 2 interfaces; scores of 62.33 (SD 20) for the mobile app and 62.5 (SD 17.7) for the website were achieved, equating to *acceptable* usability. The usability issues from end users represented 4 different themes, with the 2 most common being *recognition rather than recall* and *flexibility and efficiency of use* (Table 3). Task scenarios were also revised to target the functions the researchers expected to be most frequently used, as end users mostly had 15-45 minutes to test the app during the deployment meeting.

Table 3. HWPP[a] Assistant app usability issues reported by end users (n=23).

App function and usability feedback	Remedied	Usability theme
Log-in		
"Finger-print option would be helpful"	No	Recognition rather than recall
"Due to employer restriction on email access, made it difficult"	Yes	Match between system and the real world
Meetings		
"Need to be able to edit events"	No	Flexibility and efficiency of use
"Email addresses should auto-fill"	No	Recognition rather than recall
"Create room location on meeting app"	No	Flexibility and efficiency of use
Survey		
"Surveys should be associated with meeting"	Yes	Recognition rather than recall
Time logs		
"Add more minute options"	No	Flexibility and efficiency of use
"A bit clunky"	N/A[b]	Aesthetic and minimalist design
Other		
"People who work with people (clinical, care providers, etc) don't usually like technical things"	N/A	Match between system and the real world

[a]HWPP: Healthy Workplace Participatory Program.

[b]N/A: not applicable.

Download and Log-in

During testing, the app was still being approved by the university for distribution from the web page, so end users could not download it directly from built-in app stores. This caused some frustration, especially for iOS users, who could not always find the code in their institutional email. This was because of users not having their work email on their phone or the email address being incorrectly entered into the system, so they did not receive the install package. During the password-change task, end users requested the option of entering *the password twice to avoid mistyping it*. To improve efficiency, one user requested a *fingerprint login* function, which has been a rising feature in many apps during this time. However, this was forgone because of the cost and the fact that not all phones would have this feature.

Meeting Creation and View

For meeting creation, meetings were not always received by the intended participants. As the users were testing as a group, they were able to look at each other's phones to see whether anything was submitted. It was at this time that the site research

assistants and coaches helped troubleshoot the issue, but there were some issues reported regarding the validation of email addresses. Similar to the students, end users requested the ability to edit a meeting. Room location was requested on the app by one user, but this feature was forgone because of the additional cost of adding the feature and immense restructuring of the back end to accommodate an extra field. In addition, the DTs met consistently, which meant that meeting times and locations did not change frequently.

Survey Submission

Some issues reported were similar to those by the student testers, in that end users found that sometimes no survey was offered or received, and there were some screen freezes. Meeting-associated surveys were designed to show up after the meeting end time, but some users found that the survey for the next week showed up early.

Time Log Reporting

No functionality issues were reported for the time log function by end users, but there was feedback that it was a bit *clunky*. Similar to the student users, these users requested additional time duration options, but this was forgone because of the cost.

Chat Communications

Chats developed an issue that sometimes no text box appeared to type in. With some troubleshooting by the research assistants on site, this issue was resolved but was still reported to the developer.

General Feedback and Use

One user provided feedback that people who work for and with people, such as clinical workers or care providers, “don’t usually like technical things.”

For the 4 main functions (meetings, surveys, time logs, and chats), users in both the control and intervention periods spent the most time on meetings, on average, whereas the settings function was used the least. Users were likely to use this function to check if and when there was a meeting occurring. The median time spent on most app functions was generally higher at sites during the control periods than during intervention periods, but there were more visits to each of the pages in the intervention period than during the control periods (Table 4).

Overall, users utilized the app to answer more meeting surveys, set more meetings, and create more time log entries during the intervention period than those at the uncoached sites (Table 5).

Table 4. Time spent on HWPP[a] Assistant app functions by intervention status in the Safety and Health through Integrated, Facilitated Teams study (n=23).

App screen	Control status				Intervention status			
	Visits, n (%)	Minimum screen time, seconds	Maximum screen time, seconds	Median total screen time, seconds	Visits, n (%)	Minimum screen time, seconds	Maximum screen time, seconds	Median total screen time, seconds
Surveys	2 (13)	7	9	8	37 (20)	3.0	349	19
Home	3 (19)	9	98	10	32 (17)	1.0	1561	7.0
Meetings	3 (19)	6	147	130	30 (16)	1.0	899	23.5
Chats	4 (25)	5	10	7	24 (13)	1.0	107	6.5
Profile	1 (6)	29	29	29	18 (10)	2.0	58	5.0
Log-in	1 (6)	43	43	43	14 (7)	18	970	61
Time logs	1 (6)	3	3	3	19 (10)	1.0	299	6.0
Settings	1 (6)	3	3	3	13 (7)	1.0	9.0	2.0

[a]HWPP: Healthy Workplace Participatory Program.

Table 5. Number of user entries by HWPP[a] Assistant app function in the Safety and Health through Integrated, Facilitated Teams study (n=23).

App function	Number of user entries	
	Uncoached period, n (%)	Coached period, n (%)
Surveys	1 (33)	57 (41)
Time logs	0 (0)	42 (30)
Meetings	2 (67)	20 (14)
Chats	0 (0)	11 (8)
Chat threads	0 (0)	9 (6)

[a]HWPP: Healthy Workplace Participatory Program.

Discussion

Principal Findings

The primary objective of this study is to describe the logic and sequence of iterative usability testing that informed the development of a mobile app to track a workplace change process evaluation. Testing by students during early iterations of the app was immensely useful for problem discovery and identification of usability problems that would have led to frustration in end users and a potential loss of data in the field. Both student testers and end users mentioned concerns over flexibility and efficiency of use and suggested features related to the app's ability to recognize information rather than asking users to recall details, but only end users were able to provide important feedback on the match between their real-world occupational context and the app system. App use was aided by the on-site encouragement of the research team, but delays in fixing app issues may have led to initial users normalizing the use of paper data collection alternatives.

Students ranked the usability of the mobile app and website as *good*, whereas the end users ranked the 2 interfaces as *ok*. The most likely reason for this discrepancy was user knowledge of the context and ability to identify problems that students would not have been aware of. However, the difference in SUS scores between students and end users could also be because of student testers being watched during the tests, which may have skewed student responses and scores favorably, otherwise known as the Hawthorne effect. Another possible reason is that students completed these tests on the compensated project work time. Therefore, when problems and errors occurred that affected their ability to perform functions, it may not have caused as much inconvenience as for employees with busy work schedules who were voluntarily taking on extra responsibilities.

Some feature requests by students and end users were based on experiences with other apps or devices' capabilities. With mobile app development, users' expectations change over time; some requests may be small, whereas others require large-scale changes. It is difficult to plan for these considerations ahead of time without knowing in advance what software enhancements will become common and will be expected by users [38]. However, using participatory design during the concept stage may be helpful in understanding users' mental models. Clarifying the vision and needs of the app at an early stage between researchers, software developers, and end users is critical for the success of the app and staying within the budget, which has also been noted by others [38]. Having an additional budget for feature requests may increase users' satisfaction with the app and potentially increase use.

The higher number of uses of the meetings, surveys, and chat functions by the intervention group may have been because of encouragement by the coach during regularly scheduled meetings, whereas the control groups did not receive the same level of in-person encouragement or support with technical issues. These results are in line with another study's findings that social influences from colleagues, employers, and health care professionals can exert a strong effect on intention to use a personal health record app in a workplace setting [4].

However, the fit between technical products and the user audience must also be considered, as noted by one end user. All apps must consider the work context, culture, and characteristics of the intended user population [15,39]. When intended for a specific occupational setting, the range of educational levels and experience with new technology may vary greatly among job groups and require strategic choices about whom to design for.

The biggest strength of this study is that our app was uniquely built to document the process outcomes of a workplace change study. The findings and app evaluations from this study provide information on the usefulness of mHealth apps as a data collection method for other researchers conducting workplace interventions.

Testing iteratively was a strength of this study, as it helped the developer pinpoint problem areas, debug across platforms, and inform each stage of development. This resulted in a more refined app for our users during deployment and prevented potential loss of data. The documentation of this iterative process fills a gap noted by others that more usability studies focused on user engagement and product interaction are needed [11].

The use of the SUS is also one of its strengths, as it is a validated instrument for assessing usability, and when combined with the task scenarios, it provided qualitative feedback from users as well. This mixed methods approach provided multidimensional information to customize the app for both the researchers' and target users' needs. Future studies looking to develop an app with a similar purpose will be able to build upon what we have done and avoid potential pitfalls that may result in substantial project delays.

Limitations

One weakness of our study is that the small end user sample limited the ability to stratify by demographics, site, or other variables of interest.

The use of Google Analytics, although useful as another measure of app use, did not capture some user visits. We also did not ask specific questions on reasons for adoption and reasons for attrition, which might have provided additional information on why some users dropped out early and some dropped out later. However, there seemed to be a substantial shift in app users opting for paper surveys after some fixes took longer than expected. This delay was because of a change in the developers hired for this research project to maintain the app, which required onboarding time. Although not covered in this paper, this study depended heavily on the paper duplications of the app functions not only for nonapp users but also for when the app encountered issues, and this should be expected for the development of apps for assisting with data collection in a workplace intervention study. Future work will involve the analysis of the process data that were collected through the app for the SHIFT study.

Conclusions

End users deemed our process evaluation mobile app to be of acceptable usability, thanks to the student testers identifying a number of bugs and errors that could be fixed before deployment

to our study population. Researchers looking to develop an app for a similar purpose would benefit from early and iterative user testing. Understanding user standards for a usable app and budgeting to keep up with the pace of other apps' features could improve overall satisfaction and acceptability.

Acknowledgments

This research was funded by grant number U19-OH008857 from the National Institute of Occupational Safety and Health, Centers for Disease Control and Prevention. The contents of this manuscript are solely the responsibility of the authors and do not necessarily represent the official views of the National Institute of Occupational Safety and Health. The research team thanks the students and the end user volunteers who participated in this study. The authors also thank Serena Rice, Cesar Morocho, and Merve Armagan for assisting with the end user app introductions and troubleshooting on site.

Authors' Contributions

The SHIFT Research Team includes Winnie Szu Yun Chin, ScD, MS; Alicia Kurowski, ScD; Rebecca Gore, PhD; Laura Punnett, ScD; Serena Rice, MS; Suzanne Nobrega, MS; Cesar Morocho, MPH; Merve Armagan, MS; Yuan Zhang, PhD, RN; Mazen El Ghaziri, PhD, MPH, RN; Sundus Siddique, MBBS, MPH; Sandy Sun, MBA; and Mumtahana Nabi, MPH.

Conflicts of Interest

All authors were involved in the development of the mobile app that has been evaluated in this study.

References

1. Butterfoss FD. Process evaluation for community participation. Annu Rev Public Health 2006;27:323-340. [doi: 10.1146/annurev.publhealth.27.021405.102207] [Medline: 16533120]
2. Steckler A, Linnan L. Process Evaluation for Public Health Interventions and Research. Hoboken, New Jersey, United States: Wiley; 2002:1-432.
3. Department of Health and Human Services. Mobile health: technology and outcomes in low and middle income countries (R21). National Institutes of Health. 2013. URL: https://grants.nih.gov/grants/guide/pa-files/PAR-14-028.html [accessed 2020-05-20]
4. Park HS, Kim KI, Soh JY, Hyun YH, Jang SK, Lee S, et al. Factors influencing acceptance of personal health record apps for workplace health promotion: cross-sectional questionnaire study. JMIR Mhealth Uhealth 2020 Jun 04;8(6):e16723 [FREE Full text] [doi: 10.2196/16723] [Medline: 32496202]
5. Jimenez P, Bregenzer A. Integration of eHealth tools in the process of workplace health promotion: proposal for design and implementation. J Med Internet Res 2018 Feb 23;20(2):e65 [FREE Full text] [doi: 10.2196/jmir.8769] [Medline: 29475828]
6. Mattila E, Orsama A, Ahtinen A, Hopsu L, Leino T, Korhonen I. Personal health technologies in employee health promotion: usage activity, usefulness, and health-related outcomes in a 1-year randomized controlled trial. JMIR Mhealth Uhealth 2013;1(2):e16 [FREE Full text] [doi: 10.2196/mhealth.2557] [Medline: 25098385]
7. Jaworski B, Kohli AK, Sahay A. Market-driven versus driving markets. J Acad Mark Sci 2000 Jan 01;28(1):45-54. [doi: 10.1177/0092070300281005]
8. Ericsson K, Hoffman R, Kozbelt A, Williams A. The Cambridge Handbook of Expertise and Expert Performance. Cambridge: Cambridge University Press; 2018:1-984.
9. User-centered design basics. U.S. Department of Health & Human Services. URL: https://www.usability.gov/what-and-why/user-centered-design.html [accessed 2021-05-14]
10. Usability testing. U.S. Department of Health & Human Services. URL: https://www.usability.gov/how-to-and-tools/methods/usability-testing.html [accessed 2021-05-14]
11. Georgsson M, Staggers N. Quantifying usability: an evaluation of a diabetes mHealth system on effectiveness, efficiency, and satisfaction metrics with associated user characteristics. J Am Med Inform Assoc 2016 Jan;23(1):5-11 [FREE Full text] [doi: 10.1093/jamia/ocv099] [Medline: 26377990]
12. Hilliard ME, Hahn A, Ridge AK, Eakin MN, Riekert KA. User preferences and design recommendations for an mhealth app to promote cystic fibrosis self-management. JMIR Mhealth Uhealth 2014;2(4):e44 [FREE Full text] [doi: 10.2196/mhealth.3599] [Medline: 25344616]
13. Hudson W. Reduced empathizing skills increase challenges for user-centered design. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2009 Presented at: CHI '09: CHI Conference on Human Factors in Computing Systems; April, 2009; Boston MA USA p. 1327-1330. [doi: 10.1145/1518701.1518901]
14. Card sorting. U.S. Department of Health & Human Services. URL: https://www.usability.gov/how-to-and-tools/methods/card-sorting.html [accessed 2021-05-14]

15. Brown M, Hooper N, Eslambolchilar P, John A. Development of a web-based acceptance and commitment therapy intervention to support lifestyle behavior change and well-being in health care staff: participatory design study. JMIR Form Res 2020 Dec 30;4(11):e22507 [FREE Full text] [doi: 10.2196/22507] [Medline: 33252350]
16. Zhang D, Adipat B. Challenges, methodologies, and issues in the usability testing of mobile applications. Int J Hum-Comput Int 2005 Jul;18(3):293-308. [doi: 10.1207/s15327590ijhc1803_3]
17. Moran K. Usability testing 101. Nielsen Norman Group. 2019. URL: https://www.nngroup.com/articles/usability-testing-101/ [accessed 2021-05-05]
18. Hirschheim R. Assessing participative systems design: some conclusions from an exploratory study. Inf Manag 1983 Jan;6(6):317-327 [FREE Full text] [doi: 10.1016/0378-7206(83)90040-x]
19. Pilemalm S, Timpka T. Third generation participatory design in health informatics - making user participation applicable to large-scale information system projects. J Biomed Inform 2008 Apr;41(2):327-339 [FREE Full text] [doi: 10.1016/j.jbi.2007.09.004] [Medline: 17981514]
20. Reporting usability test results. U.S. Department of Health & Human Services. URL: https://www.usability.gov/how-to-and-tools/methods/reporting-usability-test-results.html [accessed 2021-05-05]
21. Zapata BC, Fernández-Alemán JL, Idri A, Toval A. Empirical studies on usability of mHealth apps: a systematic literature review. J Med Syst 2015 Feb;39(2):1. [doi: 10.1007/s10916-014-0182-2] [Medline: 25600193]
22. Ma X, Yan B, Chen G, Zhang C, Huang K, Drury J. A toolkit for usability testing of mobile applications. In: Mobile Computing, Applications, and Services. Berlin, Heidelberg: Springer; 2012:226-245.
23. Nielsen J, Landauer T. A mathematical model of the finding of usability problems. In: Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems. 1993 Presented at: INTERCHI93: Conference on Human Factors in Computing; April, 1993; Amsterdam, The Netherlands p. 206-213. [doi: 10.1145/169059.169166]
24. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of web-based and mobile health interventions. J Med Internet Res 2011;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
25. Goyal S, Morita PP, Picton P, Seto E, Zbib A, Cafazzo JA. JMIR Mhealth Uhealth 2016;4(1):e32 [FREE Full text] [doi: 10.2196/mhealth.4730] [Medline: 27012937]
26. De Vito Dabbs A, Myers BA, Mc Curry KR, Dunbar-Jacob J, Hawkins RP, Begey A, et al. User-centered design and interactive health technologies for patients. Comput Inform Nurs 2009;27(3):175-183 [FREE Full text] [doi: 10.1097/NCN.0b013e31819f7c7c] [Medline: 19411947]
27. Boudreaux ED, Fischer AC, Haskins BL, Zafar ZS, Chen G, Chinai SA. Implementation of a computerized screening inventory: improved usability through iterative testing and modification. JMIR Hum Factors 2016 Mar 09;3(1):e10 [FREE Full text] [doi: 10.2196/humanfactors.4896] [Medline: 27025766]
28. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
29. Bender MS, Choi J, Arai S, Paul SM, Gonzalez P, Fukuoka Y. Digital technology ownership, usage, and factors predicting downloading health apps among Caucasian, Filipino, Korean, and Latino Americans: The digital link to health survey. JMIR Mhealth Uhealth 2014;2(4):e43 [FREE Full text] [doi: 10.2196/mhealth.3710] [Medline: 25339246]
30. Punnett L, Nobrega S, Zhang Y, Rice S, Gore R, Kurowski A, SHIFT Project Research Team. Safety and Health through Integrated, Facilitated Teams (SHIFT): stepped-wedge protocol for prospective, mixed-methods evaluation of the Healthy Workplace Participatory Program. BMC Public Health 2020 Oct 29;20(1):1463 [FREE Full text] [doi: 10.1186/s12889-020-09551-2] [Medline: 32993607]
31. Robertson M, Henning R, Warren N, Nobrega S, Dove-Steinkamp M, Tibirica L, CPH-NEW Research Team. The Intervention Design and Analysis Scorecard: a planning tool for participatory design of integrated health and safety interventions in the workplace. J Occup Environ Med 2013 Dec;55(12 Suppl):86-88. [doi: 10.1097/JOM.0000000000000036] [Medline: 24284761]
32. Chin W, Kurowski A, Chen G, Gore R, Punnett L. Enhancing the usability of a mobile app for process evaluation in a participatory ergonomics healthcare intervention. In: Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018). Switzerland: Springer; 2019:523-530.
33. Dumas J, Redish J. Practical Guide to Usability Testing. Chicago, IL, U.S.A: University of Chicago Press; 1999:1-404.
34. Brooke J. SUS: a 'Quick and Dirty' usability scale. In: Usability Evaluation In Industry. Florida: CRC Press; 1996:4-7.
35. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009;4(3):114-123.
36. Sauro J. A Practical Guide to the System Usability Scale: Background, Benchmarks & Best Practices. Scotts Valley, California, US: CreateSpace Independent Publishing Platform; 2011:1-162.
37. Nielsen J. Heuristic Evaluation. New York, NY: John Wiley & Sons; 1994:1-471.
38. Roth WR, Vilardaga R, Wolfe N, Bricker JB, McDonell MG. Practical considerations in the design and development of smartphone apps for behavior change. J Contextual Behav Sci 2014 Oct 1;3(4):269-272 [FREE Full text] [doi: 10.1016/j.jcbs.2014.08.006] [Medline: 25485231]

39. Muuraiskangas S, Harjumaa M, Kaipainen K, Ermes M. Process and effects evaluation of a digital mental health intervention targeted at improving occupational well-being: lessons from an intervention study with failed adoption. JMIR Mental Health 2016 May 11;3(2):e13 [FREE Full text] [doi: 10.2196/mental.4465] [Medline: 27170553]

Abbreviations

DT: Design Team
HWPP: Healthy Workplace Participatory Program
mHealth: mobile health
SC: Steering Committee
SHIFT: Safety and Health through Integrated, Facilitated Teams
SUS: System Usability Scale

Edited by G Eysenbach; submitted 03.07.20; peer-reviewed by M Nitsch, B Zapata; comments to author 27.07.20; revised version received 31.03.21; accepted 27.04.21; published 28.10.21.

Please cite as:
Chin WSY, Kurowski A, Gore R, Chen G, Punnett L, SHIFT Research Team
Use of a Mobile App for the Process Evaluation of an Intervention in Health Care: Development and Usability Study
JMIR Form Res 2021;5(10):e20739
URL: https://formative.jmir.org/2021/10/e20739
doi:10.2196/20739
PMID:34709186

Original Paper

WhatsApp-Based Focus Groups Among Mexican-Origin Women in Zika Risk Area: Feasibility, Acceptability, and Data Quality

Elizabeth Anderson[1,2], MPH, PhD; Mary Koss[1], PhD; Ana Lucía Castro Luque[3], PhD; David Garcia[1], PhD; Elise Lopez[1], MPH, DrPH; Kacey Ernst[4], MPH, PhD

[1]Department of Health Promotion Sciences, University of Arizona, Tucson, AZ, United States

[2]International Center for Research on Women, Washington, DC, United States

[3]El Colegio de Sonora, Hermosillo, Mexico

[4]Department of Epidemiology and Biostatistics, University of Arizona, Tucson, AZ, United States

Corresponding Author:
Elizabeth Anderson, MPH, PhD
Department of Health Promotion Sciences
University of Arizona
1295 N Martin Ave
Tucson, AZ, 85721
United States
Phone: 1 5205050040
Email: andersone@email.arizona.edu

Abstract

Background: Despite unprecedented advances in worldwide access to the internet via smartphones, barriers to engaging hard-to-reach populations remain in many methods of health research. A potential avenue for conducting qualitative research is via participatory web-based media, including the free, popular social platform WhatsApp. However, despite the clear advantages of engaging with participants over a well-established web-based platform, logistical challenges remain.

Objective: This study aims to report evidence on the feasibility and acceptability of WhatsApp as a method to conduct focus groups.

Methods: A pilot focus group was conducted with Spanish-speaking women near the US–Mexico border. The content focus was knowledge and perceived risks for exposure to the Zika virus during pregnancy.

Results: Evidence was obtained regarding WhatsApp as a low-cost, logistically feasible methodology that resulted in rich qualitative data from a population that is often reticent to engage in traditional research. A total of 5 participants participated in a focus group, of whom all 5 consistently contributed to the focus group chat in WhatsApp, which was conducted over 3 consecutive days.

Conclusions: The findings are noteworthy at a time when face-to-face focus groups, the gold standard, are risky or precluded by safe COVID-19 guidelines. Other implications include more applications and evaluations of WhatsApp for delivering one-on-one or group health education interventions on sensitive topics. This paper outlines the key steps and considerations for the replication or adaptation of methods.

(JMIR Form Res 2021;5(10):e20970) doi:10.2196/20970

KEYWORDS

WhatsApp; synchronous text-based focus groups; Zika; Mexican-origin Latinas; social media; mHealth; focus groups; smartphones; mobile phone

Introduction

Background

Web-based focus groups are being increasingly used in health research to facilitate or expedite access to hard-to-reach respondents [1]. Compared with traditional, in-person focus groups, web-based platforms, specifically smartphone-based platforms, have lower costs, allow flexible time for participant responses, better protect participant confidentiality, and may increase acceptability in some populations of interest [2]. However, many populations with internet access remain hard

to reach for health behavior research, in part because researchers do not use established platforms in which the population is already literate, resulting in limited acceptability [3]. WhatsApp is a chat-based communication app used widely across the globe for one-on-one and large group conversations.

Web- and chat-based focus groups provide rich qualitative data comparable with those collected in traditional in-person focus groups [4], promote more uniform participation rates [5,6], and increase disclosure of personal views, presumably because of greater anonymity than face-to-face methods [7]. Young populations, in particular, increasingly prefer to express themselves using text-based methods sent from their phones [8]. Furthermore, SMS text message–based data collection may even result in more accurate reporting of sexual and other health behaviors than paper-based or voice formats [9]. Social media norms of emotional expression can be qualitatively documented, which mitigates the loss of nonverbal information that would be observed in in-person or video-based focus groups [10-12]. Differences between web-based and in-person focus groups are being eroded as new technologies improve group interaction and the quality of information offered by participants [4]. Despite evidence for this phenomenon in the literature on web-based communication [13], chat-based methods for qualitative data have not been updated accordingly [14,15].

Distinct subgroups of frequent web-based chat users, including young Spanish-speaking Latina women in the United States, may be more likely to participate in research via their preferred web-based medium [13]. For a pilot test of WhatsApp as a focus group platform, we recruited Latina women in a Zika virus risk area (southern Arizona) to assess their knowledge of Zika virus infection in women who are pregnant or may become pregnant. Secondarily, we queried participants' preferences for receiving health messages on the web. Mexican-origin women in southern Arizona are often difficult to engage with in research because of distrust, population transience related to migratory work patterns, and fears of immigration-related surveillance [16]. High rates of smartphone ownership and WhatsApp use (which is used by 46% of Mexicans every month [17] in lieu of conventional SMS text messaging), especially among younger people, make this group an ideal target population for testing WhatsApp for qualitative research. In addition, wide-reaching cell networks and Wi-Fi coverage in the United States reduce accessibility issues for the purpose of pilot testing.

WhatsApp is a free, globally prevalent mobile app that contributes an estimated 20% of the total time spent on smartphones [18] and allows free instant messaging to individuals or social network groups. It is used less commonly in the United States, where conventional SMS text messaging is more accessible, as are similar app-based group chat platforms, such as Facebook Messenger, GroupMe, and Viber. These apps have variable popularity among age groups, and preferences shift over time. Smartphone users consistently use multiple chat apps with equivalent functionality in idiosyncratic ways, indicating that preferred communication environments can be leveraged for different communication purposes [19]. Interactions on WhatsApp with strangers in special interest groups are commonly accepted [20] and are a key component of social networking [21,22]. Thus, using the platform to engage Mexican-origin participants in focus groups is a potential and highly acceptable method of health data collection. As desired participants likely spend a significant amount of time on WhatsApp regularly (if not daily) already, incorporating a focus group into the platform minimizes participant burden and encourages ongoing participation over several hours or days.

WhatsApp has many other technical and logistical benefits as a qualitative data collection medium. The platform is *end-to-end encrypted*, meaning that a third party cannot decrypt a message even if they are able to access shared data, which is ethically essential to both researchers and participants. The relative anonymity in a chat group with strangers likely increases willingness to discuss sensitive or embarrassing topics; participants can additionally send immediate, direct messages to a group moderator if they want to share a thought but are uncomfortable sending it to the whole group. Users can express themselves with a variety of media ranging from emojis to multiple languages to photos, small image files (eg, gifs), and videos.

Objective

Despite the potential for using WhatsApp in data collection in a variety of hard-to-reach populations, security and logistical challenges remain to be explored and documented. The purpose of this study is to pilot test the WhatsApp platform as a method of conducting focus groups with Spanish-speaking Latinas and explore its logistical feasibility for broader use. WhatsApp is being increasingly documented in the scientific literature as a useful tool [23-26]. We describe our methods in detail for replication or adaptation.

Methods

Context

Environmental conditions in southern Arizona are conducive to a future Zika virus outbreak driven by the rainy season *Aedes aegypti* mosquitos. The Mexican state of Sonora shares a long border with southern Arizona and had the highest number of Zika virus cases in Mexico in 2018 (n=349) [27]. The Zika virus is secondarily transmitted through sex; however, most public health responses have focused exclusively on mosquito-borne transmission, leaving a knowledge gap for women of childbearing age who are at most risk of negative outcomes if infected with the Zika virus. Latinas in the United States have high rates of unintended pregnancy [28], possibly related to low self-efficacy in safe sex negotiation [29] compared with non-Hispanic Whites and may be less equipped to prevent pregnancy to avoid the Zika virus or to use a condom once already pregnant. Therefore, we designed the present focus group guide and corresponding survey to qualitatively explore current Zika virus knowledge, fears about the Zika virus among women who may be pregnant during a future outbreak, and preferences for future public health communications related to Zika virus prevention.

Preparation for Implementation

Before the study began, in-depth key informant interviews based on a loose script were performed with 6 clinical and administrative members of 2 health care organizations to explore

the perceived need for and feasibility of using the WhatsApp platform for collecting information on Zika virus risk in antenatal Latina women. The 2 partnering health care organizations were located in southern Arizona and had prior research relationships with the academic team. Key informants reviewed study materials, including the consent process, proposed focus group questions, and provided feedback on the survey content. Questions were additionally informed by the current literature, as described in the context section above. Interviews were recorded, and the observing researcher (EA) took extensive notes during the interviews. Key informants confirmed that Zika virus awareness programs were ongoing at the target health care organizations, although they were mostly passive and focused on the prevention of mosquito-borne transmission (eg, promoted use of mosquito repellent) with little or no focus on sexual transmission. Key informants also recommended that in-clinic recruitment flyers be provided in English as well as Spanish.

An initial pilot test of the focus group script was conducted in English with graduate student volunteers to remediate any technical issues with WhatsApp. Volunteers additionally tested the enrollment, consent, and survey forms and provided feedback for confusing questions. Qualitative responses of the test focus group with graduate students (n=7) were not recorded except for notes on the process; student volunteers generally provided detailed, high-level commentary on Zika risk as all were health researchers themselves. Bilingual student volunteers offered instrumental feedback on the face validity and quality of interpretation of the focus group questions, enrollment forms, and consent documentation.

Pilot Study Design

The initial pilot study design involved a multistep recruitment, enrollment, and data collection process. The 2 partnering health care organizations that provided interviews with key informants distributed flyers to Spanish-speaking women of childbearing age seeking care at primary health clinics, prenatal clinics, and Women, Infants, and Children clinics in a region of southern Arizona that is predominantly home to people of Mexican origin. Inclusion criteria were the following: Latina, aged >18 years, and either pregnant or intending to become pregnant in the next year. The flyers included a general study overview, information about compensation for participation, and the WhatsApp contact number of the study coordinator (Figure 1). Clinic staff were briefed on the purpose and methods of the pilot study so that they could answer questions about participation. Approximately 100 flyers were distributed, and the distribution was not limited to women who met the inclusion criteria. Although WhatsApp focus groups may generally be advertisable via social media, this approach would have been unacceptable to the population of interest because of distrust in research as described by the key informants; the partnering clinics and the research projects they endorse are perceived as trustworthy by the population they serve. This study was approved by the University of Arizona institutional review board.

Figure 1. Recruitment flyer for a pilot test of WhatsApp as a focus group platform.

Enrollment

After the potential participants sent a WhatsApp message to the study coordinator, the coordinator responded with a link to a screening, enrollment, and consent tool on REDCap (Research Electronic Data Capture) hosted at the University of Arizona [30]. REDCap is highly adaptive to all mobile platforms and can be formatted to allow users to toggle between multiple languages on demand; it additionally allows users to save their progress and return to forms at their convenience. Participants were encouraged to consult the study coordinator via WhatsApp for assistance with the enrollment process. Participants entered their email addresses to receive a digital copy of the consent information as well as to receive a digital gift card for their participation later.

Ensuring Participant Privacy

After successful enrollment, the study coordinator messaged participants to individually counsel about the identifying information that would be collected (ie, phone number and email address). Although the information was included in the consent form, participants were again reminded that other focus group members would be able to see their phone numbers, profile pictures, and public status messages (often consisting of quotes, emojis, personalized messages, or some combination thereof). Participants were reminded to consider the privacy of other participants (ie, they were asked not to take screenshots of group content or contact other group members directly and were asked to report immediately if a phone was lost or stolen while the study was ongoing). Participants were advised to change their profile pictures and not include images of their faces and were offered personalized, step-by-step advice on how to create an anonymous WhatsApp account not linked to their phone numbers (Multimedia Appendix 1 includes instructions for both Android and iOS systems). The study coordinator additionally provided an estimated timeline until the study would begin.

Focus Group Execution

Focus groups were designed to include between 5 and 7 enrolled participants who selected the same preferred primary language (ie, Spanish or English). A natively bilingual group moderator with graduate-level training in health promotion (DG) was enlisted to conduct the groups. Both the group moderator and the study coordinator communicated with study participants using anonymous WhatsApp accounts to avoid accidental positive identification of study participants by linking to their web-based identities (eg, Facebook's *people you may know* feature, which links user data such as phone numbers that an individual has contacted to suggest new Facebook friends). The study coordinator created a messaging group within WhatsApp, which included participants, the moderator, and herself.

The group moderator interacted with the participants using a prepared script and a set of primary questions. Given the small sample size, the moderator focused on effective elicitation of the prespecified themes (discussed below) to maximize the utility of the resulting data. The study coordinator followed along and took notes but did not contribute to the group. The group moderator and study coordinator were able to send direct, private messages to one another while the focus group was proceeding, which permitted real-time diagnosis of logistical challenges as well as the ability to discuss key follow-ups and probing questions. Similarly, participants were encouraged to send direct, private messages to the group moderator if they were uncomfortable sharing a thought with the broader group.

Focus Group Themes

The prepared focus group questions were organized into three themes (general Zika virus knowledge, knowledge about sexual transmission and attitudes toward avoiding sexual transmission, and preferences for internet and WhatsApp use for health messaging), which were delivered (one theme per day) over 3 consecutive weekdays to avoid exhausting the participants. Each day, the start time varied slightly but began in the morning. The purpose of this design was to maximize the time available for each topic, as we anticipated that not all participants would spend time on the app equally. Researchers were able to see which group members had opened any given message, a timestamp for when that message was opened, and when they were last active on WhatsApp. This created data monitoring points for the group moderator to make informed judgments on whether and when to send additional follow-up probes if participants were unresponsive to a given question. At the end of the question set for each day, the group moderator informed participants that data collection for the day was finished but that participants could continue sending messages if they wished and additionally primed participants for the next day's theme.

Concluding the Focus Group

After the third day of focus group data collection, the group moderator informed participants that they would receive their gift cards by email. Participants were again encouraged to contact the study coordinator with any questions or concerns and were informed that they would be blocked from the group at the end of the study in order to protect the privacy of fellow participants. The study coordinator sent each participant a private message that consisted of a series of Zika virus infographics in either Spanish or English that were published on the web by the Centers for Disease Control and Prevention (Figure 2). The purpose of this follow-up was to ensure that any misinformation provided by other group members (eg, reported an incorrect Zika virus transmission method, such as by food) would be dispelled.

Figure 2. Debrief messages shared as images with participants in a pilot WhatsApp focus group. Source: Centers for Disease Control and Prevention.

Data Processing and Analysis

All data from the focus group were exported from WhatsApp as a text file. To deidentify the data, the study coordinator saved each participant with a code name (eg, FG1 respondent-A) as a phone contact rather than the phone number that was used. Each message was indicated with a timestamp as well as the *name* of the message writer. The file type additionally permitted any emojis to be preserved with no interprogram loss of the image. Any nontext files (eg, voice recordings, screenshots, or shared photos) were additionally included in this export. WhatsApp exports these deidentified files to a variety of web-based locations (eg, Google Drive or Dropbox, or sends them to an email address).

Two researchers (EA and DG) reviewed the transcripts jointly, discussed emergent themes, and agreed upon translations to English. Given the pilot nature of this project as a test of platform feasibility (as well as the small number of participants), the transcript was not systematically coded, and interpretation was limited to a simple thematic analysis with exemplary quotes [31]. We have reported quotes both as originally shared in Spanish as well as their English translations. The reporting of specific results was limited to general observations to protect participant privacy in such a small study.

Corresponding Survey Data Collection

Surveys on demographics, mosquito knowledge, media and technology use, and sexual relationship power were conducted using REDCap to inform the interpretation of focus group data and produce complementary sources of data. Questions were adapted and abridged where appropriate from previous literature to assess the relationship, if any, between focus group responses to similar topic areas and quantifiable knowledge and ability to avoid Zika virus transmission either via mosquitos or from a sexual partner [32], as well as internet use habits [33].

Alterations in Response to Feedback

After the initial distribution of flyers to the participating clinics, the flyers were revised based on feedback from clinic staff who were promoting the study to patients. Potential enrollees did not initially understand how the study was to be conducted as they had only heard of in-person focus groups. Clinic patients also expressed hesitation about their privacy in the study. Therefore, the flyer was revised to include more information about how the entire study would take place on WhatsApp as well as a more explicit assurance of participant privacy.

Additional changes in the planned methods were made to address key informant feedback. Although we intended to collect survey data at the time of enrollment to streamline the time spent on the REDCap platform, data were collected after the focus groups instead. Our key informants suggested that our intended participants were generally hesitant to share information with perceived authorities because of immigration- and documentation-related fears for themselves or family members. We correspondingly moved the surveys to the end of the process and split the proffered compensation to indicate that responses to the questionnaires were optional (ie, participants received a US $10 gift card for participating in the focus group and an additional US $5 gift card if they selected to complete the REDCap questionnaires, rather than a lump sum of US $15). The questionnaires were linked by REDCap to the participants' enrollment information by their email addresses.

Results

Focus Groups

Of the 7 potential participants who responded to the study coordinator for participation, 5 (71%) were included in a Spanish-language focus group. One enrollee was excluded as she was the only respondent who preferred to participate in English, and one enrollee was excluded as she joined after the first focus group was performed and no additional participants were identified. The recruitment period (September to December) corresponded with the end of the farm work season, meaning that the number of individuals seeking care at the clinic dropped off dramatically shortly after enrollment began; therefore, we terminated enrollment after one focus group was completed at the suggestion of the partnering clinic staff. None of the participants opted to use an anonymous WhatsApp account after receiving the direct message outlining the potential risks of using her personal WhatsApp number.

Respondents consistently participated in the focus group for 3 days (ie, all participants shared at least 2 distinct responses per day [range 2-8], either in response to a question from the moderator or a comment from a fellow participant). The frequency, comprehensiveness, and timing of responses to questions appeared to be highest in the first 2 hours after the group moderator began the daily session. Every respondent provided at least one response per day; the longest delay between the daily initiation and the slowest participant's response was approximately 2.5 hours. On average, of the 5 respondents, 2 (40%) answered any given question. Respondents who began responding later in the day often began by responding directly to the moderator's earlier questions and did not feel compelled to skip ahead to the questions that were currently being discussed by other group members. Questions in messages that were opened but did not result in responses were asked again by the group moderator later in the conversation. The overall tone of the conversation was casual, as evidenced by the continued use of slang and internet and text abbreviations commonly used by Mexican-origin Spanish speakers. Approximately 40% (2/5) of participants elaborated on answers using emojis such as these  to express fear (ie, worry about exposure to the Zika virus through mosquito bites) or these  to communicate the awkwardness explaining why men do not like using condoms.

The median length of responses to focus group prompts was 14 words (range 4-66, IQR 9–22). Responses were concise and demonstrated a clear understanding of the prompt; no responses strayed from the topic of a given prompt, including the longest replies. Evidence that respondents were expressing complex and thoughtful replies was observed where separate ideas or clauses related to the same prompt were shared in back-to-back messages (up to 3 in a row) by the same respondent.

Corroborating Survey Data

Overview

The external survey data shared by individual participants informed the interpretation of focus group responses as the data were linked to respondents by email address. Of the 5 participants, 4 (80%) responded to the optional surveys, indicating that the additional burden of completing the web-based click-through survey was likely not overwhelming for the population. As the sample size was very small, we summarized the most relevant demographic and response trends rather than reporting which survey responses were associated with individual participants: of the 4 respondents, 3 (75%) were currently pregnant, and 3 (75%) had at least 1 previous live birth; all 4 (100%) were married and living with their spouses; 2 (50%) participants indicated that they searched the internet *all the time* for information, including health information such as symptom searches, whereas one-third of participants reported completing such searches several times a week; 3 (75%) participants said they used social media several times a day, though no one said they knew or regularly interacted with strangers on the web; 2 (75%) respondents indicated that they had lower sexual relationship power than their primary male partners, with the other 2 (75%) indicating equal power with their partners. The median age was 29 years; all respondents had finished at least secondary school (high school), and only 1 was currently working.

Below are several examples that illustrate the capacity of WhatsApp focus groups to generate meaningful responses.

Theme 1: Zika Virus Knowledge

Focus group participants were in agreement that the Zika virus can influence babies born to infected women ("it's a virus transmitted by mosquitos and unfortunately mostly affects the baby whose blood gets infected with the virus") but had mixed knowledge as to what the effects could be (eg, Zika virus "causes paralysis in babies"). Participants knew about mosquito-borne Zika virus transmission, but none had heard of sexual transmission of the virus. There was a general perception that everyone in their southern Arizona community was worried, given that mosquito bite exposure is very common locally:

> *Respondent C: A mi en lo personal si me preocupa mucho. Creo q a todos. Por eso trato de siempre usar manga larga y lantalom [sic] aparte q siempre llevo repelente de mosquitos en mi bolso.*
>
> *Respondent C: *pantalon*
>
> *Respondent C: For me personally, yes I'm very worried. I think that's true for everyone. Therefore, I always try to wear long sleeves and lants [sic] and additionally, I always carry mosquito repellent in my bag.*
>
> *Respondent C: *pants*
>
> *Respondent A: Yo tampoco miro moskitos dentro d mi hogar y como casi no salgo evitó mucho xk vivo a un lado d un parke y evitamos dejar aguas en botes o estancadas x lo mismo k no se junten mas moscos*
>
> *Respondent A: I also don't see mosquitos in my house and since I hardly go out b/c I live next to a park, and we avoid leaving water in containers or stagnating so flies don't gather*

Participants reported using repellents and cleaning up standing water as their primary methods of avoiding exposure to the virus during pregnancy.

Theme 2: Sexual and Reproductive Health

When asked about hypothetical medical recommendations to delay becoming pregnant because of Zika virus risk, respondents were hesitant to say that they would be willing to do so for an indefinite period of time:

> *Moderator: En algunos países los doctores recomiendan que no queden embarazadas porque no hay tratamiento para el zika, que harían si su doctor les dijera que eviten quedar embarazadas?*
>
> *Moderator: In some countries doctors recommend not becoming pregnant because there is no treatment for Zika, what would you do if your doctor told you to avoid becoming pregnant?*
>
> *Respondent B: Si en verdad quisiera tener un bebé me cuidaría lo más posible de los mosquitos [...] y lo pensaría mucho para ver los pros y cons*
>
> *Respondent B: If I really wanted to have a baby I would take care of mosquitos as much as possible [...] and I would think about it a lot to see the pros and cons*
>
> *Moderator: Qué tal si dice que se esperen 6 meses?*
>
> *Moderator: What if they told you to wait 6 months?*
>
> *Respondent B: Entonces si me espero*
>
> *Respondent B: In that case, yes I'd wait*
>
> *Respondent E: Si el dr lo recomienda creo que debemos hacer caso ya que ellos son los que saben*
>
> *Respondent E: If the dr recommends it, I think we should pay attention as they are the ones who know*

Respondents said they had not spoken to their doctors about the Zika virus and that their Zika virus risk conversations with their primary sexual partners were exclusively about mosquito bites, not sexual transmission. When prompted about potentially starting condom use during pregnancy, perceived confidence was consistently high among respondents; 2 suggested that if their partners were reluctant to use a condom, they would show him photos of babies born with microcephaly after maternal Zika virus exposure to change his mind. Participants said they did not know any women whose partners would get mad or suspicious if asked to use a condom during pregnancy and additionally reported that there would be no barriers to getting condoms:

> *Respondent A: Creo k nada ps ala pareja ps no le gusta pero a yo pienso k si se trata d cuidarse y d salud eso sale sobrando primero la salud d ambos si esta embarazada ps mas d la d el bebé y si uno esta tratando d salir embarazada ps con musho mas consiencia y cuidarnos para si*
>
> *Respondent A: I don't think anything about a couple who wouldn't like it [to use condoms or other contraceptives], but I think it is about taking care of yourself and your health, that means health first, for both of you if you're pregnant, but mostly the health of the baby, and if you are trying to get pregnant to be more conscientious and take care of ourselves if we got pregnant, assuming everything goes well first*

Theme 3: Technology Use and Preferences

When discussing preferences for receiving Zika virus information from a health professional, most respondents indicated that they would prefer to have a doctor explain prevention methods in person. However, there were mixed responses as to whether some women they knew would prefer to get their information from the news, from friends, or from the internet. Participants commented on a shared concern of receiving false information over the internet and indicated a preference to speak to a health professional, although not necessarily in person:

> *Respondent A: Yo si miro o escucho d algo k se está escuchando musho o algo asi en mi siguiente sita selo comento ami doctora*
>
> *Respondent A: If I see or hear something a lot [on WhatsApp or other social media, about Zika] or something like that I just tell my doctor at my next appointment*

Discussion

Principal Findings

The pilot WhatsApp focus group consistently involved all participants over multiple days, elicited responses on sensitive topics, included participant interactions that mimicked those seen in traditional focus groups, and privately engaged a population of Spanish-speaking Latinas that is generally hesitant to participate in research. The use of a separate web-based survey to collect data on demographics, knowledge, and attitudes expanded the depth of information without high levels of attrition. Although the focus group provided initial insight into methodological best practices (Textbox 1), positive identification of participants (ie, confirmation that enrollees are who they say they are) remains a challenge. This may be of greater concern when topic areas are sensitive or when the population is more vulnerable to adverse events if their confidentiality is breached. WhatsApp shows initial promise as a focus group platform in a population that already uses the app regularly.

Textbox 1. Best practice takeaways from a pilot test of WhatsApp as a focus group platform.

- A smaller group (5-7 participants) worked well on the platform though other literature supports larger web-based groups (10-12 participants); more testing is needed.
- Having the study coordinator available to respond quickly to study inquiries and logistical questions about enrollment via WhatsApp was well received and frequently used by participants.
- Conducting the focus group over multiple days seemed to prevent participation fatigue among respondents, who were most responsive to questions early in the day compared with questions asked later in the day.
- Study staff should not link study information to their personal phone numbers or WhatsApp accounts as much as possible, as participants use their phone numbers with other web-based media (eg, Facebook), which could compromise their privacy.
- The burden of "data" use on the target population will likely vary but should be accounted for in the study design.
- It was possible to enroll in the study even if a potential participant did not have an email address; in one case, the study coordinator instructed the participant to enroll, then manually linked her to the survey at the end of the focus group and sent her gift card as a screenshot via WhatsApp, rather than by email.

Strengths and Weaknesses

In addition to the premise that a WhatsApp-based approach to focus group delivery could engage a reluctant population, we identified several advantages over the course of this pilot study (Textbox 2). For example, the total cost of the pilot group was limited to printing costs, researcher time, and the per-participant cost of gift cards (up to US $15 per person in this study); there were no costs for room rental, equipment rental, snacks, transportation, transcription, software purchases, or other components that are frequently incurred for in-person focus group discussions. However, the weaknesses we identified may merit additional consideration for future implementation on a larger scale or with more vulnerable populations. For example, when a single individual failed to respond to a prompt, it was not possible to distinguish between absence of response because of not having an opinion and not understanding the question. This problem, also present for in-person focus groups, could be mitigated by directly tagging a nonresponding participant in a follow-up prompt to check for understanding.

Although this pilot study spanned 3 days to concur with the 3 parent themes of inquiry, additional testing to identify the optimal length of a WhatsApp focus group is warranted and likely varies among populations and topics. As noted, themes organically overlapped with other days beyond their intended focus, although this would also be expected in a traditional in-person focus group. Overly vocal respondents occur in all types of focus group research; so, in a chat-based setting, it may be that the first voice heard is also the most dominant or that it is simply perceived as such because of the format. As noted, there was a delay of up to several minutes between the time multiple participants opened and read a question and the time the first reply was received, indicating that first repliers were also the most enthusiastic and corresponded with the dominant voices that emerge in traditional in-person focus groups. Chat-based conversation etiquette varies among platforms as well as among populations; therefore, ongoing observation of this phenomenon is warranted for future studies using WhatsApp as described in the present research.

Although we did not identify any safety concerns related to breaches of privacy in our pilot, some potential issues merit consideration if WhatsApp is used for focus groups in other populations. Participants themselves were tasked with determining their level of privacy in the study (eg, decision to use a profile photo of their face, allowing their phone numbers to be seen by other group members, and using the app to discuss sensitive topics on a mobile device that a nonparticipant could potentially access) as well as ensuring the privacy of other participants. The most personal disclosure of sensitive personal information in this study was the inclusion of questions about current intimate partner violence experiences; however, after participants submitted this survey, it was not possible for them (or their intimate partners) to reaccess the answers they had submitted. Although the sensitive topics discussed in this pilot did not likely create a meaningful risk of anticipated or unanticipated breaches of privacy or confidentiality, this may not be the case in smaller subpopulations where participants are more likely to know each other or where inappropriate sharing of information creates a danger to other participants.

Textbox 2. Pros and cons of WhatsApp as a focus group platform identified in a pilot study.

Pros

- It is possible to conduct focus groups without risking the health of the population as may happen in-person gatherings during increasing numbers of COVID-19 cases.
- It maintains continuity of research projects despite stay-at-home orders.
- It allows people from different geographic areas to be recruited to attend. Hard-to-reach populations, including those that are mobile, could be easier to include.
- There are low to no implementation costs.
- Participants do not need to travel, make childcare plans, or adjust their daily schedules to be active respondents to the focus group discussion.
- End-to-end encryption and no face-to-face contact increase the sense of security for those hesitant to engage in research.
- Participants already use WhatsApp and do not need to install a new program or adopt a new behavior in order to be successful, active participants.

Cons

- Participants are unfamiliar with the research approach, and additional recruiter training may be needed to ensure accurate study explanations before participants enroll.
- Interactions among the group may limit building upon others' answers, particularly for those that first comment; they may not go back to respond to new comments from participants that interact later in the day.
- Guarantee of privacy is limited to the extent of each participant's compliance with privacy guidelines, although whether this risk is greater than similar risks posed by in-person research depends on the subject area.
- Highly vulnerable individuals are least likely to own smartphones in many global populations and may be inadvertently excluded from samples.

Implications for Future Use

Recent research has increasingly identified both intentional and unintentional uses of WhatsApp for health communication [34]. Although most formal evaluations of WhatsApp groups in the literature revolve around its use among health care providers [35], a small randomized controlled trial found that moderated group text discussions in WhatsApp reduced smoking relapse compared with pamphlets alone [36]. However, WhatsApp has also been a medium for the spread of serious misinformation, including around Zika virus transmission (eg, rumors about government conspiracies [37,38]) and vaccine safety [39], indicating a gap in public health education coverage that could be positively leveraged by creating WhatsApp-specific information from demonstrably authoritative sources.

To our knowledge, our pilot study is the first to use WhatsApp for focus group discussions. The potential for expanding access to subpopulations is an important step in data collection. For some global populations, data use can be a prohibitive expense, especially where Wi-Fi is not common; however, WhatsApp uses very little *data* compared with other apps, which is partly why it has been so successful globally. Nevertheless, data use remains a significant consideration for study design, which can often be mitigated by reimbursing participants with phone credit or cash equivalents.

Future Directions

Beyond focus groups or one-on-one in-depth interviews, WhatsApp could be used to deliver structured educational information to target groups. For example, the population engaged in this pilot study may benefit from structured messages about Zika virus behavior, best practices to avoid mosquito bites, or ways to encourage a partner to consistently use a condom during pregnancy. However, participants in our WhatsApp focus group strongly preferred to receive Zika virus information directly from a health care provider, which may indicate a need to recruit trained community health workers (eg, *promotoras*) who can demonstrate their health authority to participants before intervention delivery.

Conclusions

This pilot focus group provides a template for using WhatsApp for focus group delivery, as well as initial evidence that WhatsApp is a feasible, low-cost medium for efficient qualitative data enumeration. Innovative methods for distance data collection are in high demand during COVID-19–related restrictions on in-person methods, and low availability of research funding may presage greater future use as well. Additional testing is needed with a wider range of populations and subject matter to broaden the understanding of the risks and benefits to both researchers and participants. Beyond focus groups, WhatsApp has strong potential for use in health promotion research and implementation among global populations with smartphone access, especially where health care professionals are involved.

A full translation of this paper to Spanish is available as Multimedia Appendix 2.

Acknowledgments

The authors gratefully acknowledge the contributions of Denisse Zavala Garcia in data collection.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Participant instructions for setting up an anonymous phone number in WhatsApp.
[DOCX File , 13 KB - formative_v5i10e20970_app1.docx]

Multimedia Appendix 2
Spanish-language version of the paper.
[DOCX File , 532 KB - formative_v5i10e20970_app2.docx]

References

1. Fox F, Morris M, Rumsey N. Doing synchronous online focus groups with young people: methodological reflections. Qual Health Res 2007 Apr;17(4):539-547. [doi: 10.1177/1049732306298754] [Medline: 17416707]
2. Reid DJ, Reid FJ. Online Focus Groups: An in-depth comparison of computer-mediated and conventional focus group discussions. Int J Mark Res 2018 Jan 30;47(2):131-162. [doi: 10.1177/147078530504700204]
3. Welch V, Petkovic J, Pardo J, Rader T, Tugwell P. Interactive social media interventions to promote health equity: an overview of reviews. Health Promot Chronic Dis Prev Can 2016 Apr;36(4):63-75 [FREE Full text] [doi: 10.24095/hpcdp.36.4.01] [Medline: 27077792]
4. Stewart DW, Shamdasani P. Online focus groups. J Advert 2016 Nov 16;46(1):48-60. [doi: 10.1080/00913367.2016.1252288]
5. Schneider SJ, Kerwin J, Frechtling J, Vivari BA. Characteristics of the discussion in online and face-to-face focus groups. Soc Sci Comput Rev 2016 Aug 18;20(1):31-42. [doi: 10.1177/089443930202000104]
6. Han J, Torok M, Gale N, Wong QJ, Werner-Seidler A, Hetrick SE, et al. Use of web conferencing technology for conducting online focus groups among young people with lived experience of suicidal thoughts: Mixed methods research. JMIR Ment Health 2019 Oct 04;6(10):e14191 [FREE Full text] [doi: 10.2196/14191] [Medline: 31588913]
7. Tates K, Zwaanswijk M, Otten R, van Dulmen S, Hoogerbrugge PM, Kamps WA, et al. Online focus groups as a tool to collect data in hard-to-include populations: Examples from paediatric oncology. BMC Med Res Methodol 2009 Mar 03;9(1):15 [FREE Full text] [doi: 10.1186/1471-2288-9-15] [Medline: 19257883]
8. Skierkowski D, Wood RM. To text or not to text? The importance of text messaging among college-aged youth. Comput Hum Behav 2012 Mar;28(2):744-756. [doi: 10.1016/j.chb.2011.11.023]
9. Lim MS, Sacks-Davis R, Aitken CK, Hocking JS, Hellard ME. Randomised controlled trial of paper, online and SMS diaries for collecting sexual behaviour information from young people. J Epidemiol Community Health 2010 Oct 18;64(10):885-889. [doi: 10.1136/jech.2008.085316] [Medline: 19767322]
10. Schober MF, Conrad FG, Antoun C, Ehlen P, Fail S, Hupp AL, et al. Precision and disclosure in text and voice interviews on smartphones. PLoS One 2015 Jun 10;10(6):e0128337 [FREE Full text] [doi: 10.1371/journal.pone.0128337] [Medline: 26060991]
11. Berard B. I second that emoji: The standards, structures, and social production of emoji. First Monday 2018 Sep 01;23(9). [doi: 10.5210/fm.v23i9.9381]
12. Neviarouskaya A, Prendinger H, Ishizuka M. Analysis of affect expressed through the evolving language of online communication. In: Proceedings of the 12th international conference on Intelligent user interfaces. 2007 Presented at: IUI07: 12th International Conference on Intelligent User Interfaces; January 28 - 31, 2007; Honolulu Hawaii USA p. 278-281. [doi: 10.1145/1216295.1216346]
13. Biedermann N. The use of Facebook for virtual asynchronous focus groups in qualitative research. Contemp Nurse 2018 Feb 19;54(1):26-34. [doi: 10.1080/10376178.2017.1386072] [Medline: 28975852]
14. Leavy P. The Oxford Handbook of Qualitative Research. USA: Oxford University Press; 2014.
15. Barratt M. The efficacy of interviewing young drug users through online chat. Drug Alcohol Rev 2012 Jun;31(4):566-572. [doi: 10.1111/j.1465-3362.2011.00399.x] [Medline: 22168668]
16. Diaz VA, North American Primary Care Research Group. Research involving Latino populations. Ann Fam Med 2005 Sep 01;3(5):470-471. [doi: 10.1370/afm.402] [Medline: 16189068]
17. Number of smartphone users in Mexico from 2015 to 2022. Statista. 2017. URL: https://www.statista.com/statistics/270970/number-of-smartphone-users-mexico/ [accessed 2021-09-13]
18. Montag C, Błaszkiewicz K, Sariyska R, Lachmann B, Andone I, Trendafilov B, et al. Smartphone usage in the 21st century: who is active on WhatsApp? BMC Res Notes 2015 Aug 04;8(1):331 [FREE Full text] [doi: 10.1186/s13104-015-1280-z] [Medline: 26238512]
19. Nouwens M, Griggio CF, Mackay W. Whatsapp is for family; Messenger is for friends" communication places in app ecosystems. In: Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems. 2017 Presented at: CHI '17: CHI Conference on Human Factors in Computing Systems; May 6 - 11, 2017; Denver Colorado USA p. 727-735. [doi: 10.1145/3025453.3025484]

20. Mudliar P, Rangaswamy N. Offline strangers, online friends: Bridging classroom gender segregation with whatsapp. In: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. 2015 Presented at: CHI '15: CHI Conference on Human Factors in Computing Systems; April 18 - 23, 2015; Seoul Republic of Korea p. 3799-3808. [doi: 10.1145/2702123.2702533]
21. Lin K, Lu H. Why people use social networking sites: An empirical study integrating network externalities and motivation theory. Comput Hum Behav 2011 May;27(3):1152-1161. [doi: 10.1016/j.chb.2010.12.009]
22. Coyle CL, Vaughn H. Social networking: Communication revolution or evolution? Bell Labs Tech J 2008 Aug 18;13(2):13-17. [doi: 10.1002/bltj.20298]
23. Blauvelt C, West M, Maxim L, Kasiya A, Dambula I, Kachila U, et al. Scaling up a health and nutrition hotline in Malawi: The benefits of multisectoral collaboration. Br Med J 2018 Dec 07;363:k4590 [FREE Full text] [doi: 10.1136/bmj.k4590] [Medline: 30530659]
24. Arroz JA, Candrinho BN, Mussambala F, Chande M, Mendis C, Dias S, et al. WhatsApp: A supplementary tool for improving bed nets universal coverage campaign in Mozambique. BMC Health Serv Res 2019 Feb 01;19(1):86-87 [FREE Full text] [doi: 10.1186/s12913-019-3929-0] [Medline: 30709376]
25. Jhaveri KD, Pascarelli B, Hasan A, Kozikowski A, Fishbane S, Pekmezaris R. 'WhatsApp'ening in nephrology training. Clin Kidney J 2020 Feb;13(1):8-13 [FREE Full text] [doi: 10.1093/ckj/sfz045] [Medline: 32082548]
26. Khan A, Usmani A, Khaliq T. WhatsApp an informal learning and communication tool: Perception of medical officers working in low resource setting. J Pak Med Assoc 2019:1-230. [doi: 10.5455/jpma.17802]
27. Fischer M. Update on the current status of the Zika virus epidemic. Centers for Disease Control and Prevention. 2019. URL: https://www.fda.gov/media/128251/download [accessed 2021-09-24]
28. Finer L, Kost K. Unintended pregnancy rates at the state level. Perspect Sex Reprod Health 2011 Jun;43(2):78-87. [doi: 10.1363/4307811] [Medline: 21651706]
29. Davila YR. Influence of abuse on condom negotiation among Mexican-American women involved in abusive relationships. J Assoc Nurses AIDS Care 2002 Nov;13(6):46-56. [doi: 10.1177/1055329002238025] [Medline: 12469543]
30. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
31. Corden A, Sainsbury R. Using Verbatim Quotations in Reporting Qualitative Social Research: Researchers' Views. York: The University of York; 2006.
32. Matsuda Y, McGrath JM, Jallo N. Use of the sexual relationship power scale in research: An integrative review. Hisp Hlth Care Int 2012 Dec 01;10(4):175-189. [doi: 10.1891/1540-4153.10.4.175]
33. Costa JJ, Matos AP, do Rosário Pinheiroa M, do Céu Salvado M, da Luz Vale-Diasa M, Zenha-Rela M. Evaluating use and attitudes towards social media and ICT for Portuguese youth: the MTUAS-PY scale. The European Proceedings of Social and Behavioural Sciences. 2016. URL: https://www.europeanproceedings.com/files/data/article/43/995/article_43_995_pdf_100.pdf [accessed 2021-09-15]
34. Boulos M, Giustini D, Wheeler S. Instagram and Whatsapp in health and healthcare: An overview. Future Internet 2016 Jul 26;8(3):37. [doi: 10.3390/fi8030037]
35. Mars M, Scott RE. WhatsApp in clinical practice: A literature review. Stud Health Technol Inform 2016;231:82-90. [Medline: 27782019]
36. Cheung YT, Chan CH, Lai CJ, Chan WF, Wang MP, Li HC, et al. Using Whatsapp and Facebook online social groups for smoking relapse prevention for recent quitters: A pilot pragmatic cluster randomized controlled trial. J Med Internet Res 2015 Oct 22;17(10):e238 [FREE Full text] [doi: 10.2196/jmir.4829] [Medline: 26494159]
37. Kou Y, Gui X, Chen Y, Pine K. Conspiracy talk on social media: Collective sensemaking during a public health crisis. In: Proceedings of the ACM on Human-Computer Interaction. 2017 Presented at: Association for Computing Machinery on Human-Computer Interaction; December 6, 2017; Ho Chi Minh City p. 1-21. [doi: 10.1145/3134696]
38. Vijaykumar S, Nowak G, Himelboim I, Jin Y. Managing social media rumors and misinformation during outbreaks. Am J Infect Control 2018 Jul;46(7):850. [doi: 10.1016/j.ajic.2018.03.014] [Medline: 29661633]
39. Palanisamy B, Gopichandran V, Kosalram K. Social capital, trust in health information, and acceptance of Measles–Rubella vaccination campaign in Tamil Nadu: A case–control study. J Postgrad Med 2018;64(4):212. [doi: 10.4103/jpgm.jpgm_249_17]

Abbreviations

REDCap: Research Electronic Data Capture

Edited by G Eysenbach; submitted 16.06.20; peer-reviewed by G Kernohan, A Trifan; comments to author 28.10.20; revised version received 23.11.20; accepted 31.05.21; published 28.10.21.

Please cite as:
Anderson E, Koss M, Castro Luque AL, Garcia D, Lopez E, Ernst K
WhatsApp-Based Focus Groups Among Mexican-Origin Women in Zika Risk Area: Feasibility, Acceptability, and Data Quality
JMIR Form Res 2021;5(10):e20970
URL: https://formative.jmir.org/2021/10/e20970
doi:10.2196/20970
PMID:34709185

Original Paper

Professional Care Experiences of Persons With Suicidal Ideation and Behavior: Model Development Based on a Qualitative Meta-Synthesis

Mareike Hechinger[1*], MScN, RN; André Fringer[1*], MScN, RN, PhD

Institute of Nursing, School of Health Professions, Zurich University of Applied Sciences, Winterthur, Switzerland

[*]all authors contributed equally

Corresponding Author:
André Fringer, MScN, RN, PhD
Institute of Nursing
School of Health Professions
Zurich University of Applied Sciences
Katharina-Sulzer-Platz 9
Winterthur, 8400
Switzerland
Phone: 41 58 934 64 79
Email: andre.fringer@zhaw.ch

Abstract

Background: Health care professionals (HCPs) are challenged in caring for persons with suicidal ideation or behavior. For affected persons, professional care is essential, and being interviewed about their experiences can be stressful. The experiences of persons ideating or attempting suicide are essential to designing eHealth products to support them in crises and provide continuous care.

Objective: This study aimed to synthesize published qualitative research about how persons with suicidal thoughts or behavior experience inpatient or outpatient care. A model will be derived from the meta-synthesis to guide HCPs in their work with affected persons and provide a thorough needs assessment for eHealth development.

Methods: A qualitative meta-synthesis was conducted using an inductive approach, as proposed by Sandelowski and Barroso. The inclusion criteria were studies in English and German that dealt with persons who ideated or attempted suicide. Relevant articles were identified by searching the PubMed and Cinahl databases and by hand searching relevant journals and reference lists. The findings of each study were analyzed using initial and axial coding, followed by selective coding. Finally, a conceptual model was derived.

Results: In total, 3170 articles were identified in the systematic literature search. Articles were screened independently by 2 researchers based on the eligibility criteria. Finally, 12 studies were included. The central phenomenon observed among persons ideating or attempting suicide is their process from feeling unanchored to feeling anchored in life again. During inpatient and outpatient care, they experience being dependent on the skills and attitudes of HCPs. While helpful skills and attitudes support persons ideating or attempting suicide to reach their feeling of being anchored in life again, adverse interactions are experienced negatively and might lead to prolonging or maintaining the feeling of being unanchored in life.

Conclusions: The study promotes a differentiated view of the experiences of persons ideating or attempting suicide. The derived conceptual model can guide HCPs in their work with affected persons to support affected persons during their recovery. Moreover, the conceptual model is useable as a springboard to develop eHealth solutions for crisis situations and long-term care.

(JMIR Form Res 2021;5(10):e27676) doi:10.2196/27676

KEYWORDS

nursing care; health care professionals; suicidal behavior; suicidal inclinations; suicidal ideation; inpatient; outpatient; eHealth; mHealth; mental health; suicide; stress

Introduction

Since 2000, the suicide rate worldwide has decreased. Nevertheless, in 2016, nearly 800,000 deaths were registered due to suicide, with a mortality rate of 10.6 per 100,000 inhabitants. While women attempt suicide 2 to 4 times more often than men, there are nearly 2 male suicide deaths for 1 female death [1]. However, Klonsky et al [2] state difficulties in suicide research, especially in statistics, due to the stigma associated with suicide, leading to underreported suicide rates.

The term suicide describes a "death caused by self-directed injurious behavior with an intent to die as a result of the behavior" [3]. A suicide attempt comprises self-directed, potentially injurious behavior. The suicide attempt is performed with an intent to die, though the attempt itself is nonfatal. The terms suicidal self-directed violence and suicidal behavior are used interchangeably. Thinking, considering, or planning to attempt suicide is called suicidal ideation [3].

The consequences of suicidal ideation and behavior are complex. On a personal level, people who are ideating or attempting suicide experience a crisis. Persons who attempted suicide also have difficulties coping with stigma in their interpersonal relationships [4]. For bereaved family members, it is a challenge to cope with grief and guilt. They fear social stigmatization [5]. For nurses, the consequences mainly include unfavorable attitudes towards persons who attempted suicide on the ward, as a survey indicates [6]. Nurses who experience a suicide or suicide attempt at work indicate that such an experience can lead to shock, anger, and frustration. They highlight the importance of being supported afterward [7]. Therapists describe the challenge in the limited time for direct contact with affected persons while prioritizing standardized assessments for diagnostics instead of therapeutic conversations [8]. Emergency department physicians are concerned about overlooking a serious suicide risk [9]. Studies indicate a need for increased knowledge, training, and clinical skills among physicians [9,10].

On a broader level, suicide causes high costs for society, as studies in Spain, Ireland, and Australia indicate. They recommend increased efforts for suicide prevention and health education [11-13]. Over the last few years, the internet has become another option for suicide prevention. Different applications and strategies have been developed and tested [14,15]. The COVID-19 pandemic is challenging people around the world, especially those already struggling with mental illness. The pandemic has also challenged previously established mental health services. Service providers had to find new ways to maintain physical distance while providing treatment and care [16]. Internet-based interventions were highlighted as a convenient option during the pandemic due to actual challenges [17]. Mental health providers perforce offered telemental services [16]. Although there are evidence-based strategies to prevent suicide, difficulties in realization remain prevalent [18]. Especially for telemedicine approaches, challenges arise due to missing legal regulation and concerning ethical aspects in acute situations with suicide risk. For this reason, family and social networks should be involved [16,17]. Difficulties arise during the call or video session, especially due to the patient's remote location [16,19]. In the context of adolescents, Holland et al [19] recommend confirming the physical location and determining if an adult is present. In their review, they conclude that risk assessment and safety planning via telehealth are safe and effective. Studies examining internet-based interventions for persons at risk of having suicidal thoughts or behavior have ethical and practical barriers. Nevertheless, the opinions and experiences of persons with suicidal thoughts or behavior are crucial for developing and testing internet-based applications [20].

These studies show the need for support of health care professionals (HCPs) in their work with persons who ideate or attempt suicide and for the development of evidence-based eHealth interventions derived from the experiences of persons in a suicidal crisis. Therefore, we decided to synthesize the existing qualitative research about how persons with suicidal thoughts or behavior experience inpatient or outpatient care. Our study sought to assess the experiences persons with suicidal thoughts or behavior have had with the professional care they receive in an inpatient or outpatient setting. The aim is to derive a conceptual model to guide HCPs in working with affected persons and provide support during their recovery. Furthermore, the model will be reviewed critically with current literature in the context of health services and eHealth usage of affected persons, from which conclusions will be drawn for the development and revision of eHealth applications. HCP refers to nurses with or without mental health or psychiatric specialization, psychologists, and psychiatrists.

Methods

Overview

A qualitative meta-synthesis was conducted using the approach by Sandelowski and Barroso [21] and followed the recommended six steps: (1) formulating a purpose, (2) systematically searching for and retrieving qualitative research reports, (3) appraising the research reports, (4) classifying the findings, (5) conducting a meta-summary, and (6) developing a meta-synthesis. The meta-synthesis method enables an interpretive integration of the results of all included qualitative studies, basing the results on a larger sample than would be possible from one single qualitative research study. Consequently, coherent experiences and events of the research topic can be described and explained instead of merely summarizing them [21]. Therefore, the purpose of this article was to conduct a meta-synthesis of experiences by persons with suicidal thoughts or behavior with nursing care.

Searching for and Appraising the Research Reports

The systematic literature search was performed between July and October 2016, with an update in June 2021, in PubMed, Cinahl, Medline-OVID, Embase-OVID, Psyndex-OVID, and PsycINFO-OVID. The search terms were (experience OR "lived experience" OR attitudes OR "patient perspective" OR perception) AND ("qualitative research" OR qualitative OR "qualitative design" OR "qualitative study" OR "phenomenological study" OR phenomenology OR "grounded theory") AND ("suicidal ideation" OR suicidal OR suicide OR "suicidal patient" OR "suicidal behavior" OR "suicidal

behaviour“) AND (nursing OR nurses OR hospitalization OR “inpatient care” OR outpatient). Search terms were slightly adjusted to fit the different search systems; in Cinahl, subject headings were used, and in PubMed, medical subject headings were used. Since we aimed for a sensitive search strategy, an additional manual search was conducted in 8 specialist journals (ie, Zeitschrift für Psychologie, The Journal of Crisis Intervention and Suicide Prevention, European Journal of Psychological Assessment, GeroPsych, Nordic Psychology, PRAXIS, Zeitschrift für Kinder-und Jugendpsychiatrie und Psychotherapie, and European Psychologist). Afterward, a literature search of the reference lists of the included studies was conducted.

We included studies with a qualitative study design, reported experiences of persons with suicidal thoughts or behavior regarding nursing care, and were published in English or German. Studies that reported experiences of persons with self-injuring tendencies without suicidal self-directed violence were excluded. We also excluded studies that focused solely on the experiences of family members or HCPs. Studies were excluded if a theory-based or deductive qualitative approach was used. Records were independently screened and identified for eligibility by 2 independent researchers. Discrepancies were discussed with a third researcher. Based on the title and abstract, 2 researchers read the full texts of the studies that appeared to meet the inclusion criteria. Next, the included studies were appraised using the checklist for qualitative research of the critical appraisal skills program (CASP) [22]. The CASP checklist is a useful tool for appraising qualitative studies and systematically identifying the strengths and weaknesses of the assessed studies in design and analysis [23]. The tool appraises the quality of studies but not the quality of the appraisal itself. It consists of 10 “yes” or “no,” and 2 researchers can select “can’t tell” as needed. The results were compared and discussed. The remaining discrepancies were clarified with a third researcher. We defined criteria for each point and a cut-off score that resulted in studies being excluded when they had less than seven “yes” points in the 10-point questionnaire.

Classifying the Findings, Conducting a Meta-Summary, and Developing a Meta-Synthesis

When conducting a qualitative meta-synthesis, the findings of studies should be read to assess which methods were applied and how the data were interpreted. Therefore, the included studies were classified with a recommended typology [21]. Studies were categorized as thematic surveys [24,25], conceptual or thematic descriptions [26-28], or interpretive explanations [29-35].

The basic assumption for conducting a meta-summary and developing a meta-synthesis was that the results of the included studies are interpretations of the data collected by the researchers. Consequently, the results sections of the included studies were treated as transcripts of a qualitative study and used as meta-synthesis data. The results sections of the studies were read several times and then analyzed inductively [36]. MAXQDA2018 (2018; VERBI GmbH) was used to support and manage the analysis process. As a first cycle method, initial sentence by sentence coding of the published material was used to go beyond what was said and discover deeper patterns [36]. In the second step, the results sections were then axially coded. With the method of constant comparison, similarities and differences could be identified. Both helped to identify categories and subcategories. Axial coding promoted a meta-summary reflecting the contents of the included studies. We used constant and comparison and interwove the emerging categories inductively to develop generic categories. A deeper theoretical level of abstraction could be gained with selective coding as the third cycle method and further axial coding [36]. With this interpretive approach, the main concepts emerged from the data, and a conceptual model was derived.

Results

Overview

A total of 3169 studies were identified based on a systematic literature search of databases. One additional study could be identified by a hand search in relevant journals or by screening the references of the included articles. The flowchart of the literature search is shown in Figure 1. Finally, 12 qualitative studies with a total of 208 persons were included in the meta-synthesis, including affected persons (n=176), parents (n=5) or other family members (n=2), nurses or psychiatric nurses (n=27), and psychiatrists (n=3). Affected persons received outpatient or inpatient care due to suicidal ideation or suicidal self-directed violence. They received care in various health care settings, such as emergency rooms or units, day hospitals, stationary psychiatric departments of hospitals, or outpatient psychiatric settings. Data were collected through interviews in all studies [24-35] and additional observation in 2 studies [27,34]. Minors aged 11 to 17 years were interviewed in 3 studies [27,33,34]. Children aged 11 to 14 years were observed in one study, and their parents were interviewed [27]. HCPs were interviewed in 3 studies [24,34,35], and 2 interviewed family members [27,35]. A detailed overview of the studies in the meta-synthesis is provided in Table 1.

In order to address the research question, a conceptual model was synthesized based on the analysis of all articles comprising two main concepts that reflect experiences during inpatient or outpatient care of people with suicidal ideation and behavior. The first main concept, “from suicidal ideation and behavior to feeling anchored in life,” describes the person with suicidal thoughts or behaviors in a process from feeling unanchored while experiencing a suicidal crisis to feeling anchored in life at the end of the subsequent recovery. This phenomenon includes the categories that describe the experience of suicidal ideation and behavior: (1) suicide as an option, (2) communication, and (3) transformation. Persons who ideate or attempt suicide are individually motivated to consider suicide because they cannot communicate their suffering. When speaking about suicide ideation is not taken seriously, they can experience suicide attempts as transformations. Moreover, suicide attempts serve as entrances to health care services. During the recovery process, individuals are being cared for by HCPs. The second main concept describes the “dependency on the skills and attitude of HCPs” in the recovery process. This phenomenon includes the categories (1) “adverse therapeutic

experiences" and (2) "helpful therapeutic experiences" (2). These findings reveal certain hindering and helpful skills and attitudes of HCPs. Helpful skills support affected persons in reaching their feeling of being anchored in life again. Hindering skills are experienced negatively and might lead to prolonged or maintained feelings of being unanchored in life. Textbox 1 provides an overview of the identified concepts, categories, and subcategories. The conceptual model is presented in Figure 2. The model, its inherent concepts and categories, as well as the corresponding findings, will be presented subsequently.

Figure 1. Flow diagram of the review process.

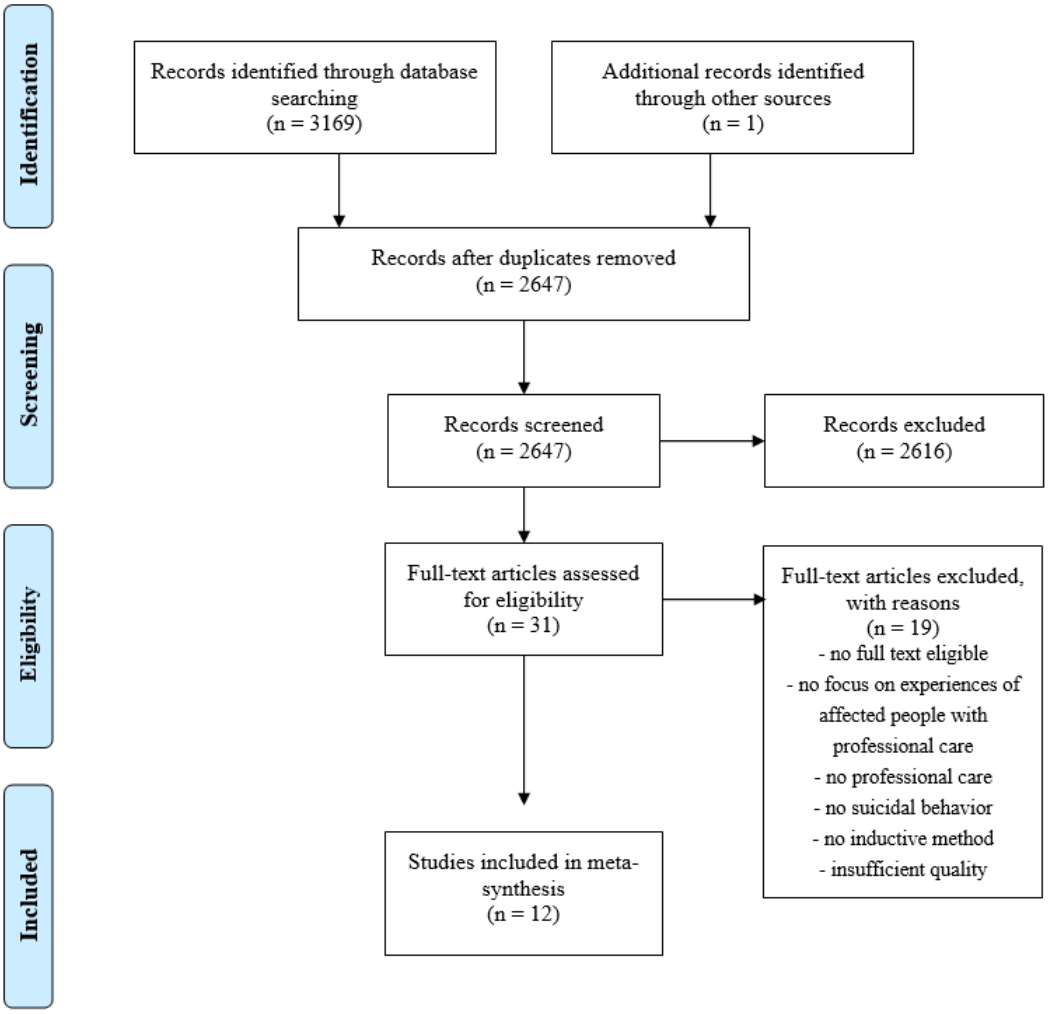


Table 1. Overview of included studies.

Author(s), year	Country of study	Sample/setting	Focus of interest	Method/analysis
Berg, Rørtveit, Walby, & Aase, 2020 [29]	Norway	18 adults (mean age 40) with suicidal ideation and behavior in psychiatric wards in hospitals.	Exploring the experiences of persons in suicidal crisis with safe clinical practice during hospitalization.	Semistructured interviews; phenomenological-hermeneutic approach
Cardell & Pitula, 1999 [26]	USA	20 adults (mean age 32 years) with suicidal ideation in psychiatric wards in hospitals.	Exploring the experiences of persons with suicidal ideation who have been constantly monitored within the last 2 weeks to determine whether this protective intervention had therapeutic benefits for the affected persons.	In-depth interviews, at least 2 times for each person; grounded theory analysis (Hutchinson)
Cutcliffe, Stevenson, Jackson, & Smith, 2006 [31]	UK	20 adults who ideated or attempted suicide. They received care from "emergency" psychiatric services as inpatients, outpatients, or day hospital patients.	Investigating whether and how psychiatric or mental health nurses provide meaningful, caring care to persons with suicidal ideation or behavior.	Semistructured interviews; analysis adhered to principles of grounded theory (Glaser)
Cutcliffe, McKenna, Keeney, Stevenson, & Jordan, 2013 [30]	Northern Ireland	36 male persons with suicidal ideation or behavior between 18 and 34 years old who had been treated in mental health inpatient or outpatient facilities.	Developing a theory on how informal and formal services can be better configured or reconfigured to respond more effectively to the needs of young men with suicidal ideation or behavior.	Semistructured interviews; analysis based on principles of grounded theory (Glaser & Strauss)
Hagen, Knizek, & Hjelmeland, 2018 [32]	Norway	5 adults with suicidal ideation or behavior between 33 and 54 years old who had been admitted to a district psychiatric center.	Exploring the experiences of former suicidal inpatients with treatment and care in psychiatric wards.	Semistructured interviews; interpretative phenomenological analysis
Holliday & Vandermause, 2015 [33]	USA	6 adolescents (15-19 years) who were treated in an emergency room after a suicide attempt.	To gain a comprehensive understanding of the experiences of adolescents who attempted suicide and were taken to an emergency room and their meaning of ideating or attempting suicide as adolescents.	Open, unstructured interviews; Heideggerian hermeneutic methodological approach, phenomenology
Lees, Procter, & Fassett, 2014 [24]	Australia	9 adults (mean age 41 years) receiving care due to suicidal crisis and 11 mental health nurses who assisted persons in suicidal crises (hospital and community setting).	Exploring the experiences and needs of individuals who had a suicidal crisis, the degree to which their needs have been met, the role of mental health nurses, and the key factors to improve quality of care.	Qualitative findings from a multimethod study. In-depth, semistructured interviews; analysis based on critical discourse, constant comparison, and classical content analysis
Montreuil, Butler, Stachura, & Pugnaire Gros, 2015 [27]	Canada	5 children (11-14 years) with suicidal risk factors and one of their parents. They were recruited from pediatric mental health inpatient, outpatient, and day hospital settings.	Assessing perceptions of children with risk factors associated with suicide and their parents regarding helpful care in a pediatric psychiatric setting.	Observations of children and semistructured interviews with parents; inductive data analysis (Colaizzi)
Samuelsson, Wiklander, Asberg, & Saveman, 2000 [25]	Sweden	18 adults (18-53 years) who attempted suicide and were treated in an inpatient psychiatric ward.	Investigating the experiences of patients in a psychiatric ward after having attempted suicide.	Semistructured interviews; content analysis (Burnard)
Sun, Long, Boore, & Tsao, 2006 [34]	Taiwan	15 persons (16-47 years) with suicidal ideation or suicidal behavior and inpatient treatment on psychiatric wards; 15 psychiatric nurses	Exploring experiences of nurses and affected persons to develop a care theory that guides the care of people with suicidal thoughts or behavior.	Semistructured interviews and observations; grounded theory (Strauss & Corbin; Eaves)
Sun, Long, Tsao, & Huang, 2014 [35]	Taiwan	14 adults (22–83 years) who attempted suicide were recruited from an outpatient clinic; 6 caregivers (family members, psychiatrists, a psychiatric nurse)	Exploring contextual and intervening conditions that influence individual healing after a suicide attempt.	Semistructured interviews; grounded theory (Strauss & Corbin)
Vatne & Nåden, 2014 [28]	Norway	10 adults (21-52 years) who ideated or attempted suicide. They were recruited in emergency psychiatric units and from a crisis resolution team.	Exploring the experiences of being suicidal and encounters with health care personnel.	Semistructured interviews; thematic analysis (Braun & Clarke)

Textbox 1. Concepts, categories, and subcategories (excerpts from data analysis).

From suicidal ideation and behavior to feeling anchored in life:

Suicide as an option:

- Being unanchored in life
- Wanting to escape
- Seeing no way out

Suicide as communication:

- Having difficulties speaking about suicidal ideations
- Hiding behind a mask
- Shouting without words

Suicide as transformation:

- Reconnecting through help
- Giving meaning to life
- Moving towards feeling anchored in life again

Dependency on the skills and attitude of health care:

Adverse therapeutic experiences:

- Having an impersonal or unempathetic attitude
- Lacking commitment and acknowledgment
- Applying coercive interventions
- Lacking time
- Not building a trusting relationship

Helpful therapeutic experiences:

- Empathetic attitude
- Acknowledging affected persons
- Appreciative communication
- Promotion of a trusting relationship
- Presenting and providing a safe environment

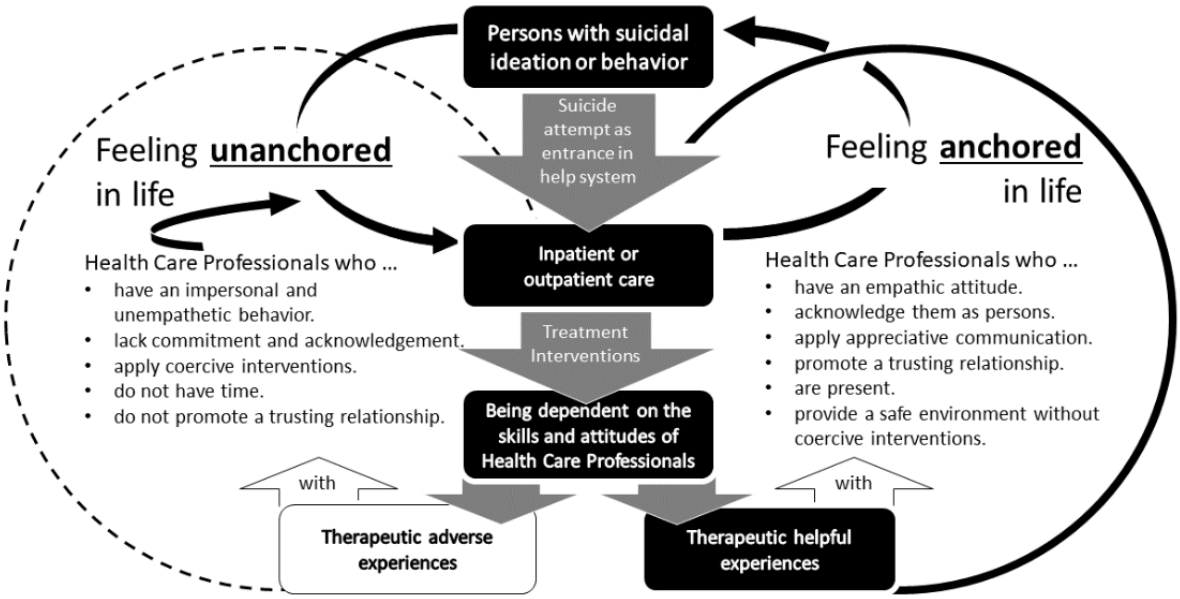
Legend:

Concept:

Category

- Subcategory 1
- Subcategory 2

Figure 2. Conceptual model of therapeutic experiences.



From Suicidal Ideation and Behavior to Feeling Anchored in Life

While having suicidal thoughts or attempting suicide, affected persons experience feelings of being unanchored in life. For different reasons, they consider "suicide as an option" and ideate or attempt suicide. The motivation for suicide has to be taken into account to understand the phenomenon of being unanchored in life. A suicidal crisis is multifactorial and is described individually [24]. For some persons who attempted suicide, the battle in their brain was a motive. They describe it as wanting to flee, having lost control of thoughts, and their mind playing tricks on them. The authors explain "that the thoughts flooded their brain making it difficult for them to 'think in reality' and that such thoughts enhanced the confusion and ambiguity" [33]. At the same time, participants describe having low self-esteem [32].

Another motive is the desire to end or escape from emotional as well as psychological pain and suffering. Painful experiences and the inability to cope with life or problems were further statements [24,33,34]. An interviewee felt "better off dead" [24]. Feelings and thoughts of powerlessness, loneliness, inability to solve problems in life, manipulation of others, or feeling down led to considering or attempting suicide as a way out [24,33,34]. Others stated mourning the loss of an important fellow human being was a motivation for suicide [24,28]. For interviewed persons from Northern Ireland, being male was one reason for suicidal ideation, as society expects great and strong men not to show their feelings and problems [30]. Some of those persons who ideated or attempted suicide reported multiple strains in everyday life. Various physical and psychological problems or illnesses, stress, interpersonal stress, social exclusion, and unpredictable life events also contribute to considering or attempting suicide [24,28,30,34].

Another subcategory describes "suicide as communication." Persons attempting suicide described that they wanted to hide their suffering and end their indescribable suffering. The suicide attempt was a cry "for" help but not only in the sense of calling attention. It was like a cry "of" pain that allowed their pain to become visible to others. Before the suicide attempt, they could not communicate their suffering to their environment through speech. The suicide thus served as a message for others. They put their pain into a social context by making others "hear" their pain [28,33]. Persons with suicidal behavior did not consider themselves ill but realized their problems and knew they needed help [25]. However, it seemed difficult for them to ask for help and access health services. They point out that they were in a vulnerable situation. They express shame for ideating and attempting suicide and the fear of forced admission [25,27,28]. They could talk about physical suffering but not about their sadness, hopelessness, suffering, and feelings. When someone talked about suicidal thoughts, that person seemed weak [28,33]. They feared being judged and stigmatized [30]. They told no one about their suicidal thoughts and wore a mask that concealed their true feelings. Suicide is still a taboo subject, and they do not want to be condemned by others. The thoughts of suicide are perceived as unusual and incomprehensible by the people around them, sometimes leading to the humiliation of the persons with suicidal self-directed violence. Thus, many of those affected hoped that someone would notice their intentions and make small gestures without verbally expressing themselves [28,29,33]. Others tried to communicate but still required the HCPs to interpret their words and understand the severity of their suicidal ideations [29].

The subcategory "suicide as transformation explains" the process of feeling unanchored during suicidal ideation and attempting suicide to recovery with professional help and feeling anchored again. Before the suicide attempt, young people describe the feeling of being disconnected. They feel lonely, rejected as outsiders, neglected, unloved, and spiraling down. They believe that no one else could understand these feelings of emotional pain [33]. Others have the feeling of "falling between two stools" and are in a struggle for life in which support from HCPs is needed [28]. One interviewee described it as follows: "You

feel like you're in the bottom of a hole and its all black and you don't' see nothing…you're just trapped" [33].

Affected persons state that without the suicide attempt, they would not have received the necessary treatment. Some consider attempting suicide when talking about suicidal ideation is not taken seriously. Although some already had treatment in the past, they described the treatment after attempting suicide as helpful and gaining better insights [25,28,33]. Consequently, the attempt can be the entrance to receiving more appropriate support from health care services. Persons with suicidal ideation or self-directed violence are given access to treatment, similar to the beginning of the transformation from a death wish to a wish to live. However, the process was perceived as challenging [28,31,33,35]. Hagen et al [32] state that affected persons are in the process of development that relates to existential, relational, and practical aspects of life: "The experiences appeared to be a part of a personal recovery process that was influenced by the support they received form [sic] mental health workers."

Cutcliffe et al [31] describe the change from death orientation to life orientation in a three-stage healing process that reconnects the person with humanity. The first stage is "reflecting an image of humanity" [31]. Those affected describe that they restore confidence in humanity through reconnecting with HCPs who function as representatives. They can give persons with suicidal ideation or behavior the feeling they are not being left alone and that they care for them, show compassion, and try to offer them help [31,34]. People with suicidal ideation or behavior need to have someone who listens and shows understanding and interest [31,33,34]. Persons with suicidal thoughts or behavior describe that HCPs can show them they are important and an individual. The relationship with HCPs is different from that with family and friends, which gives them a feeling of emancipation [31]. Most interviewed persons reported that their hospitalization was helpful. An affected person may assume he or she has ruined other persons' confidence in oneself by attempting suicide, which leads to the feeling that he or she needs "some kind of babysitter," which means inpatient observation and care [25]. Furthermore, inpatient care helps persons who attempt suicide realize that connecting with others can help in recovery, and they still have a connection to their support group. It also helps to reflect on one's constructs influenced by their hopelessness [31,33]. Affected persons need to be treated without judgment and experience acknowledgment of their narratives and acceptance from others and themselves [30,31,33,35].

In the second stage, "guiding the individual back to humanity," nurses have a more active part in their nursing role than in the first stage [31]. They support the individual in taking on new perspectives to think about one's personality, life, and opportunities [31]. Individualized treatment and care enable affected persons to disclose their thoughts and increase their self-awareness [32]. They experience a transformation of their understanding "from feeling uncared for to understanding that others have always cared for them" [33]. Persons with suicidal ideation or behavior realize that many people have similar emotional problems. They learn to accept help and have the ability to accept support, reconnect, and form new connections. They experience and accept support from the family by reconnecting with them. Those affected find more harmony with society again [30-33,35]. The inclusion of strong social support systems served as a helpful therapy. This reduced stress and made it possible to overcome difficulties. The awareness helped them realize that their families did not reject them because of their suicidal nature and instead forgives them; their families will continue to care for them, accept them, and help them in their therapy [30,35]. During the recovery process, they want to live in a friendly environment and experience being needed by others. This helped them feel valuable and accepted, enabling self-acceptance [35]. Being accepted by significant others promotes increased hope for persons with suicidal ideation or behavior [30].

In the third stage, "learning to live," persons with suicidal ideation or behavior focus on the balance between reconnecting with family, friends, and other people and disconnecting gradually from HCPs [31,32]. They successively learn to give meaning to life again, to make plans, and to set goals. They make sense of the events and gain a new understanding. Giving meaning to life and learning to live is crucial for the recovery process [30,31,33]. It is like becoming a new person and building oneself up again [30,33]. It is important for those affected to feel needed and accepted in society again [35]. Nevertheless, some people express fear of coming back to reality and feel the potential disappointment of others [33]. Feeling able to cope with symptoms and life situations is a prerequisite for discharge from the hospital [29]. Psychiatric treatment can give persons with suicidal ideation or behavior the opportunity to reflect on themselves and their lives. The journey to recovery is a long process, but it provides the opportunity to connect with other people, gain access to the social network, and establish trust [30]. The predictability of follow-up treatment after hospitalization ensures safety. In contrast, feeling unprepared can trigger a suicide attempt, as one participant notes [29].

Affected persons can re-establish internal emotional control when treated as individuals as their stressors can be addressed [29]. Through professional care, persons with suicidal ideation and self-directed violence can reach the feeling of being anchored in life again. Many realize they need support from HCPs. Professional care is central to this recovery process. Experiences during inpatient or outpatient care affect the recovery of persons with suicidal ideation or behavior [25,30,31,33,34]. These experiences with professional care are described in the second concept that follows.

Being Dependent on HCPs' Skills and Attitudes

The second concept describes experiences of persons with suicidal thoughts or behavior as dependent on the HCPs' skills and attitudes. During inpatient or outpatient care, affected persons need professional care and support. According to the HCPs' skills and attitudes, the analysis revealed that affected persons experienced therapeutically adverse or helpful aspects during their recovery process.

Therapeutically Adverse Experiences

Persons with suicidal ideation or behavior experienced these skills and attitudes as therapeutically hindering (ie, having an

impersonal or unempathetic attitude, lacking commitment and acknowledgment, applying coercive interventions, lacking time, and not building a trusting relationship).

Impersonal HCPs and those who lack empathy and commitment are experienced negatively by those who ideated or attempted suicide [25,26]. It is perceived as hindering when HCPs have judgmental attitudes, especially when combined with the affected persons' negative feelings regarding the provided care [34]. A lack of commitment is described as therapeutically adverse. Some HCPs seem to be simply doing their job, prioritizing their needs without using the available time effectively or supporting persons with suicidal ideation in the struggle for their lives [24,25]. A lack of commitment and emotional confirmation also results in persons with suicidal behavior not wanting to talk about themselves [25,29]. HCPs who act impersonally or do not show empathy can contribute to affected persons' damaged self-esteem, increased anxiety, and dysphoria [26]. When affected persons arrive in the emergency room, they experience a lack of consideration [25]. For example, after a suicide attempt in the emergency room, the physical, but not the psychological well-being, was assessed, with the affected person noting, "they just checked my heart and said everything was fine" [33]. They describe HCPs as if they were distracted, disinterested, indifferent, or uncaring. They have the impression that no one seems to care for them and as if they were a burden [25,26].

Affected persons and HCPs state a lack of knowledge [25,28,32,34]."Of particular concern was the finding that many nurses did not have the best possible attitude, education, training, or support to optimally meet the challenges and opportunities at hand, and more fully realize therapeutic engagement" [24].

These attitudes towards care resulted in people feeling lonely and not cared for, which ultimately led to fear, aggression, and a lack of trust in HCPs, although they longed for trust [24,25,28]. Some people did not feel taken seriously in their illness or received inadequate responses and information. Without their perspective being acknowledged, they sometimes feel misunderstood [25,26,28,32,34]. "When the patients did not feel that they were confirmed, it sometimes gave rise to feelings of being burdensome, a desire to go home, or even another suicide attempt" [25].

Experiencing a lack of trust leads to withdrawal from seeking help from HCPs [29,32]. Coercive interventions are unhelpful because they restrict autonomy and privacy and lead to feelings of confinement [24,26,34]. According to the person who constantly observes them, persons with suicidal ideation have uncomfortable or distressing feelings. They prefer HCPs to be supportive rather than impersonal and detached. An example of this behavior is "not responding to the participant's initiation of conversation and perceived hostile facial expressions" [26]. In the absence of engagement, affected persons experience interpersonal isolation, distress, objectification, and loss of control [24]. They sometimes feel treated like a child, which was experienced as humiliating [25,28,34]. Persons with suicidal ideation or suicidal self-directed violence describe environmental stressors as obstructive to their therapeutic experience. For example, television noise was described as annoying and caused stress and excessive demands, as participants had to stay in recreation rooms [26]. Others perceive a lack of respect when HCPs behave like guardians [25]. In situations with coercive interventions, communication is seen as very important. If HCPs failed to communicate relevant information (eg, about constant observation and the observer, or did it abruptly), affected persons stated frustration, irritability, and anxiety [26].

Some people mentioned the lack of time for care, with busy HCPs looking after many patients and having no time for the affected persons [24,32,34]. Additional adverse contextual factors included the lack of teamwork and support and inadequate professional supervision [24]. Some HCPs "did not prioritize interpersonal engagement, or thought it was sometimes inappropriate or countertherapeutic" [24]. Affected persons wanted more time for conversations to establish a close therapeutic relationship. They longed for dialogue and trust [28]. Without individual treatment, they experienced that they were a risk to themselves after discharge [29].

This also had an impact on the relationship between persons with suicidal behavior and HCPs. Instead of sharing their thoughts, they kept them to themselves if they did not feel safe with the HCP [25,28,32]. It takes time to establish a close relationship. However, sometimes this relationship was not possible due to "bad chemistry" [28]. Affected persons state a lack of therapeutic engagement with little interaction and few possibilities to talk [24]. Others experienced HCPs changing topics rather than talking about themes that persons having attempted suicide need to discuss. This leads to feelings of humiliation and annoyance [28].

Therapeutically Helpful Experiences

In addition to therapeutically adverse experiences, persons with suicidal thoughts or behavior also experience therapeutically helpful experiences. These concerns relate to an empathetic attitude, acknowledging affected persons, appreciative communication, promoting a trusting relationship, and presenting and providing a safe environment.

Persons with suicidal thoughts or behavior in the included studies describe the importance of an empathetic attitude from HCPs [24-26,34]. HCPs should be warm, pleasant, talking, listening, and understanding [27]. For those affected, it is important that HCPs honestly put themselves in their position and show concern and care [24,28,31]. They need HCPs "to be willing and able to listen intently […], to attempt to see the world 'through the eyes of the young person and be supportive" [30]. Therefore, being nonjudgmental is important [30,31,34]. HCPs with knowledge, experience, and understanding of physiological processes and stages in mental illness are seen as helpful [27,28,32]. An optimistic attitude on the part of the HCPs also provided therapeutically helpful care and increased self-esteem. The optimistic behavior of care is perceived as loving, helpful, and hopeful [26]. It is essential for those affected to be regarded as individuals, supported in their autonomy, and not condemned for their illness. This enabled them to realize that they were worth something and that they were being cared for [24,25,29,34]. Persons with suicidal ideation and behavior need acknowledgment of their person and their illness [26,35].

It was also experienced as helpful when professional care and treatment were individualized to the needs of the affected persons [27,29,32]. Persons with suicidal ideation and behavior experienced safe clinical practice when they received tailor-made treatment. This type of treatment relieved their emotional pressure as stressors were addressed [29].

Communication between HCPs and those affected plays a central role in recovery. The basis is a holistic assessment of persons with suicidal thoughts or behavior, including their needs and causes of suicidal tendencies. Some people actively seek dialogue with HCPs to talk about their suicidal thoughts and behaviors, about issues they have not yet entrusted to anyone, or answers to questions they have [24-26,28]. They value cooperation and open communication between different HCPs to understand themselves and plan their treatment [27,34]. Talking to HCPs and being understood and supported was felt by some to be vital, although talking about it was initially felt by many to be terribly painful [24,25,31]. In addition, speaking, active listening, and being taken seriously are important. The commitment of HCPs is essential to explore the complexities of suicidal ideation in affected persons [24,28,29,32]. Discussions between those affected and HCPs without judgment, and on the same level, helped many sort out their thoughts [24,25,28,31]. One interviewee put it as follows:

> *She wasn't...yes, "pitying" again then. We were two people talking together on equal terms, not prisoner and jailer...She would not divert the conversation, no matter what* [28]

It was considered important that HCPs should not scream but speak in a calm tone [27]. The direct inquiry into suicidal thoughts or plans was also positively perceived by those affected [28,34]. Some people found the denial of these thoughts impossible because HCPs often knew persons with suicidal self-directed violence well and recognized their needs and feelings through nonverbal communication [25,29].

A trusting relationship between the HCPs and their entrusted persons with suicidal thoughts or behavior served as a basis for both parties [24,25,34]. Participants noted that having an open conversation requires a "good chemistry" and connection with a health care professional [32]. Many persons with suicidal thoughts or behavior had difficulties trusting people and had no support from friends or family, so trusting caregivers was important [28,31]. A trusting relationship at the same level between the HCPs and those affected had a positive effect on suicidal tendencies [28,29]. It is essential for the therapeutic relationship that HCPs show appreciative behavior, respect, and interest for the well-being of persons with suicidal thoughts [26,32]. It is important that HCPs introduce themselves with their name when meeting each other the first time [26], and they should also know the affected persons' names [27]. Developing a caring relationship can be promoted through getting to know each other and personalized care [27]. Persons with suicidal self-directed violence want to be met on equal terms, which "is a situation whereby the parties accept each other's inherent value" [28]. The therapeutic relationship can be established with communication basics, such as showing compassion, acceptance, and appreciation and serving the healing process.

Additionally, worthiness was fundamental as it increased well-being and reduced anxiety, dysphoria, and loneliness in persons with suicidal ideation [26]. A sense of companionship provides a sense of safety and well-being and enables affected persons to disclose suicidal thoughts [32]. Therapeutic closeness served as the basis for subsequent interventions [30,32]. Promoting trust with HCPs helped the affected person reconnect with an individual as the first step before reconnecting with their social environment [31].

Another helpful experience with HCPs was present [24,26,27,29,35]. Although it sometimes felt difficult, especially for children and adolescents, to ask for help, it was important to persons with suicidal self-directed violence that someone was always available to give them security and help. The HCPs were described as very sensitive and often recognized how they felt, whether they needed time to talk or wanted to be left alone [25,27,35]. Some of the affected persons also used HCPs after discharge and called them by telephone [25,27]. Many of them describe that the greatest help was knowing that care was always present and that someone was available to take time for them [24,25,29]. Parents of minors with suicidal ideation also described this as very helpful and used the opportunity to call at their convenience [27].

Being present also affected the therapeutic environment. Experiencing protection and safety through 24-hour care of HCPs positively affects therapy and the relationship between caregivers and persons with suicidal ideation or behavior [27,29,34,35]. Protection and security within the institution have been described by many as a reason to stay, live, and take time for therapeutic interventions in this environment. However, it was not easy for many of those affected to be treated in a psychiatric institution [25,26,35]. However, many expressed that the institution would provide them with personal security, protection, peace, and no stress [29,35]. Due to the vigilant presence, constant observation, and physical presence of HCPs, suicide was hardly possible. This is also due to locked stations and the prohibition of bringing sharp objects such as knives and weapons. As a result, some persons with suicidal thoughts or behavior felt safer from their suicidal impulses thanks to the institution [26,29,34]. Additionally, a quiet, friendly atmosphere, a good climate on the ward, and a meeting on the same level had a positive effect on the therapeutic environment [28,35].

Persons with suicidal self-directed violence also experienced care without coercion as therapeutically helpful. The participants expressed that they felt relief through unconstrained hospitalization. Although hospitalization was described as a terrible feeling, HCPs stressed that the need for inpatient care was helpful if treated voluntarily. In addition, not being forced to speak but simply to be allowed to sit quietly was experienced positively [25,28].

Discussion

Principal Results

This study aimed to synthesize qualitative research to develop a model that supports HCPs in their work with persons with suicidal thoughts or behavior and provides an evidence base for

developing eHealth tools. The results of the meta-synthesis revealed two central categories shown in the conceptual model (Figure 2). The first concept describes persons with suicidal ideation or behavior in their struggle as they feel unanchored in life to feeling anchored in life again. A suicide attempt can function as an entrance to health care services to more easily achieve feelings of being anchored in life again. When receiving inpatient or outpatient care by HCPs, the affected persons are dependent on the skills and attitudes of the professionals present. Helpful therapeutic experiences stated by persons with suicidal ideation or behavior have been synthesized. HCPs' helpful skills and attitudes support persons ideating or attempting suicide to feel anchored in life again. On the other hand, the identified therapeutically adverse skills and attitudes from HCPs may hinder affected persons in proceeding with their recovery and allow them to maintain the feeling of being unanchored in life.

In the following discussion, we focus on two main aspects. First, a suicide attempt can function as an entrance to health care services. Second, adverse therapeutic experiences that hinder affected persons from feeling anchored in life again versus helpful experiences that promote their way toward being anchored. Aspects concerning eHealth are taken into account within each point.

Need for More eHealth Devices to Prevent Suicide Attempts as Entrances to Health Care Services

The meta-synthesis revealed that persons with suicidal ideation sometimes considered attempting suicide when they felt they were not being taken seriously. Through suicide attempts, they enter the health care system and receive necessary or more appropriate treatment [25,28,33]. Our results and qualitative analysis of motives show that attempting suicide is not a conscious decision but rather a complex interaction of different factors [37]. Another qualitative study focused on experiences disclosing suicidal thoughts. Persons who attempted suicide experienced difficulties disclosing primarily during and after the crisis. They did not find the right words, found themselves unable to share their thoughts, or feared being stigmatized by family, friends, or HCPs [38]. They felt ambiguous about sharing and being unable or fearing to do so. Nevertheless, one interviewee stated, "Nobody hears a silent cry for help" [38]. Difficulties in disclosing, or feeling one is not being taken seriously, can result in an initial or subsequent suicide attempt, which can be the first step in a recovery process as most unsuccessful suicide attempts gain access to appropriate health care services. Experiencing helpful skills and attitudes from HCPs is critical to feeling anchored in life again. Professional care plays a central role in helping individuals find their way back to feeling anchored in life again. The phenomenon identified in our meta-synthesis of attempting suicide as an entrance to health care services has not been found in the literature. The reason could be that we identified it from our analysis of persons who attempted suicide and reflected on their experiences. Therefore, it may not be an intention of persons with suicidal ideation but may be seen as such when reflecting it from a future perspective.

Regarding the aims of this study, using eHealth tools could potentially intervene when persons have suicidal thoughts or behavior or are at risk for self-harm [14,15,39]. Positive outcomes of utilizing eHealth applications included reducing depression, psychological distress, and self-harm [14], while others reduced suicidal ideation [15,39]. Using an app is an easy and anonymous way to deal with upcoming thoughts. Other possibilities are web-based technologies or social networks. In particular, younger people can be more easily addressed through technology-based interventions for suicide prevention [40]. A study stated that young people confronted with depression, suicidal ideation, or stress are less likely to talk to their parents about their problems and more likely to speak to no one [41]. It is challenging to identify people at risk in order to manage the crisis with close persons. Holland et al [19] propose the presence of an adult as a backup to ensure safety during calls and video sessions. A significant improvement for quality and outcome of intervention has also been found with the involvement of family or caregivers as youth and family receive coping tools and psychoeducation. One opportunity would be to detect people at risk in social networks and facilitate access to a supportive network and specialists. However, there is a lack of studies concerning suicide prevention in social networks [40]. Young persons with suicidal ideation used four types of technology-based telemental health resources: self-help, anonymous chat, crisis text lines, and online therapists or counselors [41]. Online interventions can be helpful, especially if persons with suicidal ideation or behavior have difficulties disclosing their crisis to family or friends, as studies have shown [38,41].

Nevertheless, it is vital to guide affected persons to helpful resources. A survey indicated that a search for suicide-specific themes could lead to preventive (68%) resources but also to mixed (22%) and neutral (8%) content concerning attitudes towards suicide or even prosuicide content (1%) [42]. There is a growing body of interventions for suicide prevention. In their systematic reviews, Melia et al [14] and Kreuze et al [15] conclude that there are technology-based interventions for suicide prevention. However, further research is needed to evaluate their efficacy. Difficulties also arise in the missing evaluation of self-management applications for suicidal thoughts or behavior [39,40].

Need for Integration of HCPs' Helpful Skills and Attitudes in eHealth Applications

The meta-synthesis revealed that persons with suicidal thoughts or behavior had therapeutically helpful experiences when HCPs had an empathetic attitude, acknowledged affected persons, used appreciative communication, were present, promoted a trusting relationship, and provided a safe environment. Two systematic literature reviews, a meta-synthesis, and a mixed-methods study of persons who self-harm and have suicidal ideation or behavior identified a positive relationship between patients and HCPs as crucial. They find it important that HCPs are supportive, compassionate, and ready to listen [43-46]. Although 2 of these studies are about self-harm, the focus on the importance of the relationship is the same as that shown in our results. Understanding and nonjudgmental HCPs were seen as important for future help-seeking. Studies also state the need for boundaries through a safe environment and that sometimes safety measures as special observations are

appropriate [43,45,46]. This is congruent with our result of HCPs providing a secure environment. Knowing HCPs are present creates a feeling of comfort for persons with self-harm, suicidal ideation, and behavior [44-46].

Our results show that affected persons experienced skills and attitudes of HCPs as therapeutically adverse when they have an impersonal, unempathetic attitude, lack commitment and acknowledgment, apply coercive interventions, lack time, and do not promote a trusting relationship. Similar results were found in literature about adults who self-harm, have suicidal ideation, or behavior. HCPs who have a judgmental attitude, are unsupportive, lack empathy, or exert power are seen as part of unsatisfactory care [43-46]. Negative experiences could even be a barrier for future help-seeking, while trust in services encouraged future help-seeking [43,44]. Hagen et al [8] examined therapist challenges with persons who self-harm and have suicidal thoughts or behavior. The interviewed psychiatrists and psychologists described challenges between categorizing the illness and connecting with the affected persons while following guidelines for diagnoses and treatment. Moreover, they experience challenges forming a trusting relationship due to limited time in direct care and formal obligations. Consequently, therapists spend less time with affected persons, and the greater part of direct care is done by other staff members such as psychiatric nurses. This underlines the importance of cultivating helpful skills and attitudes in nurses.

For eHealth interventions, it is a challenge to consider how affected persons benefit from the skills and attitudes of HCPs. Some eHealth resources are solely informational or unguided self-management applications, while others provide exchanges with other persons seeking help or provide virtual contact with HCPs. From the meta-synthesis results, it has to be considered how an empathetic attitude, appreciative communication, being present, and promoting a trusting relationship can be transferred into an eHealth application. The use of artificial intelligence for mental health care provides a multifaceted opportunity, although unanswered ethical questions remain [47]. A literature review identified chatbots that can be used in mental health. They are applied for different purposes, such as therapy, self-management, counseling, or diagnosis. For example, one was identified for screening symptoms of depression and suicide [48]. These chatbots could be “trained” with helpful skills and attitudes. In addition to unguided self-management applications, the focus should be on those with direct contact and crisis support. Via video conferencing, telemental, or similar tools, affected persons can communicate with HCPs directly. During the COVID-19 pandemic, a survey reported mainly positive experiences of users with telepsychiatry. Most of the participants (82.2%) found that the overall experience with telepsychiatry was excellent or good. In addition, participants either agreed or strongly agreed (63.6%) that the remote sessions were as helpful as in-person treatment [49]. Another opportunity could be a brief text message-based intervention that showed the potential to support persons who attempted suicide in connecting them with a crisis support team to reduce re-attempts [50]. Franco et al [40] stated an upward trend in using technology-based interventions for suicide prevention. However, these are mostly in English [40], presenting a barrier for persons who are not familiar with the English language.

Strengths and Limitations

The strength of this meta-synthesis is the conceptual model derived from how persons with suicidal ideation and behavior experience inpatient and outpatient care. To our knowledge, this is the first meta-synthesis of these experiences with professional care. This is important for ethical reasons, as no affected persons need to be newly involved in this meta-synthesis. The results can be used as a basic needs assessment for eHealth development and nurture an empathetic culture among HCPs. However, our findings are limited by the different study designs exploring the experiences of affected persons. Half of the studies relied on grounded theory or were analyzed with the principles of grounded theory. The other studies used content analysis, inductive data analysis, or a Heideggerian hermeneutic approach. Through the grounded theory–based analysis, we gained a high theoretical level of abstraction. Nevertheless, we could not formulate theory but could speak of a conceptual model. Including experiences from the different samples broadened the range of perspectives. However, it must be taken into account that the samples had different cultural origins. Therefore, we suppose the results could be transferred to different cultural contexts, but they must be checked beforehand.

Conclusions

We derived a conceptual model of experiences made by persons with suicidal ideation and behavior. The model showed the main helpful skills and attitudes of HCPs that can support affected persons to be anchored in life again. Conversely, we also identified hindering skills and attitudes that lead to adverse therapeutic experiences, which may prolong the recovery of persons with suicidal ideation and behavior.

We focused our research on persons who have suicidal ideation or have already attempted suicide. The discussion in previous studies showed that persons with self-harming behavior experience similar challenges and can likewise benefit from eHealth tools that address suicidal thoughts or behavior. It is useful to address these groups with one application, as the boundaries may be indistinct for affected persons.

This meta-synthesis has some practical and theoretical implications. As practical implications, the results can be used as a blueprint for technicians and HCPs to develop eHealth interventions. These could especially address younger persons, as they are more likely to use online resources or eHealth applications in cases of suicidal ideation or behavior. Especially during the COVID-19 pandemic, eHealth tools are a convenient solution. Another practical implication addresses HCPs. Our results show that suicidal ideation should be taken seriously by HCPs. They could use the conceptual model for training and education to improve professional care and improve outcomes for affected persons. HCPs need to be sensitized for the effects their skills and attitudes have on persons with suicidal ideation and suicidal self-directed violence. They should react with appreciative communication and an empathetic attitude and be present to promote a trusting relationship. Moreover, they should

ensure a safe environment to help affected persons feel anchored in life again without using a suicide attempt as another effort to benefit from health care services.

As theoretical implications, further research is needed. Research should focus on experiences made by persons with suicidal ideation or behavior from a hermeneutical perspective. The meta-synthesis with the derived conceptual model can function as a basis for developing new interventions to support affected persons. These interventions could focus on deepening the helpful skills and attitudes of HCPs in interactions with persons with suicidal thoughts or behavior. Other interventions should promote eHealth applications for affected persons, which are evaluated as to whether they accurately fit and support persons with suicidal thoughts and behavior. More research is also needed to identify helpful interventions for affected persons. Moreover, a questionnaire could be developed from the conceptual model to promote the quality of care of affected persons.

Acknowledgments

We would like to thank Tanja Lemp and Nadine Saladin for their initial work in this meta-synthesis. They did the systematic search and open coding, which formed the basis for the meta-synthesis. Saladin and AF did axial coding. MH and AF carried out the selective coding, drafted the manuscript, and approved the final version.

Conflicts of Interest

None declared.

References

1. World Health Organization. World health statistics 2019: monitoring health for the SDGs, sustainable development goals. Geneva: World Health Organization; 2019.
2. Klonsky ED, May AM, Saffer BY. Suicide, suicide attempts, and suicidal ideation. Annu Rev Clin Psychol 2016;12:307-330. [doi: 10.1146/annurev-clinpsy-021815-093204] [Medline: 26772209]
3. Crosby A, Ortega L, Melanson C. Self-directed violence surveillance: uniform definitions and recommended data elements. Atlanta (GA): Centers for Disease Control and Prevention, National Center for Injury Prevention and Control; 2011. URL: https://www.cdc.gov/violenceprevention/pdf/self-directed-violence-a.pdf [accessed 2021-09-22]
4. Rimkeviciene J, Hawgood J, O'Gorman J, De Leo D. Personal stigma in suicide attempters. Death Stud 2015;39(10):592-599. [doi: 10.1080/07481187.2015.1037972] [Medline: 26086667]
5. Gall TL, Henneberry J, Eyre M. Two perspectives on the needs of individuals bereaved by suicide. Death Stud 2014;38(6-10):430-437. [doi: 10.1080/07481187.2013.772928] [Medline: 24758212]
6. Ouzouni C, Nakakis K. Nurses' attitudes towards attempted suicide. Health Science Journal 2013;7(1):119-134 [FREE Full text]
7. Bohan F, Doyle L. Nurses' experiences of patient suicide and suicide attempts in an acute unit. Mental Health Practice 2008 Feb;11(5):12-16 [FREE Full text] [doi: 10.7748/mhp2008.02.11.5.12.c6338]
8. Hagen J, Hjelmeland H, Knizek BL. Connecting with suicidal patients in psychiatric wards: Therapist challenges. Death Stud 2017 Jul;41(6):360-367. [doi: 10.1080/07481187.2017.1284955] [Medline: 28129046]
9. Roy W, Roaten K, Downs D, Khan F, Pollio DE, North CS. Suicide risk assessment and management: real-world experience and perceptions of emergency medicine physicians. Arch Suicide Res 2017 Jul 03;21(3):365-378. [doi: 10.1080/13811118.2016.1199987] [Medline: 27309841]
10. Fernandes J, Scheuermeyer FX, Chakraborty AT, Honer WG, Barbic D. What are Canadian emergency physicians' attitudes and self-perceived competence toward patients who present with suicidal ideation? CJEM 2021 Sep 01;23(5):668-672. [doi: 10.1007/s43678-021-00157-0] [Medline: 34196944]
11. Alfonso-Sánchez JL, Martin-Moreno JM, Martinez IM, Martinez AA. Epidemiological study and cost analysis of suicide in Spain: over 100 years of evolution. Arch Suicide Res 2020;24(sup2):S356-S369. [doi: 10.1080/13811118.2019.1612802] [Medline: 31079554]
12. Kennelly B. The economic cost of suicide in Ireland. Crisis 2007;28(2):89-94. [doi: 10.1027/0227-5910.28.2.89] [Medline: 17722690]
13. Kinchin I, Doran C. The economic cost of suicide and non-fatal suicide behavior in the Australian workforce and the potential impact of a workplace suicide prevention strategy. Int J Environ Res Public Health 2017 Mar 27;14(4):347 [FREE Full text] [doi: 10.3390/ijerph14040347] [Medline: 28346379]
14. Melia R, Francis K, Hickey E, Bogue J, Duggan J, O'Sullivan M, et al. Mobile health technology interventions for suicide prevention: systematic review. JMIR Mhealth Uhealth 2020 Jan 15;8(1):e12516 [FREE Full text] [doi: 10.2196/12516] [Medline: 31939744]
15. Kreuze E, Jenkins C, Gregoski M, York J, Mueller M, Lamis DA, et al. Technology-enhanced suicide prevention interventions: A systematic review. J Telemed Telecare 2017 Jul;23(6):605-617. [doi: 10.1177/1357633X16657928] [Medline: 27377792]

16. Jobes DA, Crumlish JA, Evans AD. The COVID-19 pandemic and treating suicidal risk: The telepsychotherapy use of CAMS. Journal of Psychotherapy Integration 2020 Jun;30(2):226-237. [doi: 10.1037/int0000208]
17. Wasserman D, Iosue M, Wuestefeld A, Carli V. Adaptation of evidence-based suicide prevention strategies during and after the COVID-19 pandemic. World Psychiatry 2020 Oct;19(3):294-306 [FREE Full text] [doi: 10.1002/wps.20801] [Medline: 32931107]
18. Wasserman D, Cyranka K. Difficulties in preventing suicidal behaviours in spite of existing evidence-based preventive methods – An overvie. Arch Psych Psych 2019;21(12):7-12 [FREE Full text] [doi: 10.12740/app/104408]
19. Holland M, Hawks J, Morelli LC, Khan Z. Risk assessment and crisis intervention for youth in a time of telehealth. Contemp Sch Psychol 2021 Jan 06:1-15 [FREE Full text] [doi: 10.1007/s40688-020-00341-6] [Medline: 33425481]
20. Bailey E, Mühlmann C, Rice S, Nedeljkovic M, Alvarez-Jimenez M, Sander L, et al. Ethical issues and practical barriers in internet-based suicide prevention research: a review and investigator survey. BMC Med Ethics 2020 May 13;21(1):37 [FREE Full text] [doi: 10.1186/s12910-020-00479-1] [Medline: 32404098]
21. Sandelowski M, Barroso J. Handbook for synthesizing qualitative research. New York: Springer Publishing Company; 2007:9780826156945.
22. Critical Appraisal Skills Programme. CASP Qualitative Checklist 2018. URL: https://casp-uk.net/wp-content/uploads/2018/01/CASP-Qualitative-Checklist-2018.pdf [accessed 2021-07-27]
23. Singh J. Critical appraisal skills programme. J Pharmacol Pharmacother 2013;4(1):76. [doi: 10.4103/0976-500x.107697]
24. Lees D, Procter N, Fassett D. Therapeutic engagement between consumers in suicidal crisis and mental health nurses. Int J Ment Health Nurs 2014 Aug;23(4):306-315. [doi: 10.1111/inm.12061] [Medline: 24575883]
25. Samuelsson M, Wiklander M, Asberg M, Saveman B. Psychiatric care as seen by the attempted suicide patient. J Adv Nurs 2000 Sep;32(3):635-643. [doi: 10.1046/j.1365-2648.2000.01522.x] [Medline: 11012806]
26. Cardell R, Pitula CR. Suicidal inpatients' perceptions of therapeutic and nontherapeutic aspects of constant observation. Psychiatr Serv 1999 Aug;50(8):1066-1070. [doi: 10.1176/ps.50.8.1066] [Medline: 10445656]
27. Montreuil M, Butler KJD, Stachura M, Pugnaire Gros C. Exploring helpful nursing care in pediatric mental health settings: The perceptions of children with suicide risk factors and their Parents. Issues Ment Health Nurs 2015;36(11):849-859. [doi: 10.3109/01612840.2015.1075235] [Medline: 26631856]
28. Vatne M, Nåden D. Patients' experiences in the aftermath of suicidal crises. Nurs Ethics 2014 Mar;21(2):163-175. [doi: 10.1177/0969733013493218] [Medline: 23999663]
29. Berg SH, Rørtveit K, Walby FA, Aase K. Safe clinical practice for patients hospitalised in mental health wards during a suicidal crisis: qualitative study of patient experiences. BMJ Open 2020 Nov 06;10(11):e040088 [FREE Full text] [doi: 10.1136/bmjopen-2020-040088] [Medline: 33158829]
30. Cutcliffe JR, McKenna H, Keeney S, Stevenson C, Jordan J. 'Straight from the horse's mouth': rethinking and reconfiguring services in Northern Ireland in response to suicidal young men. J Psychiatr Ment Health Nurs 2013 Apr;20(5):466-472. [doi: 10.1111/jpm.12012] [Medline: 23145993]
31. Cutcliffe JR, Stevenson C, Jackson S, Smith P. A modified grounded theory study of how psychiatric nurses work with suicidal people. Int J Nurs Stud 2006 Sep;43(7):791-802. [doi: 10.1016/j.ijnurstu.2005.09.001] [Medline: 16289065]
32. Hagen J, Knizek BL, Hjelmeland H. Former suicidal inpatients' experiences of treatment and care in psychiatric wards in Norway. Int J Qual Stud Health Well-being 2018 Dec;13(1):1461514 [FREE Full text] [doi: 10.1080/17482631.2018.1461514] [Medline: 29652227]
33. Holliday C, Vandermause R. Teen experiences following a suicide attempt. Arch Psychiatr Nurs 2015 Jun;29(3):168-173. [doi: 10.1016/j.apnu.2015.02.001] [Medline: 26001716]
34. Sun F, Long A, Boore J, Tsao L. A theory for the nursing care of patients at risk of suicide. J Adv Nurs 2006 Mar;53(6):680-690. [doi: 10.1111/j.1365-2648.2006.03774.x] [Medline: 16553676]
35. Sun F, Long A, Tsao L, Huang H. The healing process following a suicide attempt: context and intervening conditions. Arch Psychiatr Nurs 2014 Feb;28(1):55-61. [doi: 10.1016/j.apnu.2013.10.004] [Medline: 24506988]
36. Saldaña J. The coding manual for qualitative researchers 3rd ed. Los Angeles: London SAGE; 2016:9781473902480.
37. O'Brien KHM, Nicolopoulos A, Almeida J, Aguinaldo LD, Rosen RK. Why adolescents attempt suicide: a qualitative study of the transition from ideation to action. Arch Suicide Res 2021;25(2):269-286. [doi: 10.1080/13811118.2019.1675561] [Medline: 31608796]
38. Maple M, Frey LM, McKay K, Coker S, Grey S. "Nobody hears a silent cry for help": suicide attempt survivors' experiences of disclosing during and after a crisis. Arch Suicide Res 2020;24(4):498-516. [doi: 10.1080/13811118.2019.1658671] [Medline: 31507236]
39. Witt K, Spittal MJ, Carter G, Pirkis J, Hetrick S, Currier D, et al. Effectiveness of online and mobile telephone applications ('apps') for the self-management of suicidal ideation and self-harm: a systematic review and meta-analysis. BMC Psychiatry 2017 Aug 15;17(1):297 [FREE Full text] [doi: 10.1186/s12888-017-1458-0] [Medline: 28810841]
40. Franco-Martín MA, Muñoz-Sánchez JL, Sainz-de-Abajo B, Castillo-Sánchez G, Hamrioui S, de la Torre-Díez I. A systematic literature review of technologies for suicidal behavior prevention. J Med Syst 2018 Mar 05;42(4):71. [doi: 10.1007/s10916-018-0926-5] [Medline: 29508152]

41. Toscos T, Coupe A, Flanagan M, Drouin M, Carpenter M, Reining L, et al. Teens using screens for help: impact of suicidal ideation, anxiety, and depression levels on youth preferences for Telemental Health Resources. JMIR Ment Health 2019 Jun 21;6(6):e13230 [FREE Full text] [doi: 10.2196/13230] [Medline: 31228179]
42. Thornton L, Handley T, Kay-Lambkin F, Baker A. Is a person thinking about suicide likely to find help on the internet? An evaluation of Google search results. Suicide Life Threat Behav 2017 Feb;47(1):48-53. [doi: 10.1111/sltb.12261] [Medline: 27111845]
43. Lindgren B, Svedin CG, Werkö S. A systematic literature review of experiences of professional care and support among people who self-harm. Arch Suicide Res 2018;22(2):173-192. [doi: 10.1080/13811118.2017.1319309] [Medline: 28426393]
44. Cully G, Leahy D, Shiely F, Arensman E. Patients' experiences of engagement with healthcare services following a high-risk self-harm presentation to a hospital emergency department: a mixed methods study. Arch Suicide Res 2020 Jun 24:1-21. [doi: 10.1080/13811118.2020.1779153] [Medline: 32576083]
45. Vandewalle J, Van Bos L, Goossens P, Beeckman D, Van Hecke A, Deproost E, et al. The perspectives of adults with suicidal ideation and behaviour regarding their interactions with nurses in mental health and emergency services: A systematic review. Int J Nurs Stud 2020 Oct;110:103692. [doi: 10.1016/j.ijnurstu.2020.103692] [Medline: 32682109]
46. Berglund S, Åström S, Lindgren B. Patients' experiences after attempted suicide: a literature review. Issues Ment Health Nurs 2016 Oct;37(10):715-726. [doi: 10.1080/01612840.2016.1192706] [Medline: 27327200]
47. Fiske A, Henningsen P, Buyx A. Your robot therapist will see you now: ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. J Med Internet Res 2019 May 09;21(5):e13216 [FREE Full text] [doi: 10.2196/13216] [Medline: 31094356]
48. Abd-Alrazaq AA, Alajlani M, Alalwan AA, Bewick BM, Gardner P, Househ M. An overview of the features of chatbots in mental health: A scoping review. Int J Med Inform 2019 Dec;132:103978. [doi: 10.1016/j.ijmedinf.2019.103978] [Medline: 31622850]
49. Guinart D, Marcy P, Hauser M, Dwyer M, Kane JM. Patient attitudes toward telepsychiatry during the COVID-19 pandemic: A nationwide, multisite survey. JMIR Ment Health 2020 Dec 22;7(12):e24761 [FREE Full text] [doi: 10.2196/24761] [Medline: 33302254]
50. Berrouiguet S, Larsen ME, Mesmeur C, Gravey M, Billot R, Walter M, HUGOPSY Network, et al. Toward mHealth brief contact interventions in suicide prevention: case series from the Suicide Intervention Assisted by Messages (SIAM) randomized controlled trial. JMIR Mhealth Uhealth 2018 Jan 10;6(1):e8 [FREE Full text] [doi: 10.2196/mhealth.7780] [Medline: 29321126]

Abbreviations

CASP: critical appraisal skills program
HCP: health care professionals

Edited by G Eysenbach; submitted 02.02.21; peer-reviewed by R Rosen, T Cosco; comments to author 08.06.21; revised version received 27.07.21; accepted 16.08.21; published 28.10.21.

Please cite as:
Hechinger M, Fringer A
Professional Care Experiences of Persons With Suicidal Ideation and Behavior: Model Development Based on a Qualitative Meta-Synthesis
JMIR Form Res 2021;5(10):e27676
URL: https://formative.jmir.org/2021/10/e27676
doi:10.2196/27676
PMID:34709191

Original Paper

Toward an Extended Definition of Major Depressive Disorder Symptomatology: Digital Assessment and Cross-validation Study

Nayra A Martin-Key[1], PhD; Dan-Mircea Mirea[1,2], MSc; Tony Olmert[1,3], MPhil; Jason Cooper[1,4], PhD; Sung Yeon Sarah Han[1], PhD; Giles Barton-Owen[5], MA; Lynn Farrag[5], BSc; Emily Bell[5], BA; Pawel Eljasz[1]; Daniel Cowell[5,6], BSc; Jakub Tomasik[1], PhD; Sabine Bahn[1,5], MRCPsych, MD, PhD

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

[2]Princeton Neuroscience Institute, Princeton University, Princeton, NJ, United States

[3]UC San Diego School of Medicine, University of California, San Diego, CA, United States

[4]Owlstone Medical Ltd, Cambridge, United Kingdom

[5]Psyomics Ltd, Cambridge, United Kingdom

[6]Sentinel Oncology Ltd, Cambridge, United Kingdom

Corresponding Author:
Sabine Bahn, MRCPsych, MD, PhD
Cambridge Centre for Neuropsychiatric Research
Department of Chemical Engineering and Biotechnology
University of Cambridge
Philippa Fawcett Drive
Cambridge, CB3 0AS
United Kingdom
Phone: 44 1223 334151
Email: sb209@cam.ac.uk

Abstract

Background: Diagnosing major depressive disorder (MDD) is challenging, with diagnostic manuals failing to capture the wide range of clinical symptoms that are endorsed by individuals with this condition.

Objective: This study aims to provide evidence for an extended definition of MDD symptomatology.

Methods: Symptom data were collected via a digital assessment developed for a delta study. Random forest classification with nested cross-validation was used to distinguish between individuals with MDD and those with subthreshold symptomatology of the disorder using disorder-specific symptoms and transdiagnostic symptoms. The diagnostic performance of the Patient Health Questionnaire–9 was also examined.

Results: A depression-specific model demonstrated good predictive performance when distinguishing between individuals with MDD (n=64) and those with subthreshold depression (n=140) (area under the receiver operating characteristic curve=0.89; sensitivity=82.4%; specificity=81.3%; accuracy=81.6%). The inclusion of transdiagnostic symptoms of psychopathology, including symptoms of depression, generalized anxiety disorder, insomnia, emotional instability, and panic disorder, significantly improved the model performance (area under the receiver operating characteristic curve=0.95; sensitivity=86.5%; specificity=90.8%; accuracy=89.5%). The Patient Health Questionnaire–9 was excellent at identifying MDD but overdiagnosed the condition (sensitivity=92.2%; specificity=54.3%; accuracy=66.2%).

Conclusions: Our findings are in line with the notion that current diagnostic practices may present an overly narrow conception of mental health. Furthermore, our study provides proof-of-concept support for the clinical utility of a digital assessment to inform clinical decision-making in the evaluation of MDD.

(JMIR Form Res 2021;5(10):e27908) doi:10.2196/27908

KEYWORDS

major depressive disorder; subthreshold depression, transdiagnostic symptoms; digital assessment; digital mental health; mobile phone

Introduction

Background

Major depressive disorder (MDD) is a common and heterogeneous condition representing the leading cause of disability worldwide [1]. MDD has been associated with poor global outcomes, including impaired social functioning, lower quality of life, inability to return to work, and suicide [2]. The condition is typically diagnosed in primary care settings, with most help seekers exhibiting subthreshold or subsyndromal presentations of the disorder [3,4]. Critically, recognizing diagnosable symptomatology of MDD can be particularly challenging, with any 2 individuals' meeting criteria for the condition potentially having no symptoms in common [5]. In fact, short consultation times coupled with the complexity and subjectivity of diagnosing MDD results in primary care practitioners misdiagnosing >50% of low-mood help seekers [6]. This means that many patients do not receive the most effective treatment and support.

In an attempt to improve the current diagnostic practice, the search for objective diagnostic tests and valid biomarkers for depression has received a lot of attention. However, despite substantial research expenditures and large-scale genome-wide studies, no pathognomonic biological markers of depression have been identified [7-11]. In fact, with the exception of a few neuropsychiatric disorders, not a single psychiatric diagnosis can still be validated by molecular, genetic, or imaging biomarkers [12]. Importantly, psychiatric diagnostic criteria were not conceptualized to facilitate biological differentiation [13], with extensive comorbidity across conditions being the rule rather than the exception [14] and no single condition representing a discrete entity [15-17].

Another issue pertaining to psychiatric nosology is *incomplete symptom capture* [18,19]. It has been argued that the symptom profiles described in diagnostic manuals, such as the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (*DSM-5*) [20] and the *International Statistical Classification of Diseases, Eleventh Revision* (*ICD-11*) [21], may be overly narrow, failing to capture the wide range of clinical symptoms that are endorsed by individuals with MDD [18]. For instance, although anxiety is not listed as a core symptom of the condition, many individuals with MDD experience co-occurring symptoms of anxiety and typically meet the criteria for at least one anxiety disorder [22-26].

Some authors suggest that assessing the presence of anxiety symptoms in patients with MDD is critical [27,28]. Others propose combining depression and anxiety disorders, which present with largely overlapping symptomatology. Combining these disorders may be a useful strategy for better clinical evaluation and management of patients with MDD [29]. This is important, given that estimates of prevalence rates for treatment-resistant depression range between 30% and 50% [30,31], with incomplete remission often leading to relapse [32], increased chronicity and severity of episodes [33], greater functional impairments [34], and higher risk of suicide [35]. One of the established risk factors that predispose patients to develop treatment-resistant depression are comorbid anxiety symptoms or anxiety disorders, especially generalized anxiety disorder (GAD) [31].

In this regard, a transdiagnostic view of MDD encompassing symptoms of anxiety and other commonly co-occurring disorders may improve early and accurate diagnosis, reflect biological disease understanding (eg, twin studies have shown shared genetic predisposition for MDD and GAD [36]), and allow for personalized treatment strategies. An extended definition of MDD symptomatology may also reduce the misdiagnosis of bipolar disorder (BD) as MDD [37,38], which is particularly problematic, with many individuals having to wait 8-10 years before receiving a correct diagnosis [39,40].

Critically, time is premium in primary care settings, where relying on brief symptom-count checklists, such as the Patient Health Questionnaire–9 (PHQ-9) [41], is a common practice. Importantly, the PHQ-9 may overestimate depression severity in primary care patients relative to other self- and clinician-rated scales [42], resulting in a greater reliance on medication as a first-line treatment option and an increased potential for adverse drug effects [43,44]. In addition, some researchers have suggested that the PHQ-9 may be missing the presence of symptoms that are meaningful for patients and that longer assessments may be better at capturing diagnosable levels of low mood [45]. To this end, digital technologies allow for the cost- and time-effective collection of a vast range of important patient and symptom data [46]. Such an approach offers an innovative way to improve and advance mental health care provision. In turn, the use of digital technologies could help alleviate the load on the health care system by providing individuals with subthreshold or mild MDD with self-help tips and psychoeducation. This would reserve the limited and specialized services for more severe or highly comorbid and complex patients.

Objectives of This Study

This study aims to provide evidence for an extended definition of MDD symptomatology using a digital assessment that was developed for the delta study [47]. The digital assessment was designed following an extensive analysis of existing validated questionnaires for mood disorders [48-57], the *DSM-5* [20], and the *International Statistical Classification of Diseases, Tenth Revision* (*ICD-10*) [21], as well as input from psychiatrists and a service user group. In an attempt to move away from a symptom-count approach to psychopathology, our study uses machine learning (ML) methods (advanced statistical and probabilistic techniques that automatically learn from data) to construct a data-driven view of MDD. We examined the extent to which (1) disorder-specific symptoms (ie, symptoms of depression) and (2) transdiagnostic symptoms (ie, cross-disorder symptoms) could be used to answer the following question: "when does depression become a mental disorder?" To do this, we compared individuals with MDD with those with subthreshold levels of depressive symptoms. Although this was largely an exploratory study, it was predicted that, relative to a disorder-specific model of psychopathology, an extended model would be better at identifying individuals with diagnosable levels of MDD. In particular, we predicted that symptoms of anxiety would be highly indicative of the disorder.

Methods

The Delta Study

This study used data from the delta study that was conducted by the Cambridge Centre for Neuropsychiatric Research between April 2018 and November 2019. Olmert et al [47] provided a detailed description of the delta study design and sampling procedures. In brief, the key objectives of the study were to develop and validate a diagnostic algorithm to (1) reduce the misdiagnosis of MDD and BD and (2) achieve a more accurate and earlier diagnosis of MDD in individuals presenting with depressive symptoms. The target population for the primary objective was those who had received a recent diagnosis of MDD (within the past 5 years) by a general practitioner or psychiatrist and those who were experiencing depressive symptoms at the time of recruitment. The target population for the secondary objective included those without a previous mood disorder diagnosis and who were experiencing depressive symptoms at the time of recruitment. Individuals aged between 18 and 45 years could take part in this study. This age group was selected in consultation with a practicing psychiatrist (SB) on the basis that individuals aged between 18 and 45 years are most likely to have undiagnosed BD (primary objective of the delta study). Further inclusion criteria were being a resident in the United Kingdom, not pregnant or breastfeeding, not suicidal, and a score of at least five on the PHQ-9 [41]. Information on treatment history was collected but was not deemed an inclusion or exclusion criterion. All participants provided informed consent to participate in the study, which was approved by the University of Cambridge Human Biology Research Ethics Committee (approval number HBREC 2017.11).

Over 5000 participants were recruited on the web through email, via paid Facebook (Facebook Inc) advertisements, and updates on the Cambridge Centre for Neuropsychiatric Research laboratory website. Eligible participants were invited to take part in the main study; of these, 3232 completed the digital assessment via the delta study website. The digital assessment was designed following an extensive analysis of validated questionnaires for mood disorders [48-57], the *DSM-5* [20], and the *ICD-10* [21], as well as input from psychiatrists to ensure the inclusion of a wide range of clinically meaningful and well-validated symptoms of MDD and BD and other symptoms of interest (eg, other psychiatric conditions). The assessment was further refined following advice from a service user group on features, including tone of voice and user journey. For each participant, an individualized dashboard guided their progress through the study. The digital assessment could be completed on a smartphone, laptop, tablet, or desktop computer and comprised 635 distinct questions. One question was presented at a time, with participants required to select the answer that best described their feelings and experiences (eg, from "No, not at all" to "Yes, very much"). The questions were grouped into six sections: (1) demographics and personal history; (2) bipolar and manic and hypomanic symptoms; (3) depressive symptoms; (4) personality traits; (5) medication, treatment, and substance use; and (6) other psychiatric disorders and symptoms, including GAD, social anxiety disorder, emotional instability, panic disorder, eating disorders, insomnia, and obsessive-compulsive disorder. Participants were required to press *next* to take them to the following question. Each section could be completed in 10-15 minutes, although response times varied because of the adaptive nature of the assessment, where only relevant questions were asked based on responses to previous questions.

A subgroup of the original study cohort (n=1740) consented to provide dried blood spot samples and complete a telephone interview for MDD and BD using the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition* (*DSM-IV*)–based Composite International Diagnostic Interview (CIDI [48]), with 924 participants completing both steps. Of these 924 participants, 241 (26.1%) self-reported having been diagnosed with MDD by a general practitioner or psychiatrist and were confirmed to have met the criteria for the condition by the CIDI. Of these 241 participants, 64 (26.6%) participants (11/64, 17% participants were male) met the criteria for *current* MDD (ie, symptoms present in the past month according to the CIDI). This formed the MDD group. The subthreshold depression group included 15.2% (140/924) participants (male: 35/140, 25%) who self-reported no diagnosis of MDD and whose symptoms of depression were confirmed to not meet the criteria for MDD according to the CIDI. None of the selected participants had ever experienced a manic or hypomanic episode.

Data Analytic Strategy

Participant Characteristics

Participant characteristics and comorbidities were collected via digital assessments and are shown in Table 1. Group differences in continuous variables were explored using Mann–Whitney U tests as the data were nonnormally distributed, with effect sizes reported as r (small ≥0.1; medium ≥0.3; and large ≥0.5) [58]. Group differences in categorical variables were evaluated using chi-square tests or Fisher exact test for low-frequency data (ie, values <5). Effect sizes are reported as Cramer V (ϕ_c; small ≥0.1; medium ≥0.3; and large ≥0.5) [58].

Table 1. Participant characteristics and comorbidities: major depressive disorder versus subthreshold depression group comparisons.

Characteristics	Subthreshold depression (n=140)	MDD[a] (n=64)	*U* [b]	*P* value	*r* [c]	Chi-square (*df*)	ϕ_c [d]
Age (years), mean (SD)	25.84 (6.66)	25.94 (5.7)	4209	.49	0.05	N/A[e]	N/A
BMI, mean (SD)	24.62 (4.73)	28.29 (6.8)	2901	<.001	0.28	N/A	N/A
Sex, n (%)							
Male	49 (35)	11 (17.2)	N/A	N/A	N/A	N/A	N/A
Female	91 (65)	53 (82.8)	N/A	.01	N/A	6.7 (1)	0.18
Higher education[f], n (%)							
Yes	87 (62.1)	37 (42.2)	N/A	N/A	N/A	N/A	N/A
No	53 (37.1)	27 (57.8)	N/A	.56	N/A	0.4 (1)	0.04
Employment, n (%)							
Employed	81 (57.9)	35 (54)	N/A	N/A	N/A	N/A	N/A
Unemployed	5 (3.6)	8 (12.5)	N/A	N/A	N/A	N/A	N/A
Student	54 (38.6)	21 (32.8)	N/A	.05	N/A	6.0 (2)	0.17
Support network: relationships, n (%)							
Secure and stable relationship	84 (59.6)	133 (55.4)	N/A	N/A	N/A	N/A	N/A
Insecure and unstable relationship	9 (6.4)	14 (5.8)	N/A	N/A	N/A	N/A	N/A
Single	48 (34)	93 (38.8)	N/A	.54	N/A	1.3 (2)	0.08
Support network: living alone, n (%)							
Yes	9 (6.4)	9 (14.1)	N/A	N/A	N/A	N/A	N/A
No	131 (93.6)	55 (85.9)	N/A	.07	N/A	3.2 (1)	0.13
Psychiatric history, n (%)							
GAD[g]	12 (8.6)	47 (73.4)	N/A	<.001	N/A	89.9	0.66
Personality disorder	0	3 (4.7)	N/A	N/A	N/A	N/A	N/A
OCD[h]	3 (2.1)	3 (4.7)	N/A	.28	N/A	0.4 (2)	0.07
Panic disorder	0	7 (10.9)	N/A	N/A	N/A	N/A	N/A
Social anxiety	1 (0.7)	9 (14.1)	N/A	<.001	N/A	16.8 (2)	0.29
Eating disorder	2 (1.4)	5 (7.8)	N/A	.03	N/A	5.4 (2)	0.16
Medical history, n (%)							
Thyroid disease	4 (2.9)	2 (3.1)	N/A	.99	N/A	0.0 (3)	0.01
Cardiovascular disease	1 (0.7)	0	N/A	N/A	N/A	N/A	N/A
Irritable bowel syndrome	6 (4.3)	2 (3.1)	N/A	.99	N/A	0.2 (3)	0.03
Chronic pain	30 (21.4)	14 (21.9)	N/A	.94	N/A	0.0 (3)	0.01
Migraines	53 (37.9)	31 (48.4)	N/A	.15	N/A	2.0 (3)	0.10
Current psychiatric treatment, n (%)							
SSRI[i] antidepressants	12 (8.6)	32 (50)	N/A	<.001	N/A	44.6 (3)	0.47
SNRI[j] antidepressants	0	5 (7.8)	N/A	N/A	N/A	N/A	N/A
Tricyclic antidepressants	0	3 (4.7)	N/A	N/A	N/A	N/A	N/A
Other antidepressants	1 (0.7)	8 (12.5)	N/A	<.001	N/A	14.5 (3)	0.27
Anxiety medication	4 (2.9)	10 (15.6)	N/A	.002	N/A	11.2 (3)	0.23
Antipsychotics	0	3 (4.7)	N/A	N/A	N/A	N/A	N/A
Mood stabilizers	0	2 (3.1)	N/A	N/A	N/A	N/A	N/A

Characteristics	Subthreshold depression (n=140)	MDD[a] (n=64)	U [b]	P value	r [c]	Chi-square (df)	ϕ_c [d]
Psychotherapy	4 (2.9)	16 (25)	N/A	<.001	N/A	24.4 (3)	0.35

[a]MDD: major depressive disorder.

[b]Mann–Whitney U test.

[c]Effect size (r).

[d]Effect size (Cramer V).

[e]N/A: not applicable.

[f]Undergraduate degree or equivalent and above was coded as *yes*, whereas A level or equivalent and below was coded as *no*.

[g]GAD: generalized anxiety disorder.

[h]OCD: obsessive-compulsive disorder.

[i]SNRI: serotonin–norepinephrine reuptake inhibitor.

[j]SSRI: selective serotonin reuptake inhibitor.

Model Construction and Performance

Random forest classification models were constructed in Python 3.7.4 (Python Software Foundation) using the scikit-learn library 0.21.3 to distinguish between MDD and subthreshold depression using (1) disorder-specific symptoms (ie, symptoms of depression), and (2) transdiagnostic symptoms (ie, cross-disorder symptoms). We constructed two models: a *depression model*, including 36 symptoms of depression and an *extended model*, comprising 134 symptoms (36 symptoms of depression, 12 symptoms of GAD, 19 symptoms of BD or mania, 15 symptoms of hypomania, 6 symptoms of social anxiety, 11 symptoms of emotional instability, 14 symptoms of panic disorder, nine symptoms of obsessive-compulsive disorder, two symptoms of eating disorders, and 10 symptoms of insomnia). These symptoms were manually selected based on the maximum number of available symptoms from the digital assessment. Scores per symptom ranged from 0-1, with higher scores indicating increased severity. Mean symptom severity per group can be found in Table S1, Multimedia Appendix 1.

Although some symptoms overlapped across disorders (eg, tiredness, low energy, and irritability), we did not feel that it would be appropriate to combine these as the questions were framed in the context of each condition. Furthermore, although all participants were asked about symptoms of depression, BD or mania, and hypomania, the questions for the remaining conditions were adaptive in nature, such that only relevant questions were asked based on responses to previous questions. This resulted in some participants having *missing* data. These data were imputed as zeros. Furthermore, owing to the adaptive nature of the digital mental health assessment, participants answered questions on current (eg, present in the past 2 weeks) or past symptoms of all disorders.

For each of the models, nested cross-validation (NCV) was performed to obtain the highest algorithmic accuracy while ensuring the generalizability of the models. At each iteration of NCV, the data were randomly split into three folds; two-thirds of the data were used in the inner loop for model training and validation, and one-third was used for testing the model in the outer loop. The inner loop was (further) randomly split into three folds, whereby the hyperparameters (ie, number of estimators and maximum depth) were tuned, and the best cross-validated model was selected. To do this, two of the three folds were used to tune the model parameters and train the model, which was then validated on the third fold. This procedure was repeated with the remaining combinations of training and validation folds. The final model (ie, the optimized classifier) was obtained by fitting a model with the tuned parameters to all three data folds from the inner loop and then evaluating the hold-out test data in the outer loop. This procedure was repeated 100 times with different splits of the data (into train and test sets), resulting in a total of 300 unique models for each feature set (ie, depression model vs extended model).

Model performance was evaluated by measuring the area under the receiver operating characteristic curve (AUC) for the 300 models and averaging across all models for each feature set. The AUC shows the degree of separability between two conditions (ie, MDD vs subthreshold depression) and represents the probability that a randomly selected subject with the condition is rated or ranked as more likely to have the condition than a randomly selected individual without the condition (AUC: ≥0.9=excellent; ≥0.8=good; ≥0.7=fair; ≥0.6=poor; ≥0.5=fail) [59]. Mann–Whitney U tests were used to determine significant differences in AUCs across the 300 models between the depression and transdiagnostic models.

The mean sensitivity, specificity, and accuracy scores per model were also evaluated. Here, sensitivity refers to the model's ability to classify MDD cases correctly (ie, true positives), whereas specificity refers to the model's ability to classify subthreshold depression cases correctly (ie, true negatives). Accuracy corresponds to the model's ability to classify all true cases (ie, both true positives and true negatives).

Feature Importance and Occurrence

Relative feature importances (ie, Gini impurity [60]) were calculated for the 300 models and averaged across all models for each feature set, with features with higher values showing better discrimination between MDD and subthreshold depression. Feature occurrence was calculated by summing the number of times each feature contributed to each of the 300 models and computing a percentage score per feature for each feature set, with higher values representing higher feature occurrence.

Diagnostic Performance of the PHQ-9

Finally, to establish the diagnostic performance of the PHQ-9 on the basis of its intended use, we calculated the sensitivity, specificity, and accuracy in the current sample using the standard cut-off score of ≥10 [41]. This mimics what would ordinarily happen in the clinic (ie, those scoring ≥10 would be classified as MDD).

Results

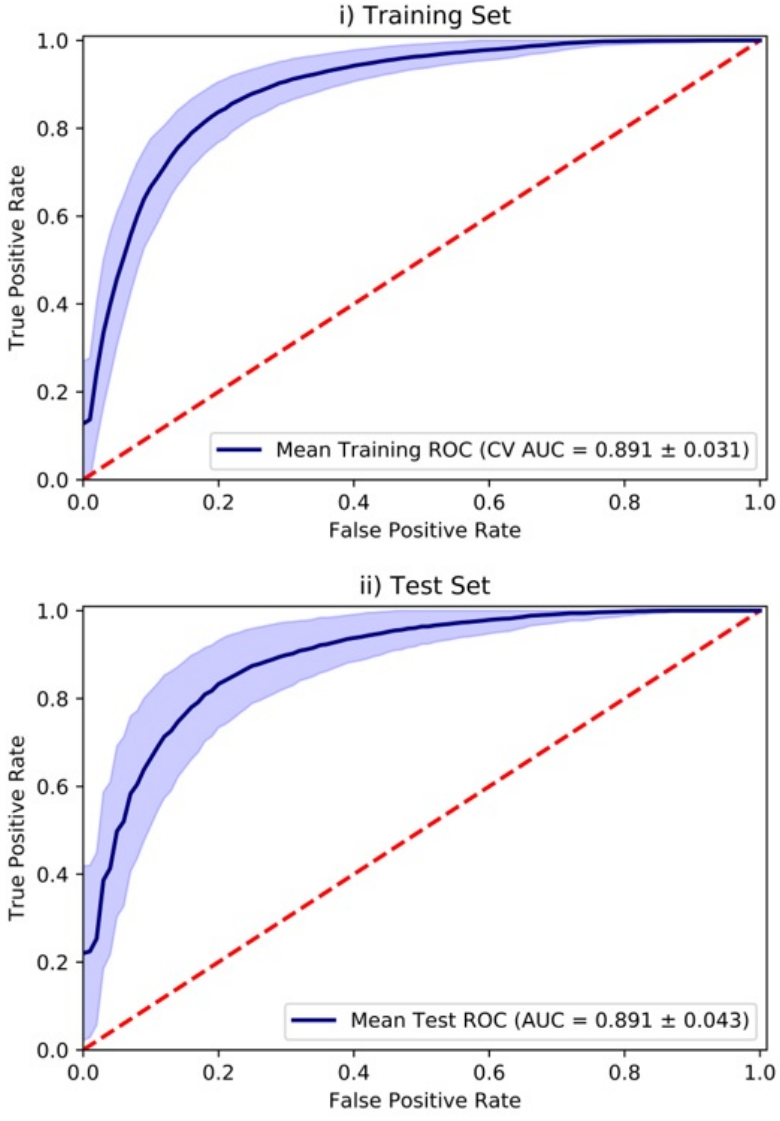
Participant Characteristics

Table 1 presents information on the characteristics of each group, with statistical comparisons. The groups did not differ significantly in age, level of education, support network (ie, relationships and living conditions), or medical history. However, the MDD group had a significantly higher proportion of women and a higher mean BMI than the subthreshold depression group. The MDD group was significantly more likely to be unemployed than the subthreshold depression group, which was more likely to be employed or in full-time education. The groups had different psychiatric histories, with the MDD group having a significantly higher proportion of individuals with comorbid GAD, social anxiety disorder, and eating disorders. Finally, relative to the subthreshold depression group, individuals with MDD were more likely to be currently taking psychiatric medication and receiving psychotherapy.

Depression Model

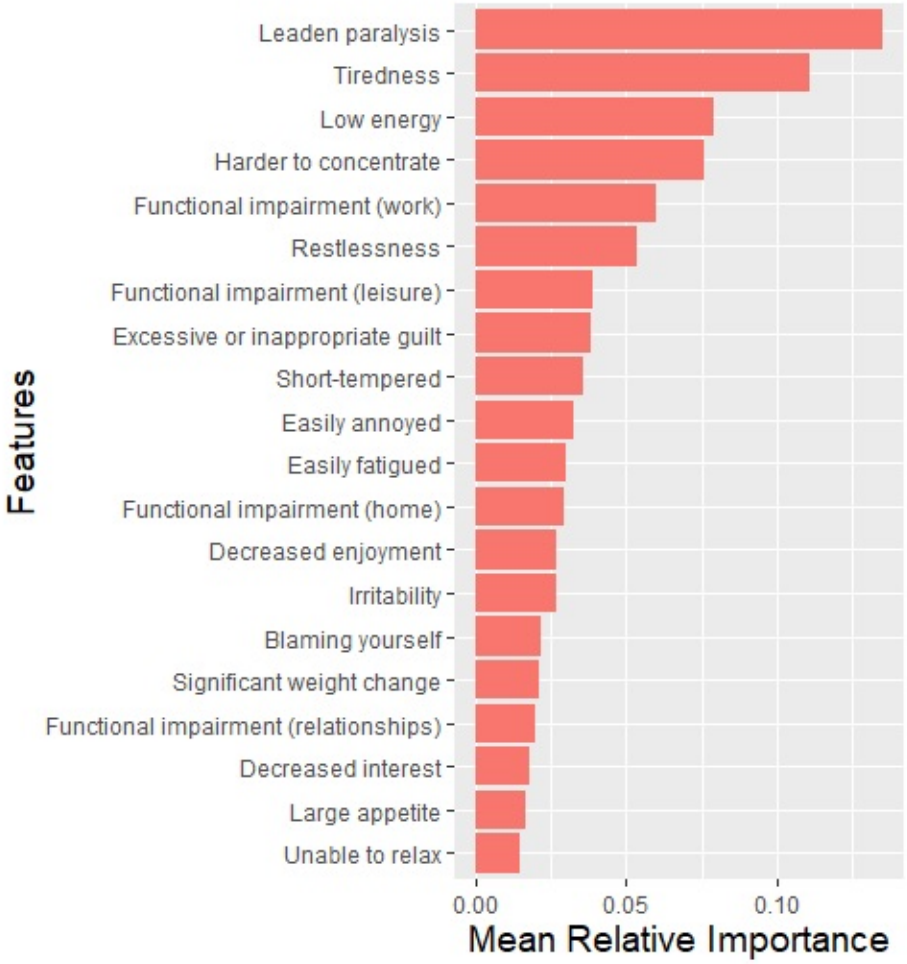
This model comprised 36 features, with analyses demonstrating good discriminatory performance on both the training (AUC=0.89±0.03) and test sets (AUC=0.89±0.04; Figure 1). Approximately 81% (52/64) of the MDD cases and 81.4% (114/140) of the subthreshold depression cases were correctly classified by the model, corresponding to the mean sensitivity and specificity scores, respectively. The mean accuracy of the model was 81.6%, corresponding to the proportion of individuals correctly classified by the model.

Figure 1. Area under the receiver operating characteristic curves showing mean predictive performance of the depression model. The models were applied to predict the probability of major depressive disorder in the: (1) training and (2) test sets. AUC: area under the receiver operating characteristic curve; CV AUC: cross-validated area under the receiver operating characteristic curve; MDD: major depressive disorder; ROC: receiver operating characteristic.



The top 20 features contributing to the depression model (averaged across all 300 models) were leaden paralysis, tiredness, low energy, harder to concentrate, functional impairment (work), restlessness, functional impairment (leisure), excessive or inappropriate guilt, short-tempered, easily annoyed, easily fatigued, functional impairment (home), decreased enjoyment, irritability, blaming oneself, significant weight change, functional impairment (relationships), decreased interest, large appetite, and unable to relax. Figure 2 shows the mean relative feature importances. These features appeared across at least 81.3% (244/300) of the models. Multimedia Appendices 2 and 3 show the relative feature importances and percentage occurrences of all 36 features.

Figure 2. Top 20 mean relative importance for the depression-specific model. Features have been ordered from most to least important.



Extended Model

Next, we added 98 features to the model, resulting in an extended model comprising 134 features. Analyses demonstrated excellent discriminatory performance on both the training (AUC=0.94±0.03) and test sets (AUC=0.94±0.04; Multimedia Appendix 4). Mann–Whitney U tests confirmed a significant improvement in model performance (ie, AUC) relative to the depression-specific model (training set: U=12922.50, P<.001; test set: U=33525.50, P<.001). Here, 83% (53/64) of MDD cases (ie, sensitivity) and 90% (126/140) of subthreshold depression cases (ie, specificity) were correctly classified by the model, whereas the ability of the model to correctly classify both MDD and subthreshold depression cases was 87.7% (ie, accuracy). Feature importances can be found in Multimedia Appendix 5, with percentage occurrences found in Multimedia Appendix 6.

On the basis of these findings, we then reran the analyses using a *truncated* version of the extended model, which only included features that appeared across at least 90.3% (271/300) of the models (Multimedia Appendix 6). The truncated model comprised 12 symptoms of depression, 11 symptoms of GAD, six symptoms of insomnia, three symptoms indicative of emotional instability, and one panic disorder symptom, resulting in a total of 33 features.

The analyses revealed a significant improvement in the model's discriminatory performance on both the training (AUC=0.95±0.02) and test sets (AUC=0.95±0.03; Figure 3) relative to the full extended model (training set: U=12856, P<.001; test set: U=322798.50, P<.001). The mean sensitivity, specificity, and accuracy scores were 86.5%, 90.8%, and 89.5%, respectively. The mean relative feature importance for the 33 features can be found in Figure 4, with percentage occurrences found in Multimedia Appendix 7.

Figure 3. Area under the receiver operating characteristic curves showing mean model performance of the truncated version of the extended model. The models were applied to predict the probability of major depressive disorder in the: (1) training and (2) test sets. AUC: area under the receiver operating characteristic curve; CV AUC: cross-validated area under the receiver operating characteristic curve; MDD: major depressive disorder; ROC: receiver operating characteristic.

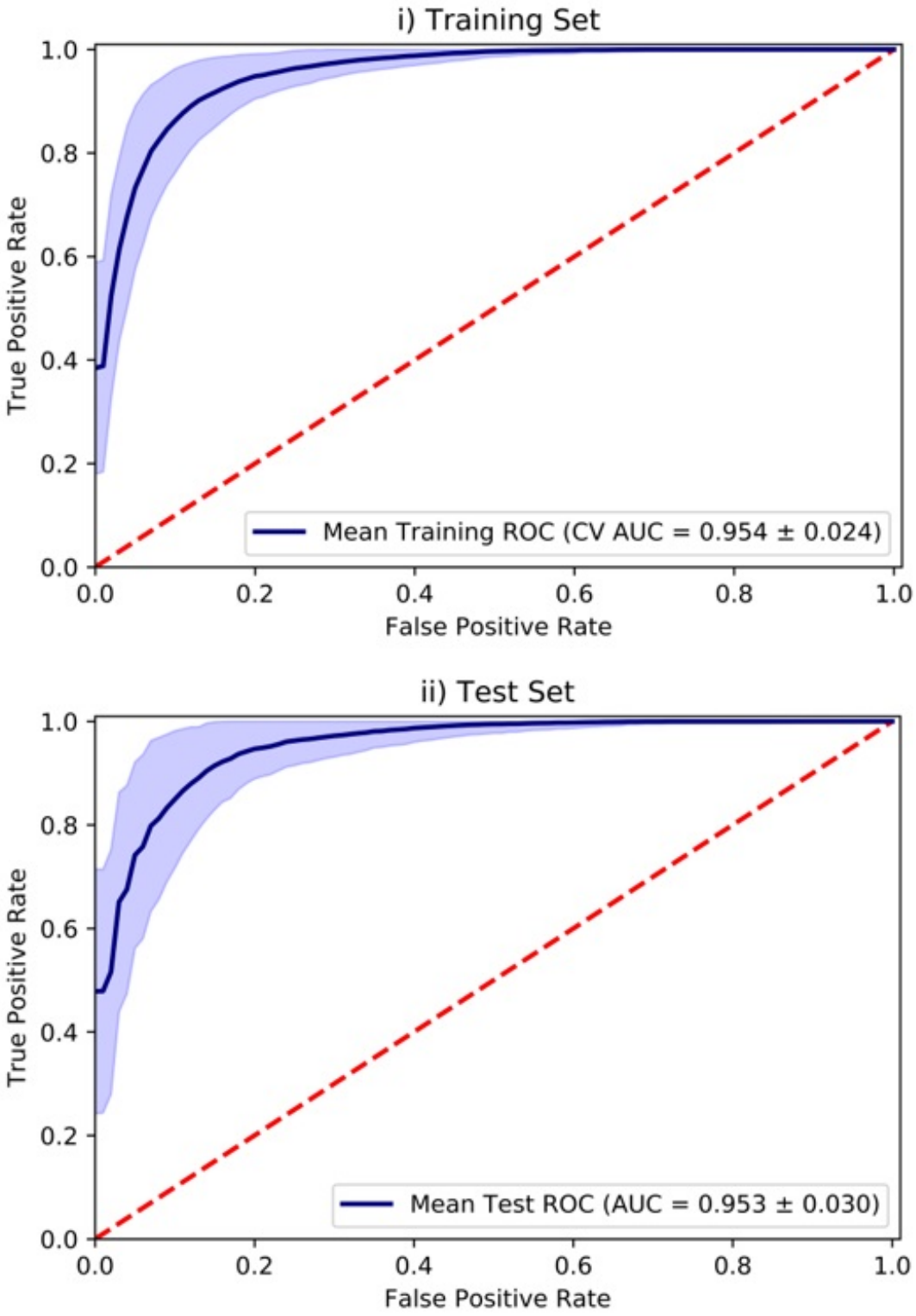
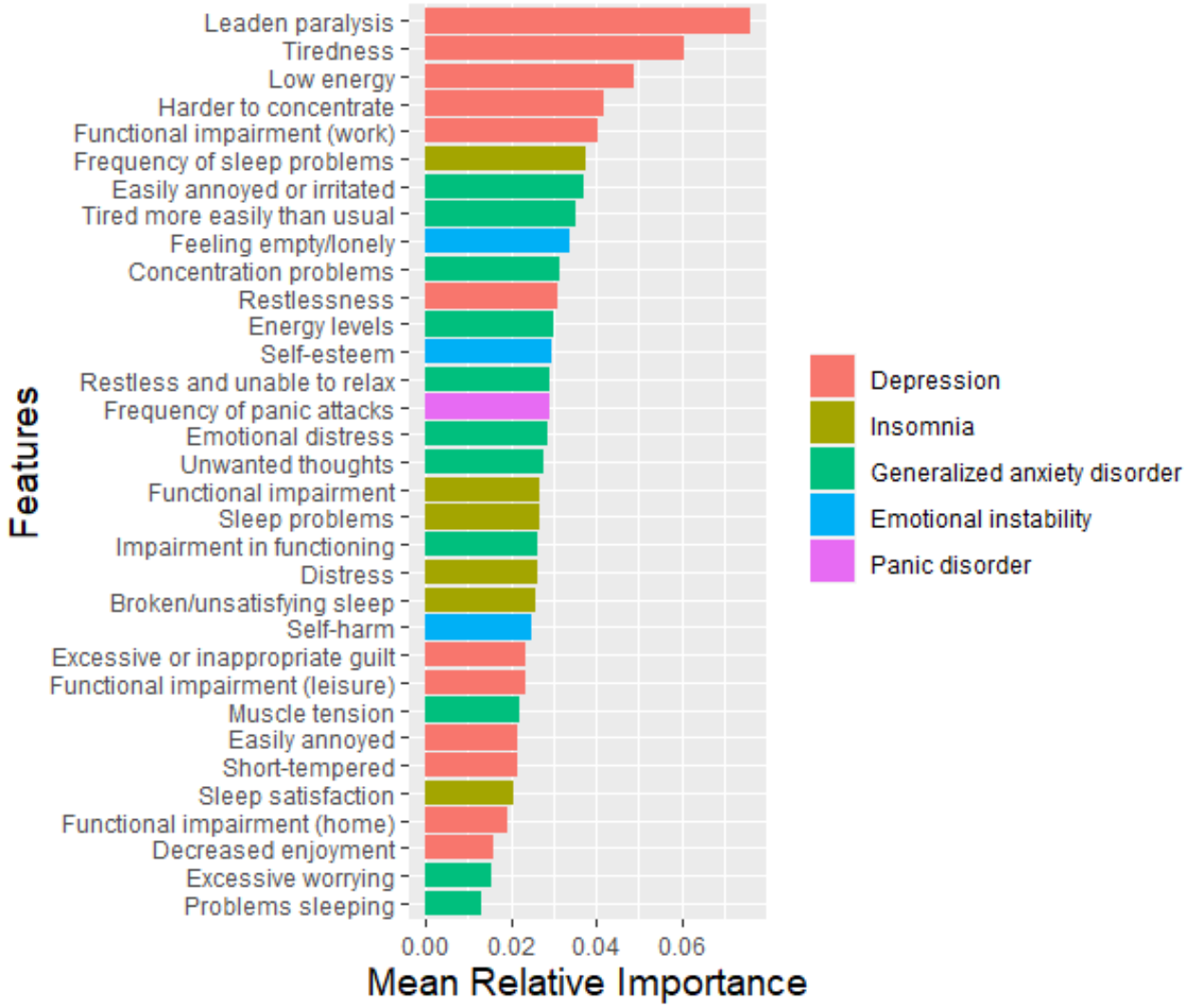


Figure 4. Mean relative importance for the 33 features in the truncated version of the extended model. Features have been ordered from most to least important and colored according to the disorder or symptom cluster they correspond to.



Diagnostic Performance of the PHQ-9

The sensitivity of the PHQ-9 for detecting MDD was 92.2%, whereas the specificity was 54.3%, and the overall diagnostic accuracy was 66.2%.

Discussion

Principal Findings

This study provides evidence for an extended definition of MDD symptomatology and supports the use of a digital assessment as an aid to clinical decision-making in the identification of MDD. Relative to a disorder-specific model of MDD psychopathology, an extended model of symptomatology was better at distinguishing between individuals with MDD and those with subthreshold levels of the disorder. In particular, a truncated version of the model, comprising symptoms of depression, GAD, insomnia, emotional instability, and panic disorder, demonstrated excellent predictive performance (AUC=0.95; sensitivity=86.5%; specificity=90.8%; and accuracy=89.5%).

Critically, although the PHQ-9 was particularly good at detecting MDD in the current sample, it tended to overdiagnose MDD in subthreshold depression and, in turn, was associated with poor overall diagnostic performance. Overdiagnosis of MDD presents a significant problem and has the potential for antidepressant overprescription and adverse drug effects in individuals who may benefit from alternative treatment options [43], such as psychotherapy or psychoeducation. Furthermore, relying on a simple cut-off score does not allow for personalized treatment plans and strategies, potentially resulting in incomplete remission rates.

Overall, the findings from our models are in line with the notion that current diagnostic practices may present a narrow conception of mental health that does not allow for the wide range of clinical signs and symptoms that are endorsed by individuals with MDD. Across our models, the most predictive symptom of MDD was leaden paralysis, which refers to an extreme form of fatigue or heavy, leaden feelings in the arms and legs. This finding is in line with a recent study by Han et al [61], whose findings revealed that leaden paralysis was a

robust and important predictor of first-onset MDD. Critically, although leaden paralysis is included in the *DSM-5* specifier for atypical depression, it is not deemed a core feature of the disorder [20]. In fact, even fatigue or loss of energy and tiredness are not considered essential symptoms according to the *DSM-5* [20]. Importantly, exhaustion, extreme tiredness, and loss of energy are typically seen in primary care settings and are often the predominant presenting complaint [62]. In fact, in a large European study comprising approximately 2000 depressed primary care patients across 6 countries, almost two-thirds of patients reported feeling tired [63]. These findings suggest that leaden paralysis and its lesser extreme variants (ie, tiredness and low energy) may be particularly important for the recognition of MDD in the primary care setting.

As expected, symptoms of GAD were among the most predictive features of MDD, with *overlapping* symptoms between GAD and depression (eg, tiredness, low energy, and irritability) being particularly indicative of the disorder. For instance, regarding the latter, irritability has been seen to occur in one-third to one-half of patients with MDD [64-66] and is associated with greater severity and chronicity, a history of suicide attempts, and reduced quality of life [64]. Other GAD-specific symptoms that were highly predictive of MDD included unwanted thoughts, excessive worrying, and emotional distress. Importantly, it could be argued that our model simply reflects the significantly higher rates of comorbidity in the MDD group relative to the subthreshold depression group, particularly with regard to GAD. Similarly, our model may capture higher levels of severity or a higher p factor in the MDD group [67] rather than important components of the condition.

Although our findings should be interpreted with caution, our view is that this should not detract from the importance of assessing for transdiagnostic symptoms of MDD, especially as these are likely to share common underlying pathophysiology and genetics [36,68]. Indeed, the evaluation of anxiety symptoms in the context of MDD is critical, particularly given the chronicity of the conditions, with research indicating that anxiety disorders may be a precursor to MDD [69]. Notably, combining clinical information with biological biomarkers, such as serum analytes, can be used to predict the development of future depressive episodes in individuals presenting with social anxiety [70] and panic disorder [71]. Identifying those who may be at a heightened risk for comorbid anxiety and depression is of clinical importance, particularly given that these individuals are likely to exhibit more pervasive and recurrent forms of illness, reduced remission rates, and increased suicidality [72-74].

Indeed, suicidality has particularly been associated with the co-occurrence of depression and panic disorder [75,76], with our findings indicating that the frequency of panic attacks was an important predictor of MDD. Importantly, comorbid panic disorder and MDD have been related to increased depression severity, an earlier age of onset, increased functional impairment, and a poorer clinical prognosis [77], suggesting that assessing for panic disorder may be important when diagnosing and treating MDD. Similarly, our findings indicate that a more in-depth evaluation of the symptoms associated with sleep problems or insomnia, including emotional distress caused by disordered sleep, warrants inclusion when diagnosing MDD. This is particularly important as sleep problems have been shown to reduce the efficacy of depression treatment [78].

Finally, symptoms of emotional instability or personality disorder, including feeling empty, low self-esteem, and self-harm, were also seen to be important when distinguishing between MDD and subthreshold depression. Feeling empty or chronic emptiness has been closely related to depression and suicidal ideation [79]. Interestingly, a recent qualitative study in adolescents with depression revealed that a partial or complete blunting of any emotion (negative or positive), with feelings of flatness, emptiness, and lack of emotions, was an important component of anhedonia [80]. It is interesting that self-esteem but not self-worth (a diagnostic criterion for MDD) was a predictor of the disorder. This finding suggests that wording may also have an important impact on individuals' subjective evaluations of their symptoms, with the concept of self-esteem perhaps being easier to grasp than that of self-worth. Finally, although diagnostic descriptions of MDD symptomatology include suicidality as a criterion for the disorder, expanding this to include self-harm may facilitate its identification.

Taken together, our findings indicate that the current diagnostic criteria for MDD may fail to evaluate relevant clinical information that is important for the diagnosis and treatment of individuals with the disorder. Although time is a luxury in the primary care setting, our study supports the use of digital technologies as a means for obtaining a more comprehensive depiction of MDD symptomatology in a time-efficient manner. Notably, related research using the same digital mental health assessment has highlighted the utility of the tool in distinguishing individuals with MDD from those with BD [81]. Indeed, digital technologies have the potential to aid in the recognition of a wide range of psychiatric conditions, allowing for more time to be spent managing and treating symptoms. In turn, digital technologies can reduce the number of in-person appointments, alleviate health care professionals' workload, and reduce the risk of burnout. The use of digital technologies also has the potential to reduce some of the barriers associated with disclosing mental health difficulties, such as discomfort as well as issues related to stigma and discrimination. Furthermore, research has demonstrated that patients are more likely to report severe symptoms on technology platforms than to a health care professional [82] and value the independence and empowerment that can be obtained from the use of a digital platform [83].

Strengths and Limitations

To our knowledge, this is the first study to provide evidence for an extended definition of MDD symptomatology using a digital assessment. Furthermore, the digital assessment was designed following an extensive analysis of existing validated questionnaires for psychiatric disorders and diagnostic manuals, as well as input from psychiatrists and a service user group. In addition, as opposed to the use of healthy controls as a reference population against which patients are compared, our subthreshold depression group represented a clinically relevant reference group. Finally, the use of ML methods meant that patterns in data could be more readily and accurately identified, whereas our NCV approach allowed us to obtain high

algorithmic accuracy while ensuring the generalizability of the models.

This study also had several limitations. First, as with any supervised ML approach to psychopathology, our analyses were limited by the *truthfulness* of the diagnostic labels (MDD vs subthreshold depression). Second, given the adaptive (nonlinear) nature of the question flow, missing data were imputed with zeros, which may have resulted in an overly artificial data set. This should be borne in mind when interpreting these findings. In addition, the MDD sample was small in size and primarily comprised women, reflecting the higher prevalence of MDD in women than in men [84] and the difficulties in recruiting males with MDD. Furthermore, given that suicidality was an exclusion criterion in this study, measures of suicidal thoughts, ideations, plans, or impulses do not appear in our list of important features of MDD. This is a key limitation of the study, as suicidality denotes an important component of the condition. Similarly, symptoms from other disorders that frequently co-occur with MDD, such as posttraumatic stress disorder, were not available for inclusion in our analyses, which may have allowed for a more comprehensive depiction of MDD symptomatology.

Conclusions

In an attempt to answer the question "when does depression become a mental disorder?", our study demonstrated that a data-driven view of MDD may improve our understanding of the condition. A more comprehensive conceptualization of the psychopathology of MDD, including symptoms of depression, GAD, insomnia, panic disorder, and emotional instability, may not only facilitate patient stratification but also allow for personalized treatment plans and strategies. Although further studies with larger sample sizes are required to replicate our findings, our study shines a positive light on the use of digital technologies as an innovative way to help develop and facilitate mental health care provision. In particular, digital technologies have the capacity to collect a vast range of key clinical information that may be important for the diagnosis and treatment of individuals with MDD.

Acknowledgments

This study was funded by the Stanley Medical Research Institute (grant number: 07R-1888). The authors thank the participants for taking part in the delta study, as well as the CIDI interviewers, without whom this work would not have been possible.

Conflicts of Interest

SB is a director of Psynova Neurotech Ltd and Psyomics Ltd. SB, DC, GBO, LF, and EB have financial interests in Psyomics, Ltd, which provided funding and support for the delta study. SB, PE, and TO could benefit financially from any product that arises from work performed in the delta study.

Multimedia Appendix 1
Mean symptom severity per group was separated by disorder or symptom clusters.
[DOCX File , 30 KB - formative_v5i10e27908_app1.docx]

Multimedia Appendix 2
Depression model: mean relative feature importance.
[DOCX File , 15 KB - formative_v5i10e27908_app2.docx]

Multimedia Appendix 3
Depression model: percentage feature occurrences.
[DOCX File , 14 KB - formative_v5i10e27908_app3.docx]

Multimedia Appendix 4
Receiver operating characteristic curves showing the mean predictive performance of the extended model. The models were applied to predict the probability of major depressive disorder in the training and test sets.
[DOCX File , 141 KB - formative_v5i10e27908_app4.docx]

Multimedia Appendix 5
Extended model: mean relative feature importance colored by disorder or symptom clusters.
[DOCX File , 33 KB - formative_v5i10e27908_app5.docx]

Multimedia Appendix 6
Extended model: percentage feature occurrences colored by disorder or symptom clusters.
[DOCX File , 32 KB - formative_v5i10e27908_app6.docx]

Multimedia Appendix 7
Truncated model: percentage of feature occurrences colored by disorder or symptom clusters.
[DOCX File , 20 KB - formative_v5i10e27908_app7.docx]

References

1. Ferrari AJ, Charlson FJ, Norman RE, Patten SB, Freedman G, Murray CJ, et al. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. PLoS Med 2013 Nov;10(11):e1001547 [FREE Full text] [doi: 10.1371/journal.pmed.1001547] [Medline: 24223526]
2. Holden C. Mental health. Global survey examines impact of depression. Science 2000 Apr 07;288(5463):39-40. [doi: 10.1126/science.288.5463.39] [Medline: 10766633]
3. Wang PS, Aguilar-Gaxiola S, Alonso J, Angermeyer MC, Borges G, Bromet EJ, et al. Use of mental health services for anxiety, mood, and substance disorders in 17 countries in the WHO world mental health surveys. Lancet 2007 Sep 8;370(9590):841-850 [FREE Full text] [doi: 10.1016/S0140-6736(07)61414-7] [Medline: 17826169]
4. Goldney RD, Fisher LJ, Dal GE, Taylor AW. Subsyndromal depression: prevalence, use of health services and quality of life in an australian population. Soc Psychiatry Psychiatr Epidemiol 2004 Apr;39(4):293-298. [doi: 10.1007/s00127-004-0745-5] [Medline: 15085331]
5. Fried EI, Nesse RM, Zivin K, Guille C, Sen S. Depression is more than the sum score of its parts: individual DSM symptoms have different risk factors. Psychol Med 2014 Jul;44(10):2067-2076 [FREE Full text] [doi: 10.1017/S0033291713002900] [Medline: 24289852]
6. Mitchell AJ, Vaze A, Rao S. Clinical diagnosis of depression in primary care: a meta-analysis. Lancet 2009 Aug 22;374(9690):609-619. [doi: 10.1016/S0140-6736(09)60879-5] [Medline: 19640579]
7. Hek K, Demirkan A, Lahti J, Terracciano A, Teumer A, Cornelis MC, et al. A genome-wide association study of depressive symptoms. Biol Psychiatry 2013 Apr 01;73(7):667-678 [FREE Full text] [doi: 10.1016/j.biopsych.2012.09.033] [Medline: 23290196]
8. Lewis CM, Ng MY, Butler AW, Cohen-Woods S, Uher R, Pirlo K, et al. Genome-wide association study of major recurrent depression in the U.K. population. Am J Psychiatry 2010 Aug;167(8):949-957. [doi: 10.1176/appi.ajp.2010.09091380] [Medline: 20516156]
9. Shi J, Potash JB, Knowles JA, Weissman MM, Coryell W, Scheftner WA, et al. Genome-wide association study of recurrent early-onset major depressive disorder. Mol Psychiatry 2011 Feb;16(2):193-201 [FREE Full text] [doi: 10.1038/mp.2009.124] [Medline: 20125088]
10. Major Depressive Disorder Working Group of the Psychiatric GWAS Consortium, Ripke S, Wray NR, Lewis CM, Hamilton SP, Weissman MM, et al. A mega-analysis of genome-wide association studies for major depressive disorder. Mol Psychiatry 2013 Apr;18(4):497-511 [FREE Full text] [doi: 10.1038/mp.2012.21] [Medline: 22472876]
11. Wray NR, Pergadia ML, Blackwood DH, Penninx BW, Gordon SD, Nyholt DR, et al. Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned. Mol Psychiatry 2012 Jan;17(1):36-48 [FREE Full text] [doi: 10.1038/mp.2010.109] [Medline: 21042317]
12. Kapur S, Phillips AG, Insel TR. Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? Mol Psychiatry 2012 Dec;17(12):1174-1179. [doi: 10.1038/mp.2012.105] [Medline: 22869033]
13. Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, et al. Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. Am J Psychiatry 2010 Jul;167(7):748-751. [doi: 10.1176/appi.ajp.2010.09091379] [Medline: 20595427]
14. Wittchen H, Beesdo K, Bittner A, Goodwin RD. Depressive episodes– evidence for a causal role of primary anxiety disorders? Eur Psychiatry 2003 Dec;18(8):384-393. [doi: 10.1016/j.eurpsy.2003.10.001] [Medline: 14680714]
15. Hankin BL, Fraley RC, Lahey BB, Waldman ID. Is depression best viewed as a continuum or discrete category? A taxometric analysis of childhood and adolescent depression in a population-based sample. J Abnorm Psychol 2005 Feb;114(1):96-110. [doi: 10.1037/0021-843X.114.1.96] [Medline: 15709816]
16. Haslam N, Holland E, Kuppens P. Categories versus dimensions in personality and psychopathology: a quantitative review of taxometric research. Psychol Med 2012 May;42(5):903-920. [doi: 10.1017/S0033291711001966] [Medline: 21939592]
17. Wright AG, Krueger RF, Hobbs MJ, Markon KE, Eaton NR, Slade T. The structure of psychopathology: toward an expanded quantitative empirical model. J Abnorm Psychol 2013 Feb;122(1):281-294 [FREE Full text] [doi: 10.1037/a0030133] [Medline: 23067258]
18. Dalgleish T, Black M, Johnston D, Bevan A. Transdiagnostic approaches to mental health problems: current status and future directions. J Consult Clin Psychol 2020 Mar;88(3):179-195 [FREE Full text] [doi: 10.1037/ccp0000482] [Medline: 32068421]
19. Kendler KS. The nature of psychiatric disorders. World Psychiatry 2016 Feb;15(1):5-12 [FREE Full text] [doi: 10.1002/wps.20292] [Medline: 26833596]
20. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (5th Edition). Washington, DC: American Psychiatric Publication; 2013.

21. The ICD-11 classification of mental and behavioural disorders: diagnostic criteria for research. World Health Organization. 2018. URL: https://www.who.int/classifications/icd/en/GRNBOOK.pdf [accessed 2021-09-14]
22. Braam AW, Copeland JR, Delespaul PA, Beekman AT, Como A, Dewey M, et al. Depression, subthreshold depression and comorbid anxiety symptoms in older europeans: results from the EURODEP concerted action. J Affect Disord 2014 Feb;155:266-272. [doi: 10.1016/j.jad.2013.11.011] [Medline: 24355647]
23. Brown TA, Campbell LA, Lehman CL, Grisham JR, Mancill RB. Current and lifetime comorbidity of the DSM-IV anxiety and mood disorders in a large clinical sample. J Abnorm Psychol 2001 Nov;110(4):585-599. [doi: 10.1037//0021-843x.110.4.585] [Medline: 11727948]
24. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. Arch Gen Psychiatry 2005 Jun;62(6):593-602. [doi: 10.1001/archpsyc.62.6.593] [Medline: 15939837]
25. Kessler RC, Merikangas KR, Wang PS. Prevalence, comorbidity, and service utilization for mood disorders in the united states at the beginning of the twenty-first century. Annu Rev Clin Psychol 2007;3:137-158. [doi: 10.1146/annurev.clinpsy.3.022806.091444] [Medline: 17716051]
26. Zimmerman M, McDermut W, Mattia JI. Frequency of anxiety disorders in psychiatric outpatients with major depressive disorder. Am J Psychiatry 2000 Aug;157(8):1337-1340. [doi: 10.1176/appi.ajp.157.8.1337] [Medline: 10910803]
27. Pedrelli P, Blais MA, Alpert JE, Shelton RC, Walker RS, Fava M. Reliability and validity of the symptoms of depression questionnaire (SDQ). CNS Spectr 2014 Dec;19(6):535-546 [FREE Full text] [doi: 10.1017/S1092852914000406] [Medline: 25275853]
28. Chelminski I, Zimmerman M. Pathological worry in depressed and anxious patients. J Anxiety Disord 2003;17(5):533-546. [doi: 10.1016/s0887-6185(02)00246-3] [Medline: 12941364]
29. Gask L, Klinkman M, Fortes S, Dowrick C. Capturing complexity: the case for a new classification system for mental disorders in primary care. Eur Psychiatry 2008 Oct;23(7):469-476. [doi: 10.1016/j.eurpsy.2008.06.006] [Medline: 18774269]
30. Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR*D report. Am J Psychiatry 2006 Nov;163(11):1905-1917. [doi: 10.1176/appi.ajp.163.11.1905] [Medline: 17074942]
31. Souery D, Oswald P, Massat I, Bailer U, Bollen J, Demyttenaere K, Group for the Study of Resistant Depression. Clinical factors associated with treatment resistance in major depressive disorder: results from a European Multicenter Study. J Clin Psychiatry 2007 Jul;68(7):1062-1070. [doi: 10.4088/jcp.v68n0713] [Medline: 17685743]
32. Paykel ES. Partial remission, residual symptoms, and relapse in depression. Dialogues Clin Neurosci 2008;10(4):431-437 [FREE Full text] [doi: 10.31887/DCNS.2008.10.4/espaykel] [Medline: 19170400]
33. Saragoussi D, Touya M, Haro JM, Jönsson B, Knapp M, Botrel B, et al. Factors associated with failure to achieve remission and with relapse after remission in patients with major depressive disorder in the PERFORM study. Neuropsychiatr Dis Treat 2017;13:2151-2165 [FREE Full text] [doi: 10.2147/NDT.S136343] [Medline: 28860772]
34. Jaffe DH, Rive B, Denee TR. The humanistic and economic burden of treatment-resistant depression in europe: a cross-sectional study. BMC Psychiatry 2019 Aug 07;19(1):247 [FREE Full text] [doi: 10.1186/s12888-019-2222-4] [Medline: 31391065]
35. Bergfeld IO, Mantione M, Figee M, Schuurman PR, Lok A, Denys D. Treatment-resistant depression and suicidality. J Affect Disord 2018 Aug 01;235:362-367. [doi: 10.1016/j.jad.2018.04.016] [Medline: 29665520]
36. Kendler KS. Major depression and generalised anxiety disorder. same genes, (partly)different environments--revisited. Br J Psychiatry Suppl 1996 Jun(30):68-75. [doi: 10.1192/S0007125000298437] [Medline: 8864151]
37. Frankland A, Cerrillo E, Hadzi-Pavlovic D, Roberts G, Wright A, Loo CK, et al. Comparing the phenomenology of depressive episodes in bipolar I and II disorder and major depressive disorder within bipolar disorder pedigrees. J Clin Psychiatry 2015 Jan;76(1):32-38. [doi: 10.4088/JCP.14m09293] [Medline: 25650671]
38. Hirschfeld RM. Differential diagnosis of bipolar disorder and major depressive disorder. J Affect Disord 2014 Dec;169 Suppl 1:12-16 [FREE Full text] [doi: 10.1016/S0165-0327(14)70004-7] [Medline: 25533909]
39. Bauer M, Andreassen OA, Geddes JR, Vedel Kessing L, Lewitzka U, Schulze TG, et al. Areas of uncertainties and unmet needs in bipolar disorders: clinical and research perspectives. Lancet Psychiatry 2018 Nov;5(11):930-939. [doi: 10.1016/S2215-0366(18)30253-0] [Medline: 30146246]
40. 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(5):e0126530 [FREE Full text] [doi: 10.1371/journal.pone.0126530] [Medline: 25992560]
41. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
42. Zimmerman M, Martinez JH, Friedman M, Boerescu DA, Attiullah N, Toba C. How can we use depression severity to guide treatment selection when measures of depression categorize patients differently? J Clin Psychiatry 2012 Oct;73(10):1287-1291. [doi: 10.4088/JCP.12m07775] [Medline: 23058983]

43. Jerant A, Kravitz RL, Garcia EF, Feldman MD, Cipri C, Nishio D, et al. Potential antidepressant overtreatment associated with office use of brief depression symptom measures. J Am Board Fam Med 2014;27(5):611-620 [FREE Full text] [doi: 10.3122/jabfm.2014.05.140038] [Medline: 25201931]
44. Mitchell AJ, Rao S, Vaze A. International comparison of clinicians' ability to identify depression in primary care: meta-analysis and meta-regression of predictors. Br J Gen Pract 2011 Feb;61(583):72-80 [FREE Full text] [doi: 10.3399/bjgp11X556227] [Medline: 21276327]
45. Malpass A, Dowrick C, Gilbody S, Robinson J, Wiles N, Duffy L, et al. Usefulness of PHQ-9 in primary care to determine meaningful symptoms of low mood: a qualitative study. Br J Gen Pract 2016 Feb;66(643):78-84 [FREE Full text] [doi: 10.3399/bjgp16X683473] [Medline: 26823268]
46. Chen M, Mao S, Liu Y. Big data: a survey. Mobile Netw Appl 2014 Jan 22;19(2):171-209. [doi: 10.1007/s11036-013-0489-0]
47. Olmert T, Cooper JD, Han SY, Barton-Owen G, Farrag L, Bell E, et al. A combined digital and biomarker diagnostic aid for mood disorders (the delta trial): protocol for an observational study. JMIR Res Protoc 2020 Aug 10;9(8):e18453 [FREE Full text] [doi: 10.2196/18453] [Medline: 32773373]
48. Kessler RC, Ustün TB. The World Mental Health (WMH) survey initiative version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). Int J Methods Psychiatr Res 2004;13(2):93-121. [doi: 10.1002/mpr.168] [Medline: 15297906]
49. Hirschfeld RM. The mood disorder questionnaire: a simple, patient-rated screening instrument for bipolar disorder. Prim Care Companion J Clin Psychiatry 2002 Feb;4(1):9-11 [FREE Full text] [doi: 10.4088/pcc.v04n0104] [Medline: 15014728]
50. Eckblad M, Chapman LJ. Development and validation of a scale for hypomanic personality. J Abnorm Psychol 1986 Aug;95(3):214-222. [doi: 10.1037//0021-843x.95.3.214] [Medline: 3745642]
51. Ghaemi SN, Miller CJ, Berv DA, Klugman J, Rosenquist KJ, Pies RW. Sensitivity and specificity of a new bipolar spectrum diagnostic scale. J Affect Disord 2005 Feb;84(2-3):273-277. [doi: 10.1016/S0165-0327(03)00196-4] [Medline: 15708426]
52. Angst J, Adolfsson R, Benazzi F, Gamma A, Hantouche E, Meyer TD, et al. The HCL-32: towards a self-assessment tool for hypomanic symptoms in outpatients. J Affect Disord 2005 Oct;88(2):217-233. [doi: 10.1016/j.jad.2005.05.011] [Medline: 16125784]
53. Leung CM, Yim CL, Yan CT, Chan CC, Xiang Y, Mak AD, et al. The bipolar II depression questionnaire: a self-report tool for detecting bipolar II depression. PLoS One 2016;11(3):e0149752 [FREE Full text] [doi: 10.1371/journal.pone.0149752] [Medline: 26963908]
54. Depue RA, Slater JF, Wolfstetter-Kausch H, Klein D, Goplerud E, Farr D. A behavioral paradigm for identifying persons at risk for bipolar depressive disorder: a conceptual framework and five validation studies. J Abnorm Psychol 1981 Oct;90(5):381-437. [doi: 10.1037//0021-843x.90.5.381] [Medline: 7298991]
55. Altman EG, Hedeker D, Peterson JL, Davis JM. The altman self-rating mania scale. Biol Psychiatry 1997 Nov 15;42(10):948-955. [doi: 10.1016/S0006-3223(96)00548-3] [Medline: 9359982]
56. First M, Williams J, Spitzer RL, Gibbon M. Structured clinical interview for DSM-IV axis I disorders (SCID). ePROVIDE - Mapi Research Trust. URL: https://eprovide.mapi-trust.org/instruments/structured-clinical-interview-for-dsm-iv-axis-i-disorders [accessed 2021-09-14]
57. Sheehan DV, Lecrubier Y, Sheehan KH, Amorim P, Janavs J, Weiller E, et al. The Mini-International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59 Suppl 20:22-33. [Medline: 9881538]
58. Cohen J. Statistical Power Analysis for the Behavioral Sciences (2nd Edition). New York: Academic Press; 1988:1-978.
59. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983 Sep;148(3):839-843. [doi: 10.1148/radiology.148.3.6878708] [Medline: 6878708]
60. Pedregosa F, Gramfort A, Michel V, Thirion B, Varoquaux G, Grisel O, et al. Scikit-learn: machine learning in python. J Mach Learn Res 2011;12:2825-2830 [FREE Full text]
61. Han SY, Cooper JD, Ozcan S, Rustogi N, Penninx BW, Bahn S. Integrating proteomic, sociodemographic and clinical data to predict future depression diagnosis in subthreshold symptomatic individuals. Transl Psychiatry 2019 Nov 07;9(1):277 [FREE Full text] [doi: 10.1038/s41398-019-0623-2] [Medline: 31699963]
62. Hamilton W, Watson J, Round A. Investigating fatigue in primary care. Br Med J 2010 Aug 24;341:c4259. [doi: 10.1136/bmj.c4259] [Medline: 20736254]
63. Tylee A, Gastpar M, Lépine JP, Mendlewicz J. DEPRES II (depression research in european society II): a patient survey of the symptoms, disability and current management of depression in the community. DEPRES Steering Committee. Int Clin Psychopharmacol 1999 May;14(3):139-151. [doi: 10.1097/00004850-199905002-00001] [Medline: 10435767]
64. Fava M, Hwang I, Rush AJ, Sampson N, Walters EE, Kessler RC. The importance of irritability as a symptom of major depressive disorder: results from the national comorbidity survey replication. Mol Psychiatry 2010 Aug;15(8):856-867 [FREE Full text] [doi: 10.1038/mp.2009.20] [Medline: 19274052]
65. Judd LL, Schettler PJ, Coryell W, Akiskal HS, Fiedorowicz JG. Overt irritability/anger in unipolar major depressive episodes: past and current characteristics and implications for long-term course. JAMA Psychiatry 2013 Nov;70(11):1171-1180. [doi: 10.1001/jamapsychiatry.2013.1957] [Medline: 24026579]

66. Posternak MA, Zimmerman M. Anger and aggression in psychiatric outpatients. J Clin Psychiatry 2002 Aug;63(8):665-672. [doi: 10.4088/jcp.v63n0803] [Medline: 12197446]
67. Caspi A, Houts RM, Belsky DW, Goldman-Mellor SJ, Harrington H, Israel S, et al. The p factor: one general psychopathology factor in the structure of psychiatric disorders? Clin Psychol Sci 2014 Mar;2(2):119-137 [FREE Full text] [doi: 10.1177/2167702613497473] [Medline: 25360393]
68. Hettema JM. What is the genetic relationship between anxiety and depression? Am J Med Genet C Semin Med Genet 2008 May 15;148C(2):140-146. [doi: 10.1002/ajmg.c.30171] [Medline: 18412101]
69. Bittner A, Goodwin RD, Wittchen H, Beesdo K, Höfler M, Lieb R. What characteristics of primary anxiety disorders predict subsequent major depressive disorder? J Clin Psychiatry 2004 May;65(5):618-626. [doi: 10.4088/jcp.v65n0505] [Medline: 15163247]
70. Gottschalk MG, Cooper JD, Chan MK, Bot M, Penninx BW, Bahn S. Discovery of serum biomarkers predicting development of a subsequent depressive episode in social anxiety disorder. Brain Behav Immun 2015 Aug;48:123-131. [doi: 10.1016/j.bbi.2015.04.011] [Medline: 25929723]
71. Gottschalk MG, Cooper JD, Chan MK, Bot M, Penninx BW, Bahn S. Serum biomarkers predictive of depressive episodes in panic disorder. J Psychiatr Res 2016 Feb;73:53-62. [doi: 10.1016/j.jpsychires.2015.11.012] [Medline: 26687614]
72. Fava M, Rush AJ, Alpert JE, Balasubramani GK, Wisniewski SR, Carmin CN, et al. Difference in treatment outcome in outpatients with anxious versus nonanxious depression: a STAR*D report. Am J Psychiatry 2008 Mar;165(3):342-351. [doi: 10.1176/appi.ajp.2007.06111868] [Medline: 18172020]
73. Penninx BW, Nolen WA, Lamers F, Zitman FG, Smit JH, Spinhoven P, et al. Two-year course of depressive and anxiety disorders: results from the netherlands study of depression and anxiety (NESDA). J Affect Disord 2011 Sep;133(1-2):76-85 [FREE Full text] [doi: 10.1016/j.jad.2011.03.027] [Medline: 21496929]
74. Trivedi MH, Rush AJ, Wisniewski SR, Nierenberg AA, Warden D, Ritz L, et al. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR*D: implications for clinical practice. Am J Psychiatry 2006 Jan;163(1):28-40. [doi: 10.1176/appi.ajp.163.1.28] [Medline: 16390886]
75. Nam Y, Kim C, Roh D. Comorbid panic disorder as an independent risk factor for suicide attempts in depressed outpatients. Compr Psychiatry 2016 May;67:13-18. [doi: 10.1016/j.comppsych.2016.02.011] [Medline: 27095329]
76. Katz C, Yaseen ZS, Mojtabai R, Cohen LJ, Galynker II. Panic as an independent risk factor for suicide attempt in depressive illness: findings from the national epidemiological survey on alcohol and related conditions (NESARC). J Clin Psychiatry 2011 Dec;72(12):1628-1635. [doi: 10.4088/JCP.10m06186blu] [Medline: 21457675]
77. Roy-Byrne PP, Stang P, Wittchen HU, Ustun B, Walters EE, Kessler RC. Lifetime panic-depression comorbidity in the national comorbidity survey. Association with symptoms, impairment, course and help-seeking. Br J Psychiatry 2000 Mar;176:229-235. [doi: 10.1192/bjp.176.3.229] [Medline: 10755069]
78. Hinkelmann K, Moritz S, Botzenhardt J, Muhtz C, Wiedemann K, Kellner M, et al. Changes in cortisol secretion during antidepressive treatment and cognitive improvement in patients with major depression: a longitudinal study. Psychoneuroendocrinology 2012 May;37(5):685-692. [doi: 10.1016/j.psyneuen.2011.08.012] [Medline: 21944955]
79. Johansen M, Karterud S, Pedersen G, Gude T, Falkum E. An investigation of the prototype validity of the borderline DSM-IV construct. Acta Psychiatr Scand 2004 Apr;109(4):289-298. [doi: 10.1046/j.1600-0447.2003.00268.x] [Medline: 15008803]
80. Watson R, Harvey K, McCabe C, Reynolds S. Understanding anhedonia: a qualitative study exploring loss of interest and pleasure in adolescent depression. Eur Child Adolesc Psychiatry 2020 Apr;29(4):489-499 [FREE Full text] [doi: 10.1007/s00787-019-01364-y] [Medline: 31270605]
81. Tomasik J, Han SY, Barton-Owen G, Mirea D, Martin-Key NA, Rustogi N, et al. A machine learning algorithm to differentiate bipolar disorder from major depressive disorder using an online mental health questionnaire and blood biomarker data. Transl Psychiatry 2021 Jan 12;11(1):41 [FREE Full text] [doi: 10.1038/s41398-020-01181-x] [Medline: 33436544]
82. Torous J, Staples P, Shanahan M, Lin C, Peck P, Keshavan M, et al. Utilizing a personal smartphone custom app to assess the patient health questionnaire-9 (PHQ-9) depressive symptoms in patients with major depressive disorder. JMIR Ment Health 2015;2(1):e8 [FREE Full text] [doi: 10.2196/mental.3889] [Medline: 26543914]
83. Knowles SE, Toms G, Sanders C, Bee P, Lovell K, Rennick-Egglestone S, et al. Qualitative meta-synthesis of user experience of computerised therapy for depression and anxiety. PLoS One 2014;9(1):e84323 [FREE Full text] [doi: 10.1371/journal.pone.0084323] [Medline: 24465404]
84. Albert PR. Why is depression more prevalent in women? J Psychiatry Neurosci 2015 Jul;40(4):219-221 [FREE Full text] [doi: 10.1503/jpn.150205] [Medline: 26107348]

Abbreviations

AUC: area under the receiver operating characteristic curve
BD: bipolar disorder
CIDI: Composite International Diagnostic Interview
DSM-IV: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition

DSM-5: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
GAD: generalized anxiety disorder
ICD-10: International Statistical Classification of Diseases, Tenth Revision
ICD-11: International Statistical Classification of Diseases, Eleventh Revision
MDD: major depressive disorder
ML: machine learning
NCV: nested cross-validation
PHQ-9: Patient Health Questionnaire–9

Edited by G Eysenbach; submitted 15.02.21; peer-reviewed by M Gottschalk, R Pine; comments to author 07.03.21; revised version received 17.03.21; accepted 01.08.21; published 28.10.21.

Please cite as:
Martin-Key NA, Mirea DM, Olmert T, Cooper J, Han SYS, Barton-Owen G, Farrag L, Bell E, Eljasz P, Cowell D, Tomasik J, Bahn S
Toward an Extended Definition of Major Depressive Disorder Symptomatology: Digital Assessment and Cross-validation Study
JMIR Form Res 2021;5(10):e27908
URL: https://formative.jmir.org/2021/10/e27908
doi:10.2196/27908
PMID:34709182

Original Paper

A Decision Support Tool for Allogeneic Hematopoietic Stem Cell Transplantation for Children With Sickle Cell Disease: Acceptability and Usability Study

Anirudh Veludhandi[1], BSc; Diana Ross[1], RN, MSN; Cynthia B Sinha[1], PhD; Courtney McCracken[2], PhD; Nitya Bakshi[1], MD; Lakshmanan Krishnamurti[1], MD

[1]School of Medicine, Emory University, Atlanta, GA, United States

[2]Center for Research and Evaluation, Kaiser Permanente, Atlanta, GA, United States

Corresponding Author:
Lakshmanan Krishnamurti, MD
School of Medicine
Emory University
100 Woodruff Circle
Atlanta, GA, 30322
United States
Phone: 1 404 727 5671
Email: lkrishn@emory.edu

Abstract

Background: Individuals living with sickle cell disease (SCD) may benefit from a variety of disease-modifying therapies, including hydroxyurea, voxelotor, crizanlizumab, L-glutamine, and chronic blood transfusions. However, allogeneic hematopoietic stem cell transplantation (HCT) remains the only nonexperimental treatment with curative intent. As HCT outcomes can be influenced by the complex interaction of several risk factors, HCT can be a difficult decision for health care providers to make for their patients with SCD.

Objective: The aim of this study is to determine the acceptability and usability of a prototype decision support tool for health care providers in decision-making about HCT for SCD, together with patients and their families.

Methods: On the basis of published transplant registry data, we developed the Sickle Options Decision Support Tool for Children, which provides health care providers with personalized transplant survival and risk estimates for their patients to help them make informed decisions regarding their patients' management of SCD. To evaluate the tool for its acceptability and usability, we conducted beta tests of the tool and surveys with physicians using the Ottawa Decision Support Framework and mobile health app usability questionnaire, respectively.

Results: According to the mobile health app usability questionnaire survey findings, the overall usability of the tool was high (mean 6.15, SD 0.79; range 4.2-7). According to the Ottawa Decision Support Framework survey findings, acceptability of the presentation of information on the decision support tool was also high (mean 2.94, SD 0.63; range 2-4), but the acceptability regarding the amount of information was mixed (mean 2.59, SD 0.5; range 2-3). Most participants expressed that they would use the tool in their own patient consults (13/15, 87%) and suggested that the tool would ease the decision-making process regarding HCT (8/9, 89%). The 4 major emergent themes from the qualitative analysis of participant beta tests include user interface, data content, usefulness during a patient consult, and potential for a patient-focused decision aid. Most participants supported the idea of a patient-focused decision aid but recommended that it should include more background on HCT and a simplification of medical terminology.

Conclusions: We report the development, acceptability, and usability of a prototype decision support tool app to provide individualized risk and survival estimates to patients interested in HCT in a patient consultation setting. We propose to finalize the tool by validating predictive analytics using a large data set of patients with SCD who have undergone HCT. Such a tool may be useful in promoting physician-patient collaboration in making shared decisions regarding HCT for SCD. Further incorporation of patient-specific measures, including the HCT comorbidity index and the quality of life after transplant, may improve the applicability of the decision support tool in a health care setting.

(JMIR Form Res 2021;5(10):e30093) doi:10.2196/30093

KEYWORDS

decision support tool; sickle cell disease; mobile application; mHealth; pediatrics; transplant; mobile phone

Introduction

Background

Sickle cell disease (SCD) is a chronic blood disorder affecting approximately 100,000 adults and children in the United States [1]. It is characterized by the inheritance of a point mutation in the β-globin gene, leading to the sickle shape of red blood cells. Complications of SCD include painful vaso-occlusive episodes, acute chest syndrome, stroke, splenic sequestration, progressive organ dysfunction, and premature mortality [2-5]. Disease-modifying therapies, such as hydroxyurea, L-glutamine [6], voxelotor [7], and crizanlizumab [8], offer the possibility of long-term amelioration of the disorder. Despite undergoing these therapies, patients with SCD have a diminished quality of life (QoL) and life expectancy, which may be 20 years less than that of the general African-American population in which SCD is most prevalent [4].

Autologous gene therapy (GT) is an emerging treatment based on genetic modification of the patient's own hematopoietic stem cells to minimize the polymerization of abnormal hemoglobin. Early phase clinical trials of GT suggest that it has the potential to result in long-term amelioration of the disease [9,10]. However, for now, allogeneic hematopoietic stem cell transplantation (HCT) remains the only nonexperimental treatment with curative intent. Although HCT offers the promise of long-term disease amelioration without maintenance medications, it is associated with substantial risks of morbidity and mortality in the short term as well as the risk of new chronic and disabling complications. The safety and efficacy of HCT have improved with advances in supportive care [11]. Clinical trials and registry-based data collected by the Center for International Blood and Marrow Transplant Research, Eurocord, and the European Society for Blood and Marrow Transplantation registries have provided estimates on the survival and risk of HCT-related morbidities for patients with SCD, such as graft failure and graft-versus-host disease (GVHD) which occurs when donor stem cells perceive the recipient's body as foreign [11-15]. These studies suggest an overall survival (OS) rate for the sampled population of >90%, with the best patient outcomes observed in younger patients with an HLA antigen–identical related donor [11,14,15]. Thus, certain patients with severe manifestations of SCD may benefit from undergoing HCT.

The unavailability of HLA antigen–identical sibling donors and the disparity between physicians' and patients' or parents' assessment of the patient's disease severity may be a barrier to HCT [16-18]. There is an increasing application of HCT for SCD with the advent of HCT from HLA antigen–matched unrelated donors, HLA antigen–haploidentical family donors, and emerging GT [15-19]. The availability of novel treatments, as well as the increased willingness of physicians and patients to consider HCT, may contribute to the increasing uptake of this treatment [20-22]. Previous studies have investigated the decision-making process and preferences of patients and families considering HCT [18,23-27].

Objectives

Although the reported HCT survival outcomes for patients with SCD are generally favorable, it may be difficult for health care providers to personalize risk factors for an individual patient or have expert knowledge of the field. In addition, although age and type of donor are the most important predictive factors of HCT outcomes, there is no available tool to individualize the risk factors for age, type of donor, type of conditioning, or other factors to provide estimates for an individual patient [11-15]. A decision support tool that incorporates published registry patient data could be used by health care providers to determine personalized risk and survival estimates post-HCT, including overall and event-free survival (EFS) and risk of GVHD, and present them to the patient and their family. The tool may help patients weigh the risks and pros and cons in the context of their own values and preferences. We are not aware of any study or decision support tool that presents personalized transplant outcomes of overall and EFS and risk of GVHD to health care providers. The objectives of this study are to: (1) create a prototype decision support tool that presents personalized risk and survival estimates related to HCT, and (2) determine the acceptability and utility of such a tool for health care providers in helping their patients with SCD make informed decisions regarding HCT.

Methods

Selection of Registry Data for Predictive Models

In preparation of creating prediction models for HCT outcomes, we analyzed registry-based studies that depicted transplant registry data reported to the Center for International Blood and Marrow Transplant Research, Eurocord, and European Society for Blood and Marrow Transplantation databases [11,14,15]. Studies were selected based on several criteria, including a large patient sample (>100 per donor type) and diversity of patients with respect to age, sex, type of donor, stem cell source, and type of conditioning regimen used for transplant. Three studies were selected to represent the 4 donor types associated with HCT: HLA antigen–matched sibling, HLA antigen–matched unrelated, haploidentical, and HLA antigen–mismatched unrelated. Patient profiles for each donor type are shown in Table 1.

Table 1. Selected patient samples from registry data.

Study and stem cell source	Patients (adults and children), n (%)
Cappelli et al [14]	
HLA antigen–matched sibling	
Age (years)	
0-5	175 (23.8)
6-15	436 (59.2)
>15	125 (17)
Gluckman et al [11]	
HLA antigen–matched sibling	1000[a] (100)
Eapen et al [15]	
HLA antigen–matched unrelated	111 (12.2)
Haploidentical	137 (15.1)
HLA antigen–mismatched unrelated	104 (11.4)

[a]Adults: 154; children: 846.

All 5 cohorts excluded patients who received uncommon conditioning regimens and inadequate follow-ups after transplant. All 5 cohorts used multivariable survival analysis to estimate hazard ratios for the events of interest. As we did not have access to the raw registry data, the published hazard ratios were extrapolated to estimate survival probabilities for combinations of patient factors.

Transplant Characteristic Variables

The registry-based studies containing data for the 5 cohorts were examined for OS, EFS, graft failure, and GVHD (acute and chronic). EFS was defined as the percentage of patients who both survived the transplant and did not experience graft failure during the follow-up period. For the purpose of the decision support tool, we only examined demographic information, including sex and age at HCT. Age was regarded as a continuous variable on transplant outcomes, and all possible ages in the analyzed cohorts were included. Transplant characteristics were also examined, including the donor type, stem cell source (bone marrow, peripheral blood, and cord blood), and conditioning regimen (myeloablative, reduced intensity, and nonmyeloablative). Race, ethnicity, and pretransplantation comorbidities such as vaso-occlusion and other chronic organ complications were not regarded for cohort patients.

Analysis of Registry Data

We conducted an initial analysis of the selected registry data by creating custom data sets in Microsoft Excel. These data sets included the number of patients for each donor type sorted by stem cell source, conditioning regimen, and sex. By assuming a uniform distribution of the data, we were able to divide these data sets into 2 or 3 subsets based on the median age of patients in each cohort.

Next, multivariate survival analysis was performed by examining the hazard ratios for each risk factor of HCT. Hazard ratios were interpreted as relative risks between subpopulations and held constant over the follow-up time after HCT (25-48 months depending on the stem cell source). We calculated the relative risk of transplant options by conducting a proportional analysis. This involved taking the age-based subset directly from the registry data, and for each transplant factor, we multiplied the percentage of patients who shared a variant by the corresponding hazard ratio. This allowed us to determine the relative risks for 2 or more variants for a transplant factor. This process was repeated until we had determined the relative risks for all combinations of transplant factors for each HCT outcome: overall and EFS, GVHD, and graft failure. We repeated this process for all age-based subsets of each of the 4 stem cell sources.

The hazard ratios used were pulled directly from registry-based studies that contained the respective cohorts. This is with the exception of the HLA antigen–matched sibling data, which combined the cohort data of Cappelli [14] with the hazard ratios of Gluckman [11] as the cohort contained more age-based subsets.

Creation of Transplant Prediction Models

We compiled all the proportional analyses that were conducted on the selected registry data and sorted them by transplant characteristics (sex, donor type, stem cell source, and conditioning regimen). This was done for all 5 HCT outcomes, including OS, EFS, graft failure, acute GVHD, and chronic GVHD. For each patient combination of transplant characteristics, we included proportional analyses for 2 or 3 age-dependent subgroups from our initial analysis. This allowed us to perform statistical regression on our data. From the several types of statistical regressions we analyzed, an exponential regression provided the highest coefficient of determination; therefore, this type of regression was performed for all possible patient combinations of transplant characteristics. We then tested each exponential regression equation with patient ages based on the original cohort with regard to donor type. Although the transplant outcomes of children matched the registry data, transplant outcomes for older adults, in many cases, were poorer

than expected; hence, we limited the older ages from being selected on the decision support tool.

Development of the Decision Support Tool

The Sickle Options Decision Support Tool for Children is a mobile-focused app that calculates survival and risk estimates for pediatric patients undergoing HCT. The tool was created to help health care providers make more informed decisions regarding their patients' management of SCD. The decision support tool is coded in HTML and JavaScript using the Monaca Onsen UI 2 framework to allow for a more native mobile user experience. Open-source JavaScript libraries such as Chart.js, which is licensed under the MIT license were incorporated to provide rich graphics for to visual aids. The app was hosted on the Heroku cloud app platform. The support tool provides a brief questionnaire that asks the user 5 basic questions regarding the characteristics of their patients. From this information, the decision support tool is able to display the survival and risk predictions of the selected patient in the form of pie charts and percentages. Help dialogs and tooltips are present to provide context to major keywords and percentages (refer to Figure 1 for screenshots of the Sickle Options Decision Support Tool for Children). Figure 1 depicts individual app screens of entering information for a patient's transplant to viewing a personalized transplant summary using the prediction model, which contains estimates for OS, EFS, and risk of GVHD.

Figure 1. Screenshots of Sickle Options, a decision support tool for children (mobile web app).

Study Design

Following the development of the Sickle Options Decision Support Tool for Children, we conducted alpha testing with a number of staff members familiar with SCD and HCT and made iterative changes in the tool. Once the design of the tool was finalized, we started enrolling participants in qualitative interviews for beta testing and surveys of acceptability and usability. Participants were asked to take part in a 20-minute phone call, starting with a brief overview of the study procedures. After providing verbal consent, participants were then provided with a link to the decision support tool to beta test on their own smartphones. They were asked to express their thoughts and feelings out loud as they proceeded with the tool. Participants were asked to select a transplant for a hypothetical patient before being provided with a personalized survival summary for the patient within the app. Specific questions were posed to the participants based on the particular app screen they were on at the time. The same questions were asked of each participant to maintain consistency. Following their initial run-through of the tool, the participants were asked additional questions pertaining to the tool and for any additional feedback before concluding the interview. After the beta test, participants were asked to complete 2 short surveys, sent via email, on acceptability and usability on REDCap (Research Electronic Data Capture). The interviews were recorded and then transcribed verbatim. Interview transcripts were analyzed using content analysis in Microsoft Excel [28]. The analysis included (1) generating codes based on participant responses to existing interview questions, (2) assessing the insight or other feedback raised by participants in the context of improving the decision support tool, and (3) comparing codes to create common themes exhibited from the data. Participant responses were divided into 4 categories related to user interface, data content, usefulness during patient consultation, and potential for a patient-focused decision aid. A second coder (CBS) reviewed the participant responses and the associated categories to ensure intercoder reliability. Emory University institutional review board provided ethical review and approval for the study (Reference ID: STUDY00000842).

Recruitment

Heal-Sickle ECHO (Extension for Community Health Care Outcomes) is a telementoring program that uses Project ECHO. Physicians and other health care providers signed up to participate in a 24-week series of fortnightly meetings with didactic presentations and case presentations related to HCT for SCD. Most participants were hematologists, HCT physician faculty, or fellowship trainees. We sent a recruitment email to approximately 65 participants in the Heal-Sickle ECHO series, inviting them to participate in a telephone interview and evaluation of Sickle Options Decision Support Tool for Children for acceptability and usability. Potential users from Heal-Sickle were included if they were (1) physicians who had experience and expertise in providing care to patients with SCD, (2) able to beta test the decision support tool on their phone, (3) willing to complete 2 surveys on acceptability and usability, and (4) willing to participate in a follow-up interview. Although the tool currently provides estimates of HCT outcomes for pediatric patients only, we intend to include estimates for adult patients once more data become available. For this reason, we recruited both pediatric and adult care providers for this study. In total, 18 physicians consented to participate in the study, indicating a response rate of 28%. The participants were compensated with a US $25 gift card for their involvement in the study.

Qualitative Interview

Study participants were asked for their feedback on the tool through a semistructured qualitative interview. The selection of questions posed to the participants is described in Textbox 1. In addition to these questions, participants were asked to give their thoughts on each screen of the decision support tool and suggest any improvements or modifications to the tool.

Textbox 1. Selections from the qualitative interviews.

Questions

1. How do you feel about the general user interface?
   - How easy is it to navigate through the app?
   - How useful are the help dialogs and tooltips?
2. What did you think about how the data was visualized?
   - What format would you prefer?
   - What other information would you like to see visualized?
3. Is there anything you specifically liked about the tool?
4. Is there anything you would like to change about the tool?
5. Do you see a decision support tool like this fitting into your consultations with patients?
   - (*If the participant indicated that the tool was suitable*) How will it assist you in decision-making about bone marrow transplant (BMT)?
   - (*If the participant indicated that the tool was unsuitable*) How do you think the tool should be changed to your satisfaction?
   - Would this information help to ease the decision-making process?
6. We are also planning to create a patient-focused decision aid version of the tool. Any thoughts?
   - What things would you like to see changed in the patient decision aid version?
   - How will a patient-focused decision aid affect your consultations with your patients?

Acceptability and Usability Questionnaires

An acceptability survey was created for the decision support tool by adapting the Ottawa Decision Support Framework (ODSF) [29]. This survey measured the tool comprehensibility, presentation of information, and overall suitability for decision-making. This was done using a mixed-scale questionnaire. The rating of sections (bone marrow transplant [BMT], evidence on BMT, and risk factors) was scored on a 4-point Likert scale ranging from 1 (poor) to 4 (excellent). The remaining questions scored the information presented in the tool on a 3-point Likert scale (more, less, and just right) and dichotomous scales of yes or no and easy or difficult. The acceptability survey was adapted from the original ODSF to include references to the tool and omissions regarding the references of information presented on the tool.

Next, a usability survey was created for the tool by adapting the mobile health app usability questionnaire (MAUQ) [30]. This survey measured the tool ease of use and overall usefulness. This comprised an 18-question survey in which participants responded to each statement on a 7-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree). The usability survey was not altered in any way from the original MAUQ.

Question items from both surveys were assessed for mean and SD. In addition, open-ended questions were analyzed qualitatively and viewed in conjunction with the participants' feedback during the beta test.

Results

Participant Characteristics

A total of 18 physicians with a background in SCD participated in this study. Of the 18 participants, 9 (50%) were male, and 9 (50%) were female. Most participants had completed their fellowship training (14/18, 78%) and had up to 15 years of medical experience following fellowship (13/18, 72%).

Qualitative Interview

User Interface and Experience

Participants were first asked to describe their experience in using the decision support tool. All participants stated that the tool was easy to use and straightforward overall. In addition, 39% (7/18) of participants who commented directly on the navigation of the tool reported smooth navigation between pages. In particular, participants pointed out the intuitive user experience:

> *Good size, easy to select stuff, easy to know what I've selected if I want to change. Looks pretty fluid.* [Participant 7]

Further discussion was focused on the individual components and dialogs of the user interface. Help dialogs were activated at the bottom of the tool screen to present additional information pertaining to the respective screen. Of the 6 participants who used these during the beta test, 3 (50%) mentioned that the dialog boxes were useful, whereas 1 (17%) participant indicated no preference. Participants reviewed the 2 transplant summary screens that contained posttransplant risk and survival estimates. Most participants (17/18, 94%) approved of the pie charts on

the transplant summary screens as an efficient way to visualize the survival summary:

> *I like the fact that you're given the percentages with the pie charts. I think that's the most important thing. It updates very quickly. This is great.* [Participant 12]

However, of the 15 participants who commented directly on the pie chart percentages, 7 (47%) expressed reservations about having to tap on the pie charts to view them. They preferred that the percentages be displayed automatically over the pie chart and suggested that it would save time and be more intuitive. Some participants also reported visual glitches during the beta test. A few participants (5/18, 28%) mentioned that the chart overlapped with the bottom navigation buttons, making it difficult to view it. Finally, 11% (2/18) of participants proposed the idea of a quick-edit menu on the summary page before the transplant summary screens. They attested that a quick-edit menu would be more usable as it would save time when modifying data for a particular patient. Modifications were made to the tool iteratively by incorporating the feedback and suggestions received.

Data Content and Significance

Participants were asked to provide feedback related to the pie charts. Of the 11 participants who commented on the information contained in the pie charts, 9 (82%) liked that the displayed pie chart information pertaining to survival, GVHD, and graft failure. In their clinical experience, they believed that these items were most likely what patients would want to know most about transplant:

> *I think it gives you an idea of living without SCD. Survival summary, and how things are potentially-like, gives you the risk.* [Participant 4]

A few participants who dealt primarily with adult patients suggested having larger age ranges when selecting a patient's age. The second transplant summary screen is analogous to the first except that it allows the user to calculate transplant outcomes after specifying the transplant delay in years. Of the 13 participants who commented directly on the delayed transplant feature of the app, 11 (85%) liked the feature of demonstrating a change in outcome when delaying a patient's transplant as being potentially clinically useful. However, a few participants suggested increasing or refining the range of years for delaying transplant. These participants maintained that some pediatric patients might want to wait until adulthood to make the decision for themselves, whereas others may have reasons to delay the transplant.

Participants provided feedback on the individual risk and survival estimates. A few participants (4/18, 22%) suggested that the GVHD risk summary should be broken down by severity (grade) to better understand a patient's prognosis after transplant. In addition, some participants (6/18, 33%) proposed that the tool should include information related to GVHD prophylaxis to better understand a patient's risk of GVHD. A few physicians stated that definitions of the types of conditioning regimens such as myeloablative are very ambiguous in the published literature. Some also indicated that their individual strategies and institutional outcomes for a specific transplant might be at variance from the registry data.

Some physicians (5/18, 28%) suggested that information on the HCT comorbidity index should be incorporated into the tool to allow for more accurate survival estimates as it would affect the success of the transplant. In addition, a few participants requested QoL assessments in the estimates and in the results screens to provide more context to a patient's posttransplant survival. Some participants indicated that the lack of information on GVHD prophylaxis, HCT comorbidity index, and QoL might limit the utility of the decision support tool to educate patients regarding HCT for SCD:

> *Quality of life, pain. I mean I'm trying to think of what my patients are going to want to know and they're going to kind of want to know is my life going to get better. Like for a 16-year-old, survival doesn't mean a whole lot.* [Participant 15]

Usefulness to Support Patient Consultations

Participants were asked about the usability of the decision support tool in their own patient consultations. Of the 15 participants who responded to this question, 13 (87%) expressed enthusiasm about the potential use of the tool. Common reasons included tool convenience, risk and survival estimates, and interactivity with patients:

> *You know, thinking about having something on hand, so I don't have to search through the literature or go online, which takes a lot of time [...] something that's easy and readily accessible. Yeah, this is very handy.* [Participant 6]

Approximately 11% (2/18) of participants expressed reservations about using the tool during patient consultations because of the unavailability of adult data and the fact that the HCT comorbidity index was not included in the current version of the tool. Others mentioned that they would use the tool in certain situations, such as HCT with matched sibling donors or not enrolling a patient for HCT in a clinical trial. Finally, of the 9 participants who were asked directly if the tool would help ease the decision-making process during the consult, 8 (89%) attested that it would. Participants discussed the tool function as a visual aid and the potential to better capture a patient's attention:

> *This is very useful for people - for physicians who are - you know doing consults on families and reviewing their results with them so I think it's a useful tool. I do like it. I think it's very user friendly and very simple which is good because it's easy to explain to the families.* [Participant 18]

Potential for a Patient-Focused Decision Aid

Of the 12 participants who commented on a potential patient decision aid, 9 (75%) indicated that the medical terminology currently used in the tool, including the type of donor and conditioning regimen, would need to be simplified in the patient decision aid version. Participants pointed to concerns about a patient's background and the potential for patients to take results out of context because of low health literacy:

> *Pie charts are a good way to accurately represent the risks associated with it. It's very visual [...] saying a bunch of numbers but [...] a lot of information in a short period of time for families that may not have high literacy can be very overwhelming.* [Participant 8]

The participants later discussed the timeline regarding when the patient decision aid should be used. Of the 6 participants, 3 (50%) expressed that the decision aid should be used before a consultation. Participants suggested that using the decision aid before consultation would result in more physician-patient interactions.

Other participants expressed doubts about the patients' ability to use a decision aid before consultation. They stated that patients might not know what donor or conditioning regimen they were eligible for until meeting with the transplant physician. They also expressed doubts about the potential of the decision aid to steer patients away from transplant.

Acceptability and Usability Questionnaires

Participant Questionnaire Responses

Analysis of participant feedback from postinterview questionnaires on acceptability and usability of the decision support tool centered on 3 major themes of discussion: (1) user interface and experience, (2) data content and significance, and (3) usefulness to support decision-making. Participant responses to the acceptability and usability questionnaires with itemized ratings are further described in Tables 2 and 3, respectively. The acceptability questionnaire uses a mixed Likert scale and a dichotomous scale, whereas the usability questionnaire uses a 7-point Likert scale with a score of 7 indicating *strongly agree*.

Table 2. Responses to acceptability questionnaire.

Statements	Value, mean (SD)
Please rate each section by circling "poor," "fair," "good," or "excellent" to show[a]:	
Bone marrow transplantation	2.94 (0.66)
Evidence about transplantation	2.94 (0.66)
Risk factors including age, donor type, stem cell source, and conditioning regimen	3.29 (0.47)
The amount of time the learning took was: 1=too long, 2=too short, and 3=just right	2.81 (0.40)
The amount of information was: 1=too much information, 2=too little information, and 3=just right	2.59 (0.51)
I found the learning: 1=slanted towards taking bone marrow transplantation, 2=slanted against taking bone marrow transplantation, and 3=balanced	2.88 (0.49)
Do you find this decision support tool useful while you are making your decision for your patient about bone marrow transplantation? 1=yes and 2=no	1 (0)
What did you think of the way to calculate risk factors with bone marrow transplantation? Was it: 1=easy to find your patient's risk level, or 2=difficult	1.06 (0.24)
Do you think we included enough information to help someone with sickle cell disease decide whether or not to start bone marrow transplantation? 1=yes, and 2=no	1.53 (0.51)
What did you like about the decision support tool?	—[b]
What suggestions do you have to improve the decision support tool?	—[b]

[a]For the rating of sections, 1=poor, 2=fair, 3=good, 4=excellent.

[b]Participants provided open-ended responses instead of numerical ratings.

Table 3. Responses to usability questionnaire

Statements	Value[a], mean (SD)
The app was easy to use.	6.71 (0.59)
It was easy for me to learn to use the app.	6.81 (0.54)
The navigation was consistent when moving between screens.	6.53 (0.87)
The interface of the app allowed me to use all the functions (such as entering information, responding to reminders, and viewing information) offered by the app.	6.29 (1.10)
Whenever I made a mistake using the app, I could recover easily and quickly.	6.41 (1.06)
I like the interface of the app.	6.35 (0.86)
The information in the app was well organized, so I could easily find the information I needed.	6.29 (0.92)
The app adequately acknowledged and provided information to let me know the progress of my action.	6.29 (0.92)
I feel comfortable using this app in clinical settings.	6.47 (0.94)
The amount of time involved in using this app has been fitting for me.	6.59 (0.62)
I would use this app again.	6.47 (1.01)
Overall, I am satisfied with this app.	5.94 (1.09)
The app would be useful for my health care practice.	6 (1.12)
This app improved my access to delivering health care services.	6.18 (0.88)
This app helped me manage my patients' health effectively.	5.76 (1.20)
This app has all the functions and capabilities I expected it to have.	5.53 (1.81)
I could use the app even when the Internet connection was poor or not available.	4.88 (1.93)
This mHealth[b] app provides an acceptable way to deliver health care services, such as accessing educational materials, tracking my own activities, and performing self-assessment.	5.35 (1.50)

[a]Cumulative mean, 6.15 (SD 0.79).

[b]mHealth: mobile health.

User Interface and Experience

As shown in Table 3, the usability of the decision support tool was shown to be high, with overall usability of 6.15 (SD 0.79; range 4.2-7) on the MAUQ. This is reflected in the high scores on statements related to ease of use (mean 6.71, SD 0.58; range 5-7), navigation (mean 6.53, SD 0.89; range 4-7), user interface pertaining to functionality (mean 6.29, SD 1.14; range 4-7) and overall look (mean 6.35, SD 0.87; range 5-7). The participants were also pleased with the length of the tool and the amount of time it took to complete. This statement was scored high on both the acceptability questionnaire (mean 2.81, SD 0.41; range 2-3; Table 2) and usability questionnaire (mean 6.59, SD 0.62; range 5-7).

Participants scored the web-based app low on question 17 of the MAUQ, which asked about using the app during a poor internet connection. This question had a mean score of 4.88 (SD 1.98; range 1-7; Table 3). This is in accordance with statements made by participants during the interview that mentioned spotty Wi-Fi at their medical office. These statements provide support for creating a downloadable app that can be used offline. In addition, a few participants indicated user interface glitches related to visual components in the suggestions portion of the acceptability questionnaire, matching the feedback given during the qualitative interviews.

Acceptability Data Content and Significance

In the acceptability questionnaire, participants suggested that the tool was not biased for or against HCT (mean 2.88, SD 0.5; range 1-3) and was easy to calculate risk factors with (mean 1.06 SD 0; range 1-2). Participants also indicated that they were satisfied with the types of information present on the tool, including BMT (mean 2.94, SD 0.63; range 2-4) and risk factors (mean 3.29, SD 0.48; range 3-4; Table 2). With regard to the amount of information presented in the tool, the participants were split. This statement was scored with a mean of 2.59 (SD 0.5; range 2-3), indicating a mix of *too little* and *just right*.

Some participants were not satisfied with the functions and capabilities of the tool, as shown in the usability questionnaire (mean 5.53, SD 1.61; range 2-7; Table 2). In the acceptability questionnaire, participants were asked to provide suggestions for improving the decision support tool. Most suggestions were focused on adding more context to the risk and survival summaries the tool presented. These include adding the time period for registry data, HCT comorbidity index, and type of chemotherapy in relation to the conditioning regimen used. All suggestions given in the acceptability questionnaire matched the feedback given in the qualitative interviews. Participant responses were mixed regarding whether the tool provided enough information as well as its overall acceptability to deliver health care services. In the acceptability questionnaire, participants were split regarding whether the tool provided

enough information to help someone with SCD decide on HCT (mean 1.53, SD 0.52; range 1-2; Table 2). This rating is analogous with an earlier statement in the acceptability questionnaire, which asked about the amount of information in the tool.

Usability

Starting with the MAUQ in Table 3, participants indicated their approval of the organization of information in the tool (mean 6.29, SD 0.95; range 4-7). As shown in Table 3, participants attested in the MAUQ that the tool would be useful for their health care practice (mean 6, SD 1.02; range 3-7), improve access to delivering health care services (mean 6.18, SD 0.86; range 4-7), and help manage their patients' health effectively (mean 5.76, SD 1.22; range 3-7). In the acceptability questionnaire, all participants indicated that the decision support tool would be useful when making the decision for their patients about HCT (mean 1, SD 0; Table 2).

In the MAUQ, participants were asked whether the tool provided an acceptable way to deliver health care services, including access to materials, activity tracking, and self-assessment performance. This statement was rated with a mean score of 5.35 (SD 1.54; range 2-7), indicating that participants *somewhat agreed* with the statement (Table 3). These lower ratings are likely representative of the desired additional features that participants requested during qualitative interviews and while answering questionnaires on acceptability and usability. Although some of these features are dependent on research being available in the public domain, such as the HCT comorbidity index affecting a specific type of HCT, others, such as adding the time period for registry data, can be addressed independently.

Discussion

Principal Findings

In this study, we report the development, usability, and acceptability of the Sickle Options Decision Support tool for HCT to clinicians in pediatric SCD. These data provide a proof-of-concept of the potential acceptability and utility of such a decision support tool for use by hematologists and HCT physicians. To the best of our knowledge, this is the first decision support tool to provide individualized and age-specific risk and survival estimates for pediatric patients considering HCT. Brazauskas et al [13] published a risk prediction model for patients with SCD based on age and donor type. The model was derived from large registry-based data sets [11,14,15]. These studies report age as a continuous variable for outcomes of HCT for SCD; however, the model does not individualize the risks of outcomes for a given patient. Age was correlated with OS, EFS, and GVHD risk. The lack of individualization of outcomes that families may consider important in decision-making regarding HCT is a barrier to decision-making.

The Sickle Options Decision Support Tool for Children was designed to provide health care providers specializing in SCD with individualized risk and survival estimates for their pediatric patients considering HCT. The tool risk prediction models were made possible by large-scale clinical studies that depicted registry data on transplant outcomes in thousands of patients with SCD [11,14,15]. With this information, the tool can present transplant outcomes based on a patient's sex, age, donor type, stem cell source, and type of conditioning regimen used in HCT. The decision support tool provides outcomes based on the current age of the patient as well as a selected future age of the patient. Caregivers may exhibit mixed perceptions of HCT for their child in relation to the risk of death or other HCT-related complications (graft failure and GVHD) when making decisions. In addition, although some patients and caregivers may have an interest in wanting to learn more about HCT, there may still be a decisional dilemma, especially when deciding whether to postpone transplant [18]. The use of the decision support tool may help fill a potential information gap by either the patient or the health care provider as well as encourage physician-patient collaboration regarding the decision-making process.

The decision support tool is available as a mobile web app. We conducted a comprehensive beta test of this mobile app and qualitative interviews with 18 hematologist–oncologists with a special interest in curative therapies for SCD. The participants found the decision support tool to be usable and acceptable based on the scoring frameworks of the MAUQ and ODSF, respectively.

Decision support tools can play a useful role in a health care setting in assisting health care providers with the delivery of individualized evidence-based care to their patients. Decision support tools may be used to present a visual display of data to educate patients regarding the risks, benefits, and outcomes of HCT. Thus, they can be used as tools to enhance patient knowledge and engagement, improve standards of care, and help clinicians analyze large sets of data quickly given the time-sensitive nature of decision-making [31]. The development of this tool was largely informed by experience with using the app iChoose Kidney, an app-based decision aid, which uses risk prediction models to help identify whether a patient would be best suited for kidney transplant or dialysis [32]. We believe that a mobile-focused decision support tool will be easily accessible for use by health care providers during their patient consultations and present clear visuals that both health care providers and patients can understand.

Participants engaged in beta testing of the decision support tool provided feedback and made several suggestions that allowed us to refine the decision support tool. We optimized the display of the pie chart by adding automatic percentages to the view and also introduced quick-edit menus to switch between different types of transplants. In addition, after transplant and when it is available, we intend to add information pertaining to QoL in future versions of the decision support tool.

Regarding the comorbidity index reported by registry studies, the index includes factors that do not overlap with the factors found to predict HCT risk and survival outcomes. Eapen et al [15] reported that patients who underwent HCT with a haploidentical donor were much more likely to have an HCT comorbidity index of >3 than patients who underwent HCT with any other type of donor. The lack of impact of the comorbidity index on post-HCT outcomes is likely reflective

of the predominantly young patient population in the registry data. However, as more adults with advanced progressions of SCD undergo HCT, it is possible that the HCT comorbidity index may be more of a factor in predicting post-HCT outcomes. We intend to include comorbidity data in future versions of the decision support tool when such a determination is made.

Limitations

We recognize several limitations of this study. Qualitative content analysis also brought to light the limitations of the current data in providing individualized risk. First, currently available data do not permit the incorporation of patient characteristics, including comorbidities, severity of SCD, organ damage or QoL, chronic pain, or psychological comorbidities into the risk prediction models of the tool. Thus, there are inherent limitations to individualizing risk estimates and predicting outcomes. The availability of such information and its impact on outcomes would provide a more complete picture of a patient's risk and survival outcomes after undergoing HCT. Currently, there is research based on the effect of the pre-HCT comorbidity index on overall after transplant [33]. On the basis of the outcomes of these studies, we plan to incorporate this information and other information regarding patient comorbidities in future versions of the decision support tool. Second, we developed this prototype app based on published data that contained registry data tables [11,14,15]. Therefore, we make some assumptions regarding the exact number of individuals with a specific donor type, stem cell source, and conditioning regimen for a given age based on cumulative data published. As a result, rounding off may present some error when providing the risk and survival summary for a specific patient, and we only present this tool as a prototype to provide a proof-of-concept for such a tool. To properly assess the tool validity, the next step would be to obtain the raw patient data so that we may construct appropriate multivariate statistical models that incorporate several risk factors of HCT, including the HCT comorbidity index. We are in the process of obtaining the study data set on HCT of 1518 patients with SCD aged <1 to 58 years transplanted in the United States from 106 transplant centers from the Biologic Specimen and Data Repositories Information Coordinating Center, which has been established by the National Heart, Lung, and Blood Institute (NHLBI). The Biologic Specimen and Data Repositories Information Coordinating Center combines the resources of the NHLBI Biologic Specimen Repository, which has been managed by the division of blood disease resources since 1975, and the NHLBI Data Repository, which has been managed by the division of cardiovascular sciences since 2000. We will perform validation studies of the app on this study data set and finalize the development of the app. Third, the registry data we used comprises transplants completed over a wide time frame—1986-2017 for matched sibling donor data [14] and 2008-2017 for all remaining donor data [15]. Gluckman et al [11] reported an era effect in the outcome of HCT for SCD. Refinements in supportive care may render much of the registry data out of date as they may not consider the latest clinical practices in supportive care and the prevention and treatment of GVHD and graft failure. Fourth, the completion of ongoing clinical trials of haploidentical donor BMT in adults and children with SCD or autologous GT may add new dimensions to the consideration of curative options. Fifth, the published registry data may not account for underreporting of outcomes, institutional volume, center experience effect, or physician expertise. Institutions that perform more transplants for SCD may see better outcome success than smaller institutions that perform fewer transplants. Thus, some physicians with experience and expertise in the field expressed hesitation in using the decision tool as they felt that the outcomes at their institutions were likely to be superior to those reported by the registry. This, of course, may also reflect an unconscious bias. These findings suggest the need for future studies examining outcomes with center experience in performing HCT for SCD. Such data will allow physicians to generate individualized estimates of outcomes in centers, such as their center, and they are more likely to use the decision support tool for determining individualized estimates of outcomes and sharing them with their patients.

Some physicians expressed a desire to see QoL pre-HCT and the anticipated impact of HCT on QoL in the risk prediction model of the decision support tool. Studies of decision-making by patients suggest that patients and families consider QoL important in decision-making about HCT [18,23-27]. Recent clinical trials have suggested that HCT can improve QoL after HCT [34-40]. Uniform and systematic collection of QoL before and after HCT in clinical trials and registries is necessary before such data can be incorporated into a decision support tool. Some participants mentioned the desirability of incorporating comparison data on anticipated outcomes with non-HCT disease-modifying therapies. This would, of course, require the availability of such comparative data in large clinical trials.

As most data on HCT for SCD deals with pediatric outcomes, we limited the decision support tool to children. However, future app versions may incorporate adult outcomes because of the rapid increase in the uptake of HCT in adults with SCD and the potential of GT to become the standard treatment.

We recruited participants exclusively from the Heal-Sickle ECHO program for the study. As these physicians have a declared interest in curative therapy and decision-making about HCT for SCD, their perspective may not be generalizable to that of hematologists and transplant physicians who may be less involved in the field. On the other hand, the experience and expertise of the participants were extremely useful in optimizing the design of the tool and identifying future areas of research. Participants in the study also made several recommendations for a proposed version of the decision support tool specifically designed for patients and families. Considering the technical nature of information regarding myeloablation, conditioning regimen, and donor types, the decision support tool may be inherently better suited for use with guidance by physicians who could then use the tool to display individualized risk to their patients. Finally, the response rate of the physicians for the study was low—at 28%. A potential reason for the low response rate may be physician preferences for more traditional modes of research, such as a paper survey rather than a web-based survey [41]. In addition, as the decision support tool was limited to pediatric patients with SCD, some physicians

who predominantly treated adult patients with SCD may have considered the app to be beyond the scope of their practice.

Conclusions

We report the development, beta testing, usability, and acceptability of a decision support tool for individualizing estimates of outcomes of HCT for patients with SCD. Refining the predictive algorithms for era and center experience, incorporating data on QoL, comparison of other disease-modifying therapies, outcomes in adults, and autologous GT offers he possibility of expanding the applicability of such a decision support tool in helping shared decision-making in HCT for SCD.

Acknowledgments

Funding for this study was provided by the National Heart, Lung, and Blood Institute OTA 1OT2HL152762-01 held by LK.

NB also received funding from the National Heart, Lung, and Blood Institute of the National Institutes of Health under award numbers K23HL140142 and K23HL140142-03S1. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of Interest

None declared.

References

1. Brousseau DC, Panepinto JA, Nimmer M, Hoffmann RG. The number of people with sickle-cell disease in the United States: national and state estimates. Am J Hematol 2010 Jan;85(1):77-78 [FREE Full text] [doi: 10.1002/ajh.21570] [Medline: 20029951]
2. Platt OS, Brambilla DJ, Rosse WF, Milner PF, Castro O, Steinberg MH, et al. Mortality in sickle cell disease– life expectancy and risk factors for early death. N Engl J Med 1994 Jun 09;330(23):1639-1644. [doi: 10.1056/nejm199406093302303]
3. Panepinto JA, Bonner M. Health-related quality of life in sickle cell disease: past, present, and future. Pediatr Blood Cancer 2012 Aug;59(2):377-385. [doi: 10.1002/pbc.24176] [Medline: 22522407]
4. DeBaun MR, Ghafuri DL, Rodeghier M, Maitra P, Chaturvedi S, Kassim A, et al. Decreased median survival of adults with sickle cell disease after adjusting for left truncation bias: a pooled analysis. Blood 2019 Feb 07;133(6):615-617 [FREE Full text] [doi: 10.1182/blood-2018-10-880575] [Medline: 30530750]
5. Kato GJ, Piel FB, Reid CD, Gaston MH, Ohene-Frempong K, Krishnamurti L, et al. Sickle cell disease. Nat Rev Dis Primers 2018 Mar 15;4(1):1-22. [doi: 10.1038/nrdp.2018.10] [Medline: 29542687]
6. Cieri-Hutcherson NE, Hutcherson TC, Conway-Habes EE, Burns BN, White NA. Systematic review of l-glutamine for prevention of vaso-occlusive pain crisis in patients with sickle cell disease. Pharmacotherapy 2019 Nov;39(11):1095-1104. [doi: 10.1002/phar.2329] [Medline: 31505045]
7. Blair HA. Voxelotor: First approval. Drugs 2020 Feb 04;80(2):209-215. [doi: 10.1007/s40265-020-01262-7] [Medline: 32020554]
8. Blair HA. Crizanlizumab: First approval. Drugs 2020 Jan 13;80(1):79-84. [doi: 10.1007/s40265-019-01254-2] [Medline: 31933169]
9. Tisdale JF, Kanter J, Mapara MY, Kwiatkowski JL, Krishnamurti L, Schmidt M, et al. Current results of lentiglobin gene therapy in patients with severe sickle cell disease treated under a refined protocol in the phase 1 Hgb-206 study. Blood 2018 Nov 29;132(1):1026. [doi: 10.1182/blood-2018-99-113480]
10. Frangoul H, Ho TW, Corbacioglu S. CRISPR-Cas9 gene editing for sickle cell disease and β-thalassemia. N Engl J Med 2021 Jun 10;384(23):91-260. [doi: 10.1056/NEJMc2103481] [Medline: 34107197]
11. Gluckman E, Cappelli B, Bernaudin F, Labopin M, Volt F, Carreras J, Eurocord, the Pediatric Working Party of the European Society for BloodMarrow Transplantation,the Center for International BloodMarrow Transplant Research. Sickle cell disease: an international survey of results of HLA-identical sibling hematopoietic stem cell transplantation. Blood 2017 Mar 16;129(11):1548-1556 [FREE Full text] [doi: 10.1182/blood-2016-10-745711] [Medline: 27965196]
12. Bernaudin F, Verlhac S, de Latour RP, Dalle JH, Brousse VF, Petras E, DREPAGREFFE Trial Investigators. Association of matched sibling donor hematopoietic stem cell transplantation with transcranial doppler velocities in children with sickle cell anemia. J Am Med Assoc 2019 Jan 22;321(3):266-276 [FREE Full text] [doi: 10.1001/jama.2018.20059] [Medline: 30667500]
13. Brazauskas R, Scigliuolo GM, Wang HL, Cappelli B, Ruggeri A, Fitzhugh CD, et al. Risk score to predict event-free survival after hematopoietic cell transplant for sickle cell disease. Blood 2020 Jul 30;136(5):623-626 [FREE Full text] [doi: 10.1182/blood.2020005687] [Medline: 32518950]
14. Cappelli B, Volt F, Tozatto-Maio K, Scigliuolo GM, Ferster A, Dupont S, Eurocord, the Cellular Therapy and Immunobiology Working Party (CTIWP) and the Paediatric Diseases Working Party (PDWP) of the EBMT. Risk factors and outcomes

according to age at transplantation with an HLA-identical sibling for sickle cell disease. Haematologica 2019 Dec;104(12):543-546 [FREE Full text] [doi: 10.3324/haematol.2019.216788] [Medline: 31018975]
15. Eapen M, Brazauskas R, Walters MC, Bernaudin F, Bo-Subait K, Fitzhugh CD, et al. Effect of donor type and conditioning regimen intensity on allogeneic transplantation outcomes in patients with sickle cell disease: a retrospective multicentre, cohort study. Lancet Haematol 2019 Nov 01;6(11):585-596. [doi: 10.1016/s2352-3026(19)30154-1]
16. Majhail NS, Omondi NA, Denzen E, Murphy EA, Rizzo JD. Access to hematopoietic cell transplantation in the United States. Biol Blood Marrow Transplant 2010 Aug 01;16(8):1070-1075 [FREE Full text] [doi: 10.1016/j.bbmt.2009.12.529] [Medline: 20036337]
17. Jae GA, Lewkowitz AK, Yang JC, Shen L, Rahman A, Del Toro G. Barriers to conceiving sibling donors for sickle cell disease: perspectives from patients and parents. Ethn Health 2011 Aug 01;16(4-5):431-445. [doi: 10.1080/13557858.2011.558619] [Medline: 21797728]
18. Bakshi N, Katoch D, Sinha CB, Ross D, Quarmyne M, Loewenstein G, et al. Assessment of patient and caregiver attitudes and approaches to decision-making regarding bone marrow transplant for sickle cell disease: a qualitative study. JAMA Netw Open 2020 May 01;3(5):e206742 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.6742] [Medline: 32469414]
19. Stenger E, Abraham A, Travers CD, Rangarajan HG, Hsu D, Chaudhury S, et al. Development of a multi-center cohort to evaluate long-term outcomes and late effects following hematopoietic cell transplantation for sickle cell disease: A STAR initiative. Blood 2017;130(1):3271. [doi: 10.1182/blood.V130.Suppl_1.3271.3271]
20. de la Fuente J, Dhedin N, Koyama T, Bernaudin F, Kuentz M, Karnik L, et al. Haploidentical bone marrow transplantation with post-transplantation cyclophosphamide plus thiotepa improves donor engraftment in patients with sickle cell anemia: results of an international learning collaborative. Biol Blood Marrow Transplant 2019 Jun;25(6):1197-1209 [FREE Full text] [doi: 10.1016/j.bbmt.2018.11.027] [Medline: 30500440]
21. Bolaños-Meade J, Cooke KR, Gamper CJ, Ali SA, Ambinder RF, Borrello IM, et al. Effect of increased dose of total body irradiation on graft failure associated with HLA-haploidentical transplantation in patients with severe haemoglobinopathies: a prospective clinical trial. Lancet Haematol 2019 Apr;6(4):183-193. [doi: 10.1016/s2352-3026(19)30031-6]
22. Bakshi N, Sinha CB, Ross D, Khemani K, Loewenstein G, Krishnamurti L. Proponent or collaborative: Physician perspectives and approaches to disease modifying therapies in sickle cell disease. PLoS One 2017 Jul 20;12(7):e0178413 [FREE Full text] [doi: 10.1371/journal.pone.0178413] [Medline: 28727801]
23. O'Brien SH, Hankins JS. Decision analysis of treatment strategies in children with severe sickle cell disease. J Pediatr Hematol Oncol 2009 Nov;31(11):873. [doi: 10.1097/mph.0b013e3181b83cab]
24. van Besien K, Koshy M, Anderson-Shaw L, Talishy N, Dorn L, Devine S, et al. Allogeneic stem cell transplantation for sickle cell disease. A study of patients' decisions. Bone Marrow Transplant 2001 Sep 12;28(6):545-549. [doi: 10.1038/sj.bmt.1703208] [Medline: 11607766]
25. Kodish E, Lantos J, Stocking C, Singer PA, Siegler M, Johnson FL. Bone marrow transplantation for sickle cell disease. N Engl J Med 1991 Nov 07;325(19):1349-1353. [doi: 10.1056/nejm199111073251905]
26. Khemani K, Ross D, Sinha C, Haight A, Bakshi N, Krishnamurti L. Experiences and decision making in hematopoietic stem cell transplant in sickle cell disease: patients' and caregivers' perspectives. Biol Blood Marrow Transplant 2018 May;24(5):1041-1048 [FREE Full text] [doi: 10.1016/j.bbmt.2017.11.018] [Medline: 29196076]
27. Sinha CB, Bakshi N, Ross D, Loewenstein G, Krishnamurti L. Primary caregiver decision-making in hematopoietic cell transplantation and gene therapy for sickle cell disease. Pediatr Blood Cancer 2021 Jan;68(1):e28749 [FREE Full text] [doi: 10.1002/pbc.28749] [Medline: 33034129]
28. Vaismoradi M, Jones J, Turunen H, Snelgrove S. Theme development in qualitative content analysis and thematic analysis. J Nurs Edu Pract 2016 Jan 15;6(5):6-7. [doi: 10.5430/jnep.v6n5p100]
29. O'Connor AM, Cranney A. User manual-acceptability. Ottawa Hospital Research Institute, Ottawa. 1996. URL: https://decisionaid.ohri.ca/docs/develop/user_manuals/um_acceptability.pdf [accessed 2021-09-13]
30. Zhou L, Bao J, Setiawan IM, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): development and validation study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500 [FREE Full text] [doi: 10.2196/11500] [Medline: 30973342]
31. Graham MM, James MT, Spertus JA. Decision support tools: realizing the potential to improve quality of care. Can J Cardiol 2018 Jul;34(7):821-826. [doi: 10.1016/j.cjca.2018.02.029] [Medline: 29861205]
32. Patzer RE, Basu M, Larsen CP, Pastan SO, Mohan S, Patzer M, et al. iChoose kidney: a clinical decision aid for kidney transplantation vs. dialysis treatment. Transplantation 2016 Mar;100(3):630-639. [doi: 10.1097/tp.0000000000001019]
33. Thakar MS, Broglie L, Logan B, Artz A, Bunin N, Burroughs LM, et al. The Hematopoietic Cell Transplant Comorbidity Index predicts survival after allogeneic transplant for nonmalignant diseases. Blood 2019 Feb 14;133(7):754-762 [FREE Full text] [doi: 10.1182/blood-2018-09-876284] [Medline: 30545834]
34. Bhatia M, Kolva E, Cimini L, Jin Z, Satwani P, Savone M, et al. Health-related quality of life after allogeneic hematopoietic stem cell transplantation for sickle cell disease. Biol Blood Marrow Transplant 2015 Apr;21(4):666-672 [FREE Full text] [doi: 10.1016/j.bbmt.2014.12.007] [Medline: 25559691]

35. Gallo AM, Patil C, Adeniyi T, Hsu LL, Rondelli D, Saraf S. Health-related quality of life and personal life goals of adults with sickle cell disease after hematopoietic stem cell transplantation. West J Nurs Res 2019 Apr 01;41(4):555-575 [FREE Full text] [doi: 10.1177/0193945918768277] [Medline: 29624126]
36. Gallo AM, Patil CL, Knafl KA, Angst DA, Rondelli D, Saraf SL. The experience of adults with sickle cell disease and their HLA-matched adult sibling donors after allogeneic hematopoietic stem cell transplantation. J Adv Nurs 2019 Nov;75(11):2943-2951. [doi: 10.1111/jan.14152] [Medline: 31287187]
37. Hastings B, Patil C, Gallo AM. The experience and health-related quality of life after haploidentical stem cell transplantation for adults with sickle cell disease. West J Nurs Res 2019 Dec 1;41(12):1829-1844. [doi: 10.1177/0193945919870828] [Medline: 31423943]
38. Kelly MJ, Pennarola BW, Rodday AM, Parsons SK, Journeys to Recovery Study, HSCT-CHESS™ Study. Health-related quality of life (HRQL) in children with sickle cell disease and thalassemia following hematopoietic stem cell transplant (HSCT). Pediatr Blood Cancer 2012 Oct;59(4):725-731 [FREE Full text] [doi: 10.1002/pbc.24036] [Medline: 22183952]
39. Krishnamurti L, Neuberg DS, Sullivan KM, Kamani NR, Abraham A, Campigotto F, et al. Bone marrow transplantation for adolescents and young adults with sickle cell disease: Results of a prospective multicenter pilot study. Am J Hematol 2019 Apr;94(4):446-454 [FREE Full text] [doi: 10.1002/ajh.25401] [Medline: 30637784]
40. Badawy SM, Beg U, Liem RI, Chaudhury S, Thompson AA. A systematic review of quality of life in sickle cell disease and thalassemia after stem cell transplant or gene therapy. Blood Adv 2021 Jan 26;5(2):570-583 [FREE Full text] [doi: 10.1182/bloodadvances.2020002948] [Medline: 33496753]
41. Aitken C, Power R, Dwyer R. A very low response rate in an on-line survey of medical practitioners. Aust N Z J Public Health 2008 Jun;32(3):288-289. [doi: 10.1111/j.1753-6405.2008.00232.x] [Medline: 18578832]

Abbreviations

BMT: bone marrow transplant
ECHO: Extension for Community Health Care Outcomes
EFS: event-free survival
GT: gene therapy
GVHD: graft-versus-host disease
HCT: allogeneic hematopoietic stem cell transplantation
MAUQ: mobile health app usability questionnaire
NHLBI: National Heart, Lung, and Blood Institute
ODSF: Ottawa Decision Support Framework
OS: overall survival
QoL: quality of life
REDCap: Research Electronic Data Capture
SCD: sickle cell disease

Edited by G Eysenbach; submitted 30.04.21; peer-reviewed by L Hsu, T McCall, A Johnson; comments to author 14.06.21; revised version received 26.06.21; accepted 27.07.21; published 28.10.21.

Please cite as:
Veludhandi A, Ross D, Sinha CB, McCracken C, Bakshi N, Krishnamurti L
A Decision Support Tool for Allogeneic Hematopoietic Stem Cell Transplantation for Children With Sickle Cell Disease: Acceptability and Usability Study
JMIR Form Res 2021;5(10):e30093
URL: https://formative.jmir.org/2021/10/e30093
doi:10.2196/30093
PMID:34709190

Original Paper

Pilot Evaluations of Two Bluetooth Contact Tracing Approaches on a University Campus: Mixed Methods Study

Tyler Shelby[1,2], MPhil; Tyler Caruthers[1], MPH; Oren Y Kanner[3], MPhil; Rebecca Schneider[3], BA; Dana Lipnickas[3], BFA; Lauretta E Grau[1], PhD; Rajit Manohar[4], PhD; Linda Niccolai[1], PhD

[1]Epidemiology of Microbial Diseases Department, Yale School of Public Health, Yale University, New Haven, CT, United States

[2]Yale School of Medicine, Yale University, New Haven, CT, United States

[3]Information and Technology Services, Yale University, New Haven, CT, United States

[4]Yale School of Engineering and Applied Science, Yale University, New Haven, CT, United States

Corresponding Author:
Tyler Shelby, MPhil
Epidemiology of Microbial Diseases Department
Yale School of Public Health
Yale University
60 College Street
New Haven, CT
United States
Phone: 1 6202284003
Email: tyler.shelby@yale.edu

Abstract

Background: Many have proposed the use of Bluetooth technology to help scale up contact tracing for COVID-19. However, much remains unknown about the accuracy of this technology in real-world settings, the attitudes of potential users, and the differences between delivery formats (mobile app vs carriable or wearable devices).

Objective: We pilot tested 2 separate Bluetooth contact tracing technologies on a university campus to evaluate their sensitivity and specificity, and to learn from the experiences of the participants.

Methods: We used a convergent mixed methods study design, and participants included graduate students and researchers working on a university campus during June and July 2020. We conducted separate 2-week pilot studies for each Bluetooth technology. The first was for a mobile phone app (“app pilot”), and the second was for a small electronic “tag” (“tag pilot”). Participants validated a list of Bluetooth-identified contacts daily and reported additional close contacts not identified by Bluetooth. We used these data to estimate sensitivity and specificity. Participants completed a postparticipation survey regarding appropriateness, usability, acceptability, and adherence, and provided additional feedback via free text. We used tests of proportions to evaluate differences in survey responses between participants from each pilot, paired *t* tests to measure differences between compatible survey questions, and qualitative analysis to evaluate the survey’s free-text responses.

Results: Among 25 participants in the app pilot, 53 contact interactions were identified by Bluetooth and an additional 61 by self-report. Among 17 participants in the tag pilot, 171 contact interactions were identified by Bluetooth and an additional 4 by self-report. The tag had significantly higher sensitivity compared with the app (46/49, 94% vs 35/61, 57%; $P<.001$), as well as higher specificity (120/126, 95% vs 123/141, 87%; $P=.02$). Most participants felt that Bluetooth contact tracing was appropriate on campus (26/32, 81%), while significantly fewer participants felt that using other technologies, such as GPS or Wi-Fi, was appropriate (17/31, 55%; $P=.02$). Most participants preferred technology developed and managed by the university rather than a third party (27/32, 84%) and preferred not to have tracing apps on their personal phones (21/32, 66%), due to “concerns with privacy.” There were no significant differences in self-reported adherence rates across pilots.

Conclusions: Convenient and carriable Bluetooth technology may improve tracing efficiency while alleviating privacy concerns by shifting data collection away from personal devices. With accuracy comparable to, and in this case, superior to, mobile phone apps, such approaches may be suitable for workplace or school settings with the ability to purchase and maintain physical devices.

(JMIR Form Res 2021;5(10):e31086) doi:10.2196/31086

KEYWORDS

mHealth; digital contact tracing; Bluetooth; COVID-19; mixed methods

Introduction

Background

Following its identification in Wuhan, China in December 2019, SARS-CoV-2 rapidly spread across the globe, resulting in millions of infections and deaths due to COVID-19 [1]. As health organizations throughout the world worked to develop adequate pharmaceutical therapies and vaccines, many public health agencies relied on nonpharmaceutical interventions to reduce community transmission of SARS-CoV-2. In particular, the world relied on mass screening [2], lockdowns [2], physical distancing [3], mask wearing [4], and contact tracing [5]. While large-scale lockdowns and comprehensive masking interventions are less commonly seen in public health interventions, contact tracing is a traditional intervention that has proven effective in many other contexts [6-8]. However, the implementation of contact tracing for SARS-CoV-2 has faced many challenges due to high incidence rates, even among asymptomatic individuals [9], presymptomatic transmission [10], and, in many places, a lack of staffing and infrastructure [11]. These challenges made it difficult in many settings to achieve the yield (proportion of cases and contacts interviewed, isolated, and/or quarantined) and timeliness (time from symptom onset or testing to isolation for cases, and time from exposure to quarantine for contacts) thought to be required for effectiveness [12,13].

These challenges shifted the focus of many health agencies to mitigation (rather than containment) and led many to propose contact tracing innovations designed to make tracing more feasible [14]. While traditional contact tracing relies on interviewing cases and contacts in-person or by telephone, several countries augmented data collection using individual-level GPS data [15], Bluetooth technology [16], and other personalized data sources [17]. One technology in particular, Bluetooth, gained widespread attention in both the press [18] as well as scientific literature [19]. Despite the theoretical benefits of Bluetooth-assisted contact tracing and its implementation in various countries [16], the public health and lay communities are far from reaching consensus regarding the appropriateness [20] and effectiveness [21,22] of this innovation, largely due to 2 reasons.

First, many have raised concerns about the loss of individual privacy associated with automated data collection methods such as Bluetooth-assisted tracing [23,24]. In many countries, mandating participation in Bluetooth-assisted contact tracing is not feasible, and the effectiveness of this approach relies on a high user uptake among the population [22]. Implementation of Bluetooth-assisted tracing apps in nonmandated settings has so far been met with low uptake [25,26], and therefore, a better understanding of potential users' perceptions and privacy concerns is needed. Second, while research in other contexts has found various technologies, including radio frequency detectors, Wi-Fi, and Bluetooth, to be helpful in the detection of contact interactions [27-29], there are few studies evaluating the overall impact and effectiveness of Bluetooth-assisted tracing in the context of COVID-19 [30,31]. Although it seems intuitive that Bluetooth-assisted data collection may lead to an increase in the total number of identified COVID-19 "close contacts" (defined by the Centers for Disease Control and Prevention [CDC] as in-person interactions within 6 feet for at least 15 minutes) and more rapid identification of these individuals, there is little real-world data to directly verify this or to evaluate the accuracy of Bluetooth data [21,30].

Goal of This Study

Together, doubts about the appropriateness and acceptability of Bluetooth-assisted contact tracing and the accuracy and reliability of the data pose challenges to implementation and adoption. Due to low vaccine uptake [32,33] and breakthrough transmission by variant strains [34], overcoming these challenges is critical as contact tracing will remain a core part of the public health response to COVID-19, even in the postvaccine phase of the pandemic. To address these knowledge gaps, we pilot tested 2 different Bluetooth-assisted tracing technologies on a university campus, one which collected Bluetooth data using a mobile phone app and another that used a separate carriable device ("tag") with Bluetooth functionality. Using a convergent mixed methods design, we measured the sensitivity and specificity of each Bluetooth technology and assessed participant perceptions regarding appropriateness, usability, acceptability, and adherence, using a quantitative survey and qualitative free-text analysis.

Methods

Study Setting and Population

We conducted 2 separate pilot studies in June to July 2020 at a medium-sized private university in the US Northeast. During this time, only essential personnel and select individuals were allowed on campus with prior approval. Campus-wide precautions included mask wearing, physical distancing, daily symptom assessments, and testing. Study participants included graduate students and researchers working during this period; graduate students or researchers working from home were ineligible for participation. We recruited participants by emailing faculty members and lab supervisors who subsequently forwarded our recruitment emails to their students and research staff. We then selected labs with the highest acceptance rates. We also prioritized enrollment from labs that shared workspaces with other recruited labs. Due to the focused nature of the pilots, we did not collect demographic data from participants. Each of the sequential pilots lasted 2 weeks (14 days) starting on a Monday, and different labs participated in the separate pilots. Sample size was determined by the availability of required study devices. The collected data were stored on secure university servers throughout the study and analysis period.

Pilot 1: Mobile Phone–Based Bluetooth

Technology

In the first pilot (hereafter referred to as the "app pilot"), we evaluated a mobile phone app developed by the university's information technology services staff (Multimedia Appendix 1). It functioned by detecting Bluetooth signals emitted by other phones that had the same app downloaded and activated. The app estimated the distance between mobile phones based on signal strength while recording the duration of the interaction. The app also had functionality for users to enter a date of symptom onset or positive test; however, this function was not used during the pilot. Data were automatically sent to a centralized server. The university provided Android phones to participants for the duration of the study, so that they did not have to download the app on their personal devices.

Setting and Data Collection

All app pilot participants were provided with written instructions describing how to install and use the mobile app and how to validate and report new contact interactions, as well as contact information for technical support if needed. Participants were asked to carry the study phone while on campus. At the end of each day, participants reviewed an online spreadsheet of their Bluetooth-identified close contacts and confirmed or denied each interaction. We also asked participants to identify additional contacts that were not detected by Bluetooth, and we subsequently removed any self-reported contacts who were not study participants. Participants were asked to use their best judgment when estimating the length of each interaction.

Pilot 2: Tag-Based Bluetooth

Technology

In the second pilot (hereafter referred to as the "tag pilot"), we evaluated a carriable device ("tag") equipped with Bluetooth functionality, designed by the author RM (Multimedia Appendix 2 and Multimedia Appendix 3). The tags recorded Bluetooth signals emitted from other tags, using signal strength to determine distance while recording the duration of interactions. Data were stored locally on the tags and routinely synced to a central server by study participants using a mobile app that paired with the participant's tag. The app only used Bluetooth to communicate with the tag while syncing and otherwise did not collect any additional data or use Bluetooth to communicate with any nonpaired tags or other devices. The tag software additionally allows for contact interactions to be encrypted when recorded and stored in the central server, thereby anonymizing the data. When this feature is active, decrypting the data requires the user to provide permission by submitting a decryption token through the app. However, this feature was not enabled during the study, so that we could determine all contact records for the purpose of evaluating the system's efficacy. Additional details regarding the tag's development can be found elsewhere [35]. The university provided participants with Android phones for the duration of the pilot to facilitate syncing of tag data. Participants were asked to use their best judgment when estimating the length of each interaction.

Setting and Data Collection

All tag pilot participants were provided with written instructions describing how to install and use the mobile syncing app, how to pair it with their Bluetooth tag, and how to validate and report new contact interactions, as well as contact information for technical support if needed. Participants were asked to carry the tag while on campus and to sync their Bluetooth data after each shift. At the end of each day, participants reviewed a list of their Bluetooth-identified close contacts and confirmed or denied each interaction using an online web interface. We also presented participants with the estimated duration of each recorded interaction and asked participants to report if the duration was underestimated or overestimated. Similar to the app pilot, we asked participants to identify additional contacts not detected by Bluetooth and subsequently removed those who were not study participants.

Postparticipation Survey

Following each pilot, we sent a survey to participants focusing on their experiences using the pilot technology, as well as their perceptions regarding the appropriateness of technology-assisted tracing on campus (see Table 1 for survey domains). We adapted this survey from a previously validated mHealth usability questionnaire [36]. Most questions used a 7-point Likert scale ranging from *strong agreement* to *strong disagreement*, including a *neutral* response option. The survey also contained a free-text question asking participants to provide any additional comments about their experience or suggestions about the technology. We used Cronbach alpha to measure the reliability of our adapted scale after aligning the directionality of question responses. We excluded the free-text response and 2 other scale items from the reliability measurement that asked participants to select various ways in which they carried the devices or reasons why they were not carried.

Table 1. Postparticipation survey overview.

Domain and subdomain	Goals within the domain/subdomain
Appropriateness	To measure participant perceptions about the appropriateness of Bluetooth contact tracing and the use of certain types of data (Bluetooth, GPS, Wi-Fi, etc)
Usability	
Ease of use	To measure the ease with which participants install, learn to use, and use the apps
Interface and satisfaction	To measure participant experiences and satisfaction with the design and interface of the app
Acceptability	
Usefulness	To evaluate participant beliefs surrounding the usefulness of the tracing technology
Coherence	To evaluate participants' understanding of how data are collected and protected by the technology
Social influence	To measure the presence of social influence from peers or supervisors regarding uptake of technology-assisted tracing
Setting	To measure perceptions about available assistance for the use of the apps and/or devices and individual agency in uptake
Adherence	To measure adherence and participant preferences with regard to carrying the study devices

Analysis Plan

Quantitative Study Outcomes and Measurements

We used participants' daily contact validation responses to estimate the sensitivity and specificity of the 2 technologies (see Table 2 for outcome and measure definitions) and used 2-tailed tests of proportions to compare these values between pilots. We also described the postparticipation survey by presenting proportions of participants agreeing with each Likert question or selecting responses from other categorical questions, as well as means for responses to continuous questions. We measured differences in survey responses between participants from different pilot groups using 2-tailed tests of proportions for Likert agreement and categorical questions, and unpaired 2-tailed *t* tests for continuous questions. Additionally, we used paired tests of proportions to measure differences between agreement with several comparable survey questions, including (1) appropriateness of Bluetooth vs location data (GPS and/or Wi-Fi) for contact tracing, (2) peer vs supervisor vocal support of study technology, and (3) peer vs supervisor vocal concern about the study technology.

Table 2. Definitions of Bluetooth measures and outcomes.

Measures/outcomes	Definition
Measures[a]	
True positive	Bluetooth-identified contact that is confirmed by the participant
True negative	No contacts detected, confirmed by the participant
False positive	Bluetooth-identified contact denied by the participant
False negative	Participant-recalled contact that was not detected by Bluetooth
Outcomes	
Sensitivity	True positive/(true positive + false negative)
Specificity	True negative/(true negative + false positive)

[a]15 minutes of interaction within 6 feet required to meet the definition of "close contact." In addition to confirming/denying each close contact interaction, participants from the tag pilot were asked to comment on the underestimation or overestimation of the recorded contact duration. We allowed a 5-minute window of error, within which a contact's measurement type could be altered. For example, a contact detected for 15-19 minutes would be designated as a false positive if the study participant noted that the interaction length was overestimated, while a contact detected for 10-14 minutes would be designated as a false negative if the study participant noted that the interaction length was underestimated.

Qualitative Analysis of Free-Text Responses

The coding team (TS and LG) used a codebook that was deductively based upon the survey topics. TS coded the free-text responses, and the coding team met regularly to review the coded text and reach agreement on all coding decisions. The coding team also refined code definitions and generated new codes when applicable throughout the coding process. "RADaR," a rapid qualitative analysis approach [37], was used, in which the coding and analysis were done in Microsoft Excel (Microsoft Corp) rather than in a traditional qualitative analysis software. We synthesized the qualitative and quantitative aspects as part of the mixed methods analysis [38,39] by identifying quotes that provided greater context or deeper understanding for the findings from the quantitative survey analyses. Selected quotes are presented alongside the quantitative findings within the relevant survey domains.

Study Approval

This study was approved by the Yale Human Subjects Committee, and written consent was obtained from participants prior to enrollment. We did not offer incentives for participation.

Results

Study Participants, Number of Shifts, and Frequencies of Contact Interactions

We invited 33 participants from 7 labs for the app pilot, of which 30 agreed to participate, and 25 completed the 2-week period of follow-up. Overall, 53 contact interactions were identified via Bluetooth, and an additional 61 were reported by participant recall. We invited 24 participants from 2 labs for the tag pilot, of which 17 agreed to participate, and all completed the 2-week period of follow-up. A defect was identified in the tag cases at the end of the first week of data collection that rendered the data unusable. The cases were then replaced, and only the data from the second study week were further analyzed. In the second week of data collection, 171 contact interactions were identified by Bluetooth and an additional 4 were reported by participant recall.

Sensitivity and Specificity

We present estimates of sensitivity and specificity, and counts of true/false positives and negatives in Table 3, stratified by pilot. The tag pilot had significantly higher sensitivity compared to the app pilot (46/49, 94% vs 35/61, 57%; $P<.001$), as well as higher specificity (120/126, 95% vs 123/141, 87%; $P=.02$). Of note, 3 participants in the tag pilot reported leaving their tags on their desks during days on which they were not on campus, resulting in false recordings of contact interactions. When these interactions were removed from the data set, sensitivity and specificity became 93% (43/46) and 100% (111/111), respectively.

Table 3. Counts of true/false positives and negatives, and estimates of sensitivity and specificity.

Measures/outcomes	App pilot	Tag pilot
Measures, n		
True positive	35	46
True negative	123	120
False positive	18	6
False negative	26	3
Outcomes, %		
Sensitivity	57%	94% (93%[a])
Specificity	87%	95% (100%[a])

[a]Adjusted values after removing erroneous contact records from tags left on participants' desks when they were not on campus.

Postparticipation Survey

Twenty participants from the app pilot and 12 participants from the tag pilot completed the postparticipation survey (Cronbach $\alpha=.90$). Below, we present the quantitative results from each section alongside qualitative findings when applicable.

Appropriateness

Overall, there were no differences in perceived appropriateness of technology-assisted tracing among participants between pilot groups (Table 4). Most participants felt that contact tracing via Bluetooth was appropriate but felt that the use of additional location data such as GPS or Wi-Fi was less appropriate (26/32, 81% approval for Bluetooth vs 17/31, 55% approval for GPS/Wi-Fi; $P=.02$). Most participants also preferred technology developed and managed by the university rather than a third party (27/32, 84%) and preferred to not download apps on their personal devices (21/32, 66%).

Regardless of the approach, most participants (24/32, 75%), though not all, reported concerns about how their privacy would be protected, and these concerns were expanded upon in the free-text data.

> *One [lab member] voiced concerns about how individual GPS contact data might be used against individuals (such as by police in the case of protests) - sadly, similar to what actually happened with a Mayor releasing names publicly recently....I think if the privacy aspect is addressed VERY clearly and intentionally it might increase the acceptance.* [App pilot, Participant #3]

> *I do have some concerns with privacy, but I am not sophisticated enough in this topic to articulate my concerns or to understand if I should be concerned or not. I think the data from a school-wide system does have the potential to be abused, but I think an effective contact tracing system should/could significantly increase the safety of students, faculty, and staff on campus.* [Tag pilot, Participant #17]

Table 4. Postparticipation survey: appropriateness domain.

Questions	Total percentage agreement[a] (N=32)[b], % (n/N)	App percentage agreement[a] (N=20)[b], % (n/N)	Tag percentage agreement[a] (N=12)[b], % (n/N)	*P* value[c]
It is appropriate for the university to use Bluetooth apps to monitor interactions on campus in order to more efficiently perform contact tracing.	81 (26/32)	80 (16/20)	83 (10/12)	.82
It is appropriate to use location information such as GPS and Wi-Fi connection data for contact tracing.	55 (17/31)	58 (11/19)	50 (6/12)	.67
I would prefer to use a contact tracing app on a university-owned device as opposed to downloading the app on my personal phone.	66 (21/32)	65 (13/20)	67 (8/12)	.92
I would prefer to use an app developed and owned by the university as opposed to an app developed and owned by an independent third party.	84 (27/32)	85 (17/20)	83 (10/12)	.90
I have concerns about how using this app, or an app like it, could affect my privacy.	75 (24/32)	70 (14/20)	83 (10/12)	.40

[a]Percentage agreement was calculated by dividing the number of Likert responses indicating agreement by the total number of Likert responses for each question.

[b]Some questions were not answered by all participants; exact counts of agreement and total responses are shown in parentheses for each question.

[c]*P* values obtained using tests of proportions to evaluate differences between pilots.

Usability

There were no observed differences between pilot groups regarding app usability (Table 5), and most participants from both pilots felt their respective apps were easy to install (25/31, 81%) and use (31/32, 97%). They also reported moderate levels of satisfaction with the app interfaces (21/32, 66%) and feedback from the apps (18/31, 58%). The amount of time required to use the apps was acceptable to most (29/32, 91%), and overall satisfaction was high (26/32, 81%). However, several participants from both pilots described difficulties downloading and installing the apps, syncing tags to mobile devices for uploading data, discerning how the app was responding to the user due to unclear feedback from the app, or experiencing other technological glitches.

> *[The app] would switch tracking off by itself.* [App pilot, Participant #13]

> *When I first obtained the phone, there was no contact tracing app on it, and I could not find a way to download it…When I tried syncing the tag to the phone, there was never a message telling me that the tag was synced, only "connecting" and "communicating."* [Tag pilot, Participant #19]

Table 5. Postparticipation survey: usability domain.

Subdomains and questions	Total percentage agreement[a] (N=32)[b], % (n/N)	App percentage agreement[a] (N=20)[b], % (n/N)	Tag percentage agreement[a] (N=12)[b], % (n/N)	*P* value[c]
Ease of use				
It was easy for me to install the app on the device.	81 (25/31)	84 (16/19)	75 (9/12)	.53
It was easy for me to learn to use the app.	97 (31/32)	95 (19/20)	100 (12/12)	.43
The app was easy to use.	97 (31/32)	95 (19/20)	100 (12/12)	.43
Interface and satisfaction				
I like the interface of the app.	66 (21/32)	65 (13/20)	66 (8/12)	.92
The information in the app was well organized, so I could easily find the information I needed.	71 (22/31)	63 (12/19)	83 (10/12)	.23
The app adequately acknowledged and provided information to let me know the progress of my action.	58 (18/31)	53 (10/19)	66 (8/12)	.44
The amount of time involved in using the app is acceptable.	91 (29/32)	85 (17/20)	100 (12/12)	.16
I would use this system again.	78 (25/32)	70 (14/20)	92 (11/12)	.15
Overall, I am satisfied with this system.	81 (26/32)	80 (16/20)	83 (10/12)	.82

[a]Percentage agreement was calculated by dividing the number of Likert responses indicating agreement by the total number of Likert responses for each question.

[b]Some questions were not answered by all participants; exact counts of agreement and total responses are shown in parentheses for each question.

[c]*P* values obtained using tests of proportions to evaluate differences between pilots.

Acceptability

Most participants felt that their respective app or tag would be useful for contact tracing (25/31, 81%), though a lack of consistency between recalled interactions and Bluetooth data diminished some participants' confidence in the technology.

> *The device initially failed to detect other devices, and therefore I'm worried about the efficiency of the app.* [App pilot, Participant #7]

> *I think that when it worked, it was great. There were times, such as my first day, where it didn't detect anyone even though I was well within 6 feet.* [Tag pilot, Participant #15]

Most participants understood how their respective device collected (27/32, 84%) and protected their data (22/32, 69%) (Table 6). With regard to social influence and study setting, there were no significant differences between pilot environments. Across both pilots, participants more frequently reported vocal support for the technology from supervisors than from peers (21/26, 81% from supervisors vs 10/27, 37% from peers; *P*=.001). The opposite was true regarding vocal concern, with participants more frequently reporting vocal concern from peers compared to supervisors (13/29, 45% from peers vs 2/25, 8% from supervisors; *P*=.003). Within the study environment, most participants felt that adequate technical assistance was available when needed (20/28, 71%), and also felt that, should the university adopt such technology, they would maintain individual agency over whether or not they used the devices (26/31, 84%).

Table 6. Postparticipation survey: acceptability domain.

Subdomains and questions	Total percentage agreement[a] (N=32)[b], % (n/N)	App percentage agreement[a] (N=20)[b], % (n/N)	Tag percentage agreement[a] (N=12)[b], % (n/N)	*P* value[c]
Usefulness				
The system would be useful for contact tracing.	81 (25/31)	74 (14/19)	92 (11/12)	.22
The app has all the functions and capabilities I expected it to have.	58 (18/31)	42 (8/19)	83 (10/12)	.02
Coherence				
I understand how data collected with this system would be used for contact tracing.	84 (27/32)	80 (16/20)	92 (11/12)	.38
I understand how this system currently protects my privacy.	69 (22/32)	65 (13/20)	75 (9/12)	.56
Social influence				
Peers whose opinions I value have vocalized their support for this system.	37 (10/27)	24 (4/17)	60 (6/10)	.06
Supervisors in my workplace have vocalized their support for this system.	81 (21/26)	83 (15/18)	75 (6/8)	.62
Peers whose opinions I value have voiced concerns about using this system.	45 (13/29)	50 (9/18)	36 (4/11)	.47
Supervisors in my workplace have voiced concerns about using this system.	8 (2/25)	12 (2/17)	0 (0/8)	.31
Setting				
Technical assistance was available when needed.	71 (20/28)	71 (12/17)	73 (8/11)	.90
The decision to use or not use this system will remain under my control.	84 (26/31)	79 (15/19)	92 (11/12)	.35

[a]Percentage agreement was calculated by dividing the number of Likert responses indicating agreement by the total number of Likert responses for each question.

[b]Some questions were not answered by all participants; exact counts of agreement and total responses are shown in parentheses for each question.

[c]*P* values obtained using tests of proportions to evaluate differences between pilots.

Adherence

There was no difference between pilots in overall adherence rates based on self-reported percentages of shifts during which the study device was carried (mean 87%) (Table 7), although participants in the tag pilot more commonly reported that their study device was convenient to carry than did participants from the app pilot (tag pilot: 11/12, 92% vs app pilot: 11/20, 55%; *P*=.03). While some participants from the app pilot reported leaving the device at home (2/13, 15%), participants from both pilots reported that the most common reason for not carrying the devices was forgetting it at a workstation (17/23, 74%). App pilot participants also reported inabilities to carry the study device into certain lab environments (app pilot: 5/13, 38% vs tag pilot: 0/10, 0%; *P*=.03), while tag pilot participants reported that charging the device interfered with adherence (tag pilot: 3/10, 30% vs app pilot: 0/13, 0%; *P*=.03).

Many participants from the app pilot used the free-text response to note the inconvenience of carrying an additional phone and suggested that a smaller device be used. A minority suggested that they be allowed to download the tracing app directly on their personal phones. Gender-specific difficulties in carrying the app pilot study phone were also noted by 1 participant, while a separate participant from the tag pilot noted the relative ease of carrying the tag.

> *The only problem I found with this [study phone] is that it is big. For women it just does not fit in the front pocket of the jeans and in the summer, you are not wearing a jacket under your lab coat. So, the only place left is the pocket of the jeans in the back. And that is a bit uncomfortable when you sit down, or you are scared it might fall out. I also do not feel good putting it in the pockets of my lab coat because I consider them "dirty" and I do not want to have lab dirt in my home, or touch it without gloves. So, it would be much more convenient if it would be a bracelet or a watch or something around those lines.* [App pilot, Participant #12]

> *The shape of [the tag] is pretty clunky to carry around, but as long as you wear pants with pockets it's easy enough to just wear in your back pocket.* [Tag pilot, Participant #16]

The vast majority of participants from the app pilot reported that they would be more likely to carry a Bluetooth device if it were smaller than a phone (19/20, 95%), while no participants from the tag pilot (0/12, 0%) agreed that increasing the size of

the tag would increase adherence (P<.001), indicating an overall preference for smaller devices.

Table 7. Postparticipation survey: adherence domain.

Questions	Total percentage agreement[a,b] (N=32)[c], % or % (n/N)	App percentage agreement[a,b] (N=20)[c], % or % (n/N)	Tag percentage agreement[a,b] (N=12)[c], % or % (n/N)	P value[d]
Over the course of the 2-week study period, for what proportion of your total work shifts did you have the device either on you or within arms' reach?	87[e]	91[e]	81[e]	.06
The device was convenient to carry with me throughout my work shifts.	69 (22/32)	55 (11/20)	92 (11/12)	.03
How did you carry the device with you throughout your workday? (tag only)				
Pocket	N/A[f]	N/A	92 (11/12)	N/A
Bag	N/A	N/A	0 (0/12)	N/A
Belt/lanyard	N/A	N/A	8 (1/12)	N/A
Left at workspace	N/A	N/A	8 (1/12)	N/A
What were the most common reasons why you would not carry the device with you during a work shift?				
Forgot at home	9 (2/23)	15 (2/13)	0 (0/10)	.19
Intentionally left at home	0 (0/23)	0 (0/13)	0 (0/10)	N/A
Forgot at desk/workstation	74 (17/23)	69 (9/13)	80 (8/10)	.56
Intentionally left at desk/workstation	9 (2/23)	15 (2/13)	0 (0/10)	.19
Unable to carry it into certain lab environments	22 (5/23)	38 (5/13)	0 (0/10)	.03
Left it to charge	13 (3/23)	0 (0/13)	30 (3/10)	.03
I would be more likely to carry the device with me if it were smaller (for instance, the size of a thumb drive that could be attached to a lanyard). (app only)	N/A	95 (19/20)	N/A	<.001
I would be more likely to carry the tag with me if it were larger (for instance, the size of a phone). (tag only)	N/A	N/A	0 (0/12)	<.001

[a]Unless otherwise specified.

[b]Percentage agreement was calculated by dividing the number of Likert or binary responses indicating agreement by the total number of responses for each question.

[c]Some questions were not answered by all participants; exact counts of agreement and total responses are shown in parentheses for each question.

[d]P values obtained by tests of proportions for differences in percentage agreement and by unpaired t tests for differences in means.

[e]Mean response.

[f]N/A: not applicable.

Discussion

Principal Findings and Implications

Incomplete vaccine uptake [32,33] and potential for breakthrough transmission due to new variants [34] suggest that contact tracing will remain an important tool in the ongoing response to COVID-19. However, its use thus far in the pandemic has revealed many challenges to scaling-up traditional contact tracing [40-43] and identified a need to improve upon existing methods. Digital contact tracing tools offer many opportunities to improve the impact of contact tracing [44], and increasing our understanding of how different technologies may be applied for this purpose is critical. In our dual-pilot evaluation of 2 novel contact tracing technologies, we found that Bluetooth contact tracing was perceived as appropriate to the majority of study participants, adherence to device carrying was high, and participants were largely satisfied with their experiences. However, most participants still reported concerns about privacy, and both technologies encountered occasional technical glitches. Importantly, we also found that the tag-based device was easier to carry and had superior sensitivity and specificity. These increased performance metrics may have been due to differences between the Bluetooth signal strength settings of the technologies or in how participants carried the different study devices, as reflected in the postparticipation survey.

Our findings are similar to a recent study [45] that compared a Bluetooth mobile app to a wearable, radio frequency-based, real-time locator device within a health care setting. The researchers found the wearable device to be superior to Singapore's "TraceTogether" app with regard to sensitivity and specificity, and also found that the app's performance was worse on iPhones compared to Android devices. In a similar study, a wearable device was compared to electronic medical record-assisted tracing and was again found to be superior [46].

Our study builds upon these findings by evaluating similar app-based technology in a new university setting, while also comparing it directly to a novel Bluetooth tag device, rather than a radio frequency-based device.

Although most proximity-based contact tracing technologies offer similar benefits, such as the ability to identify unknown contacts or customize detection thresholds based on evolving knowledge of transmission dynamics [47], different approaches (eg, app vs carriable device) offer certain additional benefits and drawbacks. Below, we discuss key differences while paying heed to the importance of context. While traditional contact tracing focuses on community and population transmission, COVID-19 has led many closed-door environments, such as workplaces, schools, universities, and hospitals, to conduct contact tracing independently from, or in partnership with, local public health systems [48,49]. The differences between community tracing and closed-door tracing are important when comparing app-based and tag-based systems, as different contexts are often coupled with different funding capacities, thresholds for acceptable uptake of tracing technology, and user privacy concerns.

Deploying Bluetooth tracing technology to communities or populations at large is likely only feasible using an app-based system. App-based tracing technologies, such as those developed by Apple and Google, have already been deployed throughout the globe [16], including in many US states [26], with relatively little cost to distribution beyond social marketing. Meanwhile, it would not be logistically or financially feasible to deploy a similar number of tag devices throughout the population, as each tag costs approximately US $10. Furthermore, while updating apps is relatively seamless, updating hardware poses a greater challenge, as we encountered in this study when we discovered a defect in our tag cases. Despite these potential drawbacks, tags and similar approaches may be more feasible in closed-door environments that have available funds to spend on the protection of a much smaller population.

Acceptable thresholds for uptake may also differ between environments, making the logistical concerns noted above more or less important across different settings. Public health officials in many countries are often hesitant or unable to mandate participation in health interventions, as demonstrated with mask policies in response to COVID-19 [50]. Public health programs also frequently lack funding to properly incentivize participation. As a result, population-wide uptake of app-based technology for tracing will likely always be limited. Closed-door environments, on the other hand, may face greater pressure to standardize and ensure the safety of all staff, students, or workers, and therefore may prioritize, or mandate, comprehensive uptake, as demonstrated by many universities requiring vaccination for all students [51]. However, reaching such a high uptake of digital contact tracing without diminishing individual agency or ignoring privacy concerns poses a challenge.

Privacy concerns are often related to the types of information collected as well as the organization or government collecting the data [23,24], and may be heightened in the context of a pandemic [52]. Notably, our study participants felt that using Bluetooth data for tracing was more appropriate than GPS or Wi-Fi data. While technologies, such as blockchain, may increase the security of app-based approaches [53] and further reduce the risks of data leakage, effectively communicating such methods and establishing trust with potential users may remain difficult as long as data collection relies on personal devices, as reflected by our participants' preferences against using apps on their phones. This provides several arguments for shifting data collection away from personal devices and onto organization-owned tracing tags when possible. First, the tag-based system offers users in closed-door environments the opportunity to participate in contact tracing without requiring data collection on their phones. While our study still relied on an app to sync the tag's data, the provision of "syncing stations" throughout closed-door environments could eliminate the need for an app entirely and further reduce concerns about leakage of personal phone data. Second, the use of organization-owned tags addresses concerns about governments or third-party companies accessing personal data [23,52], which was reflected in our participants' preferences against third party apps. Ultimately, these features offer the potential to reduce privacy concerns and increase uptake within closed-door environments.

There are several key strengths to this study, including the use and evaluation of novel technologies developed directly in response to the COVID-19 pandemic. Second, the setting in which the study was conducted is typical of some other environments, in particular schools and universities, that have struggled to perform contact tracing throughout the pandemic, making this study increasingly relevant to public health practitioners or researchers operating in similar environments. Lastly, the use of mixed methods, including sensitivity and specificity estimations, survey analyses, and qualitative analysis, allowed us to triangulate our findings and present a layered evaluation of the technologies' performance metrics as well as the users' experiences.

There are also several important limitations in this study. First, the sample size was relatively small, increasing the risk of type II errors. Second, the recruitment of different labs and participants for each pilot created some uncertainty about the mechanisms driving the observed differences in Bluetooth performance metrics and user experiences or perceptions. However, the lack of significant differences in survey responses regarding setting and social influences, and the baseline similarities in lab environments selected for the study minimize this risk. Third, the lack of a true "gold standard" measurement for close contact interactions introduces the potential for bias in the estimations of sensitivity and specificity. In particular, recall bias may have led to misreporting of self-report contacts, and the lack of precise measurements for the length of self-report interactions between participants may have introduced additional uncertainties. However, participants' daily review and validation of contact interactions likely minimized the potential for recall bias, which would have been more severe if the data were collected less frequently. Furthermore, these potential biases likely affected each pilot similarly, which lessens the degree to which such biases may have affected the comparisons between pilots. Fourth, based upon the participant-initiated method of qualitative data collection (optional free-text box vs traditional

interview queries), it is doubtful that meaning saturation [54] was achieved and likely that themes would have been better explicated and perhaps more abundant if a traditional approach to qualitative interviewing had been used. Nonetheless, the study provides preliminary evidence about the relative merits of the 2 technologies that can inform larger studies in the future. Fifth, demographic data were not collected from participants at the time of recruitment, limiting our ability to evaluate differences across participant characteristics. Considering the small sample size and short timeframe of the pilots, we lacked statistical power to evaluate differences across participant characteristics and therefore did not include this as a study goal. Last, the relative homogeneity of the study sample may limit the generalizability of our findings to other nonuniversity contexts, which may feature differences in behavior, familiarity with technology, and/or attitudes [55].

Conclusion

As vaccine uptake remains noncomprehensive and new variants appear, contact tracing will remain a pillar of the public health response to COVID-19. Increasing the efficiency of contact tracing through adoption of technologies, such as those evaluated here, may improve its impact and ability to prevent or control outbreaks. This is among the first studies to directly evaluate the performance metrics of novel Bluetooth technologies when used for COVID-19 contact tracing in conjunction with evaluations of user experiences. Our participants found Bluetooth-assisted tracing to be appropriate, and we noted several key differences between app-based and tag-based approaches. The benefits of the app-based system include its low cost and theoretical ease of mass distribution, and the drawbacks include increased privacy concerns of users. The benefits of the tag system include its superior sensitivity and specificity, the ease of carrying the tag, and the potential to alleviate user privacy concerns, and the drawbacks include its reliance on hardware that may be less feasible to deploy in certain settings.

Acknowledgments

We would like to acknowledge the university administration for facilitating these pilots, as well as the graduate laboratory supervisors, staff, and students for their participation. This study was funded by Yale University.

Authors' Contributions

TS, OK, DL, RM, and LN contributed to the study design. OK, RS, DL, and RM contributed to the development of study devices. TS, TC, OK, RS, DL, and RM contributed to data collection. TS, TC, RM, LG, and LN contributed to study analyses. TS and LN contributed to initial manuscript drafting. All other authors contributed to manuscript editing and revision. LN provided study oversight.

Conflicts of Interest

The authors disclose that Yale University, University of California, Los Angeles, and Carnegie Mellon University have a patent pending for the Bluetooth tag device, and the author RM has a personal financial interest through the standard patent policy at Yale University. The authors also disclose that LN is a member of the Scientific Advisory Board for Moderna, and TS is part of a COVID-19 support contract between the State of Connecticut Department of Public Health and Yale School of Public Health.

Multimedia Appendix 1
Screenshot of the app pilot mobile app.
[DOCX File , 43 KB - formative_v5i10e31086_app1.docx]

Multimedia Appendix 2
Bluetooth device (“tag”) used in the tag pilot.
[DOCX File , 652 KB - formative_v5i10e31086_app2.docx]

Multimedia Appendix 3
Screenshot of the tag pilot mobile syncing app.
[DOCX File , 44 KB - formative_v5i10e31086_app3.docx]

References

1. Keni R, Alexander A, Nayak PG, Mudgal J, Nandakumar K. COVID-19: Emergence, Spread, Possible Treatments, and Global Burden. Front Public Health 2020 May 28;8:216-216 [FREE Full text] [doi: 10.3389/fpubh.2020.00216] [Medline: 32574299]
2. Johanna N, Citrawijaya H, Wangge G. Mass screening vs lockdown vs combination of both to control COVID-19: A systematic review. J Public Health Res 2020 Oct 14;9(4):2011-2531 [FREE Full text] [doi: 10.4081/jphr.2020.2011] [Medline: 33409247]

3. Chu D, Akl EA, Duda S, Solo K, Yaacoub S, Schünemann HJ, COVID-19 Systematic Urgent Review Group Effort (SURGE) study authors. Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis. Lancet 2020 Jun 27;395(10242):1973-1987 [FREE Full text] [doi: 10.1016/S0140-6736(20)31142-9] [Medline: 32497510]
4. Van Dyke ME, Rogers TM, Pevzner E, Satterwhite CL, Shah HB, Beckman WJ, et al. Trends in County-Level COVID-19 Incidence in Counties With and Without a Mask Mandate - Kansas, June 1-August 23, 2020. MMWR Morb Mortal Wkly Rep 2020 Nov 27;69(47):1777-1781 [FREE Full text] [doi: 10.15585/mmwr.mm6947e2] [Medline: 33237889]
5. Yalaman A, Basbug G, Elgin C, Galvani AP. Cross-country evidence on the association between contact tracing and COVID-19 case fatality rates. Sci Rep 2021 Jan 25;11(1):2145 [FREE Full text] [doi: 10.1038/s41598-020-78760-x] [Medline: 33495491]
6. Ooi PL, Lim S, Chew SK. Use of quarantine in the control of SARS in Singapore. Am J Infect Control 2005 Jun;33(5):252-257 [FREE Full text] [doi: 10.1016/j.ajic.2004.08.007] [Medline: 15947741]
7. Mahachi N, Muchedzi A, Tafuma TA, Mawora P, Kariuki L, Semo B, et al. Sustained high HIV case-finding through index testing and partner notification services: experiences from three provinces in Zimbabwe. J Int AIDS Soc 2019 Jul 19;22 Suppl 3(S3):e25321-e25330 [FREE Full text] [doi: 10.1002/jia2.25321] [Medline: 31321918]
8. Fox GJ, Barry SE, Britton WJ, Marks GB. Contact investigation for tuberculosis: a systematic review and meta-analysis. Eur Respir J 2013 Jan;41(1):140-156 [FREE Full text] [doi: 10.1183/09031936.00070812] [Medline: 22936710]
9. Al-Sadeq DW, Nasrallah GK. The incidence of the novel coronavirus SARS-CoV-2 among asymptomatic patients: A systematic review. Int J Infect Dis 2020 Sep;98:372-380 [FREE Full text] [doi: 10.1016/j.ijid.2020.06.098] [Medline: 32623083]
10. Bi Q, Wu Y, Mei S, Ye C, Zou X, Zhang Z, et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. Lancet Infect Dis 2020 Aug;20(8):911-919 [FREE Full text] [doi: 10.1016/S1473-3099(20)30287-5] [Medline: 32353347]
11. Clark E, Chiao E, Amirian E. Why Contact Tracing Efforts Have Failed to Curb Coronavirus Disease 2019 (COVID-19) Transmission in Much of the United States. Clin Infect Dis 2021 May 04;72(9):e415-e419 [FREE Full text] [doi: 10.1093/cid/ciaa1155] [Medline: 32761123]
12. Prioritizing COVID-19 Contact Tracing Mathematical Modeling Methods and Findings. CDC. URL: https://www.cdc.gov/coronavirus/2019-ncov/php/contact-tracing/contact-tracing-plan/prioritization/mathematicalmodeling.html [accessed 2021-03-21]
13. Ferretti L, Wymant C, Kendall M, Zhao L, Nurtay A, Abeler-Dörner L, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 2020 May 08;368(6491):1-7 [FREE Full text] [doi: 10.1126/science.abb6936] [Medline: 32234805]
14. He W, Zhang Z, Li W. Information technology solutions, challenges, and suggestions for tackling the COVID-19 pandemic. Int J Inf Manage 2021 Apr;57:102287 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2020.102287] [Medline: 33318721]
15. Contact Transmission of COVID-19 in South Korea: Novel Investigation Techniques for Tracing Contacts. Osong Public Health Res Perspect 2020 Feb;11(1):60-63 [FREE Full text] [doi: 10.24171/j.phrp.2020.11.1.09] [Medline: 32149043]
16. Kleinman RA, Merkel C. Digital contact tracing for COVID-19. CMAJ 2020 Jun 15;192(24):E653-E656 [FREE Full text] [doi: 10.1503/cmaj.200922] [Medline: 32461324]
17. Jian S, Cheng H, Huang X, Liu D. Contact tracing with digital assistance in Taiwan's COVID-19 outbreak response. Int J Infect Dis 2020 Dec;101:348-352 [FREE Full text] [doi: 10.1016/j.ijid.2020.09.1483] [Medline: 33035674]
18. Fowler G. A covid-fighting tool is buried in your phone. Turn it on. The Washington Post. URL: https://www.washingtonpost.com/technology/2020/11/18/coronavirus-app-exposure-alerts/ [accessed 2021-03-29]
19. Abeler J, Bäcker M, Buermeyer U, Zillessen H. COVID-19 Contact Tracing and Data Protection Can Go Together. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e19359 [FREE Full text] [doi: 10.2196/19359] [Medline: 32294052]
20. Maccari L, Cagno V. Do we need a contact tracing app? Comput Commun 2021 Jan 15;166:9-18 [FREE Full text] [doi: 10.1016/j.comcom.2020.11.007] [Medline: 33235399]
21. Cebrian M. The past, present and future of digital contact tracing. Nat Electron 2021 Jan 25;4(1):2-4. [doi: 10.1038/s41928-020-00535-z]
22. Almagor J, Picascia S. Exploring the effectiveness of a COVID-19 contact tracing app using an agent-based model. Sci Rep 2020 Dec 17;10(1):22235 [FREE Full text] [doi: 10.1038/s41598-020-79000-y] [Medline: 33335125]
23. Altmann S, Milsom L, Zillessen H, Blasone R, Gerdon F, Bach R, et al. Acceptability of App-Based Contact Tracing for COVID-19: Cross-Country Survey Study. JMIR Mhealth Uhealth 2020 Aug 28;8(8):e19857 [FREE Full text] [doi: 10.2196/19857] [Medline: 32759102]
24. Garrett PM, White JP, Lewandowsky S, Kashima Y, Perfors A, Little DR, et al. The acceptability and uptake of smartphone tracking for COVID-19 in Australia. PLoS One 2021 Jan 22;16(1):e0244827 [FREE Full text] [doi: 10.1371/journal.pone.0244827] [Medline: 33481841]
25. Garza A. Contact Tracing Apps Were Big Tech's Best Idea for Fighting COVID-19. Why Haven't They Helped? Time. URL: https://time.com/5905772/covid-19-contact-tracing-apps/ [accessed 2021-03-29]

26. State Approaches to Contact Tracing during the COVID-19 Pandemic. The National Academy for State Health Policy. URL: https://www.nashp.org/state-approaches-to-contact-tracing-covid-19/ [accessed 2021-03-29]
27. Smieszek T, Barclay VC, Seeni I, Rainey JJ, Gao H, Uzicanin A, et al. How should social mixing be measured: comparing web-based survey and sensor-based methods. BMC Infect Dis 2014 Mar 10;14(1):136-113 [FREE Full text] [doi: 10.1186/1471-2334-14-136] [Medline: 24612900]
28. Eagle N, Pentland A, Lazer D. Inferring friendship network structure by using mobile phone data. Proc Natl Acad Sci U S A 2009 Sep 08;106(36):15274-15278 [FREE Full text] [doi: 10.1073/pnas.0900282106] [Medline: 19706491]
29. Stehlé J, Voirin N, Barrat A, Cattuto C, Isella L, Pinton J, et al. High-resolution measurements of face-to-face contact patterns in a primary school. PLoS One 2011 Aug 16;6(8):e23176 [FREE Full text] [doi: 10.1371/journal.pone.0023176] [Medline: 21858018]
30. Braithwaite I, Callender T, Bullock M, Aldridge RW. Automated and partly automated contact tracing: a systematic review to inform the control of COVID-19. Lancet Digit Health 2020 Nov;2(11):e607-e621 [FREE Full text] [doi: 10.1016/S2589-7500(20)30184-9] [Medline: 32839755]
31. Digital Contact Tracing Tools. CDC. URL: https://www.cdc.gov/coronavirus/2019-ncov/php/contact-tracing/contact-tracing-plan/digital-contact-tracing-tools.html [accessed 2021-05-26]
32. Harrison EA, Wu JW. Vaccine confidence in the time of COVID-19. Eur J Epidemiol 2020 Apr 22;35(4):325-330 [FREE Full text] [doi: 10.1007/s10654-020-00634-3] [Medline: 32318915]
33. See How Vaccinations Are Going in Your County and State. The New York Times. URL: https://www.nytimes.com/interactive/2020/us/covid-19-vaccine-doses.html [accessed 2021-06-06]
34. Rubin R. COVID-19 Vaccines vs Variants-Determining How Much Immunity Is Enough. JAMA 2021 Apr 06;325(13):1241-1243. [doi: 10.1001/jama.2021.3370] [Medline: 33729423]
35. Manohar N, Manohar P, Manohar R. HABIT: Hardware-Assisted Bluetooth-based Infection Tracking. Cryptology ePrint Archive. URL: https://eprint.iacr.org/2020/949 [accessed 2021-10-09]
36. Zhou L, Bao J, Setiawan IMA, Saptono A, Parmanto B. The mHealth App Usability Questionnaire (MAUQ): Development and Validation Study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e11500 [FREE Full text] [doi: 10.2196/11500] [Medline: 30973342]
37. Watkins DC. Rapid and Rigorous Qualitative Data Analysis. International Journal of Qualitative Methods 2017 Jun 08;16(1):160940691771213. [doi: 10.1177/1609406917712131]
38. Creswell JW, Clark VLP. Designing and Conducting Mixed Methods Research. Thousand Oaks, CA: SAGE Publications, Inc; 2017.
39. Guetterman TC, Fetters MD, Creswell JW. Integrating Quantitative and Qualitative Results in Health Science Mixed Methods Research Through Joint Displays. Ann Fam Med 2015 Nov 09;13(6):554-561 [FREE Full text] [doi: 10.1370/afm.1865] [Medline: 26553895]
40. Lash RR, Donovan CV, Fleischauer AT, Moore ZS, Harris G, Hayes S, Contact Tracing Assessment Team, et al. COVID-19 Contact Tracing in Two Counties - North Carolina, June-July 2020. MMWR Morb Mortal Wkly Rep 2020 Sep 25;69(38):1360-1363 [FREE Full text] [doi: 10.15585/mmwr.mm6938e3] [Medline: 32970654]
41. Sachdev DD, Brosnan HK, Reid MJA, Kirian M, Cohen SE, Nguyen TQ, et al. Outcomes of Contact Tracing in San Francisco, California-Test and Trace During Shelter-in-Place. JAMA Intern Med 2021 Mar 01;181(3):381-383. [doi: 10.1001/jamainternmed.2020.5670] [Medline: 33136116]
42. Lash RR, Moonan PK, Byers BL, Bonacci RA, Bonner KE, Donahue M, COVID-19 Contact Tracing Assessment Team. COVID-19 Case Investigation and Contact Tracing in the US, 2020. JAMA Netw Open 2021 Jun 01;4(6):e2115850 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.15850] [Medline: 34081135]
43. Lewis D. Why many countries failed at COVID contact-tracing - but some got it right. Nature 2020 Dec 14;588(7838):384-387. [doi: 10.1038/d41586-020-03518-4] [Medline: 33318682]
44. Zeng K, Bernardo SN, Havins WE. The Use of Digital Tools to Mitigate the COVID-19 Pandemic: Comparative Retrospective Study of Six Countries. JMIR Public Health Surveill 2020 Dec 23;6(4):e24598 [FREE Full text] [doi: 10.2196/24598] [Medline: 33302255]
45. Huang Z, Guo H, Lee Y, Ho EC, Ang H, Chow A. Performance of Digital Contact Tracing Tools for COVID-19 Response in Singapore: Cross-Sectional Study. JMIR Mhealth Uhealth 2020 Oct 29;8(10):e23148 [FREE Full text] [doi: 10.2196/23148] [Medline: 33006944]
46. Ho HJ, Zhang ZX, Huang Z, Aung AH, Lim W, Chow A. Use of a Real-Time Locating System for Contact Tracing of Health Care Workers During the COVID-19 Pandemic at an Infectious Disease Center in Singapore: Validation Study. J Med Internet Res 2020 May 26;22(5):e19437 [FREE Full text] [doi: 10.2196/19437] [Medline: 32412416]
47. Bazant MZ, Bush JWM. A guideline to limit indoor airborne transmission of COVID-19. Proc Natl Acad Sci U S A 2021 Apr 27;118(17):e2018995118 [FREE Full text] [doi: 10.1073/pnas.2018995118] [Medline: 33858987]
48. Case Investigation and Contact Tracing in Non-healthcare Workplaces: Information for Employers. CDC. URL: https://www.cdc.gov/coronavirus/2019-ncov/community/contact-tracing-nonhealthcare-workplaces.html [accessed 2021-03-29]

49. Considerations for Case Investigation and Contact Tracing in K-12 Schools and Institutions of Higher Education (IHEs). CDC. URL: https://www.cdc.gov/coronavirus/2019-ncov/community/schools-childcare/contact-tracing.html [accessed 2021-03-29]
50. Guzman-Cottrill JA, Malani AN, Weber DJ, Babcock H, Haessler SD, Hayden MK, SHEA Board of Trustees. Local, state and federal face mask mandates during the COVID-19 pandemic. Infect Control Hosp Epidemiol 2021 Apr 05;42(4):455-456 [FREE Full text] [doi: 10.1017/ice.2020.1403] [Medline: 33397525]
51. Queen CS, Allen J. 100 U.S. colleges will require vaccinations to attend in-person classes in the fall. The New York Times. URL: https://www.nytimes.com/2021/04/29/us/colleges-vaccinations-enrollment.html [accessed 2021-05-26]
52. Chan EY, Saqib NU. Privacy concerns can explain unwillingness to download and use contact tracing apps when COVID-19 concerns are high. Comput Human Behav 2021 Jun;119:106718 [FREE Full text] [doi: 10.1016/j.chb.2021.106718] [Medline: 33526957]
53. Idrees SM, Nowostawski M, Jameel R. Blockchain-Based Digital Contact Tracing Apps for COVID-19 Pandemic Management: Issues, Challenges, Solutions, and Future Directions. JMIR Med Inform 2021 Feb 09;9(2):e25245 [FREE Full text] [doi: 10.2196/25245] [Medline: 33400677]
54. Hennink MM, Kaiser BN, Marconi VC. Code Saturation Versus Meaning Saturation: How Many Interviews Are Enough? Qual Health Res 2017 Mar 26;27(4):591-608. [doi: 10.1177/1049732316665344] [Medline: 27670770]
55. Bente BE, van 't Klooster JWJR, Schreijer MA, Berkemeier L, van Gend JE, Slijkhuis PJH, et al. The Dutch COVID-19 Contact Tracing App (the CoronaMelder): Usability Study. JMIR Form Res 2021 Mar 26;5(3):e27882 [FREE Full text] [doi: 10.2196/27882] [Medline: 33724198]

Edited by G Eysenbach; submitted 09.06.21; peer-reviewed by T Alessa, H Mehdizadeh, R Aldekhyyel; comments to author 13.08.21; revised version received 28.08.21; accepted 27.09.21; published 28.10.21.

Please cite as:
Shelby T, Caruthers T, Kanner OY, Schneider R, Lipnickas D, Grau LE, Manohar R, Niccolai L
Pilot Evaluations of Two Bluetooth Contact Tracing Approaches on a University Campus: Mixed Methods Study
JMIR Form Res 2021;5(10):e31086
URL: https://formative.jmir.org/2021/10/e31086
doi:10.2196/31086
PMID:34586078

Original Paper

Young Adults' Responses to an African and US-Based COVID-19 Edutainment Miniseries: Real-Time Qualitative Analysis of Online Social Media Engagement

Venetia Baker[1], BA, MSc; Georgia Arnold[2], BA; Sara Piot[2], MA; Lesedi Thwala[2], BA; Judith Glynn[1], PhD; James Hargreaves[3], PhD; Isolde Birdthistle[1], PhD

[1]Department of Population Health, Faculty of Epidemiology and Population Health, The London School of Hygiene & Tropical Medicine, London, United Kingdom

[2]The MTV Staying Alive Foundation, London, United Kingdom

[3]Faculty of Public Health and Policy, The London School of Hygiene & Tropical Medicine, London, United Kingdom

Corresponding Author:
Venetia Baker, BA, MSc
Department of Population Health
Faculty of Epidemiology and Population Health
The London School of Hygiene & Tropical Medicine
Keppel Street
London, WC1E 7HT
United Kingdom
Phone: 44 20 7636 8636
Email: venetia.baker1@lshtm.ac.uk

Abstract

Background: In April 2020, as cases of the novel COVID-19 spread across the globe, MTV Staying Alive Foundation created the educational entertainment miniseries *MTV Shuga: Alone Together*. In 70 short episodes released daily on YouTube, *Alone Together* aimed to disseminate timely and accurate information to increase young people's knowledge, motivation, and actions to prevent COVID-19.

Objective: We sought to identify *Alone Together* viewer's perspectives on the global COVID-19 pandemic and national lockdowns by examining the words, conversations, experiences, and emotions expressed on social media in response to the *Alone Together* episodes. We also assessed how viewers used the series and its online community as a source of support during the global pandemic.

Methods: A total of 3982 comments and 70 live chat conversations were extracted from YouTube between April and October 2020 and analyzed through a data-led inductive thematic approach. Aggregated demographic and geographical data were collected using YouTube Analytics.

Results: The miniseries had a global reach across 5 continents, with a total of 7.7 million views across MTV Shuga platforms. The series had over 1 million views over 70 episodes on YouTube and an average of 5683 unique viewers per episode on YouTube. The dominant audience was adults under the age of 35 years and women. Across diverse countries such as Nigeria, Ghana, the United States, and the UK, viewers believed that COVID-19 was serious and expressed that it was socially responsible to follow public health measures. Storylines of the series about the impact of self-isolation on mental health, exposure to violence in lockdowns, and restricted employment opportunities due to the pandemic resonated with young viewers. Tuning in to the miniseries provided viewers with reliable information, entertainment, and an online community during an isolating, confusing, and worrying time.

Conclusions: During the first wave of COVID-19, viewers from at least 53 countries connected on social media via the MTV miniseries. The analysis showed how digitally connected people under the age of 35 years, predominantly women, felt compelled to follow COVID-19 safety measures despite the pandemic's impact on their social, educational, and financial needs. Viewers used social media to reach out to fellow viewers for advice, solace, support, and resources. Organizations, governments, and individuals have been forced to innovate during the pandemic to ensure people can access services safely and remotely. This analysis showed that women under 35 years of age were especially receptive to receiving support from online communities and media services. Peer influence and support online can be a powerful public health tool as people have a great capacity to influence

each other and shape norms around public health. However, online services are not accessible to everyone, and COVID-19 has increased disparities between digitally connected and unconnected younger adults.

(JMIR Form Res 2021;5(10):e30449) doi:10.2196/30449

KEYWORDS

COVID-19; adolescents; young people; social media; edutainment

Introduction

The quick dissemination of accurate information has been an essential priority to educate the global public about the transmission and prevention of the novel SARS-CoV-2. Mass media is a cost-effective method for reaching and informing large populations quickly, like the TV and radio dramas used to educate captive audiences about the HIV epidemic [1-5]. During the COVID-19 pandemic, social media platforms used algorithms to direct its over 2 billion users toward trusted sources of information and flag misinformation, though it has also been responsible for fueling rumors, hoaxes, and misinformation about COVID-19 [6,7]. Younger adults (aged <35 years) rely particularly heavily on social media and online news platforms during the COVID-19 pandemic compared with older adults who additionally rely on close social partners and more traditional media [8].

In April 2020, as the novel COVID-19 spread across the globe, MTV Staying Alive Foundation (SAF) quickly produced an online "edutainment" miniseries that aimed to disseminate timely and accurate information to increase young people's knowledge, motivation, and actions to reduce the spread of SARS-CoV-2. The series was called *MTV Shuga: Alone Together* and included storylines about the pandemic's impact on employment, sexual and intimate partner violence, and mental health. MTV Shuga is MTV SAF's flagship behavior change campaign, including a multiseason TV drama series, which integrates health messaging into storylines that reflect the lives and experiences of youth audiences. Social media engagement is a crucial element of the MTV Shuga campaign.

The *Alone Together* series featured characters from previous MTV Shuga series experiencing challenges with COVID-19 and lockdown measures across 6 countries: Cote d'Ivoire, Kenya, Nigeria, South Africa, Botswana, and the United States. MTV Shuga producers and writers formed a real-time collaboration with epidemiologists and social science researchers to create scripts and storylines that incorporated the most current COVID-19 science and country-specific guidance. Five-minute episodes were released daily, freely available on the internet using social media channels, where viewers could make comments as they watched.

Here we will present an analysis of the online viewer engagement with the *MTV Shuga: Alone Together* miniseries on YouTube. The analysis had 2 main objectives. First, by examining how viewers reacted to the miniseries storylines, we explored people's perspectives on the burgeoning COVID-19 pandemic and national lockdown measures happening in real life. Second, we examined how viewers used an online community of virtual peers and fictional characters as a source of support during the pandemic.

Methods

Data Extraction

MTV SAF released 70 short episodes in daily instalments on YouTube, the MTV Shuga website, and Facebook Live. We extracted data from the YouTube comment section and the live chat stream within 2 weeks after episodes had aired during the first phase of data extraction. We then conducted the second data collection stage of extraction in October 2020, where 3982 comments and 70 live chats were extracted from YouTube as the series steadily gained more viewers and engagement. The series was broadcast in English; French subtitles were provided on a separate MTV Shuga YouTube channel. Comments that were not in English were excluded so no comments from the episodes with French subtitles were selected.

Consent and Ethics

All comments and live chats extracted are public content, freely available on the internet, and we did not seek research ethics approval. We sought to treat the material responsibly, avoiding the publication of identifying information (ie, usernames) and deductive disclosure. We did not seek to identify the demographic data of individuals who posted in the comment section and live chat discussions; however, we collected aggregate demographic data from the series viewers who consented to YouTube's terms and conditions.

Analysis

We used NVivo software version 12 (QSR International) for data extraction, management, and analysis of the YouTube comments and live chats. As the data were unstructured, we chose to conduct data-led research using an inductive thematic approach [9]. In the first stage of data extraction and analysis, we analyzed the data in batches and compiled weekly reports to describe themes and codes emerging from the online comments. The reports were discussed weekly with the production and research team to identify viewer reactions and areas in which subsequent scripts and social media engagement could maximize the series' impact on viewers. This timeline captured audience participation while allowing a rapid analysis that encapsulated the evolving global trajectory and epidemiological understanding of the COVID-19 pandemic, impacting country responses and public health messaging.

The comments from the first phase of analysis fell into 5 different categories (Textbox 1).

Textbox 1. Categories of comments.

Hooks

Comments about the entertainment elements of the show that made viewers want to engage in *Alone Together*, including music, actors, country representation, drama, and suspense.

COVID-19 Storylines

Comments about COVID-19 and lockdown-related storylines, including stories about COVID-19, isolation, social distancing, domestic violence in lockdown, and economic hardships caused by COVID-19 restrictions.

Non–COVID-19 Storylines

Comments about storylines that were not COVID-19 or lockdown specific, including stories about romance, racial justice, and family conflicts.

Viewers Experiences

Comments that revealed viewers' personal experience with the pandemic and their coping mechanism.

Irrelevant Data

Comments about self-promotion, spam, or random topic that did not relate to *Alone Together* or the COVID-19 pandemic.

Once the series was complete, we conducted a more detailed data analysis with a narrower scope. Using data extracted between April and October 2020, we selected data that fell into the categories of comments about COVID-19–related storylines and viewers' personal experiences and coping mechanisms during the pandemic. Using a thematic analysis, we identified patterns in viewers' comments that appeared across episodes. We familiarized ourselves with the data by reading through the comments and live chats from all episodes, then coded the data, and grouped codes into themes. Themes were finalized based on their reoccurrence throughout the data set and if they fell within the scope of the second analysis. We compiled a coding frame to demonstrate overarching patterns and categories across all episodes and then summarized and selected quotes to illustrate each theme.

We used YouTube Analytics to collect data about *Alone Together* viewers' ages, gender, and locations in aggregate, and the number of views and minutes watched for each episode.

Results

Viewers' Characteristics

As of April 2021, the *Alone Together* series had 1 million views over 70 episodes on YouTube, with an average of 5683 unique viewers per episode and over 6.7 million views on Facebook Live (Multimedia Appendix 1). On YouTube, the series had a large reach across 5 continents and was most popular in Nigeria, Ghana, the United States, the United Kingdom, Kenya, Canada, and South Africa. The majority of the audience on the English YouTube channel (Table 1) were aged 18-24 years (529,856/974,000, 54.40%, viewers) or 25-34 years old (394,470/974,000, 40.50%, viewers), and most were female (861,990/974,000, 88.50%). The gender breakdown was consistent through the series.

On a separate YouTube channel, the series provided episodes with French subtitles. Overall, the series with subtitles generated over 27,000 views and had a similar gender and age breakdown to the English series. Cote d'Ivoire was the most prevalent reported location of the audience (Table 2).

Table 1. Reported characteristics of *Alone Together* YouTube viewers: English channel (N=974,000).

Reported characteristics (English)	Viewers
Gender, n (%)[a]	
Female	861,990 (88.50)
Male	112,010 (11.50)
Age group (years), n (%)[a]	
13-17	22,403 (2.30)
18-24	529,856 (54.40)
25-34	394,470 (40.50)
35-44	26,298 (2.70)
44+	974 (0.10)
Geographical location, n (%)	
Nigeria	336,030 (34.50)
Ghana	73,050 (7.50)
United States of America	61,362 (6.30)
United Kingdom	30,194 (3.10)
Canada	29,220 (3.00)
Kenya	13,636 (1.40)
South Africa	11,688 (1.20)
46 Other countries	36,038 (3.70)
Country unknown	382,782 (39.30)

[a]Absolute number of viewers is estimated based on the total viewer population of 974,000 and percentage breakdown provided by YouTube Analytics (percentage of missing data were unavailable).

Table 2. Reported characteristics of *Alone Together* YouTube viewers: French channel (N=27,552).

Reported characteristics (French)	Viewers
Gender, n (%)[a]	
Female	23,805 (86.40)
Male	3747 (13.60)
Age group (years), n (%)[a]	
13-17	0 (0)
18-24	14,933 (54.20)
25-34	12,619 (45.80)
35-44	0 (0)
44+	0 (0)
Geographical location, n (%)	
Côte d'Ivoire	5621 (20.40)
France	771 (2.80)
États-Unis (United States of America)	110 (0.40)
Italie (Italy)	110 (0.40)
Maroc (Morocco)	55 (0.20)
Sénégal (Senegal)	28 (0.10)
Botswana	28 (0.10)
18 other countries	165 (0.60)
Country unknown	20,664 (75.00)

[a]Absolute number of viewers is estimated based on the total viewer population of 27,552 and percentage breakdown provided by YouTube Analytics (percentage of missing data were unavailable).

Themes

Overview

There were 3985 comments posted in the YouTube comment section, and all 70 episodes had viewers engaging in live chat conversations. By examining reactions to *Alone Together* storylines and the opinions viewers expressed online, we identified the following perspectives on the COVID-19 pandemic and national lockdown measures expressed by the predominantly young female audience: (1) people have a social responsibility to follow COVID-19 measures; (2) lockdowns are creating significant economic challenges for people in their community; (3) social isolation is necessary but affects mental health; and (4) lockdowns make it difficult for people who experience intimate partner violence to seek help. We also identified ways in which watching the *Alone Together* series and engaging with MTV Shuga producers and fellow viewers online helped younger adults to cope with the pandemic by (1) forming online communities through shared experience; (2) asking for resources and information; (3) using the series as a link to the real world; and (4) enjoying entertainment together. We extracted comments to illustrate these themes and have presented them below. To accurately represent the viewers, we have presented comments using the spelling and slang that viewers use on YouTube.

People Have a Social Responsibility to Follow COVID-19 Measures

From the earliest episodes in April 2020, many commenters expressed that COVID-19 was serious (“COVID19 is definitely no joke”) and that everyone should follow public health measures to control the “virus” spread. “I just hope people take it seriously for the sake of the vulnerable amongst us.” Commenters reacted with anger when *Alone Together* characters chose to disregard social distancing rules. For example, they rebuked the character “Tobi” who planned an exclusive party during the countrywide lockdown in Nigeria.

> *This Tobi guy, he wants to infect about 100 people exclusively. It's alright. community service and fines day await.*
>
> *[Reply] T you have no sense* 
>
> *[Reply] Right? Like how dumb is that idea?' 'You'd think corona skips exclusive parties?*

The audience reacted similarly to “Lemo” in South Africa when he chose to meet friends despite being exposed to his aunt, who had tested positive for COVID-19. “Bruh I’m this close to put my hands on that little boy Lemo like wtf  .” Commenters also criticized characters who did not wear masks or spread misinformation online and appealed to their online peers to keep themselves and others safe. They viewed individuals as selfish who were not taking COVID-19 seriously or did not recognize

their role in protecting vulnerable people. This was most evident in response to the character "Joe Massive," the Instagram influencer who contracts COVID-19 after encouraging his fans not to wear masks.

> *Hopefully Massive does another [Instagram] live video saying [COVID-19 is] real so that he can correct his wrong information. Not only positive vibes should be shared.*
>
> *[Reply] Joe Massive is a wakeup call to everyone #Lets all stay safe.*
>
> *[Reply] Massive please learn your lesson…. COVID is not a joke.*
>
> *[Reply] Yes wear your mask please.*

Many found these scenes relatable, and they mentioned people in their own lives who were not following social distancing measures.

> *Definitely going to parties and into large gatherings rn is not advised at all*
>
> *[Reply] Right....my family just had a big wedding. I stayed home.*

In the first 2 weeks of the show, there were a few comments in which viewers were distrustful or skeptical with COVID-19 health messages around lockdowns and social distancing.

> *No offence it's two COVID cases in Nigeria, why y'all in lockdown?*
>
> *The propagandists have polluted the whole world with this economy killing lockdown BS […] this sh*t is toxic and detrimental to black folks. unsubscribed.*

Although many viewers reacted very strongly to characters who were not adhering to COVID-19 regulations, viewers did not engage with comments from their peers online who resisted COVID-19 measures. Comments opposing COVID-19 measures became less frequent as the series progressed.

Lockdowns Are Creating Significant Economic Challenges for People in Their Community

Many episodes prompted viewers to comment on the economic hardship that COVID-19 mitigation measures created for many people.

> *Covid 19 is definitely not a holiday, it's really money sapping and the lockdown is limiting.*

Commenters reminded themselves and their online peers to be compassionate and grateful.

> *I love how @MTV Shuga brought the issue of 'hunger virus', people are literally going through hell in times like this. A little kindness will save a life.*

For example, in one episode, the character "Leo" complains about being bored in lockdown. His friend "Zamo" tells him he is being selfish and is lucky to live in a comfortable, luxurious home, and commenters agree.

> *Leo was a bit selfish.*
>
> *[Reply] Zamo gonna be real with him.*
>
> *[Reply] He is maybe a little bit…he should be supportive [to Zamo].*

One commenter reinforces this message of awareness and gratitude by openly reflecting on their privilege of access to the internet and electricity, which her online peers, who can watch *Alone Together* online also share.

> *I remember chatting with my brother in Nigeria that I am bored and he said I have electricity, internet and food, why was I complaining. I felt bad*

Audiences reacted to the storylines of "Wasiu" (based in Lagos) and "Zamo" (in Johannesburg), who were both without income during lockdowns in Nigeria and South Africa, by discussing how to navigate a challenging economic situation during the pandemic. Commenters acknowledged the difficulty of finding work during the lockdown or making money safely and responsibly. At a point in the series when "Wasiu" seeks to supplement his rent by letting strangers with possible COVID-19 symptoms sleep in his home, commenters weighed the risk in his decision as they discussed his potential exposure to COVID-19 versus his economic situation.

> *Wasiu shouldn't have allowed [people to live with him].*
>
> *[Reply] But he needs the money.*
>
> *[Reply] He needs to find a job, a real one.*
>
> *[Reply] Wasiu has collected the money for subletting so he either returns the money or he deals with the consequences.*
>
> *[Reply] Where is the job in this lockdown?*

In this exchange, 1 commenter links this scenario to the real world, noting the scarcity of work many people face during the lockdown. Commenters demonstrated shared creative ways for the characters to make money within the constraints of lockdown, possibly providing advice to their peers who are struggling in real life.

> *Zamo can give beauty tutorials classes online and charge a registration fee. She can make money that way and not have to leave her house, she can even link up with Leo to help her promote it.*

Social Isolation Is Necessary but Affects Mental Health

Social isolation was another topic in the series that resonated strongly with commenters: "Truth is we need human interaction everyday." Many were sympathetic to the series' characters "Leo" and "Daniel" who lived alone during the lockdowns in Kenya and Cote d'Ivoire, respectively.

> *Daniel doesn't look ok at all.*
>
> *[Reply] He definitely doesn't. Tough time to be living on one's own, away from family.*

Commenters explained how loneliness and isolation impacted their mental health. They described being stuck at home alone as depressing, unnaturally sad, draining, limiting, and boring.

> *Leo, it always depressed- but I understand that being stuck in the house 24/7 isn't easy.*
>
> *[Reply] Yes it isn't easy.*

[Reply] It can drain you mentally.

[Reply] Leo it's ok not to be ok.

They also explained how being unable to control their loneliness due to social distancing measures and lockdowns worsened their situation.

Let us not forget persons living on their own (from the young adults to the elderly). It's one thing to be alone, but to be lonely and not be able to do much about it, is sad.

One viewer sympathized with the character Mbali, whose mental health is suffering while isolating due to a COVID-19–positive test. She comments that Mbali should go see a friend because of the toll isolation is taking on her mental health. Other viewers responded that Mbali should continue her isolation.

Mbali looks a mess, Jesus. She should go hang out if it'll make her feel better, lol.

[Reply] No, I think she should just avoid the risk and stay safe at home.

[Reply] Personally, I don't think she should.

[Reply] She definitely needs to isolate.

[Reply] Mbali should not go, I sense danger.

[Reply] Mbali looks fine - She just needs to heal.

Commenters expressed the challenges for those who are away from their family at this time. "Being without my family all this time is draining." One viewer shared their painful experience of losing their father and being unable to travel to attend his funeral. Online peers offered support, empathy, and condolences.

I lost my dad this period and they buried him, I can't even go and pay the last respect… I'm so much in pain 

[Reply] So so sorry for your loss. I believe he's resting.

[Reply] May his soul rest in peace amen.

[Reply] Hard luck dear, it is so painful.

[Reply] Ooh sorry dear...may the good Lord comfort you.

Lockdowns Make It Difficult for People Who Experience Intimate Partner Violence to Seek Help

Many commenters appreciated a storyline about partner violence as they felt it reflected current reality, especially during the pandemic.

My heart breaks for the women who actually go through this especially now during the lockdown. My mom left my abusive father this year and my heart is at peace finally cause I know the pain she felt.

The dramatic arc in the storyline for the character "Dineo"—in which it becomes evident that she is living with a violent partner during the lockdown—generated many responses and discussion. Although some commenters blamed Dineo for her situation, "I don't feel sorry for her because when a relationship is toxic u should leave," online peers explained why it was difficult for women to leave violent relationships, even more so during a pandemic.

Dineo can't just leave during this pandemic if she got no place to go.

Look at [Dineo's] situation. She can't leave even to go buy something, he won't let her. The most she can do is call the police or a hotline but she isn't even mentally capable cause he is fooling and abusing her.

I've heard experiences of ladies in Dineo's shoes! Once you're in a relationship with a manipulative abuser, it's hard to leave them.

Commenters agreed that despite the COVID-19 lockdown measures, it was important that Dineo finds a safe way to leave home and escape the violence.

Dineo get out! Pandemic or not this situation isn't good for your physical or mental health! Talk to someone anyone!

Most comments on episodes about intimate partner violence were oversimplified calls for Dineo to leave her relationship "Dineo, run". Only a few viewers offered strategies for leaving the violent relationship.

Dineo.....gurrrrll you better order a pizza. 911 child!

This having a code word [to signify to a friend you are in trouble] is definitely a great thing.

People! Know where your friends live! Code word for what if you can't find them when they're in trouble. Sha!

Creating an Online Community Through Shared Experience

The comment section reflected a desire from viewers to connect with their peers by sharing advice, support, and their own experiences during the pandemic. Conversations on the *Alone Together* online platform offered a socially distant way to reactivate community engagement, while in-person connections were restricted. The shared experiences and hardships created by the COVID-19 pandemic, enacted by *Alone Together* characters, helped develop a sense of community. One common way viewers utilized the comments was to thank frontline workers.

Shout out to all health workers out there.

Salute to all our frontline health workers 

They also urged their peers to think communally and raised awareness about those making sacrifices or struggling during the pandemic.

I appreciate the introduction of the hospital cleaners and other non-medic staff in hospitals as high-risk workers. We should remember them as we pray and celebrate our doctors, nurses and lab technicians etc.

Prayer for single mothers out there  *through this pandemic* 

Viewers also used the online community to build camaraderie with other viewers and encourage optimism and resolve despite the challenges of the pandemic and lockdowns.

> *Wait, am I the only one seeing this loneliness brings love? I guess there's something positive about COVID-19.*
>
> *[Reply] True.*
>
> *[Reply] Yh, like you have time for others.*

They expressed how their engagement with *Alone Together* went beyond watching the series and helped to connect them across countries and cultures.

> *Been watchin Shuga since day one from a French country, first time to comment but I just wanted to say this is a real family. From different places but a real family* 

A Link to Resources From Alone Together

Alone Together raised awareness about available mental health resources, abortion care, domestic violence, and COVID-19. Viewers used the comment section to connect directly with the *Alone Together* team to receive links to resources that fit their individual need.

> *I actually need these links to talk to these mental health organisations, MTV Shuga if it's possible please.*
>
> *[Reply] MTVShuga: "Hi, please let us know what country you're in at the moment, or visit www.mtvshugaalonetogether.com to access the helplines for the countries currently available. Thank you.*

Viewers also used the comment section to ask *Alone Together* producers for more education and awareness about specific scenarios that people might be facing during COVID-19. Some of the requests included characters (some from past series of MTV Shuga) struggling with family support, pregnancy, and addiction.

> *@MTVShuga you should also address addiction, especially now.*
>
> *@MTVShuga, I want to see [the character from MTV Shuga, Down South] Sol, a lot going there, losing his mum, taking care of his sis and then Covid-19*  *... a lot.*

Used the Series as a Link to the "Real World"

Commenters expressed that they found the series realistic and reflected on what is happening in real life. As viewers grappled with isolation and disconnection from their communities, *Alone Together* served as a link to unite viewers to fictional but relatable community members. "This is like a glimpse into the outside world..." Watching the series raised their awareness of how COVID-19 might be affecting certain groups disproportionately:

> *Hmmm cannot imagine being a child and having to deal with a sick parent or guardian with a contagious disease. Great job again MTV Shuga.*
>
> *Single mothers must be going through alot RN.*
>
> *[Reply] Honestly they are...I wonder how they are coping?*
>
> *[Reply] A day at a time is what I am doing.*

Commenters were empathetic toward characters who were sick with COVID-19 or had a loved one who was ill. Additionally, the series made them aware of the working conditions in overwhelmed hospitals and the sacrifices frontline workers made for their community.

> *I legit started crying when [the character] Dr. Sope started talking about corona virus patients.*
>
> *[Reply] It takes a lot of gut to be a nurse or a doctor.*

Enjoying Entertainment Together

People tuned into the series as a source of support during the pandemic. Commenters applauded the episodes' short and daily format, which served as a consistent and enjoyable moment in their day during an unstable time. For some people, watching *Alone Together* "kept them sane" and lifted their mood:

> *Love love this show... It's amazing how I look forward to my 5 exciting minutes 5 days a week… hum not bad huh? Looking forward to Monday* 

On the season finale, commenters shared that tuning into the series made lockdown more enjoyable and expressed that the series had been a source of comfort to them over the past months.

> *I appreciated what you guys did as a way to occupy our minds and attention during the pandemic. All in all I want to say thank you for #alonetogether it was entertaining for me as well as being informative #alonetogether #staymasked #covidisreal #SENDINGLOVEFROMTHEBAHAMAS*
>
> *Am so proud of you all for putting all this beautiful content together through this hard, scary and stressful moment. Love all your content. ALONE TOGETHER*  *. All this will be something of the past by God's grace.*
>
> *Thank you so much MTV Shuga for keeping me informed and entertained, you guys make me wanna do better and I'm going to miss 'Alone together'.*

Discussion

Principal Findings

Our findings showed that the new miniseries based on COVID-19—*MTV Shuga: Alone Together*—reached dominantly female audiences who understood the importance of individual behaviors to prevent COVID-19, and encouraged their online peers to follow COVID-19 mitigation measures. However, the audience also empathized and related to those suffering the most from COVID-19 lockdowns and were significantly concerned about the psychosocial and economic implications of strict

COVID-19 mitigation measures. Thus, there appeared to be a struggle between social responsibility (following COVID-19 prevention rules for the social good) and the isolation and hardship that this entailed for people. In the absence of face-to-face meet-ups due to social distancing measures, *Alone Together*'s YouTube page became a place where people could safely connect and discuss current events and their personal experiences. It became a valued support system, particularly for adult women under 35 years of age, during the first wave of the COVID-19 pandemic. Viewers used the *Alone Together* social media platform to share support, decision making, and coping strategies to face challenges created by the pandemic measures, uniting viewers across diverse places and cultures.

Alone Together reached at least 53 countries, with the largest reported audience in Nigeria, where 4 previous seasons of the MTV Shuga series were filmed and have an established MTV Shuga fanbase. The United States, UK, and Canada also had high reported viewership, which is likely aided by very high-speed internet access in these countries. Episodes with French subtitles had the most reported viewers in Cote d'Ivoire, which is the location for the latest series of MTV Shuga. Although MTV Shuga creates its content for young people (ages 15-24 years), 43.37% (434,361/1,001,552 viewers) of French and English viewers were above the age of 24 years. MTV Shuga also aims to develop gender-balanced content; however, overall, the MTV Shuga series are more popular with a female audience on YouTube. *Alone Together*, which likely gathered viewers from fans who had watched the previous series of MTV Shuga, was more successful at attracting female viewers from the premiering episode. Young men appear less receptive to watching or are unaware of the MTV Shuga series on YouTube. Although men in the analysis are unrepresented, it allows us to understand the views, perceptions, and experiences of women under the age of 35 years, primarily in countries across Africa and the United States, Canada, and UK, during the pandemic.

The social and economic challenges of the COVID-19 pandemic have had an immense impact on the daily lives of younger adults. Viewers shared how national lockdowns had exposed themselves and people they knew to economic uncertainty. People in their 20s and 30s are more likely than other age groups to work in sectors heavily affected by the pandemic. In particular, women were more likely to lose employment and take on unpaid care responsibilities [10]. Oxfam International estimated that COVID-19 has cost women US $800 billion in lost income globally in 2020 and has widened both gender and economic inequalities [11,12]. Viewers also discussed their isolation and loneliness and being away from family and friends during this time. A UK study showed that adults aged 18-30 years and adults living alone experienced a greater risk of loneliness during the COVID-19 pandemic than other groups [13]. Young people with existing mental health illnesses have reported that COVID-19 made their condition worse, and for some, it restricted their access to mental health services [14]. Additionally, the national lockdowns across the world caused spikes in household violence, leaving girls and women particularly vulnerable [15]. Viewers struggled to find solutions for the character in the series that was stuck in a violent relationship. This issue played out in real life as women worldwide reported difficulties accessing essential gender-based violence services because they could not leave their homes or seek help online or over the phone [14,16].

Social media provided viewers with a space to raise awareness about the challenges they face during the pandemic and ask for the resources they need. For many younger adults, online connections are as much a part of their social life as in-person peers and have possibly become more vital during lockdowns and social distancing. If channeled in positive and constructive ways, social connection online can buffer anxiety and depression induced by isolation in the pandemic and offer support in the absence of in-person contact [17]. Social media can also be a useful public health tool for establishing prosocial public health behaviors. Content on social media influences people's perception of their peers' behavior, affecting their own intentions for that behavior [18]. Young people are especially susceptible to peer influence and have a great capacity to influence one another [19]. Viewers who engaged with or consumed *Alone Together*'s social media content saw their online peers imploring others to follow COVID-19 measures despite the psychosocial impact and to show empathy, compassion, and concern for people struggling during the pandemic. Commenters mainly were united in these expressions, establishing social norms about the appropriate and ethical behavior their peers should demonstrate during the first waves of the COVID-19 pandemic.

Limitations and Strengths

Those who watched *Alone Together* represent individuals with access to media (the internet and a device). Their experiences of COVID-19 may be different from those who lack such access. This audience is more likely to have access to other resources and may experience fewer challenges due to COVID-19. In such situations, the choice to act with social responsibility may come with fewer personal consequences (eg, than somebody who is facing hunger and homelessness). Furthermore, those who commented on the series comprised only a small percentage of viewers and may not represent all viewers. Demographic information extracted by YouTube Analytics represents all viewers, not just those who post comments. YouTube analytic data are self-reported, so viewers can select their gender and age. Young people under 18 years might be motivated to set their age to over 18, so their content is not restricted. A large number of participants chose not to share their location information with YouTube, causing the location data to be incomplete. Additionally, by excluding non-English comments, the voices and experiences of viewers from French-speaking countries will not be represented in the data. Viewers were free to choose if and how they engaged in topics, so commenters may reflect viewers who are more confident, experienced (with digital media), have better internet access, or are more engaged and supportive of the show. This may explain in part why most comments were positive and supportive, and relatively few commenters challenged the content or themes raised in the series. Also, those who were not receptive to *Alone Together*'s messages could have unsubscribed, stopped watching the series, or stopped posting comments.

Although social media's anonymous nature can enable candid, honest comments and viewpoints, it is also possible that viewers

can misrepresent themselves [20]. For example, in their YouTube account, viewers could list themselves as a different age or gender. Their comments could present a persona or opinions that differ from those they offer in person. By contrast, viewers who are truthful about their identity could have potentially shied away from sharing opinions, personal experiences, and asking for help due to fears of judgment, identification, or cyberbullying. These factors could influence the comments' validity and the risk of treating social media posts as empirical data.

Despite these limitations, most comments were validated through the agreement of fellow viewers, and it was clear (from responses to posts) that viewers found value and truth in what their peers posted. There was little reason or incentive for viewers to misrepresent themselves, and we have assumed that most are sharing honest views. The comments represent viewers' own choices about what to share, and thus reflect the themes and experiences that resonated most for them. Our study result offers a valuable first-person perspective from young people facing a dramatic global event. As future waves of COVID-19 infections and lockdown measures occur, their reactions provide insights and lessons to help support young people through further challenges.

Conclusions

This analysis has shown how younger adults across the globe responded to the first wave of the COVID-19 pandemic and the strict lockdowns implemented in most countries. These viewers—predominantly women under 35 years of age who were digitally connected and engaged—recognized the critical role they played in keeping other people safe, even at the expense of their own economic and social needs. They recognized the immediate and lasting impact that COVID-19 restrictions were having on younger adults' mental health, livelihoods, and relationships. As future waves of infection and lockdown occur, especially in low- and middle-income countries with limited COVID-19 vaccine availability, it will be increasingly difficult for younger adults to continue to make such sacrifices. Support systems are needed, and within those edutainment supported by social media can offer connectedness and linkages to resources to cope with the COVID-19 crisis. Viewers who commented on the series showed that they are receptive to receiving support from online communities and media services. Online, people have a great capacity to influence each other and establish social norms. Peer influence and support online can be a powerful public health tool, particularly during a pandemic where individual behaviors need to shift rapidly to prevent disease transmission.

Acknowledgments

We thank Alistair Chase for comments on an early draft of this manuscript. We also thank Tinashe Venge for his contributions to discussions and analytic summaries of the social media data. Finally, we are grateful to Laetitia Kouassi Pessek and Bolade Ogunfuye who assisted us with YouTube Analytics data extraction. This research was funded by Unitaid. Unitaid is a hosted partnership of the World Health Organization.

Conflicts of Interest

GA, SP, and LT work for the MTV Staying Alive Foundation which produces *MTV Shuga: Alone Together*. The series is freely available online so there are no financial conflicts of interest. This study does not aim to evaluate or promote the *MTV Shuga: Alone Together* series, so GA, SP, and LT's contribution to this manuscript does not interfere with the objective presentation of this research.

Multimedia Appendix 1
Final Digital Report of <italic>MTV Shuga: Alone Together</italic> 16 April - 25 September, 2020.
[PPTX File , 12501 KB - formative_v5i10e30449_app1.pptx]

References

1. Jones R, Hoover DR. Reduction in a high-risk sex script among young urban women in the Love, Sex, & Choices web video HIV prevention intervention study. Res Nurs Health 2018 Dec;41(6):535-543 [FREE Full text] [doi: 10.1002/nur.21912] [Medline: 30281811]
2. Vidanapathirana J, Abramson MJ, Forbes A, Fairley C. Mass media interventions for promoting HIV testing. Cochrane Database Syst Rev 2005 Jul 20(3):CD004775. [doi: 10.1002/14651858.CD004775.pub2] [Medline: 16034948]
3. Cassidy J. The soap opera that saves lives. BMJ 2008 May 14;336(7653):1102-1103. [doi: 10.1136/bmj.39570.422778.3a]
4. Banerjee A, La Ferrara E, Orozco-Olvera VH. The Entertaining Way to Behavioral Change: Fighting HIV With MTV.: National Bureau of Economic Research; 2019 Jul. URL: https://www.nber.org/papers/w26096 [accessed 2021-10-13]
5. Shahmanesh M, Mthiyane N, Chimbindi N, Zuma T, Dreyer J, Birdthistle I. 'MTV shuga': mass media communication, HSV2 and sexual health in adolescent girls and young women in rural South Africa. Sexually Transmitted Infections 2019. London, UK: BMJ Publishing Group Ltd; 2019 Jul. URL: https://sti.bmj.com/content/95/Suppl_1/A196.2 [accessed 2021-10-13]
6. Rathinaswamy J, Durairaj B, Waran K. Social Media Reigned by Information or Misinformation About COVID-19: A Phenomenological Study. SSRN Journal 2021 (forthcoming) [FREE Full text] [doi: 10.2139/ssrn.3596058]

7. Spandana S, Bagchi K. How internet platforms are combating disinformation and misinformation in the age of COVID-19. The New America. 2020. URL: https://www.newamerica.org/oti/reports/how-internet-platforms-are-combating-disinformation-and-misinformation-age-covid-19/youtube/ [accessed 2021-10-13]
8. Chu L, Fung HH, Tse DCK, Tsang VHL, Zhang H, Mai C. Obtaining Information From Different Sources Matters During the COVID-19 Pandemic. Gerontologist 2021 Feb 23;61(2):187-195 [FREE Full text] [doi: 10.1093/geront/gnaa222] [Medline: 33388758]
9. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
10. Power K. The COVID-19 pandemic has increased the care burden of women and families. Sustainability: Science, Practice and Policy 2020 Jun 21;16(1):67-73. [doi: 10.1080/15487733.2020.1776561]
11. COVID-19 cost women globally over $800 billion in lost income in one year Oxfam International.: Oxfam International; 2021. URL: https://www.oxfam.org/en/press-releases/covid-19-cost-women-globally-over-800-billion-lost-income-one-year [accessed 2021-10-13]
12. UN Secretary-General's Policy Brief: The Impact of COVID-19 on Women.: World Health Organization; 2020. URL: https://www.unwomen.org/en/digital-library/publications/2020/04/policy-brief-the-impact-of-covid-19-on-women [accessed 2021-10-13]
13. Bu F, Steptoe A, Fancourt D. Who is lonely in lockdown? Cross-cohort analyses of predictors of loneliness before and during the COVID-19 pandemic. Public Health 2020 Sep;186:31-34 [FREE Full text] [doi: 10.1016/j.puhe.2020.06.036] [Medline: 32768621]
14. Lee J. Mental health effects of school closures during COVID-19. Lancet Child Adolesc Health 2020 Jun;4(6):421 [FREE Full text] [doi: 10.1016/S2352-4642(20)30109-7] [Medline: 32302537]
15. Humphreys KL, Myint MT, Zeanah CH. Increased Risk for Family Violence During the COVID-19 Pandemic. Pediatrics 2020 Jul 21;146(1):e20200982 [FREE Full text] [doi: 10.1542/peds.2020-0982] [Medline: 32317306]
16. Kottasová I, Di DV. Women are using code words at pharmacies to escape domestic violence during lockdown. CNN World. 2020. URL: https://edition.cnn.com/2020/04/02/europe/domestic-violence-coronavirus-lockdown-intl/index.html [accessed 2021-10-13]
17. Stuart J, O'Donnell K, O'Donnell A, Scott R, Barber B. Online Social Connection as a Buffer of Health Anxiety and Isolation During COVID-19. Cyberpsychol Behav Soc Netw 2021 Aug;24(8):521-525. [doi: 10.1089/cyber.2020.0645] [Medline: 33601944]
18. Young SD, Jordan AH. The influence of social networking photos on social norms and sexual health behaviors. Cyberpsychol Behav Soc Netw 2013 Apr;16(4):243-247 [FREE Full text] [doi: 10.1089/cyber.2012.0080] [Medline: 23438268]
19. Andrews JL, Foulkes L, Blakemore S. Peer Influence in Adolescence: Public-Health Implications for COVID-19. Trends Cogn Sci 2020 Aug;24(8):585-587 [FREE Full text] [doi: 10.1016/j.tics.2020.05.001] [Medline: 32444200]
20. Mislove A, Jørgensen SL, Ahn YY, Onnela JP, Rosenquist JN. Understanding the demographics of Twitter users. In: Proceedings of the Fifth International AAAI Conference on Weblogs And Social Media. Menlo Park, CA: The AAAI Press; 2011 Presented at: The Fifth International AAAI Conference on Weblogs And Social Media; July 17-21, 2011; Barcelona, Spain p. 554-557 URL: http://www.ccs.neu.edu/home/amislove/publications/Twitter-ICWSM.pdf

Abbreviations

MTV SAF: The MTV Staying Alive Foundation

Edited by G Eysenbach; submitted 17.05.21; peer-reviewed by D Bell, J Woods; comments to author 10.07.21; revised version received 29.07.21; accepted 18.09.21; published 29.10.21.

Please cite as:
Baker V, Arnold G, Piot S, Thwala L, Glynn J, Hargreaves J, Birdthistle I
Young Adults' Responses to an African and US-Based COVID-19 Edutainment Miniseries: Real-Time Qualitative Analysis of Online Social Media Engagement
JMIR Form Res 2021;5(10):e30449
URL: https://formative.jmir.org/2021/10/e30449
doi:10.2196/30449
PMID:34596568

Original Paper

Knowledge, Attitudes, and Practices Regarding COVID-19 Among Health Care Workers in Public Health Facilities in Eastern Ethiopia: Cross-sectional Survey Study

Alinoor Mohamed Farah[1], MPH; Tahir Yousuf Nour[1], MPH; Muse Obsiye[1], MPH; Mowlid Akil Aden[1], MPH; Omar Moeline Ali[1], MPH; Muktar Arab Hussein[1], MPH; Abdullahi Bedel Budul[1], MPH; Muktar Omer[1], MPH; Fentabil Getnet[2], PhD

[1]Department of Public Health Nutrition, School of Public Health, College of Medicine and Health Sciences, Jigjiga University, Jigjiga, Ethiopia

[2]Department of Epidemiology and Biostatistics, School of Public Health, College of Medicine and Health Sciences, Jigjiga University, Jigjiga, Ethiopia

Corresponding Author:
Alinoor Mohamed Farah, MPH
Department of Public Health Nutrition
School of Public Health, College of Medicine and Health Sciences
Jigjiga University
CoMHS Building, 2nd Floor
Jigjiga, 1020
Ethiopia
Phone: 251 911053913
Email: alinuriana@yahoo.com

Abstract

Background: On March 13, 2020, Ethiopia reported the first confirmed case of COVID-19 in Addis Ababa. COVID-19 is likely to overwhelm an already-fragile health care delivery system and reduce the availability of essential health services. This analysis of data from the Somali Region of Eastern Ethiopia on health care workers' (HCWs) knowledge, attitudes, and practices regarding the prevention and control of COVID-19 may be used in planning health education programs about the emerging viral disease.

Objective: This study aimed to investigate the knowledge, attitudes, and practices of HCWs regarding COVID-19 infection.

Methods: This cross-sectional study was conducted among HCWs in three public health facilities in the Somali Region, Eastern Ethiopia. A self-administered questionnaire was shared with all HCWs working at the public health facilities. A total of 15 knowledge questions were scored as 1 or 0 for correct or incorrect responses, respectively. A total of 14 practice questions were scored on a 3-point scale from 1 ("always") to 3 ("never"). A total of six attitude questions were rated on a 5-point Likert scale, in a negative dimension, as follows: 1 ("strongly agree"), 2 ("agree"), 3 ("neutral"), 4 ("disagree"), and 5 ("strongly disagree"). Mean scores were calculated and used as a cut point to dichotomize the outcome variables (>13.7 indicated good knowledge, <18.8 indicated good practices, and ≤10.5 indicated favorable attitudes). We used *t* tests and analyses of variance (ie, *F* tests) to analyze the mean score differences of knowledge, attitudes, and practices between the independent variables. Spearman correlation was used to assess the relationship between mean knowledge and attitude scores.

Results: Of the 686 HCWs approached, a total of 434 HCWs responded (63.3% response rate). The mean age of the participants was 27.6 (SD 5.3) years, and the majority of the participants were male (293/434, 67.5%). The mean knowledge score was 13.7 (SD 2.6), and 73.3% (318/434) of participants had sufficient knowledge. The mean attitude score was 10.5 (SD 4.1), and 54.8% (238/434) of the participants had a good attitude toward COVID-19. The mean practice score was 18.8 (SD 5.8), and 61.5% (267/434) of the participants practiced precautionary measures to prevent COVID-19. There was a negative correlation between knowledge and attitude scores (r=–0.295, P<.001) and between knowledge and practice scores (r=–0.298, P<.001).

Conclusions: The overall levels of knowledge and practice were relatively better than the attitude level. This highlights the need to implement strategies that enhance the positive attitudes and safe practices of the HCWs for better containment of the pandemic and supporting of essential health care services.

(JMIR Form Res 2021;5(10):e26980) doi:10.2196/26980

KEYWORDS

COVID-19; knowledge; attitude; practice; health care workers; Eastern Ethiopia

Introduction

COVID-19 is an emerging, rapidly changing global health challenge affecting all sectors, including the health sector [1]. In Ethiopia and in the Somali Region, in particular, there is already a limited number of health care workers (HCWs) in the health sector, which puts extreme strain on the capacity to serve patients, especially for nonemergency care [2]. Thus, COVID-19 is likely to overwhelm an already-fragile health care delivery system and reduce the availability of services for endemic health concerns, such as malaria and diarrheal diseases including cholera [3]. The proportion of Ethiopia's water supply that is safe is 30%, which means the water supply serves as the leading cause of communicable diseases in the country. As a result, pre-existing inadequate hygiene practices, insufficient coverage in water and sanitation services, and overcrowding living situations all contribute to the virus's occurrence and transmission. The pandemic is also adding to the burden of endemic infectious diseases that prevail in many countries with an ongoing humanitarian response; these diseases include cholera, measles, malaria, HIV, and tuberculosis.

To date, no specific antiviral treatment has been confirmed to be effective against COVID-19 [4]. Regarding infected patients with COVID-19, it has been recommended to apply appropriate symptomatic treatment and supportive care [5,6]. Therefore, applying preventive measures to reduce the spread of disease is of the utmost importance.

In this regard, airborne precautions and other protective measures have been discussed and proposed for prevention. Recommended preventive measures include regular handwashing with soap, social and physical distancing, and respiratory hygiene (ie, covering the mouth and nose while coughing or sneezing) [5,7].

The World Health Organization (WHO) also issued guidelines on the use of face masks in different settings, including in the community, in home-based care, and in the health care settings of COVID-19 [8]. In this guideline, HCWs are recommended to use face masks, such as certified N95 respirator masks, when performing aerosol-generating procedures and to use medical masks when providing care to suspected or confirmed COVID-19 cases. However, these effective prevention and control practices depend on awareness and compliance among HCWs at all levels.

A poor level of knowledge has been implicated in the rapid spread of infections in health facilities [9,10] and delay of treatment [11], which may put patients' lives at risk. On the other hand, HCWs are not only at the forefront of the fight against this highly contagious infectious disease, but are also directly or indirectly affected by it, and the likelihood of acquiring this disease is higher among HCWs compared to the general population [12]. Therefore, it is crucial that HCWs across Ethiopia and in the Somali Region have adequate knowledge about all aspects of the disease, including clinical manifestation, diagnosis, proposed treatment, and established prevention strategies.

Thus, this study aimed to investigate the knowledge, attitudes, and practices of HCWs regarding COVID-19 infection. The findings may be useful in recommending any remedial measures and additional interventions in the study area to improve awareness, attitudes, and practices among HCWs.

Methods

Study Design, Setting, and Period

An institution-based cross-sectional study was conducted from June to August 2020 at Jigjiga University Teaching Hospital and two health centers in Jigjiga City, Somali Region, State of Ethiopia. There are two hospitals in the town; one is a multidisciplinary specialized teaching hospital with 351 inpatient beds and the other one is a general hospital with 193 inpatient beds. Currently, they provide health services to more than one million inhabitants in the catchment area. During data collection, Karamara General Hospital was turned into a COVID-19 treatment center and was not accessible for the study. Most of its staff were reassigned to other health facilities to support the provision of basic health services.

Sample Size Determination and Sampling

Employees that were invited to participate in this study were all HCWs employed by the public health facilities in Jigjiga City. We included HCWs who were believed to have patient care or specimen contact, such as physicians, nurses, midwives, health officers, laboratory professionals, x-ray professionals, pharmacists and druggists, anesthetists, and biomedical professionals.

All public health facilities in the city, except for the one fully dedicated to COVID-19, were included. We contacted the human resources units to receive the lists of HCWs who have direct patient and specimen contact in the respective facilities. Finally, we came up with a total of 686 HCWs, which constituted the final sample size.

Eligibility Criteria

Only full-time employees (ie, HCWs) who were potentially at high risk of COVID-19 infection (ie, physicians, medical laboratory technologists, nurses, and midwives), were available during the data collection period, and were ready to take part in the study were included.

Data Collection and Quality Control

A structured self-administered questionnaire was used to collect the data. The questionnaire was designed in reference to the study by Asaad et al regarding knowledge and attitudes toward the Middle East respiratory syndrome coronavirus, and it was adapted from the current interim guidance and information for HCWs published by the US Centers for Disease Control and Prevention (CDC) [5,13]. The questionnaire consists of six sociodemographic questions and 35 questions on knowledge,

attitudes, and infection control practices related to COVID-19 in the health care setting. A pretest was done among 5.0% (34/686) of the study participants to estimate the duration required to complete the survey, ensure clarity of the questions, avoid potential bias, and validate the internal consistency of the items in attitude measurement. The Cronbach α value was .72, which was acceptable. The questionnaire has a satisfactory level of construct validity and internal consistency, according to this result.

The questionnaires were collected daily and checked for completeness; in the case of an incomplete questionnaire, the respondent was contacted in order to complete it. In addition, timely supervision of the data collection process was done by the investigators. During collection of the questionnaires, precautionary measures, such as wearing of masks and physical distancing, were observed. Data quality was ensured during data collection, and regular supervision and follow-up were performed.

Measurements of Knowledge, Attitudes, and Practices

Knowledge questions were scored as 1 or 0 for correct or incorrect responses, respectively. The total knowledge score ranged from 0 (with no correct answer) to 15 (for all correct answers); a mean score of ≤13.7 indicated poor knowledge, and a mean score of >13.7 indicated good knowledge. There were 14 practice questions, which were scored on a 3-point scale from 1 (“always”) to 3 (“never”). Total practice scores ranged from 14 to 42; a mean score of >18.8 (answering “never” or “occasionally”) indicated poor practices, and a score of ≤18.8 (answering “always”) indicated good practices. Thus, the lower the practice scores were, the higher the probability of good practices, and vice versa. Attitude questions were scored on a 5-point Likert scale as follows: 1 (“strongly agree”), 2 (“agree”), 3 (“neutral”), 4 (“disagree”), and 5 (“strongly disagree”). There were six attitude questions. The total score ranged from 6 to 30; a mean score of >10.5 (answering “disagree,” “strongly disagree,” or “neutral”) indicated a negative attitude, and a score of ≤10.5 (answering “strongly agree” or “agree”) indicated a positive attitude. Therefore, the lower the attitude scores were, the higher the probability of positive attitudes, and vice versa.

Data Analysis

Data were coded and entered into Epi Info software (version 3.5.1; CDC) and exported into Stata software (version 14.1; StataCorp LP) for analysis. Summary statistics, such as frequencies and proportions, were computed as appropriate. We conducted *t* tests and analyses of variance to analyze the relationship between the dependent (ie, knowledge, attitudes, and practices) and independent variables (ie, demographic characteristics of the participants). Spearman correlation was used to assess the relationship between mean knowledge and attitude scores. The differences in estimated variables were considered statistically significant at $P<.05$.

Ethical Considerations

Ethical clearance and support letters were obtained from the Ethical Review Committee of the College of Medicine and Health Sciences, Jigjiga University. The support letter was then submitted to the public health facilities. Permission was then obtained from the health facilities’ director and department or section heads. Study participants were informed about the purpose and importance of the study through written informed consent before the data collection process. In addition, participants who were unwilling to take part in the study and those who needed to quit their participation at any stage were informed that they could do so without any restriction.

Results

Sociodemographic Characteristics

Of the 686 HCWs approached, a total of 434 responded (63.3% response rate). The mean age of the participants was 27.6 (SD 5.3) years, and the majority of the participants were male (293/434, 67.5%) and below 40 years of age (423/434, 97.5%). The majority of the participants were working in the referral hospital (345/434, 79.5%), were nurses (322/434, 74.2%), and had more than 5 years of experience (307/434, 70.7%). The main sources of information about COVID-19 were the WHO followed by government health authorities and media (Table 1).

Table 1. Sociodemographic characteristics of health care workers in Jigjiga City.

Variable	Value (N=434)
Sex, n (%)	
Male	293 (67.5)
Female	141 (32.5)
Age in years	
Mean, SD	27.6 (5.3)
18-39, n (%)	423 (97.5)
≥40, n (%)	11 (2.5)
Occupation, n (%)	
Nurse (including midwives)	322 (74.2)
Physician	36 (8.3)
Pharmacist	26 (6.0)
Dentist	7 (1.6)
Laboratory technologist	31 (7.1)
X-ray physician	12 (2.8)
Place of work, n (%)	
Referral hospital	345 (79.5)
Karamara General Hospital	33 (7.6)
Ablele Health Center	44 (10.1)
Ayrdage Health Center	12 (2.8)
Work experience in years	
Mean, SD	8.3 (9.1)
0-4, n (%)	127 (29.3)
≥5, n (%)	307 (70.7)
Source of COVID-19 information, n (%)	
World Health Organization	241 (55.5)
Government site and media	105 (24.2)
Social media[a]	52 (12.0)
Other news media	30 (6.9)
Others	6 (1.4)

[a]Social media platforms include WhatsApp, Facebook, and Telegram.

Knowledge

Table 2 shows the details of the responses given by the health professionals for each knowledge question dealing with COVID-19 signs and symptoms, potential admission criteria required to identify patients at risk, approaches to prevent transmission in hospitals, and possible supportive treatment for patients with COVID-19.

Of the total HCWs who participated in the study, 73.3% (318/434) had sufficient knowledge (Table 3). Almost all HCWs were able to correctly identify COVID-19 key symptoms. Data from this study revealed that 95.4% (414/434) of respondents fully understood the common signs and symptoms of COVID-19. A total of 89.9% (390/434) were also aware of factors likely to be associated with severity of the disease.

The majority of the respondents understood the dynamics of COVID-19 infectiousness: 91.5% (397/434) of respondents were aware of the possibility that one could infect others before the onset of symptoms and 90.8% (394/434) of HCWs responded “true” to the question about droplets as a major transmission route. A considerable proportion of respondents (404/434, 93.1%) knew that the incubation period is not constant and could vary from 2 to 14 days.

Table 2. Health care workers' (HCWs) knowledge regarding COVID-19.

Statements about knowledge	Responses by HCWs (N=434), n (%)		
	True	False	I don't know
The common symptoms of COVID-19 are fever, fatigue, and dry cough.	414 (95.4)	13 (3.0)	7 (1.6)
The causative agent of COVID-19 is coronavirus.	410 (94.4)	12 (2.8)	12 (2.8)
The incubation period of COVID-19 is 2 to 14 days.	404 (93.1)	21 (4.8)	9 (2.1)
Not all persons with COVID-19 will develop severe cases. Only those who are elderly and have chronic illnesses are likely to be severe cases.	390 (89.9)	33 (7.6)	11 (2.5)
COVID-19 can be transmitted through direct contact of respiratory droplets when infected persons cough or sneeze.	394 (90.8)	28 (6.4)	12 (2.8)
The disease can be transmitted from asymptomatic patients.	397 (91.5)	26 (6.0)	11 (2.5)
Training and observation of standard precautionary measures are required by caregiving personnel in suspected and probable cases of COVID-19 infection.	392 (90.3)	30 (6.9)	12 (2.8)
Visitors to patients with suspected, probable, and confirmed cases of COVID-19 infection should be limited, both in hospital and at home.	406 (93.6)	22 (5.1)	6 (1.3)
The disease can be spread by people touching a surface or object that has the virus on it and then touching their own mouth or nose or possibly their eyes.	401 (92.4)	22 (5.1)	11 (2.5)
Isolation and treatment of people who are infected with the COVID-19 virus are effective ways to reduce the spread of the virus.	404 (93.1)	23 (5.3)	7 (1.6)
People who have contact with someone infected with the COVID-19 virus should be immediately isolated in a proper place. In general, the observation period is 14 days.	400 (92.2)	24 (5.5)	10 (2.3)
A person with mild symptoms of COVID-19 must remain at home until resolution of clinical symptoms.	398 (91.7)	30 (6.9)	6 (1.4)
Standard precautions should be followed by health care providers in dealing with suspected, probable, and confirmed cases of COVID-19 infection.	395 (91.0)	29 (6.7)	10 (2.3)
Oxygen therapy should be given to all cases of severe COVID-19 with acute respiratory infection.	369 (85.0)	54 (12.5)	11 (2.5)
Ventilation with an endotracheal tube must be carried out in patients with confirmed and/or suspected COVID-19 with clinical manifestations of acute respiratory distress symptoms.	379 (87.3)	40 (9.2)	15 (3.5)

Table 3. Knowledge, attitude, and practice levels among health care workers.

Domain	Good level (N=434), n (%)	Poor level (N=434), n (%)
Knowledge	318 (73.3)	116 (26.7)
Attitude	238 (54.8)	196 (45.2)
Practice	267 (61.5)	167 (38.5)

With respect to prevention of transmission from known or suspected patients, HCWs knew most of the preventive measures. The majority of HCWs (400/434, 92.2%) responded that isolation could be one of the possible ways to prevent COVID-19. The majority of HCWs (395/434, 91.0%) responded that standard precautions should be followed by health care providers when dealing with suspected, probable, and confirmed cases of COVID-19 infection (Table 2).

Attitude

Of the total of HCWs who participated in the study, 54.8% (238/434) had positive attitudes (Table 3). Only 48.8% (212/434) were confident that COVID-19 can be controlled by public health institutes. More than half of the HCWs (226/434, 52.1%) agreed that this disease can be prevented, and less than half of the HCWs (214/434, 49.3%) agreed that it is imperative to use surgical masks when working with patients with COVID-19; unfortunately, 52.1% (226/434) of the HCWs were not confident in dealing with patients with COVID-19 (Table 4).

Table 4. Health care workers' (HCWs) attitudes toward COVID-19.

Statements about attitudes	Responses by HCWs (N=434), n (%)				
	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
Caring for patients with COVID-19 infection may be a threat to health care personnel.	226 (52.1)	147 (33.9)	28 (6.4)	24 (5.5)	9 (2.1)
Public health agencies, like the Ethiopia Public Health Institute, can control the outbreak of COVID-19.	212 (48.8)	121 (27.9)	42 (9.7)	50 (11.5)	9 (2.1)
COVID-19 can have a negative effect on the economies of the countries involved.	231 (53.2)	135 (31.1)	24 (5.5)	36 (8.3)	8 (1.9)
It is important to report suspected cases to health authorities.	301 (69.3)	104 (24.0)	11 (2.5)	13 (3.0)	5 (1.2)
COVID-19 is preventable.	226 (52.1)	136 (31.3)	33 (7.6)	23 (5.3)	16 (3.7)
It is imperative to use surgical masks when working with patients with COVID-19.	214 (49.3)	136 (31.3)	23 (5.3)	26 (6.0)	35 (8.1)

Practice

Of the 434 HCWs who participated in the study, 61.5% (n=267) practiced precautionary measures to prevent COVID-19 (Table 3). Practical measures put in place by HCWs in order to protect themselves and their families are presented in Table 5. A total of 72.4% (n=314) of the 434 HCWs reported that they always wore personal protective equipment when coming into contact with the patients, and 67.5% (n=293) washed their hands before and after touching each patient. Other measures observed included the following: 72.8% (n=316) of the HCWs always avoided going to crowded places and 71.4% (n=310) always avoided shaking hands, hugging, or kissing. In addition, 71.7% (n=311) of the HCWs always kept a safe distance and 71.0% (n=308) covered their mouths when sneezing or coughing. Unfortunately, as high as 72.6% (n=315) of the participants had avoided patients with symptoms suggestive of COVID-19.

Table 5. Health care workers' (HCWs) practices regarding COVID-19.

Statements about practices	Responses by HCWs (N=434), n (%)		
	Always	Occasionally	Never
I always wear the personal protective equipment (PPE) when handling a patient.	314 (72.4)	101 (23.3)	19 (4.4)
I wash my hands with water and soap before putting on the gloves.	302 (69.6)	105 (24.2)	27 (6.2)
For residual protection, I use hand sanitizer after washing my hands with soap and water.	303 (69.8)	103 (23.7)	26 (6.5)
I wash my hands between patients (before and after examining).	293 (67.5)	105 (24.2)	36 (8.3)
I wash my hands whenever they are soiled.	298 (68.7)	113 (26.0)	23 (5.3)
I wash my hands after blowing my nose or covering a sneeze.	289 (66.6)	114 (26.3)	31 (7.1)
I wash my hands with soap and water and apply hand sanitizer when leaving the health facility.	208 (71.0)	107 (24.6)	19 (4.4)
When I sneeze or cough, I cover my nose or mouth with a tissue or clean cloth or do so into my elbow.	308 (71.0)	106 (24.4)	20 (4.6)
I keep at least 1 meter of distance from a sick person.	311 (71.7)	108 (24.9)	15 (3.4)
I avoid shaking hands with, hugging, or kissing colleagues and patients.	310 (71.4)	110 (25.4)	14 (3.2)
I avoid touching my mouth, nose, and eyes with my hands if I have not washed them with water and soap.	307 (70.7)	117 (27.0)	10 (2.3)
I avoid going to crowded places.	316 (72.8)	107 (24.7)	11 (2.5)
I safely dispose of the used PPE items when I finish the service.	315 (72.6)	100 (23.0)	19 (4.4)
I avoid patients with signs and symptoms suggestive of COVID-19.	315 (72.6)	91 (21.0)	28 (6.4)

Distribution of Knowledge, Attitudes, and Practices

The relationships between sociodemographic characteristics and knowledge, attitudes, and practices regarding COVID-19 are demonstrated in Table 6. The mean knowledge, attitude, and practice scores were 13.7 (SD 2.6), 10.5 (SD 4.1), and 18.8 (SD 5.8), respectively. Male participants had a higher mean knowledge score than female participants (13.9, SD 2.3, and 13.3, SD 3.1, respectively; *P*=.01). Nurses had the highest knowledge scores, followed by pharmacists and other professions (*P*=.04). No difference in knowledge was found with respect to age and work experience. Mean knowledge scores were significantly related to the place of work as well as to the source of knowledge. Participants from the referral

hospital had a significantly higher mean knowledge score compared to those working in other health facilities (*P*<.001). Additionally, significantly increased knowledge scores were observed in those getting information from WHO websites followed by government media (*P*<.001).

Similarly, male participants had more positive attitudes than female participants (10.3 , SD 4.1, and 10.8, SD 4.1, respectively) with no statistically significant difference. No difference in attitude was found with respect to qualification, work experience, and age. A positive attitude was significantly related to the place of work as well as to the source of knowledge. Participants from the referral hospital had a significantly lower mean attitude score compared to those working in other health facilities (*P*<.001). Additionally, a significantly decreased attitude score was detected among those getting information from social media, followed by WHO websites (*P*=.02).

Likewise, male participants tended to practice precautionary measures more than female participants (18.6, SD 5.8, and 19.3, SD 6.3, respectively) with no statistically significant difference. No difference in practice was found with respect to qualification and work experience. Good practice was significantly related to age, place of work, and source of knowledge. Older participants (≥40 years) tended to practice precautionary measures more than those in younger age groups (*P*=.04). Additionally, significantly better practices were observed among participants working at Ayrdage Health Center (*P*=.01) and those getting information from government sites and media (*P*=.02).

Table 6. Distribution of knowledge, attitude, and practice scores among health care workers.

Variable	Knowledge score[a], mean (SD)	*t* test[b] (*df*=432)	*F* test[c] (*df*=433)	*P* value	Attitude score[d], mean (SD)	*t* test (*df*=432)	*F* test (*df*=433)	*P* value	Practice score[e], mean (SD)	*t* test (*df*=432)	*F* test (*df*=433)	*P* value
Overall	13.7 (2.6)	__[f]	—	—	10.5 (4.1)	—	—	—	18.8 (5.8)	—	—	—
Sex												
Male	13.9 (2.3)	2.49	—	.01	10.3 (4.1)	–1.02	—	.31	18.6 (5.8)	–1.18	—	.24
Female	13.3 (3.1)	—	—	—	10.8 (4.1)	—	—	—	19.3 (6.3)	—	—	—
Age in years												
18-39	13.7 (2.6)	0.014	—	.99	10.5 (4.1)	0.24	—	.81	18.9 (6.0)	2.01	—	.04
≥40	13.7 (1.8)	—	—	—	10.2 (2.9)	—	—	—	15.3 (2.7)	—	—	—
Occupation												
Nurse (including midwives)	13.9 (2.2)	—	2.39	.04	10.3 (3.8)	—	1.45	.21	19.2 (6.1)	—	1.61	.16
Physician	12.9 (3.9)	—	—	—	11.0 (4.4)	—	—	—	19.3 (5.9)	—	—	—
Pharmacist	13.2 (2.4)	—	—	—	10.7 (3.8)	—	—	—	16.5 (3.5)	—	—	—
Dentist	12.2 (3.5)	—	—	—	9.9 (2.3)	—	—	—	18.3 (4.5)	—	—	—
Laboratory technologist	13.0 (3.5)	—	—	—	10.9 (5.6)	—	—	—	17.3 (5.9)	—	—	—
X-ray physician	13.6 (3.1)	—	—	—	13.2 (6.7)	—	—	—	17.4 (4.4)	—	—	—
Place of work												
Referral hospital	14.3 (1.8)	—	40.5	<.001	10.1 (4.1)	—	5.93	<.001	18.8 (6.0)	—	3.87	.01
Karamara General Hospital	12.3 (3.4)	—	—	—	12.0 (4.2)	—	—	—	17.6 (4.4)	—	—	—
Ablele Health Center	10.6 (4.0)	—	—	—	12.3 (3.8)	—	—	—	20.9 (6.6)	—	—	—
Ayrdage Health Center	12.6 (2.4)	—	—	—	11.5 (2.4)	—	—	—	15.3 (1.8)	—	—	—
Work experience in years												
0-4	13.8 (2.3)	0.41	—	.68	9.9 (3.4)	–1.78	—	.08	18.9 (6.0)	0.54	—	.59
≥5	13.7 (2.7)	—	—	—	10.7 (4.4)	—	—	—	15.3 (2.7)	—	—	—
Source of COVID-19 information												
World Health Organization	14.0 (2.0)	—	5.94	<.001	10.4 (3.9)	—	2.96	.02	19.2 (6.1)	—	2.88	.02
Government site and media	13.9 (2.2)	—	—	—	11.1 (4.4)	—	—	—	17.2 (4.5)	—	—	—
Social media	13.3 (3.1)	—	—	—	9.1 (3.3)	—	—	—	19.1 (6.1)	—	—	—
Other news media	12.6 (3.9)	—	—	—	11.6 (4.4)	—	—	—	19.7 (7.0)	—	—	—
Other	9.8 (6.9)	—	—	—	12.2 (7.0)	—	—	—	22.2 (11.2)	—	—	—

[a]Knowledge questions were scored as 1 or 0 for correct or incorrect responses, respectively. The total knowledge score ranged from 0 (with no correct

answer) to 15 (for all correct answers).

[b]The *t* test was used to compare two mean scores within demographic characteristic categories with two subitems (eg, sex) for knowledge, attitude, and practice.

[c]The analysis of variance (ie, *F* test) was used to compare three or more mean scores within demographic characteristic categories with three or more subitems (eg, occupation) for knowledge, attitude, and practice.

[d]Attitude questions were scored on a 5-point Likert scale from 1 (“strongly agree”) to 5 (“strongly disagree”). The total score ranged from 6 to 30.

[e]Practice questions were scored on a 3-point scale from 1 (“always”) to 3 (“never”). Total practice scores ranged from 14 to 42.

[f]Values are reported in the top row of each category only when the corresponding statistical test (ie, *t* test or *F* test) was conducted.

Correlation Between Knowledge, Attitude, and Practice Domains

Table 7 presents the correlation coefficients between the domains of knowledge, attitude, and practice. It was found that the attitude (r=–0.295, P<.001) and practice (r=–0.298, P<.001) domains were inversely associated with the knowledge score. A significant positive correlation was found between attitude and practice (r=0.173, P<.001). The lower the attitude and practice scores were, the higher the probability of positive attitudes and good practices, while the higher the knowledge scores were, the higher the probability of good knowledge. Therefore, good knowledge of COVID-19 was directly associated with a positive attitude and good practices. Similarly, a positive attitude was directly associated with good practices.

Table 7. Correlation analysis (Pearson *r* and two-tailed *P* value) among the research variables.

Variable	Knowledge	Attitude	Practice
Knowledge			
r	1	–0.295[a]	–0.298
P value	—[b]	<.001	<.001
Attitude			
r	–0.295	1	0.173
P value	<.001	—	<.001
Practice			
r	–0.298	0.173	1
P value	<.001	<.001	—

[a]The correlation is significant at a significance level of .05 (two-tailed).

[b]Not applicable.

Availability of Data and Materials

The data set for this study is available from the corresponding author on reasonable request.

Discussion

Principal Findings

This study assessed the knowledge, attitudes, and practices related to COVID-19 among HCWs working at public health facilities. In this study, responses from 434 HCWs regarding sociodemographic characteristics, knowledge, attitudes, and infection prevention practices were analyzed. We found that the majority of HCWs had good knowledge and that over half of them had positive attitudes and good practices regarding COVID-19.

Our study found that most of the HCWs were well informed about COVID-19–related knowledge during the outbreak. Among these HCWs, the level of knowledge about COVID-19 was higher among male participants and nurses. From our study, we found that 73.3% (318/434) of the HCWs had sufficient knowledge about COVID-19, which is almost the same as the values reported by studies conducted in Uganda and Northern Ethiopia [14,15], where 69% and 74% of HCWs had sufficient knowledge, respectively. The proportion of HCWs with sufficient knowledge in this study was lower than in studies conducted in Vietnam and China [16,17].

This study explored the overall mean knowledge score, which was 13.7 (SD 2.6). These results were higher compared to a study from Vietnam that reported a mean knowledge score of 8.17 (SD 1.3). Their findings showed that HCWs had a high level of knowledge and a positive attitude regarding the COVID-19 outbreak [16]. In contrast, our results showed a lower knowledge level and a less positive attitude than in studies conducted in Egypt [18].

In general, having sufficient knowledge may reflect the successful dissemination of information about COVID-19 by different media. In this regard, most of the participants in this study used information from international and governmental media (ie, websites and verified social media pages). Our study suggests that the majority of HCWs used WHO websites as a source of information for COVID-19. This could be explained by the fact that the WHO website is regularly updated with facts. This suggests that such media should be frequently used to disseminate information on COVID-19 by the stakeholders involved in the COVID-19 response.

About 52.1% (226/434) of HCWs believed that caring for patients with COVID-19 may be a threat to HCWs, which is similar to findings by a study conducted in China [17] that showed that around 85% of the surveyed HCWs were afraid of getting infected while caring for their patients. HCWs help patients during their routine tasks, such as patient consultation, infusion, dressing changes, and surgery. They must also deal with many other emergency situations, and they may become infected with the virus if they are not cautious. Similarly, about 14.1% (61/434) of HCWs believed that wearing general medical masks may not protect the spread COVID-19, contrary to findings by Ng et al, which showed adequate protection [19].

Our study revealed that more than two-fifths of HCWs had negative attitudes toward COVID-19, which is in congruence with a knowledge, attitude, and practice study on COVID-19 in Uganda among HCWs [15], but in contrast to studies about COVID-19 in China and Ethiopia [17,20]. This study also explored the overall mean attitude score, which was 10.5 (SD 4.1). These results were lower than those in a study from Egypt that reported a mean knowledge score of 13.3 (SD 2.1) [18].

However, a Spearman analysis found a significant negative correlation between the mean knowledge and attitude scores of HCWs regarding COVID-19 (r=–0.295, P<.001), which is consistent with studies conducted in Vietnam and Egypt [16,18]. In other words, the lower the attitude scores were, the higher the probability of positive attitudes, while the higher the knowledge scores were, the higher the probability of good knowledge and practice. Therefore, good knowledge of COVID-19 was directly associated with a positive attitude and good practices. Similarly, a good attitude was positively associated with good practices.

Our study showed that the majority of HCWs had good COVID-19 prevention practices, which is in line with findings by other studies on COVID-19 [14,15,17]. The majority of the HCWs were following key infection prevention and control practices recommended by the Ministry of Health Ethiopia and the WHO. These included regular hand hygiene, social distancing, and wearing of personal protective equipment when in high-risk situations. A total of 72.4% (314/434) and 67.5% (293/434) of HCWs reported wearing personal protective equipment when in contact with patients as well as washing hands before and after handling patients, respectively. These are vital practices to prevent transfer of COVID-19 from patient to patient and to the HCWs themselves. This finding is in line with a study conducted in Ethiopia [20], but the values are lower than in many other studies [21-23]. However, up to 72.6% (315/434) of HCWs admitted having avoided patients with symptoms suggestive of COVID-19. This can be attributed to a global shortage of personal protective equipment [24,25].

It was also observed that occupation was significantly associated with knowledge. Nurses and pharmacists showed relatively better knowledge than other health professionals. This finding is in line with a study conducted in Vietnam that showed that pharmacists had better knowledge than other health professionals [16]. Similarly, a significant increase in the knowledge score was detected in those getting information from WHO websites, followed by government media. This finding corroborates a report from Egypt that found that HCWs who gained information about the disease from the internet, either through social media or the WHO website, had better knowledge scores [18]. This could be explained by the fact that younger, highly educated persons tend to use the internet more than older, less educated persons. In contrast, a study conducted in Uganda showed that HCWs who gained information from the traditional news media, such as television, radio, and newspapers, had more knowledge [15].

Our results also found that good knowledge was significantly associated with positive attitudes. This is in line with several studies that found an association between the COVID-19 knowledge level of HCWs and their attitudes [16,21]. Knowledge of HCWs is a very important prerequisite for positive attitudes and for promoting positive practices. Attitudes toward COVID-19 were also better among HCWs who got information from social media, followed by WHO websites. This is in line with a study that showed that social media exposure to COVID-19 information influences the adoption of preventive attitudes and behaviors through shaping risk perception [26]. Thus, understanding the role of social media during the pandemic could help policy makers and communicators to develop better communication strategies that enable HCWs to adopt appropriate attitudes and behaviors.

We also observed that older participants tended to practice precautionary measures more than younger participants; our findings support a report from Bangladesh that showed that a notable proportion of young adults did not have good preventive practices regarding COVID-19 [27]. This could be explained by the fact that younger participants considered themselves less at risk compared to older participants.

Limitations

Our study has some limitations. Firstly, no standardized tool for assessing knowledge, attitudes, and practices regarding COVID-19 has been previously validated. We adapted and modified a previously published tool for assessment of knowledge, attitudes, and practices regarding the prevention of respiratory tract infections. Secondly, only HCWs in public health facilities in parts of the Somali Region were surveyed, and the results of this study may not reflect the knowledge, attitudes, and practices of HCWs in the private sector. A similar study may be extended to the community.

Conclusions

This study highlights that more than 73% of HCWs had sufficient knowledge on the transmission, diagnosis, and prevention of COVID-19; more than half had positive attitudes toward COVID-19; and two-thirds had good practices regarding COVID-19 precautionary measures. There was a statistically significant difference in the level of knowledge about COVID-19 among HCWs. In a nutshell, the overall levels of knowledge and good practice were better compared to the positive attitude level. Hence, efforts to enhance the capacity of HCWs can be very helpful for the containment of the pandemic through enhancing positive attitudes and good practices.

Acknowledgments

We would like to extend a special thanks to Jigjiga University for funding research studies on COVID-19, which may be used in planning health programs about the emerging viral disease. This research was funded by Jigjiga University.

Authors' Contributions

AMF conceived the study, prepared the proposal, analyzed the data, interpreted the findings, and wrote the manuscript. TYN, M Obsiye, MAA, OMA, MAH, ABB, M Omer, and FG were involved in developing the study proposal, data analysis, and reviewing the manuscript.

Conflicts of Interest

None declared.

References

1. Kassema JJ. COVID-19 outbreak: Is it a health crisis or economic crisis or both? Case of African counties. SSRN 2020 Mar 23;9(51):16 [FREE Full text] [doi: 10.2139/ssrn.3559200]
2. COVID-19 Weekly Bulletin. Ethiopian Public Health Institute (EPHI), Ministry of Health (Ethiopia). URL: https://ephi.gov.et/?s=covid-19+weekly+bulletin [accessed 2020-11-22]
3. Poole DN, Escudero DJ, Gostin LO, Leblang D, Talbot EA. Responding to the COVID-19 pandemic in complex humanitarian crises. Int J Equity Health 2020 Mar 21;19(1):41 [FREE Full text] [doi: 10.1186/s12939-020-01162-y] [Medline: 32199457]
4. Cascella M, Rajnik M, Aleem A, Dulebohn SC, Di Napoli R. Features, Evaluation, and Treatment of Coronavirus (COVID-19). Treasure Island, FL: StatPearls Publishing; Jan 2021.
5. Healthcare workers: Information on COVID-19. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-nCoV/hcp/index.html [accessed 2020-10-14]
6. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 2020 Feb 15;395(10223):497-506 [FREE Full text] [doi: 10.1016/S0140-6736(20)30183-5] [Medline: 31986264]
7. Infection Prevention and Control During Health Care When Novel Coronavirus (nCoV) Infection Is Suspected. Interim Guidance. Geneva, Switzerland: World Health Organization; 2020 Mar 19. URL: https://www.who.int/publications/i/item/10665-331495 [accessed 2020-10-14]
8. Advice on the Use of Masks in the Community, During Home Care and in Health Care Settings in the Context of the Novel Coronavirus (2019-nCoV) Outbreak: Interim Guidance. Geneva, Switzerland: World Health Organization; 2020 Jan 29. URL: https://apps.who.int/iris/handle/10665/330987 [accessed 2020-10-10]
9. Selvaraj SA, Lee KE, Harrell M, Ivanov I, Allegranzi B. Infection rates and risk factors for infection among health workers during Ebola and Marburg virus outbreaks: A systematic review. J Infect Dis 2018 Nov 22;218(suppl_5):S679-S689 [FREE Full text] [doi: 10.1093/infdis/jiy435] [Medline: 30202878]
10. McCloskey B, Heymann DL. SARS to novel coronavirus - Old lessons and new lessons. Epidemiol Infect 2020 Feb 05;148:e22 [FREE Full text] [doi: 10.1017/S0950268820000254] [Medline: 32019614]
11. Hoffman SJ, Silverberg SL. Delays in global disease outbreak responses: Lessons from H1N1, Ebola, and Zika. Am J Public Health 2018 Mar;108(3):329-333. [doi: 10.2105/AJPH.2017.304245] [Medline: 29345996]
12. Ehrlich H, McKenney M, Elkbuli A. Protecting our healthcare workers during the COVID-19 pandemic. Am J Emerg Med 2020 Jul;38(7):1527-1528 [FREE Full text] [doi: 10.1016/j.ajem.2020.04.024] [Medline: 32336585]
13. Asaad A, El-Sokkary R, Alzamanan M, El-Shafei M. Knowledge and attitudes towards Middle East respiratory sydrome-coronavirus (MERS-CoV) among health care workers in south-western Saudi Arabia. East Mediterr Health J 2020 Apr 16;26(4):435-442 [FREE Full text] [doi: 10.26719/emhj.19.079] [Medline: 32338362]
14. Tadesse D, Gebrewahd GT, Demoz GT. Knowledge, attitude, practice and psychological response toward COVID-19 among nurses during the COVID-19 outbreak in northern Ethiopia, 2020. New Microbes New Infect 2020 Nov;38:100787 [FREE Full text] [doi: 10.1016/j.nmni.2020.100787] [Medline: 33072339]
15. Olum R, Chekwech G, Wekha G, Nassozi DR, Bongomin F. Coronavirus disease-2019: Knowledge, attitude, and practices of health care workers at Makerere University Teaching Hospitals, Uganda. Front Public Health 2020;8:181 [FREE Full text] [doi: 10.3389/fpubh.2020.00181] [Medline: 32426320]
16. Huynh G, Nguyen TH, Tran V, Vo K, Vo V, Pham L. Knowledge and attitude toward COVID-19 among healthcare workers at District 2 Hospital, Ho Chi Minh City. Asian Pac J Trop Med 2020 Mar 20;13(6):260. [doi: 10.4103/1995-7645.280396]
17. Zhang M, Zhou M, Tang F, Wang Y, Nie H, Zhang L, et al. Knowledge, attitude, and practice regarding COVID-19 among healthcare workers in Henan, China. J Hosp Infect 2020 Jun;105(2):183-187 [FREE Full text] [doi: 10.1016/j.jhin.2020.04.012] [Medline: 32278701]

18. Abdel Wahed WY, Hefzy EM, Ahmed MI, Hamed NS. Assessment of knowledge, attitudes, and perception of health care workers regarding COVID-19, a cross-sectional study from Egypt. J Community Health 2020 Dec;45(6):1242-1251 [FREE Full text] [doi: 10.1007/s10900-020-00882-0] [Medline: 32638199]
19. Ng K, Poon BH, Kiat Puar TH, Shan Quah JL, Loh WJ, Wong YJ, et al. COVID-19 and the risk to health care workers: A case report. Ann Intern Med 2020 Jun 02;172(11):766-767 [FREE Full text] [doi: 10.7326/L20-0175] [Medline: 32176257]
20. Jemal B, Aweke Z, Mola S, Hailu S, Abiy S, Dendir G, et al. Knowledge, attitude, and practice of healthcare workers toward COVID-19 and its prevention in Ethiopia: A multicenter study. SAGE Open Med 2021;9:1-10 [FREE Full text] [doi: 10.1177/20503121211034389] [Medline: 34377469]
21. Zhang M, Zhou M, Tang F, Wang Y, Nie H, Zhang L, et al. Knowledge, attitude, and practice regarding COVID-19 among healthcare workers in Henan, China. J Hosp Infect 2020 Jun;105(2):183-187 [FREE Full text] [doi: 10.1016/j.jhin.2020.04.012] [Medline: 32278701]
22. Zhong B, Luo W, Li H, Zhang Q, Liu X, Li W, et al. Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: A quick online cross-sectional survey. Int J Biol Sci 2020;16(10):1745-1752 [FREE Full text] [doi: 10.7150/ijbs.45221] [Medline: 32226294]
23. Kumar J, Katto MS, Siddiqui AA, Sahito B, Jamil M, Rasheed N, et al. Knowledge, attitude, and practices of healthcare workers regarding the use of face mask to limit the spread of the new coronavirus disease (COVID-19). Cureus 2020 Apr 20;12(4):e7737 [FREE Full text] [doi: 10.7759/cureus.7737] [Medline: 32455057]
24. Bauchner H, Fontanarosa PB, Livingston EH. Conserving supply of personal protective equipment-A call for ideas. JAMA 2020 May 19;323(19):1911. [doi: 10.1001/jama.2020.4770] [Medline: 32196543]
25. Ranney ML, Griffeth V, Jha AK. Critical supply shortages - The need for ventilators and personal protective equipment during the COVID-19 pandemic. N Engl J Med 2020 Apr 30;382(18):e41. [doi: 10.1056/NEJMp2006141] [Medline: 32212516]
26. Zeballos Rivas DR, Lopez Jaldin ML, Nina Canaviri B, Portugal Escalante LF, Alanes Fernández AMC, Aguilar Ticona JP. Social media exposure, risk perception, preventive behaviors and attitudes during the COVID-19 epidemic in La Paz, Bolivia: A cross sectional study. PLoS One 2021;16(1):e0245859 [FREE Full text] [doi: 10.1371/journal.pone.0245859] [Medline: 33481945]
27. Banik R, Rahman M, Sikder MT, Rahman QM, Pranta M. Knowledge, attitudes, and practices related to the COVID-19 pandemic among Bangladeshi youth: A web-based cross-sectional analysis. Z Gesundh Wiss 2021 Jan 16:1-11 [FREE Full text] [doi: 10.1007/s10389-020-01432-7] [Medline: 33489718]

Abbreviations

CDC: Centers for Disease Control and Prevention
HCW: health care worker
WHO: World Health Organization

Edited by G Eysenbach; submitted 30.01.21; peer-reviewed by A Bhagavathula, M Dobbins; comments to author 02.04.21; revised version received 23.04.21; accepted 24.08.21; published 01.10.21.

Please cite as:
Farah AM, Nour TY, Obsiye M, Aden MA, Ali OM, Hussein MA, Budul AB, Omer M, Getnet F
Knowledge, Attitudes, and Practices Regarding COVID-19 Among Health Care Workers in Public Health Facilities in Eastern Ethiopia: Cross-sectional Survey Study
JMIR Form Res 2021;5(10):e26980
URL: https://formative.jmir.org/2021/10/e26980
doi:10.2196/26980
PMID:34477559

Original Paper

Using a Constraint-Based Method to Identify Chronic Disease Patients Who Are Apt to Obtain Care Mostly Within a Given Health Care System: Retrospective Cohort Study

Yao Tong[1], MSc; Zachary C Liao[1], MD, MPH; Peter Tarczy-Hornoch[1,2,3], MD; Gang Luo[1], DPhil

[1]Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, United States

[2]Department of Pediatrics, Division of Neonatology, University of Washington, Seattle, WA, United States

[3]Department of Computer Science and Engineering, University of Washington, Seattle, WA, United States

Corresponding Author:
Gang Luo, DPhil
Department of Biomedical Informatics and Medical Education
University of Washington
UW Medicine South Lake Union
850 Republican Street, Building C, Box 358047
Seattle, WA, 98195
United States
Phone: 1 206 221 4596
Fax: 1 206 221 2671
Email: gangluo@cs.wisc.edu

Abstract

Background: For several major chronic diseases including asthma, chronic obstructive pulmonary disease, chronic kidney disease, and diabetes, a state-of-the-art method to avert poor outcomes is to use predictive models to identify future high-cost patients for preemptive care management interventions. Frequently, an American patient obtains care from multiple health care systems, each managed by a distinct institution. As the patient's medical data are spread across these health care systems, none has complete medical data for the patient. The task of building models to predict an individual patient's cost is currently thought to be impractical with incomplete data, which limits the use of care management to improve outcomes. Recently, we developed a constraint-based method to identify patients who are apt to obtain care mostly within a given health care system. Our method was shown to work well for the cohort of all adult patients at the University of Washington Medicine for a 6-month follow-up period. It is unknown how well our method works for patients with various chronic diseases and over follow-up periods of different lengths, and subsequently, whether it is reasonable to perform this predictive modeling task on the subset of patients pinpointed by our method.

Objective: To understand our method's potential to enable this predictive modeling task on incomplete medical data, this study assesses our method's performance at the University of Washington Medicine on 5 subgroups of adult patients with major chronic diseases and over follow-up periods of 2 different lengths.

Methods: We used University of Washington Medicine data for all adult patients who obtained care at the University of Washington Medicine in 2018 and PreManage data containing usage information from all hospitals in Washington state in 2019. We evaluated our method's performance over the follow-up periods of 6 months and 12 months on 5 patient subgroups separately—asthma, chronic kidney disease, type 1 diabetes, type 2 diabetes, and chronic obstructive pulmonary disease.

Results: Our method identified 21.81% (3194/14,644) of University of Washington Medicine adult patients with asthma. Around 66.75% (797/1194) and 67.13% (1997/2975) of their emergency department visits and inpatient stays took place within the University of Washington Medicine system in the subsequent 6 months and in the subsequent 12 months, respectively, approximately double the corresponding percentage for all University of Washington Medicine adult patients with asthma. The performance for adult patients with chronic kidney disease, adult patients with chronic obstructive pulmonary disease, adult patients with type 1 diabetes, and adult patients with type 2 diabetes was reasonably similar to that for adult patients with asthma.

Conclusions: For each of the 5 chronic diseases most relevant to care management, our method can pinpoint a reasonably large subset of patients who are apt to obtain care mostly within the University of Washington Medicine system. This opens the door to building models to predict an individual patient's cost on incomplete data, which was formerly deemed impractical.

International Registered Report Identifier (IRRID): RR2-10.2196/13783

(JMIR Form Res 2021;5(10):e26314) doi:10.2196/26314

KEYWORDS

asthma; chronic kidney disease; chronic obstructive pulmonary disease; data analysis; diabetes mellitus; emergency department; health care system; inpatients; patient care management

Introduction

Background

Care management is widely used to improve the outcomes of patients with chronic diseases [1]. Typically, a model is built to predict an individual patient's cost [1-5]. For a patient predicted to incur high costs in the future, we enroll the patient in a care management program for preemptive interventions. Then a care manager will call the patient regularly to check the patient's status and help arrange health and related services [1-5]. Proper use of care management can lower costs by up to 15%, can reduce hospital visits (emergency department visits and inpatient stays) by up to 40%, and has many other benefits [4,6-13]. Care management is typically used for managing several chronic diseases including asthma, chronic obstructive pulmonary disease, chronic kidney disease, and diabetes, as these diseases fulfill 3 conditions that allow a care management program to be economically feasible for implementation: (1) The disease has a high prevalence rate. (2) If not treated appropriately, the disease can result in acute exacerbations, which are associated with high expenses. (3) Relatively low-cost and effective interventions within the patient's control are available for the disease [6,14].

In the United States, a patient often obtains care from several health care systems such as academic medical centers and private physician groups. Therefore, the patient's medical data are spread across these health care systems, and none has complete medical data for the patient. Our prior work [15] showed that less than one-third of hospital visits by adult patients at the University of Washington Medicine (UWM) took place within the UWM in a 6-month follow-up period from April to October 2017. Other researchers showed similar evidence of care fragmentation for adult hospital visits in Massachusetts [16] and for emergency department visits in Indiana [17]. Typical models for forecasting an individual patient's cost presume complete historical data [14,18,19]. These models cannot be used for a health care system with incomplete data, resulting in many patients with future high costs that could be predicted being missed by care management interventions and having poor outcomes.

Recently, we developed the first constraint-based method to pinpoint a reasonably large subset of patients who are apt to obtain care mostly within a given health care system [15]. For a 6-month follow-up period from April to October 2017, we showed that this constraint-based method worked well for the cohort of all adult patients at the UWM [15]. However, we do not yet know how well our method works for patients with various chronic diseases and over follow-up periods of different lengths. If our method performs well in these cases, for the subset of patients with chronic diseases that is pinpointed by our method and for which the health care system has more complete data, it would then be possible to build a model to predict an individual patient's cost. This would be better than the current practice of not using any cost prediction model to facilitate care management for this health care system at all.

Objectives

To understand the potential of our constraint-based method at enabling building models to predict an individual patient's cost using incomplete medical data, we aimed to assess our method's performance at the UWM for 5 subgroups of adult patients and over follow-up periods of 2 different lengths. Each subgroup corresponds to one of the 5 major chronic diseases for which care management is used—asthma, chronic kidney disease, chronic obstructive pulmonary disease, type 1 diabetes, and type 2 diabetes.

Methods

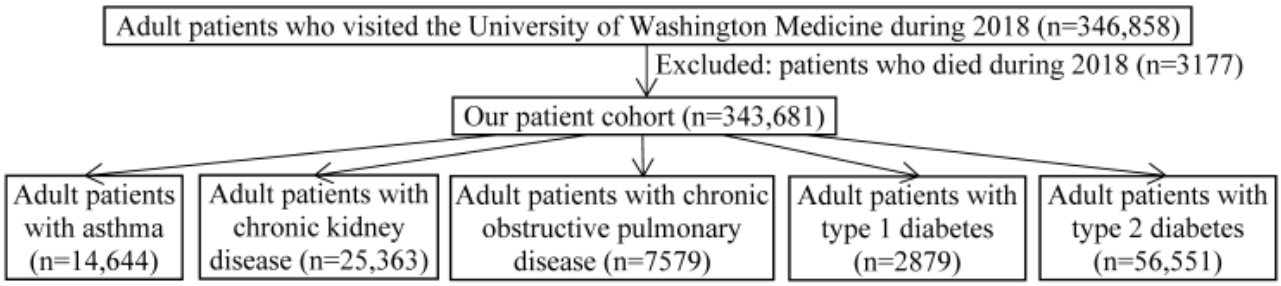
Ethics Approval

The UWM's institutional review board approved this retrospective cohort study (STUDY00000118).

Patient Population

As the largest academic health care system in Washington state, the UWM provides both clinic-based and hospital-based care for adults. The patient cohort (Figure 1) included adult patients (age ≥18 years) who visited the UWM during 2018 and who had information stored in the UWM's enterprise data warehouse. Unless explicitly specified as a particular type of visit, a visit can be of any type (outpatient visit, emergency department visit, or inpatient stay) in this paper. Patients who died during 2018 were excluded from the cohort.

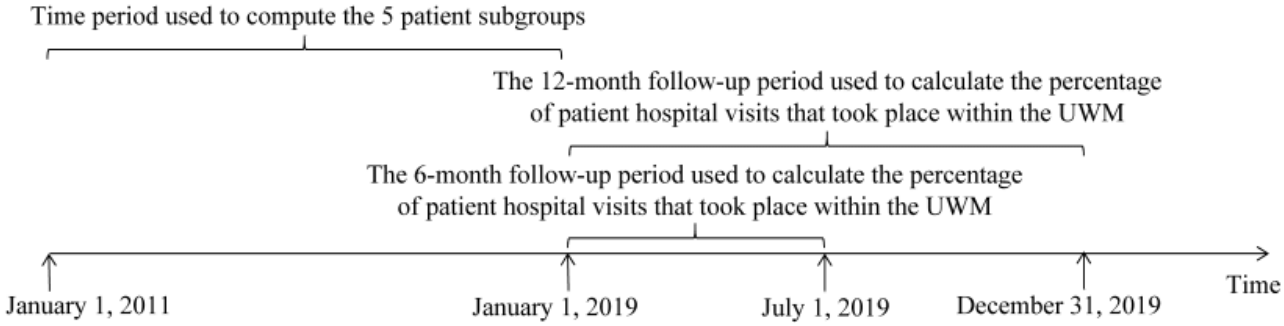
Figure 1. The patient cohort and the 5 patient subgroups.



Data Set

We used clinical and administrative data for the period from 2011 to 2018 stored in the UWM's enterprise data warehouse. The data set included information on demographics, visits, diagnoses, laboratory tests, medications, and primary care physicians for patients in our cohort. We also used data of UWM patients from 2019 collected in the commercial product PreManage (Collective Medical Technologies Inc). PreManage contains diagnosis and visit data of hospital visits (emergency department visits and inpatient stays) at all hospitals in Washington state as well as those from many hospitals in other US states [20]. We used January 1, 2019 as the index date to separate the subsequent and prior periods for our analysis task (Figure 2).

Figure 2. The time periods used to compute the patient subgroups and the percentages of patient hospital visits that took place within the UWM. UWM: University of Washington Medicine.



Patient Subgroups

Overview

We considered 5 patient subgroups that comprised patients with a specific major chronic disease in our patient cohort in 2018. One subgroup was created for each of 5 major chronic diseases: asthma, chronic kidney disease, chronic obstructive pulmonary disease, type 1 diabetes, and type 2 diabetes.

Asthma

A patient was deemed to have asthma in 2018 if the patient had ≥1 International Classification of Diseases, Ninth Revision (ICD-9) or Tenth Revision (ICD-10) diagnosis code for asthma (ICD-9: 493.0x, 493.8x, 493.1x, 493.9x; ICD-10: J45.x) in 2018 [21-23].

Chronic Kidney Disease

A patient was deemed to have chronic kidney disease if the patient had an estimated glomerular filtration rate (eGFR) <60 mL/min/1.73m^2 or proteinuria in 2 measurements that were ≥3 months apart [24,25]. The UWM computes eGFR using the Modification of Diet in Renal Disease equation: eGFR (mL/min/1.73m^2) = 175 × age$^{-0.203}$ × serum creatinine$^{-1.154}$ × 0.742 (if female) × 1.212 (if Black or African American) [26]. Proteinuria was detected by urine dipstick test result for protein ≥ 1+ (ie, ≥30 mg/dL) [24].

Chronic Obstructive Pulmonary Disease

By adjusting the criteria adopted by the National Quality Forum and the Centers for Medicare and Medicaid Services [27-29], we encompassed emergency department and outpatient visit data [30] to identify patients with chronic obstructive pulmonary disease. A patient was deemed to have chronic obstructive pulmonary disease if the patient was ≥40 years and fulfilled any of these 4 conditions: (1) an outpatient visit diagnosis code of chronic obstructive pulmonary disease (ICD-9: 491.22, 491.21, 491.9, 491.8, 493.2x, 492.8, 496; ICD-10: J42, J41.8, J44.*, J43.*) followed by ≥1 prescription of long-acting muscarinic antagonist (aclidinium, glycopyrrolate, tiotropium, and umeclidinium) within 6 months, (2) ≥1 emergency department or ≥2 outpatient visit diagnosis codes of chronic obstructive pulmonary disease (ICD-9: 491.22, 491.21, 491.9, 491.8, 493.2x, 492.8, 496; ICD-10: J42, J41.8, J44.*, J43.*), (3) ≥1 inpatient stay discharge having a principal diagnosis code of chronic obstructive pulmonary disease (ICD-9: 491.22, 491.21, 491.9, 491.8, 493.2x, 492.8, 496; ICD-10: J42, J41.8, J44.*, J43.*), and (4) ≥1 inpatient stay discharge having a principal diagnosis code of respiratory failure (ICD-9: 518.82, 518.81, 799.1, 518.84; ICD-10: J96.0*, J80, J96.9*, J96.2*, R09.2) and a secondary diagnosis code of acute chronic obstructive pulmonary disease exacerbation (ICD-9: 491.22, 491.21, 493.22, 493.21; ICD-10: J44.1, J44.0).

Type 1 and Type 2 Diabetes

We used Nichols et al's method [31] to identify patients with diabetes. A patient was deemed to have diabetes if the patient had ≥1 inpatient stay diagnosis code for diabetes (ICD-9: 250.x, 357.2, 362.0x, 366.41; ICD-10: E10.x, E11.x) or if any 2 of the following events occurred on the patient within 2 years of each other: (1) hemoglobin A_{1c} (HbA_{1c}) ≥6.5%, (2) random plasma glucose ≥200 mg/dL, (3) fasting plasma glucose ≥126 mg/dL, (4) an outpatient visit diagnosis code of diabetes (ICD-9: 250.x, 357.2, 362.0x, 366.41; ICD-10: E10.x, E11.x), and (5) a prescription of antihyperglycemic medication (α-glucosidase inhibitor, amylin analogue, biguanide, dipeptidyl peptidase-4 inhibitor, incretin mimetic, insulin, meglitinide, sulfonylurea, and thiazolidinedione). Two events of the same type, such as 2 instances of HbA_{1c} ≥6.5%, qualified as separate events if they occurred on 2 different days. We did not count 2 prescriptions of metformin or thiazolidinedione with no other manifestation of diabetes, as metformin (a biguanide) and thiazolidinedione could be used for other diseases. We also excluded events that occurred during a pregnancy.

We used Klompas et al's method [32,33] to distinguish type 1 diabetes and type 2 diabetes. Using all diagnosis codes, laboratory test results, and medication prescriptions during the period 2011 to 2018, we deemed a patient with diabetes to have type 1 diabetes if the patient fulfilled any of the following 4 conditions: (1) the number of type 1 diabetes diagnosis codes (ICD-9: 250.x3, 250.x1; ICD-10: E10.x) was greater than the number of type 2 diabetes diagnosis codes (ICD-9: 250.x2, 250.x0; ICD-10: E11.x) and there was a prescription of glucagon, (2) the number of type 1 diabetes diagnosis codes (ICD-9: 250.x3, 250.x1; ICD-10: E10.x) was greater than the number of type 2 diabetes diagnosis codes (ICD-9: 250.x2, 250.x0; ICD-10: E11.x) and there were no prescriptions of oral hypoglycemic medications other than metformin, (3) a negative C-peptide laboratory test result, and (4) a positive diabetes autoantibody laboratory test result. A patient with diabetes was deemed to have type 2 diabetes if the patient was not deemed to have type 1 diabetes.

Constraint-Based Method for Identifying Patients

We looked at 3 UWM hospitals whose clinical and administrative data are kept in the UWM's enterprise data warehouse: University of Washington Medical Center, Harborview Medical Center, and Northwest Hospital (all are in Seattle, Washington). To identify patients who are apt to obtain care mostly within the UWM, we used the parameterized primary care physician constraint developed in our recent paper [15]: the patient has a UWM primary care physician and resides within *d* km of at least 1 of the 3 UWM hospitals. The distance between a UWM hospital and a patient's home is the ellipsoid great-circle distance computed by the distVincentyEllipsoid function contained in R's geosphere package (version 1.5-5 [34]). *d* is a parameter. For all UWM adult patients and the follow-up period of 6 months, we showed that the optimal value of *d* is approximately 8 km (5 miles) [15].

Data Analysis

We considered 2 follow-up periods: the subsequent 6 months (January 1, 2019 to June 30, 2019) and the subsequent 12 months (January 1, 2019 to December 31, 2019) (Figure 2). The 6-month follow-up period was chosen to be consistent with the duration of the follow-up period used in our prior paper [15]. The 12-month follow-up period was chosen because, to facilitate care management, typically a minimum of 1 year of historical data is needed to build models that predict an individual patient's cost [14]. For each of the 5 patient subgroups and each of the 2 follow-up periods, we computed our method's performance in identifying patients who are apt to obtain care mostly within the UWM. We employed administrative data in the UWM's enterprise data warehouse to assess whether a patient fulfilled the parameterized primary care physician constraint. For each of the 5 patient subgroups, we calculated the percentage of patients identified by the constraint = $n_0/m_0 \times 100\%$. Here, n_0 is the number of patients in the subgroup fulfilling the constraint. m_0 is the number of patients in the subgroup. For all patients in the subgroup fulfilling the constraint, we used PreManage data to calculate

1. the percentage of their hospital visits taking place within the UWM in the subsequent 6 months = $n_1/m_1 \times 100\%$, where n_1 is the number of their hospital visits taking place within the UWM in the subsequent 6 months, and m_1 is the number of their hospital visits taking place anywhere in the subsequent 6 months; and
2. the percentage of their hospital visits taking place within the UWM in the subsequent 12 months = $n_2/m_2 \times 100\%$, where n_2 is the number of their hospital visits taking place within the UWM in the subsequent 12 months, and m_2 is the number of their hospital visits taking place anywhere in the subsequent 12 months.

Since an average hospital visit costs much more than an average visit of another type, this percentage signifies the proportion of those patients' care obtained from the UWM.

To determine the optimal value of the distance threshold parameter *d*, we balanced 2 goals. (1) The proportion of hospital visits taking place within the UWM should be as large as possible for patients fulfilling the constraint. This will maximize the completeness of UWM medical data and minimize bias in the results of analyses done on those data. As outpatient visits are often handled by primary care physicians and these patients each have a UWM primary care physician, we expect most of their outpatient visits to occur within the UWM in the subsequent 12 months. (2) The percentage of patients fulfilling the constraint should be as large as possible. This will help maximize the impact of the application using UWM medical data.

To show how our method performs for every UWM hospital, for all patients in the subgroup fulfilling the constraint, we employed PreManage data to calculate (1) the percentage of their hospital visits taking place at the UWM hospital in the subsequent 6 months = $n_3/m_1 \times 100\%$, where n_3 is the number of their hospital visits taking place at the UWM hospital in the

subsequent 6 months, and m_1 is the number of their hospital visits taking place anywhere in the subsequent 6 months, and (2) the percentage of their hospital visits taking place at the UWM hospital in the subsequent 12 months = $n_4/m_2 \times 100\%$, where n_4 is the number of their hospital visits taking place at the UWM hospital in the subsequent 12 months, and m_2 is the number of their hospital visits taking place anywhere in the subsequent 12 months.

Results

The cohort of adult patients who visited UWM facilities during 2018 with information stored in the UWM's enterprise data warehouse comprised 343,681 patients (Table 1).

Figures 3 and 4 present the percentage of patients fulfilling the parameterized primary care physician constraint for each of the 5 patient subgroups. The percentage rises with increase in d, at first swiftly when d is small and then at a slower pace when d grows larger.

Table 1. Demographic and clinical characteristics of the study cohort.

Characteristic	Patients (n=343,681), n (%)
Age	
18 to <40 years	120,422 (35.04)
40 to 65 years	149,418 (43.48)
>65 years	73,841 (21.49)
Gender	
Male	159,964 (46.54)
Female	183,701 (53.45)
Unknown or not reported	16 (<.01)
Race	
Black or African American	25,513 (7.42)
American Indian or Alaska native	4795 (1.40)
Asian	34,474 (10.03)
Native Hawaiian or other Pacific islander	2843 (0.83)
Multiple races	1 (<0.01)
Unknown or not reported	45,094 (13.12)
White	230,961 (67.20)
Ethnicity	
Non-Hispanic	271,582 (79.02)
Hispanic	21,718 (6.32)
Unknown or not reported	50,381 (14.66)
Insurance	
Private	163,908 (47.69)
Public (Medicare and Medicaid)	160,026 (46.56)
Self-paid or charity	19,747 (5.75)
Disease	
Asthma	14,644 (4.26)
Chronic kidney disease	25,363 (7.38)
Chronic obstructive pulmonary disease	7579 (2.21)
Type 2 diabetes	56,551 (16.45)
Type 1 diabetes	2879 (0.84)

Figure 3. The percentage of patients in each of the 5 patient subgroups fulfilling the parameterized primary care physician constraint. CKD: chronic kidney disease. COPD: chronic obstructive pulmonary disease. T1D: type 1 diabetes. T2D: type 2 diabetes.

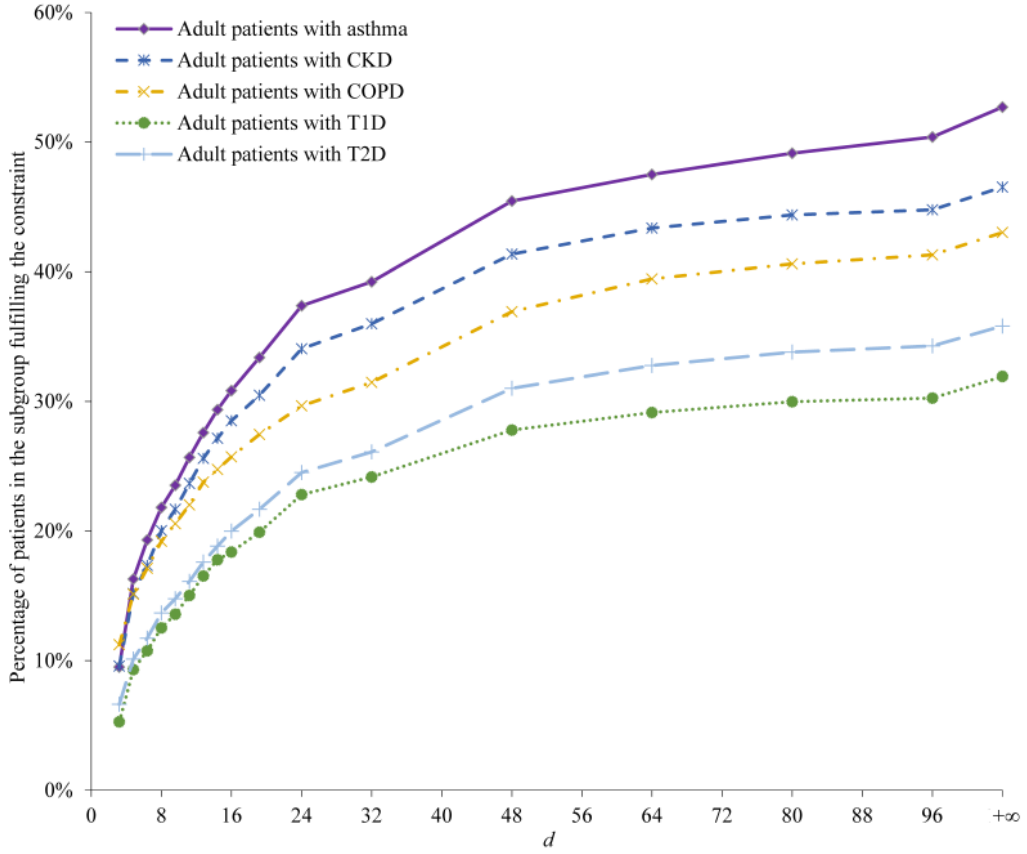
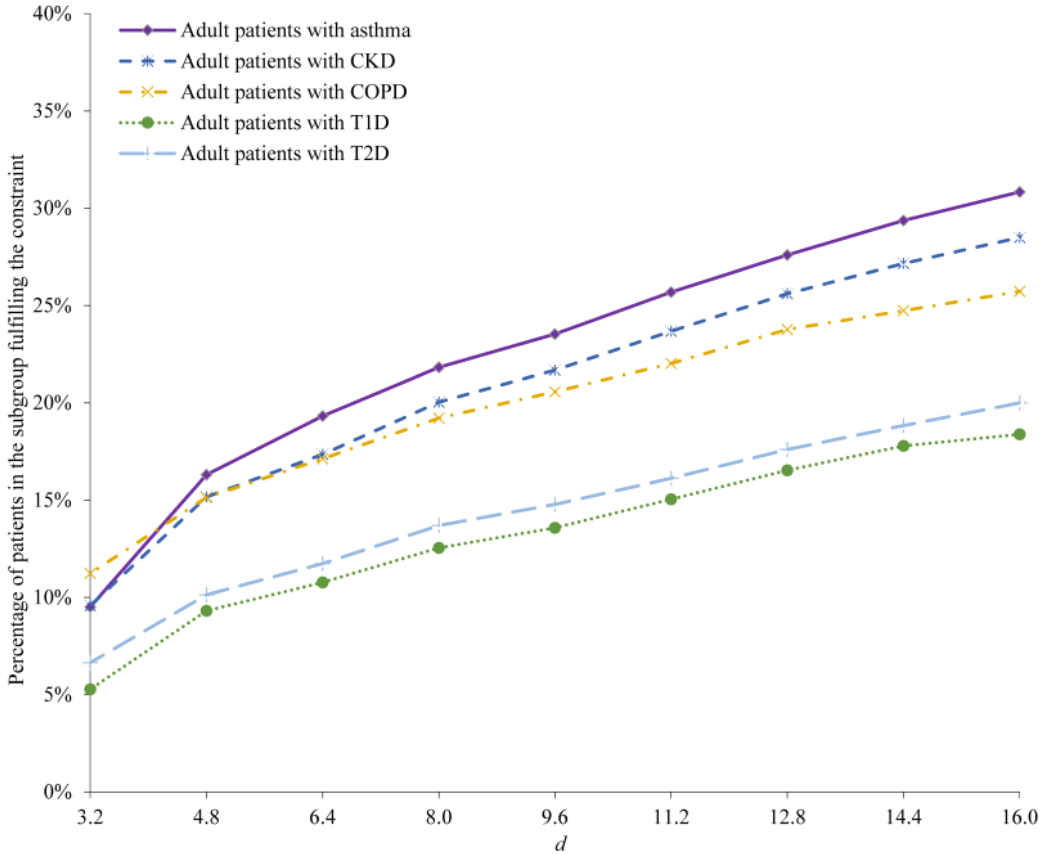


Figure 4. The percentage of patients in each of the 5 patient subgroups fulfilling the parameterized primary care physician constraint, when *d* is ≤10. CKD: chronic kidney disease. COPD: chronic obstructive pulmonary disease. T1D: type 1 diabetes. T2D: type 2 diabetes.



For all patients in each of the 5 patient subgroups fulfilling the parameterized primary care physician constraint, Figures 5 and 6 present the percentages of their hospital visits taking place within the UWM in the subsequent 6 months and in the subsequent 12 months. Except for a few cases at small values of *d*, the percentage decreases with increasing *d*, swiftly when *d* is small and then slowly when *d* is large.

We chose *d*=8 km as the optimal value to use for each patient subgroup and each follow-up period. Table 2 shows the corresponding performance measures of our constraint-based method.

For every UWM hospital and all patients in each of the 5 patient subgroups fulfilling the parameterized primary care physician constraint, Multimedia Appendix 1 shows the percentages of their hospital visits taking place at the UWM hospital in the subsequent 6 months and in the subsequent 12 months.

Figure 5. For all patients in each of the 5 patient subgroups fulfilling the parameterized primary care physician constraint, the percentages of their hospital visits taking place within the University of Washington Medicine (UWM) in the subsequent 6 months and in the subsequent 12 months. CKD: chronic kidney disease. COPD: chronic obstructive pulmonary disease. T1D: type 1 diabetes. T2D: type 2 diabetes.

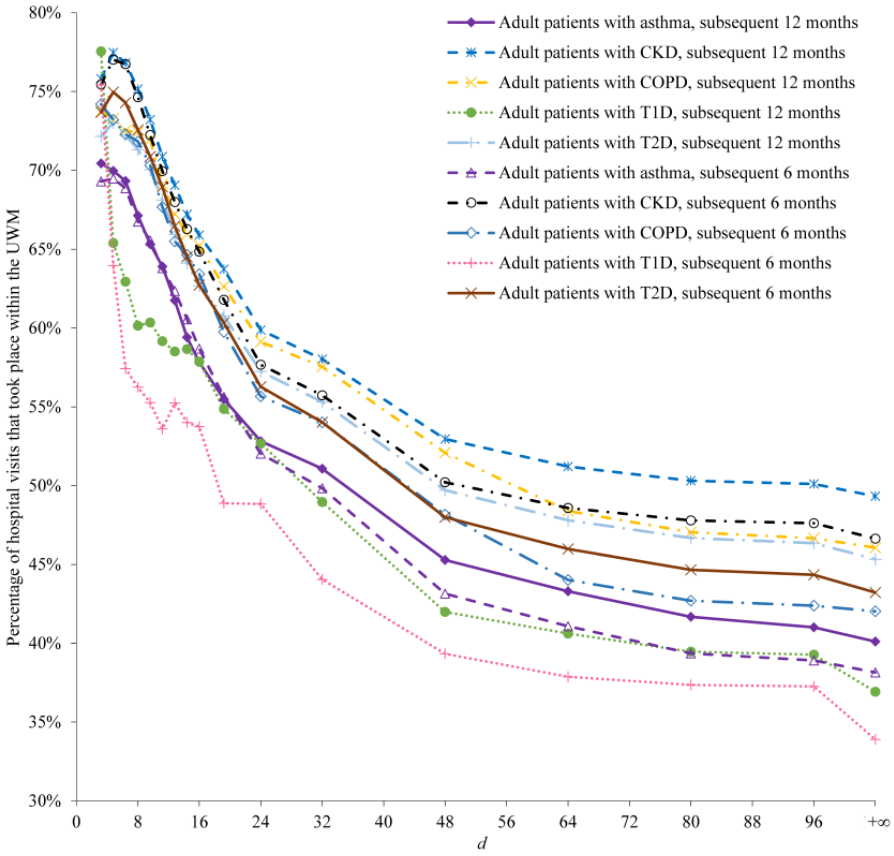


Figure 6. For all patients in each of the 5 patient subgroups fulfilling the parameterized primary care physician constraint when *d* is ≤10, the percentages of their hospital visits taking place within the University of Washington Medicine (UWM) in the subsequent 6 months and in the subsequent 12 months. CKD: chronic kidney disease. COPD: chronic obstructive pulmonary disease. T1D: type 1 diabetes. T2D: type 2 diabetes.

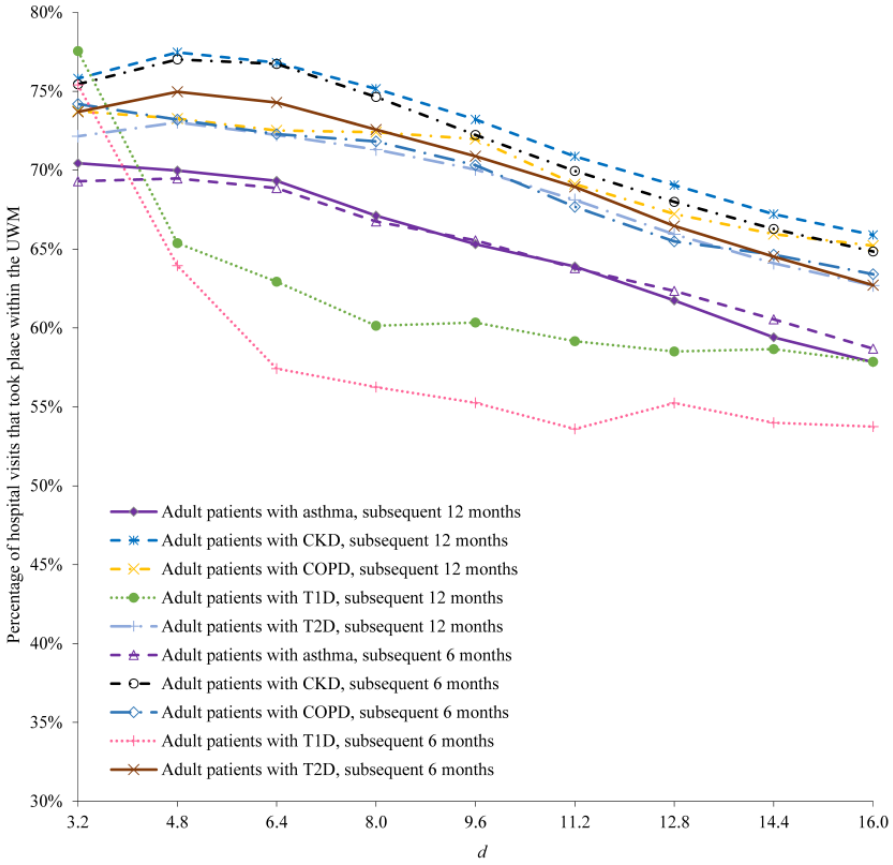


Table 2. For *d*=8 km and for each of the 5 patient subgroups, the percentage of patients fulfilling the parameterized primary care physician constraint, the percentages of patient hospital visits taking place at the University of Washington Medicine (UWM) in the subsequent 6 months and in the subsequent 12 months, and the percentages of hospital visits by patients fulfilling the constraint that took place within the UWM in the subsequent 6 months and in the subsequent 12 months.

Patient subgroup	Patients fulfilling the parameterized primary care physician constraint, n/N (%)	Patient hospital visits taking place within the UWM[a] in the subsequent 6 months, n/N (%)	Hospital visits by patients fulfilling the constraint that took place within the UWM in the subsequent 6 months, n/N (%)	Patient hospital visits taking place within the UWM in the subsequent 12 months, n/N (%)	Hospital visits by patients fulfilling the constraint that took place within the UWM in the subsequent 12 months, n/N (%)
Adult patients with asthma	3194/14,640 (21.81)	2648/7135 (37.11)	797/1194 (66.75)	6857/18,206 (37.66)	1997/2975 (67.13)
Adult patients with chronic kidney disease	5081/25,363 (20.03)	7503/18,404 (40.77)	2178/2918 (74.64)	19,558/45,994 (42.52)	5634/7496 (75.16)
Adult patients with chronic obstructive pulmonary disease	1456/7579 (19.21)	2587/6659 (38.85)	831/1157 (71.82)	7026/16,941 (41.47)	2179/3009 (72.42)
Adult patients with type 1 diabetes	361/2879 (12.54)	317/1333 (23.78)	63/112 (56.25)	845/3330 (25.38)	169/281 (60.14)
Adult patients with type 2 diabetes	7744/56,551 (13.69)	10,926/30,707 (35.58)	2847/3923 (72.57)	29,272/79,775 (36.69)	7177/10,065 (71.31)

[a]UWM: University of Washington Medicine.

Discussion

Principal Results

For each of the 5 major chronic diseases most relevant to care management (asthma, chronic obstructive pulmonary disease, chronic kidney disease, type 1 diabetes, and type 2 diabetes), our constraint-based method with a properly chosen value of the parameter *d* can pinpoint a reasonably large subset of patients who are apt to obtain care mostly within the UWM. Using our method to pinpoint a subset of UWM adult patients with asthma, we roughly doubled the percentage of patient hospital visits taking place within the UWM in the subsequent 6 months from 37.11% (2648/7135) to 66.75% (797/1194), and the corresponding percentage for the subsequent 12 months from 37.66% (6857/18,206) to 67.13% (1997/2975). The results for adult patients with chronic kidney disease, adult patients with chronic obstructive pulmonary disease, adult patients with type 1 diabetes, and adult patients with type 2 diabetes are relatively similar. As the patients fulfilling the constraint all have a UWM primary care physician, we expect a majority of their outpatient visits to happen within the UWM in the subsequent 12 months, although we did not examine this in our study.

Explanation of the Results Shown in Figure 5

UWM primary care physicians are inclined to refer within the UWM. Thus, intuitively, patients with a UWM primary care physician are apt to obtain a larger percentage of their care from the UWM than other patients. All else being equal, the UWM tends to provide a larger portion of a patient's care when the patient resides closer to UWM hospitals. This is reflected in Figure 5. When *d*=+, distance is no longer relevant for identifying patients. No matter how small *d* is, for all patients in each of the 5 patient subgroups fulfilling the parameterized primary care physician constraint, the percentage of their hospital visits taking place within the UWM in the subsequent 6 months (or in the subsequent 12 months) never becomes 100%, partially because patients can also visit multiple non-UWM hospitals within 1.6 km of some of the UWM hospitals. For each positive *d*, the percentage is relatively similar across the 5 patient subgroups and the 2 follow-up periods.

Comparison With our Prior Work

The findings in this paper are relatively similar to those of our previous study [15]—for the group of all adult patients and the 6-month follow-up period from April to October 2017. The optimal value of *d*=8 km found in this study is the same as that chosen in our previous study. In our previous study, using our constraint-based method with a parameter value of *d*=8 km to pinpoint 16.01% (55,707/348,054) of the UWM adult patients, we roughly doubled the percentage of patient hospital visits taking place within the UWM in the subsequent 6 months from 31.80% (39,171/123,162) to 69.38% (10,501/15,135).

Differences in the Results for Patients With Type 1 Diabetes and Patients With Type 2 Diabetes

Type 1 diabetes tends to occur in younger people than type 2 diabetes. There are many young adults who are students at the University of Washington and several other universities in the Seattle metropolitan area. During the summer and other university breaks, many of these students return to their hometowns outside the Seattle metropolitan area that the UWM primarily serves. The hospital visits that they incur during these periods are likely to be outside of the UWM system. Partly due to this, as shown in Table 2, the percentage of hospital visits by UWM adult patients with type 1 diabetes that took place within the UWM in each follow-up period is approximately 30% less than the corresponding percentage for UWM adult patients with type 2 diabetes. For patients fulfilling the parameterized primary care physician constraint with *d*=8 km (5 miles), the percentage of hospital visits by patients with type 1 diabetes taking place within the UWM in each follow-up period ranges from 15% to 30% less than the corresponding percentage for patients with type 2 diabetes.

Possible Use of our Results

We showed that for each of 5 major chronic diseases most relevant to care management, the UWM offers most of the care and has decently complete medical data for patients fulfilling the parameterized primary care physician constraint with *d*=8 km (5 miles). For these patients, we can build a predictive model to identify future high-cost patients and intervene preemptively via care management to avert poor outcomes [1-5]. As patients residing farther from the 3 UWM hospitals were inclined to obtain a smaller percentage of their care from the UWM, the UWM could consider adopting differing preventive interventions for patients residing at different distances from the UWM hospitals, which could help care management gain better results. For patients obtaining only a small percentage of their care from the UWM, it is hard for the UWM to adopt costly preventive interventions in an economic way.

Possible Ways to Assess our Method's Performance for Other Health Care Systems That Have No Access to PreManage Data

Like many other health care systems, the UWM has no complete claims data on its patients' health care use outside of the UWM. If a health care system has complete claims data on its patients' outside health care use, we could employ claims data instead of PreManage data to perform a similar study.

A health care system with no access to PreManage data could also adopt our method. Without using PreManage data, one could assess our method's performance by asking some patients of the health care system about care obtained elsewhere.

Limitations

This study has 2 limitations, which could be interesting topics for future work.

This study assessed the performance of our constraint-based method for 5 chronic diseases at the UWM, which primarily serves an urban region. To know how well our method generalizes to other health care systems, we need to redo our analysis at other health care systems, such as those providing care to rural regions or primarily serving urban regions. Residence are more concentrated in urban regions than in rural regions. For a health care system primarily serving rural regions, we expect *d*>8 km for the optimal value.

For a health care system having incomplete medical data for its patients, we can employ our method to identify a subset of patients who are apt to obtain care mostly within the health care system and assess the degree of data incompleteness for this subset. Analyzing incomplete data could lead to biased results, which are better than no result if the degree of bias is small. At present, we know neither the exact relationship between data incompleteness and bias nor the extent of data incompleteness that can be tolerated before the results of data analysis become invalid. This is a gap. To assess whether our method could safely enable the data analysis task in such a health care system, we could obtain a more complete data set from Kaiser Permanente or any other similar healthcare system, remove different portions of the data set, and assess the effect on analysis results.

Conclusions

Our constraint-based method to pinpoint a reasonably large subset of patients who are apt to obtain care mostly within a given health care system opens the door to building models to predict an individual patient's cost on incomplete data, which was formerly deemed infeasible.

Acknowledgments

We thank Hye-Eun Yoon for useful discussions. GL was partially supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health (R01HL142503). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' Contributions

YT performed this work at the University of Washington as a visiting PhD student. YT participated in data analysis, performing literature review, and writing the paper's first draft. GL conceptualized and designed the study, participated in data analysis and performing literature review, and wrote most of the paper. ZCL and PT-H provided feedback on various medical issues, contributed to conceptualizing the presentation, and revised the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Results for every University of Washington Medicine hospital.
[PDF File (Adobe PDF File), 81 KB - formative_v5i10e26314_app1.pdf]

References

1. Mays GP, Claxton G, White J. Managed care rebound? Recent changes in health plans' cost containment strategies. Health Aff (Millwood) 2004;Suppl Web Exclusives:W4-427-W4-436 [FREE Full text] [doi: 10.1377/hlthaff.w4.427] [Medline: 15451964]
2. Curry N, Billings J, Darin B, Dixon J, Williams M, Wennberg D. Predictive risk project literature review. King's Fund. 2005. URL: http://www.kingsfund.org.uk/sites/files/kf/field/field_document/predictive-risk-literature-review-june2005.pdf [accessed 2021-09-06]
3. Vogeli C, Shields AE, Lee TA, Gibson TB, Marder WD, Weiss KB, et al. Multiple chronic conditions: prevalence, health consequences, and implications for quality, care management, and costs. J Gen Intern Med 2007 Dec;22 Suppl 3:391-395 [FREE Full text] [doi: 10.1007/s11606-007-0322-1] [Medline: 18026807]
4. Caloyeras JP, Liu H, Exum E, Broderick M, Mattke S. Managing manifest diseases, but not health risks, saved PepsiCo money over seven years. Health Aff (Millwood) 2014 Jan;33(1):124-131. [doi: 10.1377/hlthaff.2013.0625] [Medline: 24395944]
5. Nelson L, Congressional Budget Office. Lessons from Medicare's demonstration projects on disease management and care coordination. 2012. URL: https://www.cbo.gov/sites/default/files/112th-congress-2011-2012/workingpaper/WP2012-01_Nelson_Medicare_DMCC_Demonstrations_1.pdf [accessed 2021-09-05]
6. Fishbane S, Agoritsas S, Bellucci A, Halinski C, Shah HH, Sakhiya V, et al. Augmented nurse care management in CKD stages 4 to 5: a randomized trial. Am J Kidney Dis 2017 Oct;70(4):498-505. [doi: 10.1053/j.ajkd.2017.02.366] [Medline: 28396108]
7. Greineder DK, Loane KC, Parks P. A randomized controlled trial of a pediatric asthma outreach program. J Allergy Clin Immunol 1999 Mar;103(3 Pt 1):436-440. [Medline: 10069877]
8. Kelly CS, Morrow AL, Shults J, Nakas N, Strope GL, Adelman RD. Outcomes evaluation of a comprehensive intervention program for asthmatic children enrolled in Medicaid. Pediatrics 2000 May;105(5):1029-1035. [Medline: 10790458]
9. Axelrod RC, Zimbro KS, Chetney RR, Sabol J, Ainsworth VJ. A disease management program utilizing life coaches for children with asthma. J Clin Outcomes Manag 2001;8(6):38-42.

10. Rubin RJ, Dietrich KA, Hawk AD. Clinical and economic impact of implementing a comprehensive diabetes management program in managed care. J Clin Endocrinol Metab 1998 Aug;83(8):2635-2642. [doi: 10.1210/jcem.83.8.5075] [Medline: 9709924]
11. Dorr DA, Wilcox AB, Brunker CP, Burdon RE, Donnelly SM. The effect of technology-supported, multidisease care management on the mortality and hospitalization of seniors. J Am Geriatr Soc 2008 Dec;56(12):2195-2202. [doi: 10.1111/j.1532-5415.2008.02005.x] [Medline: 19093919]
12. Beaulieu N, Cutler DM, Ho K, Isham G, Lindquist T, Nelson A, et al. The business case for diabetes disease management for managed care organizations. Forum Health Econ Policy 2006;9(1):1-37. [doi: 10.2202/1558-9544.1072]
13. Levine SH, Adams J, Attaway K, Dorr DA, Leung M, Popescu P, et al. Predicting the financial risks of seriously ill patients.: California Healthcare Foundation; 2011. URL: https://www.chcf.org/wp-content/uploads/2017/12/PDF-PredictiveModelingRiskStratification.pdf [accessed 2021-09-06]
14. Duncan I. Healthcare Risk Adjustment and Predictive Modeling, 2nd ed. New Hartford, CT: ACTEX Learning; 2018.
15. Luo G, Tarczy-Hornoch P, Wilcox AB, Lee ES. Identifying patients who are likely to receive most of their care from a specific health care system: demonstration via secondary analysis. JMIR Med Inform 2018 Nov 05;6(4):e12241 [FREE Full text] [doi: 10.2196/12241] [Medline: 30401670]
16. Bourgeois FC, Olson KL, Mandl KD. Patients treated at multiple acute health care facilities: quantifying information fragmentation. Arch Intern Med 2010 Dec 13;170(22):1989-1995. [doi: 10.1001/archinternmed.2010.439] [Medline: 21149756]
17. Finnell JT, Overhage JM, Grannis S. All health care is not local: an evaluation of the distribution of emergency department care delivered in Indiana. AMIA Annu Symp Proc 2011;2011:409-416 [FREE Full text] [Medline: 22195094]
18. Ash A, McCall N. Risk assessment of military populations to predict health care cost and utilization. RTI International. 2005. URL: http://www.rti.org/pubs/tricare_riskassessment_final_report_combined.pdf [accessed 2021-09-06]
19. Iezzoni LI. Risk Adjustment for Measuring Health Care Outcomes, 4th ed. Chicago, IL: Health Administration Press; 2013.
20. Collective Medical Technologies Inc. Collective Medical and Consonus Healthcare announce partnership to improve postacute transitions of care. 2018. URL: https://collectivemedical.com/resources/press-release/collective-medical-and-consonus-healthcare-announce-partnership-to-improve-post-acute-transitions-of-care [accessed 2021-09-03]
21. Schatz M, Cook EF, Joshua A, Petitti D. Risk factors for asthma hospitalizations in a managed care organization: development of a clinical prediction rule. Am J Manag Care 2003 Aug;9(8):538-547 [FREE Full text] [Medline: 12921231]
22. Desai JR, Wu P, Nichols GA, Lieu TA, O'Connor PJ. Diabetes and asthma case identification, validation, and representativeness when using electronic health data to construct registries for comparative effectiveness and epidemiologic research. Med Care 2012 Jul;50 Suppl:S30-S35. [doi: 10.1097/MLR.0b013e318259c011] [Medline: 22692256]
23. Wakefield DB, Cloutier MM. Modifications to HEDIS and CSTE algorithms improve case recognition of pediatric asthma. Pediatr Pulmonol 2006 Oct;41(10):962-971. [doi: 10.1002/ppul.20476] [Medline: 16871628]
24. Lee Y, Park S, Lee S, Kim Y, Kang MW, Cho S, et al. Lipid profiles and risk of major adverse cardiovascular events in CKD and diabetes: a nationwide population-based study. PLoS One 2020;15(4):e0231328 [FREE Full text] [doi: 10.1371/journal.pone.0231328] [Medline: 32271842]
25. Levey AS, Becker C, Inker LA. Glomerular filtration rate and albuminuria for detection and staging of acute and chronic kidney disease in adults: a systematic review. JAMA 2015 Feb 24;313(8):837-846 [FREE Full text] [doi: 10.1001/jama.2015.0602] [Medline: 25710660]
26. Levey AS, Coresh J, Greene T, Stevens LA, Zhang YL, Hendriksen S, Chronic Kidney Disease Epidemiology Collaboration. Using standardized serum creatinine values in the modification of diet in renal disease study equation for estimating glomerular filtration rate. Ann Intern Med 2006 Aug 15;145(4):247-254. [doi: 10.7326/0003-4819-145-4-200608150-00004] [Medline: 16908915]
27. NQF #1891 Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization. National Quality Forum. URL: http://www.qualityforum.org/Projects/n-r/Pulmonary_Endorsement_Maintenance/1891_30_Day_RSRR_COPD.aspx [accessed 2021-09-06]
28. Nguyen HQ, Chu L, Amy Liu LI, Lee JS, Suh D, Korotzer B, et al. Associations between physical activity and 30-day readmission risk in chronic obstructive pulmonary disease. Ann Am Thorac Soc 2014 Jun;11(5):695-705. [doi: 10.1513/AnnalsATS.201401-017OC] [Medline: 24713094]
29. Lindenauer PK, Grosso LM, Wang C, Wang Y, Krishnan JA, Lee TA, et al. Development, validation, and results of a risk-standardized measure of hospital 30-day mortality for patients with exacerbation of chronic obstructive pulmonary disease. J Hosp Med 2013 Aug;8(8):428-435. [doi: 10.1002/jhm.2066] [Medline: 23913593]
30. Cooke CR, Joo MJ, Anderson SM, Lee TA, Udris EM, Johnson E, et al. The validity of using ICD-9 codes and pharmacy records to identify patients with chronic obstructive pulmonary disease. BMC Health Serv Res 2011 Feb 16;11:37 [FREE Full text] [doi: 10.1186/1472-6963-11-37] [Medline: 21324188]
31. Nichols GA, Schroeder EB, Karter AJ, Gregg EW, Desai J, Lawrence JM, SUPREME-DM Study Group. Trends in diabetes incidence among 7 million insured adults, 2006-2011: the SUPREME-DM project. Am J Epidemiol 2015 Jan 1;181(1):32-39. [doi: 10.1093/aje/kwu255] [Medline: 25515167]

32. Klompas M, Eggleston E, McVetta J, Lazarus R, Li L, Platt R. Automated detection and classification of type 1 versus type 2 diabetes using electronic health record data. Diabetes Care 2013 Apr;36(4):914-921 [FREE Full text] [doi: 10.2337/dc12-0964] [Medline: 23193215]
33. Schroeder EB, Donahoo WT, Goodrich GK, Raebel MA. Validation of an algorithm for identifying type 1 diabetes in adults based on electronic health record data. Pharmacoepidemiol Drug Saf 2018 Oct;27(10):1053-1059 [FREE Full text] [doi: 10.1002/pds.4377] [Medline: 29292555]
34. R's geosphere package homepage. 2020. URL: https://cran.r-project.org/web/packages/geosphere/index.html [accessed 2021-09-06]

Abbreviations

eGFR: estimated glomerular filtration rate
HbA_{1c}: hemoglobin A_{1c}
ICD-9: International Classification of Diseases, Ninth Revision
ICD-10: International Classification of Diseases, Tenth Revision
UWM: University of Washington Medicine

Edited by G Eysenbach; submitted 29.12.20; peer-reviewed by B Zarowitz, C Jeong; comments to author 09.08.21; accepted 24.08.21; published 07.10.21.

Please cite as:
Tong Y, Liao ZC, Tarczy-Hornoch P, Luo G
Using a Constraint-Based Method to Identify Chronic Disease Patients Who Are Apt to Obtain Care Mostly Within a Given Health Care System: Retrospective Cohort Study
JMIR Form Res 2021;5(10):e26314
URL: https://formative.jmir.org/2021/10/e26314
doi:10.2196/26314
PMID:34617906

Original Paper

Web-Based Information Seeking Behaviors of Low-Literacy Hispanic Survivors of Breast Cancer: Observational Pilot Study

Francisco Iacobelli[1], PhD; Ginger Dragon[1], BSc; Giselle Mazur[1], BSc; Judith Guitelman[2], BSc

[1]Computer Science Department, Northeastern Illinois University, Chicago, IL, United States

[2]ALAS-WINGS, Chicago, IL, United States

Corresponding Author:
Francisco Iacobelli, PhD
Computer Science Department
Northeastern Illinois University
5500 N. St. Louis Ave.
Chicago, IL, 60625
United States
Phone: 1 7734424728
Email: f-iacobelli@neiu.edu

Abstract

Background: Internet searching is a useful tool for seeking health information and one that can benefit low-literacy populations. However, low-literacy Hispanic survivors of breast cancer do not normally search for health information on the web. For them, the process of searching can be frustrating, as frequent mistakes while typing can result in misleading search results lists. Searches using voice (dictation) are preferred by this population; however, even if an appropriate result list is displayed, low-literacy Hispanic women may be challenged in their ability to fully understand any individual article from that list because of the complexity of the writing.

Objective: This observational study aims to explore and describe web-based search behaviors of Hispanic survivors of breast cancer by themselves and with their caregivers, as well as to describe the challenges they face when processing health information on the web.

Methods: We recruited 7 Hispanic female survivors of breast cancer. They had the option to bring a caregiver. Of the 7 women, 3 (43%) did, totaling 10 women. We administered the Health LiTT health literacy test, a demographic survey, and a breast cancer knowledge assessment. Next, we trained the participants to search on the web with either a keyboard or via voice. Then, they had to find information about 3 guided queries and 1 free-form query related to breast cancer. Participants were allowed to search in English or in Spanish. We video and audio recorded the computer activity of all participants and analyzed it.

Results: We found web articles to be written for a grade level of 11.33 in English and 7.15 in Spanish. We also found that most participants preferred searching using voice but struggled with this modality. Pausing while searching via voice resulted in incomplete search queries, as it confused the search engine. At other times, background noises were detected and included in the search. We also found that participants formulated overly general queries to broaden the results list hoping to find more specific information. In addition, several participants considered their queries satisfied based on information from the snippets on the result lists alone. Finally, participants who spent more time reviewing articles scored higher on the health literacy test.

Conclusions: Despite the problems of searching using speech, we found a preference for this modality, which suggests a need to avoid potential errors that could appear in written queries. We also found the use of general questions to increase the chances of answers to more specific concerns. Understanding search behaviors and information evaluation strategies for low-literacy Hispanic women survivors of breast cancer is fundamental to designing useful search interfaces that yield relevant and reliable information on the web.

(JMIR Form Res 2021;5(10):e22809) doi:10.2196/22809

KEYWORDS

low literacy; health literacy; online searches; Hispanic breast cancer survivors

Introduction

Background

Internet searching has become an increasingly popular tool for patients to find health-related information [1], and it has been linked to improved health outcomes and greater patient engagement [2]. Despite efforts to mitigate the challenges that Hispanics face when seeking information on the web (guided user searches [3] and video, audio, and simplified text [4-7]), Hispanics still do not use web-based health information at the same rate as non-Hispanic White individuals do.

In this paper, we report an observational study that explores web-based information seeking behaviors (search and selection of results) of low-literacy Hispanic survivors of breast cancer and some of their caregivers when using either a voice- or text-based search engine. We describe behaviors that stress the difficulty of processing web-based information by this population as well as attributes inherent to the interface (traditional search engine or voice search engine) that make this task even more difficult and provide recommendations for future search interface designs.

Health Information Search Behaviors and Hispanics

The Health Information National Trend Survey has shown for several years that the internet is the most used source of health information [8]. Searching for health information on the web has been linked with improved health outcomes and greater patient engagement [2]. A study on US adults found that using the internet for health information has been strongly correlated with self-reports of very good or excellent health status, and the largest increase in health status has been observed in adults without a high school diploma [9]. However, education is also highly correlated with the kind of information favored by individuals. For example, adults with a high school diploma use more text-based sources (the dominant modality of web-based sources), whereas adults without a high school diploma use more verbal sources [9]. Consistent with this, another study of US adults' trends over 4 years revealed that education level was positively correlated with using the internet for health information and negatively correlated with using friends, family, and coworkers for health information [8].

Although this suggests that populations with lower literacy should benefit the most from web-based health information seeking, they do not search for web-based health information frequently.

As health information migrates to digital formats, Hispanics and other minorities are at a disadvantage, as more of them do not report using the internet as their source of information (except through surrogates) [10]. In particular, researchers found that in recent years (2011-2016), US-born Mexicans and foreign-born Hispanics do not use the internet for health information seeking or for sending emails to health providers as US-born non-Hispanic White people do [11,12]. Overall, Hispanic participants are more likely to use health care professionals as a source of health information compared with non-Hispanic participants. Moreover, being older, having low internet skills, and being Hispanic were determinants of using a health care provider or traditional media, such as print and magazines, as a source of health information versus using the web. Being Hispanic and having a history of cancer is highly correlated with using health care professionals as a primary source of health information [8]. These studies suggest that Hispanic survivors of breast cancer do not use the internet to find information.

Barriers to Accessing Web-Based Health Information

Hispanics of low socioeconomic status may be at an even further disadvantage of using web-based health information. Research that has tried to explain the barriers that low socioeconomic status individuals may encounter while trying to seek health information on the web has found that spotty internet access as well as frustration with the information search process are detrimental to seeking information on the web [13]. Web-based searching is a challenging task for low-literacy Hispanics. Over a decade ago, Birru et al [14] described problems with formulating queries, selecting and understanding results by low-literacy Hispanics who searched for information independently. As to what mode of internet search is favored by low-literacy adults, they tend to prefer voice searching (dictating search queries) to written searches when given the option [15]. This can be a strategy to mitigate common mistakes such as misspelling, misappropriation of words (writing lymphoma when they mean lymphedema), and incomplete search queries that can result in inadequate results and misleading information [14].

In addition, the complexity of information on the web is difficult to process for individuals with low literacy. Most internet health content is written at a level that is above the average reading level of adults in the United States [16-20]. For example, Walsh and Volsko [19] analyzed the reading levels of 100 publicly accessible articles related to the top leading causes of death in the United States—heart disease, stroke, cancer, chronic obstructive pulmonary disease, and diabetes—using three different readability assessment tools: Flesch-Kincaid, simple measure of gobbledygook, and frequency of gobbledygook. They found that the reading levels and comprehension of the articles consistently surpassed the average reading level in the United States, which is between seventh and eighth grade. Leroy et al [20] examined information from WebMD and MEDLINE and reported similar findings, with readability levels above 12th grade. This is also the case for health websites with information written in Spanish [21]. More specifically, research on information finding about breast cancer survivorship shows that overly complicated web-based information can negatively affect patients' care seeking and treatment decisions [22].

This is problematic, particularly for Hispanics. Hispanic adults in the United States have significantly lower literacy scores when compared with White adults with the same educational level. In terms of reading comprehension, Hispanics scored the lowest of any ethnic group in the United States [23]. When processing web-based information, research shows that, in general, low literacy and low health literacy are detrimental to an individual's ability to evaluate health information [17].

In general, many researchers state that strategies to increase Hispanics' access to internet health information will likely help

them become empowered and educated consumers, potentially having a favorable impact on health outcomes [24]. However, internet content has changed little to make this possible, and we believe it is important to understand the internet health information search behaviors of Hispanics to effect change.

Given the research cited here, Hispanic survivors of breast cancer fall into a segment of the population that tends to turn away from internet searches. To the best of our knowledge, the present pilot study is the first that does not rely on surveys or interviews to study Hispanics' health information search behaviors on the web but instead relies on observation.

Methods

Recruitment and Study Design

We recruited 7 women from a support group for Hispanic patients and survivors of breast cancer and from a pool of Hispanic women who had participated in other mobile health studies related to cancer education [25]. The women were in remission after diagnosis and treatment of breast cancer. As many of these women rely on their caregivers for information, we asked them to bring their caregivers to the experiment if desired; 43% (3/7) of women brought their caregivers. This resulted in 3 survivor-caregiver dyads and 4 individual survivors.

We asked the participants to talk about their experiences searching for information and the importance of web-based content and search abilities. After this short conversation, we proceeded to have them complete demographic information, the McArthur social mobility ladder [26], and the Health LiTT health literacy questionnaire. This is a short questionnaire that has been used previously in web-based settings with Hispanic women and was designed to address important attributes recommended by the Medical Outcomes Trust for multi-item measures of latent traits [27]. Finally, we asked them to fill out a 16-item breast cancer knowledge questionnaire used in previous mobile health interventions [25]. These questionnaires were given in the language of preference of the participant (English or Spanish).

Following previous research methodology [14], we proceeded to ask them to search for information they thought was relevant on several topics. Each search was given a maximum time of 10 minutes. In addition, we showed participants that they could search using their voice (using Google Chrome and the Google search engine with the option of voice search). For those who preferred to search in Spanish, we configured their search engines to understand Spanish and conducted the whole session in Spanish. After each search, we asked participants to switch to a note-taking application and write a sentence or two about the information they found interesting regarding the topic they were searching.

The first search was free form, and participants were directed to search for any topic they thought was important. We encouraged them to pick topics that may have come up in the questionnaires they had just answered. The purpose of this search was to allow participants to become accustomed to searching on the computers we provided and to switch back and forth from the web browser to the note-taking application. As research suggests that Hispanics prefer voice searches to written ones [15], participants had the chance to search via voice and via text. Once this task was completed and the participants felt comfortable using the computers, voice and written queries, and note-taking applications, the researcher proceeded to ask them to perform three more searches. At this point, the participants could choose whether to use voice searches or written ones. The topics of these additional searches were based on those that are highly correlated with the quality of life of Hispanic survivors of breast cancer [28] and that have come up on surveys and user studies [4]. The topics to search were (1) maintaining good spirits as a survivor of breast cancer, (2) affording treatment and medication, and (3) breast cancer and most common treatments. Consistent with previous research [14], we observed that our first 3 participants were formulating their searches almost verbatim from the researcher's prompt. Therefore, we added a fourth, more free-form search for the remaining participants: (4) search for any lingering issues they had related to survivorship.

After participants had finished searching, we debriefed and asked about their thoughts regarding their experience searching for these topics. These responses were audio recorded, whereas all computer screen activities were video and audio recorded.

Analysis

To analyze the participants' search activities and behaviors, we created a coding scheme with codes divided into four categories. (1) Web activity: in this category, we recorded clicks, clicks on advertisements, images, and clicks on a result. Tracking where users click and the number of clicks it takes a person to find information have traditionally been good indicators of search proficiency and interest in results [29,30]. In this category, we also tracked whether the searches were made by voice or typed as research indicates a different mindset—expectations and behaviors are associated with the modality of search [15]. (2) Participant behavior: here, we recorded whether participants read aloud, whether dyads discussed or talked to each other, or whether there were durations of silence without action; behaviors such as these are mechanisms frequently used by beginner or nonproficient readers for memorization and comprehension [31,32]. (3) Content: under this category, we coded the text of the query, the text of a note, and the webpage URLs they accessed. These allowed for qualitative examination of the search queries and notes taken and allowed us to trace and find information such as the readability level of the websites visited, trustworthiness, and whether they found answers to their queries. Tracking this is important, as readability is important to understand obstacles to low literacy information seekers [33,34], and it can lead to an information *rabbit hole* where users would not find answers to their original query [35]. (4) Information retrieval–related issues: In this category, we tracked misspellings, misappropriations (the wrong word for a given term; for example, lymphomas instead of lymphedema), and speech recognition misunderstandings. All these have been documented as barriers for low-literacy populations when searching on the web [14]. We used the initial free-form searches to train our coders and establish the reliability of the coding scheme. We obtained interrater reliability of k=0.89.

However, because of the small sample size, we mostly report descriptive statistics.

Results

Participant Characteristics

The average age of the participants and caregivers was 57.7 (SD 9.9) years. Of the 10 participants, 3 (30%) participants had less than high school education, 2 (20%) had high school diplomas or equivalent, and 5 (50%) had some college education. Of the 10 participants, 3 (30%) were caregivers: 2 (20%) were daughters of the survivor, and 1 (10%) was a friend. None of the participants had a college diploma or higher. In terms of health literacy scores in the Health LiTT test, measured as the proportion of items correct, the minimum score obtained by a participant was 21.4%, the maximum was 50%, and the mean was 37.5% (SD 9.9%). All the women considered themselves Hispanic. Approximately 80% (8/10) of women reported that they felt very comfortable speaking Spanish, and only 40% (4/10) felt very comfortable in English.

In terms of knowledge about breast cancer, in one dyad, the caregiver knew more about breast cancer than the survivor. The average score on the knowledge of breast cancer questionnaire among these women was 66.88%. The 2 caregivers who were daughters of the survivors scored above the mean score, whereas the caregiver who was a friend of the survivor scored very low. Table 1 shows the scores of each participant in the breast cancer knowledge test.

Table 1. Breast cancer knowledge scores of survivors and caregivers.

Dyad per participant	Caregiver score, % (relationship to survivor)	Survivor score (%)
1	75 (daughter)	50
2	43.8 (friend)	75
3	N/A[a]	62.5
4	N/A	62.5
5	N/A	93.8
6	N/A	68.8
7	68.8 (daughter)	68.8

[a]N/A: individual participated alone.

Query Formulation

For each of the searches, we tracked whether they were done via speech (spoken query) or writing (written query). Tables 2-5 show the attempts by our users and whether the search was spoken or written. When a cell has multiple lines, it denotes multiple queries for the same search task. The modality of each query is at the end of each query (S: spoken query; W: written query). The queries in Spanish have their corresponding translations in italics and were provided by one of the researchers. All the original text has been maintained as typed or as transcribed by the speech-to-text Google engine.

As can be seen from these searches, most participants (and dyads) use speech to search for one point or another. In five of the seven sessions, the participants used spoken searches several times before reaching the desired results. Figure 1 shows the number of written searches versus voice searches per participant or dyad.

Table 2. Search queries used by participants on the first search.

ID	Searches about breast cancer and most common treatments	Individual or dyad
1	• What is breast cancer (S[a]); • The whole thing (S); • What is breast cancer in the most common treatment (S); • What is breast cancer and the most common treatments (W[b])	Dyad
2	• What is breast cancer and better treatment (S)	Dyad
3	• Que es el cancer de seno y sus mejores tratamientos? (W; what is breast cancer and its best treatments?)	Individual
4	• What is breast cancer and the most common (S) • What is breast cancer and the most common treatment(S)	Individual
5	• Qué es el cáncer de mama (S; what is breast cancer)	Individual
6	• What is breast cancer and the most common treatments (S); • What is breast cancer (S)	Individual
7	• Cancer de mama y sus tratamientos mas comunes (W; breast cancer and its most common treatments)	Dyad

[a]S: spoken query.

[b]W: written query.

Table 3. Search queries used by participants on the second search.

ID	Searches about maintaining good spirits as a breast cancer survivor	Individual or dyad
1	• No surviving breast cancer (S[a]); • Positive outlook for breast cancer survival (W[b]); • How to have a positive attitude about cancer (S); • omo mantener el animo positivo (W; [how] to maintain good spirits) • Como mantener el animo positivo sobreviviente (W; how to maintain good spirits survivor) • Como mantener el animo positivo sobreviviente de cancer de mama (W; how to maintain good spirits survivor breast cancer) • Como animar a alguien que tuvo cancer (W; how to cheer up someone who had cancer) • Maintaining a positive outlook after cancer (W)	Dyad
2	• Breast cancer survivor (S); • How to maintain my humor (S); • How to maintain a good humor after being a breast cancer survivor (S)	Dyad
3	• Cómo mantener un buen ánimo Después (S; how to maintain good spirits after) • Cómo mantener el ánimo después del cáncer de seno (S; how to maintain good spirits after breast cancer) • Cómo mantener el ánimo después de los del cáncer de seno; (S; how to maintain good spirits after of the breast cancer) • Cómo mantener el ánimo después del cáncer de seno (S; how to maintain good spirits after breast cancer)	Individual
4	• How to maintain good spirits (S); • Here's a summary from URMC[c] University (S); • How to maintain good spirits as a breast cancer survivor (S)	Individual
5	• Cómo tener buen ánimo para cel- (S; how to keep good spirits for cel-[sic]) • Cómo es mantener buen ánimo para ser sobreviviente de cáncer de mama (S; how is it to maintain good spirits to be a survivor of breast cancer)	Individual
6	• How do I maintain a good spirit after breast cancer (S)	Individual
7	• Como mantener buen animo siendo sobreviviente de mama (W; how to maintain good spirits being a breast survivor)	Dyad

[a]S: spoken query.

[b]W: written query.

[c]URMC: University of Rochester Medical Center.

Table 4. Search queries used by participants on the third search.

ID	Searches about affording treatment and medication	Individual or dyad
1	• How to afford cancer treatment and medication (W[a]) • Cancer patient assistance programs (W)	Dyad
2	• How could someone (S[b]) • Height of someone (S); • How could someone afford (S); • How can someone afford (S); • How can someone afford medication (S); • Half of someone afford (S); • How can someone afford medication (S); • How can someone afford medication cancer and treatment (W)	Dyad
3	• Como pagar tratamientos y medicamentos para el cancer de seno? (W; how to afford treatment and medication for breast cancer?)	Individual
4	• How can someone afford treatment and medication (S)	Individual
5	• Hay organisaciones que ayunan para el tratamiento de mama (W; are there organizations [misspelled] that fast [misspelled help] for the treatment of breast)	Individual
6	• Is there an affordable way (S); • Is there an affordable way for breast cancer treatment (S)	Individual
7	• Que opciones hay para pagar tratamientos del cancer de mama (W; what options are there to afford treatment of breast cancer)	Dyad

[a]W: written query.

[b]S: spoken query.

Table 5. Search queries used by participants on the fourth search.

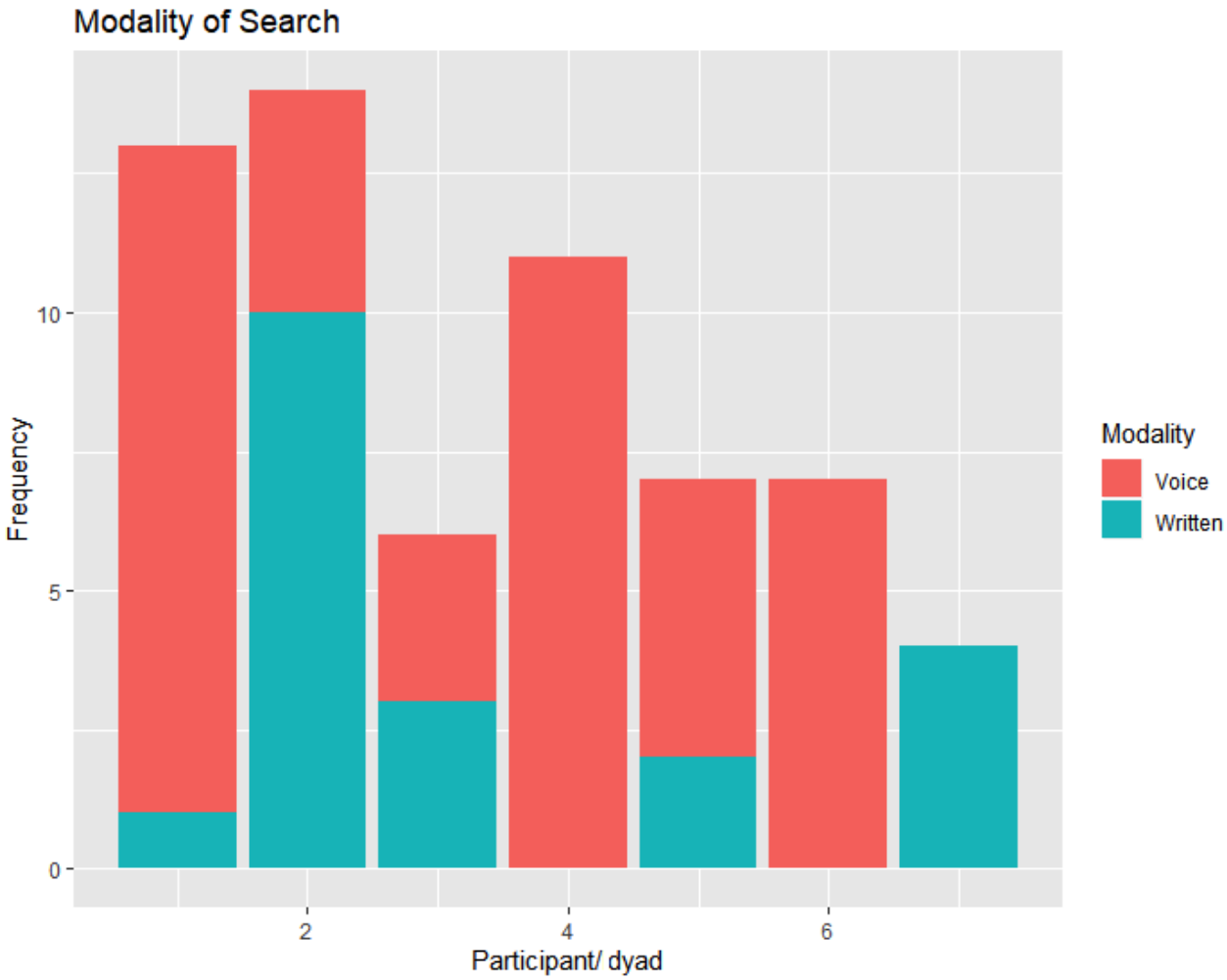
ID	Searches about any lingering issues they had related to survivorship	Individual or dyad
1	• N/A[a]	Dyad
2	• N/A	Dyad
3	• N/A	Individual
4	• After 5 years of a breast cancer survivor, can breast cancer come back (S[b])	Individual
5	• Como yo puedo saber que mi cancer no va alvolver (W[c]; how can I know if my cancer is not coming back [misspelled])	Individual
6	• After cancer treatment (S); • After tonsil cancer treatment (S); • Here's some information for the treatment of tonsil cancer according to cancer research UK (S); • Tonsil cancer treatment after surgery combined with chemotherapy (S); • Once cancer treatment is over are you considered in (S)	Individual
7	• Me podria regresar el cancer (W; could my cancer come back?)	Dyad

[a]N/A: not applicable.

[b]S: spoken query.

[c]W: written query.

Figure 1. Number of voice versus written search queries per participant per dyad.



We also noticed several search queries that lacked content. For example, “How can someone afford treatment and medication (S)” or “How to maintain good spirits survivor (W).” The results were not necessarily focused on breast cancer because of the lack of context. Moreover, on several occasions, participants clicked to search by voice, and the search engine picked up speech that was not part of the intended search. These included (1) spoken information from a previous result set that Google was still reading when the participants started a new search (eg, “Here’s a summary from URMC University” and “here’s some information for treatment of tonsil cancer according to cancer research UK”); (2) making comments while the computer was listening (eg, “the whole thing”); or (3) not completing their utterance before the search engine started retrieving results, which led to several searches being performed until a query was completely articulated (eg, participant 2 [dyad] in two occasions; Table 4).

Participants reformulated their queries by adding search terms, by switching between spoken or written searches, or, in one case (participant 1 on search 2; Table 3), switching from English to Spanish, trying to obtain adequate results.

It is important to note the behavior of participant or dyad 7. They used only written searches and only one attempt at searching. This was a dyad where the caregiver told us that the survivor was unable to read and trusted her caregiver with finding information. They also arrived late at the experiment and were somewhat constrained by time.

Readability of Websites Chosen

We kept track of the websites where users obtained information to later take notes they considered important. We submitted the text of the websites in English to a Fleish-Kinkaid analysis. However, Fleisch-Kinkaid does not reflect a correct grade level in Spanish texts, mainly because of the difference in the average number of syllables in a word between English and Spanish. Therefore, for websites in Spanish, we used the Gilliam et al [36] adaptation of the Fry graph for readability (FGR), which has been validated in previous research [21]. As the length and number of sentences are one of the main components of the FGR, we excluded titles and lists that make sentences artificially short and selected a sample of the first two paragraphs of each article in Spanish. The average grade level of the websites in Spanish was 7.15 (SD 0.83). Table 6 shows the websites visited and their average reading grade levels. Websites in Spanish are marked next to their readability grade level. The average is shown separately for the English and Spanish websites.

Table 6. Readability scores for websites visited.

Website	Reading grade level
Cancer.net	12.22
Cancer.org	12.08
Komen.org	10.35
Cdc.gov	9.44
Verywellhealth.com	12.51
Lynparzahcp.com	19.28
Ibtimes.com	14.1
Webmd.com	7.78
Chemocare.com	11.2
Prescriptionhope.com	10.56
Cancer.gov	9.07
Mayoclinic.org	8[a]
Roche.com	8[a]
Cancer.gov	6[a]
Vidaysalud.com	7[a]
Cancer.org	7[a]
Unimiamihealth.org	8[a]
Medlineplus.gov	6[a]
Cdc.gov	7[a]
Overall total, mean (SD)	
English	11.33 (3.01)
Spanish	7.15 (0.83)[a]

[a]Websites in Spanish.

The mean grade level readability was 11.33 (SD 3.01) for all the articles in English. All but one of the articles had a readability score below the sixth-grade level. The average readability exceeded the recommended readability (sixth grade) by 5.33 grade levels. Moreover, the readability of the articles exceeded the eighth-grade level by an average of 3.33 grade levels. The only article with readability below the sixth-grade level was one of the articles from WebMD with readability (F-K)=5.8.

The mean grade level readability of the Spanish language articles was 7.15 (SD 0.83), with articles ranging from sixth to eighth grade. This suggests a higher grade level readability for articles in English. Moreover, in one article from cdc.gov, we found an English version and its manual translation into Spanish. The Spanish version from cdc.gov [37] yielded a grade level of seventh grade readability using the Gilliam et al adaptation of the FGR. However, when accessing the translation in English [38], it yielded 14th grade level using the FGR.

Answers to Participants' Questions

When analyzing participants' notes on each search, we found that although many typed pertinent answers, some copied verbatim from snippets in the results lists, resulting in notes with mixed contextual information such as, "Breast cancer is a tumor or mass. Treaty [*sic*] by chemo, mastectomy or Lumpectomy. Radiation therapy." Others copied and pasted, resulting in notes including URLs and characters that had to do with the formatting of the web pages or notes ending with the start of a new topic: "[...] despues de cinco anos de estar libre de cancer, el viaje aun no termina>> tener los cuidados necesarios, a largo plazo" ("[...] after five years of being cancer free, the journey is not over yet>>having appropriate care in the long term").

In addition, 2 participants did not find any information with respect to the questions they posed. However, the other participants who visited the same website did. Finally, our coders determined whether the participants answered the questions they posed and found that on eight occasions, the participants did not. Most notably, none of the participants answered their original question on the fourth search, which was free form. For example, participant 4's question was, "After 5 years of breast cancer survivor, can breast cancer come back," and her written answer was, "You should keep doing breast

self-exams checking the treated area and your other breast exams."

Satisfactory information was mostly found after clicking, on average, between 1 and 2 websites. However, some users found the information they needed straight from the snippets of text in the results list. In particular, participant 6 never visited a website to answer her questions. Table 7 shows the number of websites visited before the participants wrote their notes to answer their queries.

Table 7. Number of websites visited before writing down useful information.

Participant	Search 1[a]	Search 2[b]	Search 3[c]	Search 4[d]
1	2	3	2	N/A[e]
2	1	1	1	N/A
3	4	3	1	N/A
4	2	1	1	1
5	1	0	0	1
6	0	0	0	0
7	0	1	1	1

[a]Search 1: average 1.4 (SD 1.39).

[b]Search 2: average 1.3 (SD 1.25).

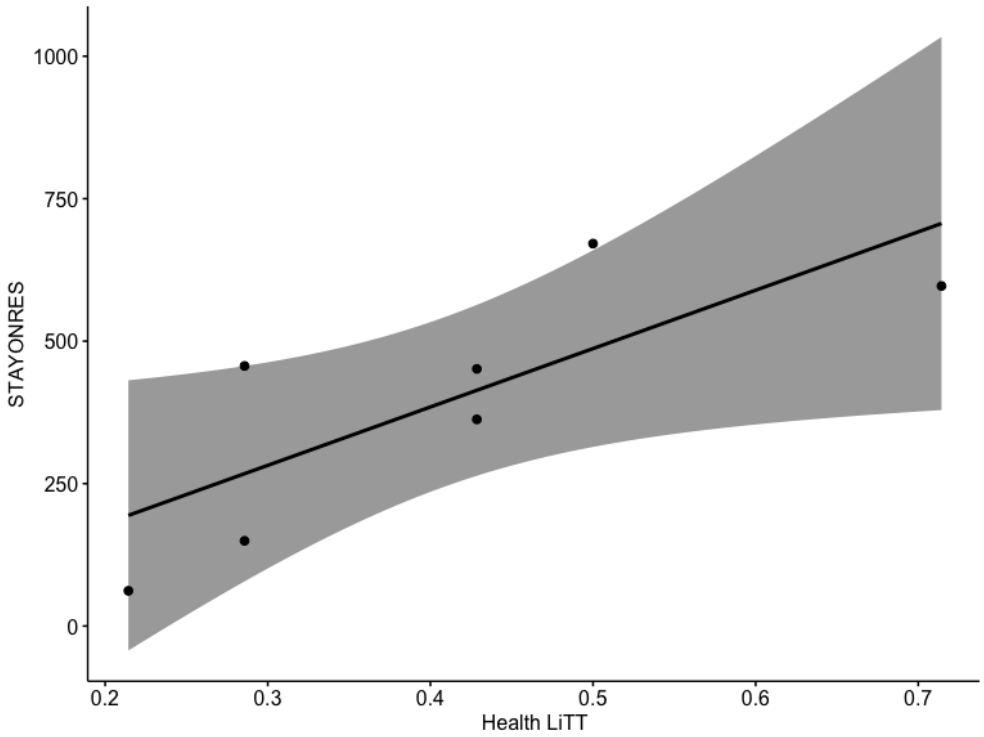
[c]Search 3: average 0.9 (SD 0.7).

[d]Search 4: average 0.75 (SD 0.5).

[e]N/A: not applicable.

Our participants spent, on average, 25.5 seconds browsing result lists and 52.9 seconds on actual articles (STAYONRES). However, in 3 of the 7 sessions, participants spent more time browsing the search's result list than reading information pages. To further explore whether health literacy could be related to the amount of time spent on individual articles, which are comprised of significantly more health-related content than snippets in a list of results, we plotted the Health LiTT scores versus STAYONRES. Figure 2 suggests a potential correlation with this small sample size, which may be worth exploring in future work.

Figure 2. Exploratory correlation of Health LiTT scores as predictors of time (in seconds) spent on individual articles (STAYONRES); R=0.78; *P*=.04. STAYONRES: staying on the results page; Health Litt: Health Literacy Assessment Using Talking Touchscreen Technology.



Discussion

Principal Findings

In this observational study, we set out to review the search behaviors of Hispanic survivors of breast cancer when examining health-related content. To our knowledge, this study is unique in that it focuses on low-literacy Hispanic survivors of breast cancer and examines searching (1) in their language of preference, (2) using voice as an alternative to writing search queries, and (3) with regard to health literacy and prior knowledge of breast cancer.

None of our participants used isolated search terms (keywords) to formulate queries, as was done in previous studies [14,39]. Instead, they all attempted to use full sentences, whereas on a few occasions inserting the terms *cancer* or *breast cancer* for context. Often, participants searched for a given problem without specifying the context of the search and, thus, obtained less than satisfactory results lists. In particular, and as Birru et al [14] found, search queries were often a verbatim transcription of the prompt the researchers gave the participants. However, in this study, the last search performed was such that the participants needed to find any information interesting to them as naturally as they would search at home. This resulted in mostly fully formulated questions, as opposed to isolated search terms. Perhaps the familiarity with technology has increased in our population to the point that they understand that the interfaces now respond well to full natural language queries.

In terms of results, we found that in agreement with studies from over 15 years ago [21,39], the grade level of the information displayed on the web is far above what most participants are prepared to read.

To find the results, in several instances, participants simply grabbed content from the lists of results. This can be because of their familiarity with their condition and as they may already know some of the answers. However, it can also be as the snippets in the list of results are simpler to read than full-fledged articles, given their low literacy and health literacy levels. This may pose a danger of finding snippets of information without the appropriate context to interpret them adequately. Despite the information being readily available in the articles visited, some participants still struggled to find something useful even when others did. This, again, can be because of a lack of comprehension or novelty of the information. That is, as the patient may already know the basic answers displayed in the results, they may have determined that the information displayed was not useful. It is interesting to note that on the fourth search, which was free form, all patients asked about their cancer coming back but took notes that did not answer the question directly. All searches asked whether cancer could come back. However, their notes were about the continuing care they should take as survivors; that is, their notes were related to the question but not to a direct answer. This could indicate that they have difficulty formulating a question that captures their exact concerns; instead, they ask a more general question hoping to find a detailed answer that resonates with their direct concern. Perhaps their intention was to know how to monitor and prevent the recurrence of breast cancer (which is what all the notes were about).

Search Modality

Although some participants preferred written searches, most used the spoken search capabilities to a large extent. However, when using the spoken searches, the search engine detected pauses (as to think what to say) as a sign that the query had finished and retrieved results with an incomplete query. For example, “how could someone” was a search term in which the results were not related to breast cancer survivorship. At other times, when faced with a spoken search that retrieved no good results at first, participants wrote the same query hoping that they would get a different results list. However, overall, participants persisted in searching via voice. In conversations after their experiment, they all expressed a desire for it to work and said it was very useful.

Limitations and Future Work

The main limitation of this study is the number of participants. More participants will certainly add strength to some of our intuitions and may result in strong patterns. For any quantitative analysis to find correlations and significant statistical effects, a power analysis reveals that for the correlations to be ≥0.5, with 80% power and P=.05, we would need 18 participants and 18 dyads (to account for correlations of dyads only). A second limitation is that although each participant received training on the use of the computer, technological fluency was not assessed or controlled for. A third limitation is that the notes taken are not necessarily a direct reflection on the comprehension of the texts. Perhaps more subtle measurements of comprehension can be obtained after each search, or we could simulate an urgency to find appropriate results, as this has been documented to enhance and focus on the use of search engines [40]. Finally, not all participants used these kinds of tools on the web (voice search or internet search). Instead, some let their caregivers search on the web and tell them what to do. Therefore, it is important to have an adequate number of caregivers in future research.

Acknowledgments

This research was funded by the Student Center for Science Engagement and the Community Organized Research programs at Northeastern Illinois University, The Chicago Cancer Health Equity Collaborative (Chicago) and National Institutes of Health Diversity Supplement 3U54CA202995.

Conflicts of Interest

None declared.

References

1. Tonsaker T, Bartlett G, Trpkov C. Health information on the Internet: gold mine or minefield? Can Fam Physician 2014 May;60(5):407-408 [FREE Full text] [Medline: 24828994]
2. Suziedelyte A. How does searching for health information on the internet affect individuals' demand for health care services? Soc Sci Med 2012 Nov;75(10):1828-1835. [doi: 10.1016/j.socscimed.2012.07.022] [Medline: 22884947]
3. Bickmore TW, Utami D, Matsuyama R, Paasche-Orlow MK. Improving access to online health information with conversational agents: a randomized controlled experiment. J Med Internet Res 2016 Jan 04;18(1):e1 [FREE Full text] [doi: 10.2196/jmir.5239] [Medline: 26728964]
4. Iacobelli F, Adler RF, Buitrago D, Buscemi J, Corden ME, Perez-Tamayo A, et al. Designing an mHealth application to bridge health disparities in Latina breast cancer survivors: a community-supported design approach. Design Health (Abingdon) 2018 Apr 02;2(1):58-76 [FREE Full text] [doi: 10.1080/24735132.2018.1452871] [Medline: 30506017]
5. Baik SH, Oswald LB, Buscemi J, Buitrago D, Iacobelli F, Perez-Tamayo A, et al. Patterns of use of smartphone-based interventions among Latina breast cancer survivors: secondary analysis of a pilot randomized controlled trial. JMIR Cancer 2020 Dec 08;6(2):e17538 [FREE Full text] [doi: 10.2196/17538] [Medline: 33289669]
6. Ginossar T, Shah S, West A, Bentley J, Caburnay C, Kreuter M, et al. Content, usability, and utilization of plain language in breast cancer mobile phone apps: a systematic analysis. JMIR Mhealth Uhealth 2017 Mar 13;5(3):e20 [FREE Full text] [doi: 10.2196/mhealth.7073] [Medline: 28288954]
7. Magnani JW, Schlusser CL, Kimani E, Rollman BL, Paasche-Orlow MK, Bickmore TW. The atrial fibrillation health literacy information technology system: Pilot assessment. JMIR Cardio 2017 Dec 12;1(2):e7 [FREE Full text] [doi: 10.2196/cardio.8543] [Medline: 29473644]
8. Jacobs W, Amuta AO, Jeon KC. Health information seeking in the digital age: An analysis of health information seeking behavior among US adults. Cogent Soc Sci 2017 Mar 13;3(1):1302785. [doi: 10.1080/23311886.2017.1302785]
9. Feinberg I, Frijters J, Johnson-Lawrence V, Greenberg D, Nightingale E, Moodie C. Examining associations between health information seeking behavior and adult education status in the U.S.: An analysis of the 2012 PIAAC data. PLoS One 2016;11(2):e0148751 [FREE Full text] [doi: 10.1371/journal.pone.0148751] [Medline: 26882339]
10. Massey PM. Where do U.S. adults who do not use the internet get health information? Examining digital health information disparities from 2008 to 2013. J Health Commun 2016 Nov 23;21(1):118-124. [doi: 10.1080/10810730.2015.1058444] [Medline: 26166484]
11. Gonzalez M, Sanders-Jackson A, Wright T. Web-based health information technology: Access among Latinos varies by subgroup affiliation. J Med Internet Res 2019 Apr 16;21(4):e10389 [FREE Full text] [doi: 10.2196/10389] [Medline: 30990462]
12. Gonzalez M, Sanders-Jackson A, Emory J. Online health information-seeking behavior and confidence in filling out online forms among Latinos: a cross-sectional analysis of the California Health Interview Survey, 2011-2012. J Med Internet Res 2016 Jul 04;18(7):e184 [FREE Full text] [doi: 10.2196/jmir.5065] [Medline: 27377466]
13. McCloud RF, Okechukwu CA, Sorensen G, Viswanath K. Beyond access: barriers to internet health information seeking among the urban poor. J Am Med Inform Assoc 2016 Nov;23(6):1053-1059 [FREE Full text] [doi: 10.1093/jamia/ocv204] [Medline: 27206459]
14. Birru M, Monaco V, Charles L, Drew H, Njie V, Bierria T, et al. Internet usage by low-literacy adults seeking health information: an observational analysis. J Med Internet Res 2004 Sep 03;6(3):e25 [FREE Full text] [doi: 10.2196/jmir.6.3.e25] [Medline: 15471751]
15. Guy I. The characteristics of voice search. ACM Trans Inf Syst 2018 Apr 25;36(3):1-28. [doi: 10.1145/3182163]
16. Neuhauser L, Kreps GL. Online cancer communication: meeting the literacy, cultural and linguistic needs of diverse audiences. Patient Educ Couns 2008 Jun;71(3):365-377. [doi: 10.1016/j.pec.2008.02.015] [Medline: 18424046]
17. Diviani N, van den Putte B, Giani S, van Weert JC. Low health literacy and evaluation of online health information: a systematic review of the literature. J Med Internet Res 2015 May 07;17(5):e112 [FREE Full text] [doi: 10.2196/jmir.4018] [Medline: 25953147]
18. Macias W, Lee M, Cunningham N. Inside the mind of the online health information searcher using think-aloud protocol. Health Commun 2018 Dec 27;33(12):1482-1493. [doi: 10.1080/10410236.2017.1372040] [Medline: 28952804]
19. Walsh TM, Volsko TA. Readability assessment of internet-based consumer health information. Respir Care 2008 Oct;53(10):1310-1315 [FREE Full text] [Medline: 18811992]
20. Leroy G, Helmreich S, Cowie JR, Miller T, Zheng W. Evaluating online health information: beyond readability formulas. AMIA Annu Symp Proc 2008 Nov 06:394-398 [FREE Full text] [Medline: 18998902]
21. Berland GK, Elliott MN, Morales LS, Algazy JI, Kravitz RL, Broder MS, et al. Health information on the Internet: accessibility, quality, and readability in English and Spanish. J Am Med Assoc 2001;285(20):2612-2621 [FREE Full text] [doi: 10.1001/jama.285.20.2612] [Medline: 11368735]
22. Fefer M, Lamb CC, Shen AH, Clardy P, Muralidhar V, Devlin PM, et al. Multilingual analysis of the quality and readability of online health information on the adverse effects of breast cancer treatments. JAMA Surg 2020 Aug 01;155(8):781-784 [FREE Full text] [doi: 10.1001/jamasurg.2020.1668] [Medline: 32520317]

23. Program for the International Assessment of Adult Competencies (PIAAC). National Center for Education Statistics. 2017. URL: https://nces.ed.gov/surveys/piaac/current_results.asp [accessed 2020-02-27]
24. Peña-Purcell N. Hispanics' use of Internet health information: an exploratory study. J Med Libr Assoc 2008 Apr;96(2):101-107 [FREE Full text] [doi: 10.3163/1536-5050.96.2.101] [Medline: 18379664]
25. Yanez BR, Buitrago D, Buscemi J, Iacobelli F, Adler RF, Corden ME, et al. Study design and protocol for My Guide: An e-health intervention to improve patient-centered outcomes among Hispanic breast cancer survivors. Contemp Clin Trials 2018 Feb;65:61-68 [FREE Full text] [doi: 10.1016/j.cct.2017.11.018] [Medline: 29198729]
26. Adler N, Stewart J. The MacArthur Scale of Subjective Social Status. The Psychosocial Working Group. 2007. URL: http://www.macses.ucsf.edu/research/psychosocial/subjective.php [accessed 2020-02-27]
27. Hahn E, Choi S, Griffith J, Yost K, Baker D. Health literacy assessment using talking touchscreen technology (Health LiTT): a new item response theory-based measure of health literacy. J Health Commun 2011;16 Suppl 3:150-162 [FREE Full text] [doi: 10.1080/10810730.2011.605434] [Medline: 21951249]
28. Lopez-Class M, Perret-Gentil M, Kreling B, Caicedo L, Mandelblatt J, Graves KD. Quality of life among immigrant Latina breast cancer survivors: realities of culture and enhancing cancer care. J Cancer Educ 2011 Dec;26(4):724-733 [FREE Full text] [doi: 10.1007/s13187-011-0249-4] [Medline: 21706194]
29. Downey D, Dumais S, Liebling D, Horvitz E. Understanding the relationship between searchers' queries and information goals. In: Proceedings of the 17th ACM Conference on Information and Knowledge Management. 2008 Presented at: 17th ACM Conference on Information and Knowledge Management; Oct. 26, 2008; New York p. 449-458. [doi: 10.1145/1458082.1458143]
30. Saba J. More readers skimming Google headlines than going directly to newspaper web sites? Editor and Publisher Magazine. 2010 Jan. URL: https://www.editorandpublisher.com/stories/more-readers-skimming-google-headlines-than-going-directly-to-newspaper-web-sites,139278? [accessed 2021-09-12]
31. Forrin ND, MacLeod CM. This time it's personal: the memory benefit of hearing oneself. Memory 2018 Apr 02;26(4):574-579. [doi: 10.1080/09658211.2017.1383434] [Medline: 28969489]
32. Hale AD, Skinner CH, Williams J, Hawkins R, Neddenriep CE, Dizer J. Comparing comprehension following silent and aloud reading across elementary and secondary students: Implication for curriculum-based measurement. Behav Anal Today 2007;8(1):9-23. [doi: 10.1037/h0100101]
33. Johnson AR, Doval AF, Egeler SA, Lin SJ, Lee BT, Singhal D. A multimetric evaluation of online Spanish health resources for lymphedema. Ann Plast Surg 2019 Mar;82(3):255-261. [doi: 10.1097/SAP.0000000000001762] [Medline: 30730864]
34. Friedman D, Hoffman-Goetz L. A systematic review of readability and comprehension instruments used for print and web-based cancer information. Health Educ Behav 2006 Jun;33(3):352-373. [doi: 10.1177/1090198105277329] [Medline: 16699125]
35. Makri S, Buckley L. Down the rabbit hole: Investigating disruption of the information encountering process. J Assoc Inf Sci Technol 2019 Apr 11;71(2):127-142. [doi: 10.1002/asi.24233]
36. Gilliam B, Peña S, Mountain L. The fry graph applied to Spanish readability. Read Teach 1980 Jan;33(4):426-430 [FREE Full text]
37. ¿Cómo se trata el cáncer de mama? Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/spanish/cancer/breast/basic_info/treatment.htm [accessed 2020-02-10]
38. How is breast cancer treated? Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/cancer/breast/basic_info/treatment.htm [accessed 2020-02-01]
39. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. Br Med J 2002 Mar 09;324(7337):573-577 [FREE Full text] [doi: 10.1136/bmj.324.7337.573] [Medline: 11884321]
40. Case DO, Given LM. In: Case DO, editor. A Survey of Research on Information Seeking, Needs, and Behavior. 4th Ed. United Kingdom: Emerald Group Publishing; Apr 29, 2016:1-528.

Abbreviations

FGR: Fry graph for readability

Edited by G Eysenbach; submitted 23.07.20; peer-reviewed by T Ginossar, C Bakker; comments to author 24.09.20; revised version received 05.01.21; accepted 01.08.21; published 27.10.21.

Please cite as:
Iacobelli F, Dragon G, Mazur G, Guitelman J
Web-Based Information Seeking Behaviors of Low-Literacy Hispanic Survivors of Breast Cancer: Observational Pilot Study
JMIR Form Res 2021;5(10):e22809
URL: https://formative.jmir.org/2021/10/e22809
doi:10.2196/22809
PMID:34704952

Original Paper

Peer-to-Peer Social Media Communication About Dietary Supplements Used for Weight Loss and Sports Performance Among Military Personnel: Pilot Content Analysis of 11 Years of Posts on Reddit

Kendall J Sharp[1,2], BA, MPH; Julia A Vitagliano[2], BA; Elissa R Weitzman[2,3], MSci, SCD; Susan Fitzgerald[2], RN, MSN, PPCNP-BC; Suzanne E Dahlberg[2,3], PhD; S Bryn Austin[1,2,3], SCD

[1]Department of Social and Behavioral Sciences, Harvard T. H. Chan School of Public Health, Boston, MA, United States

[2]Division of Adolescent and Young Adult Medicine, Boston Children's Hospital, Boston, MA, United States

[3]Department of Pediatrics, Harvard Medical School, Boston, MA, United States

Corresponding Author:
Kendall J Sharp, BA, MPH
Department of Social and Behavioral Sciences
Harvard T. H. Chan School of Public Health
677 Huntington Ave
Boston, MA, 02115
United States
Phone: 1 5129837561
Email: kendall.sharp@utsouthwestern.edu

Abstract

Background: Over 60% of military personnel in the United States currently use dietary supplements. Two types of dietary supplements, weight loss and sports performance (WLSP) supplements, are commonly used by military personnel despite the associated serious adverse effects such as dehydration and stroke.

Objective: To understand peer-to-peer communication about WLSP supplements among military personnel, we conducted a pilot study using the social media website, Reddit.

Methods: A total of 64 relevant posts and 243 comments from 2009 to 2019 were collected from 6 military subreddits. The posts were coded for year of posting, subreddit, and content consistent with the following themes: resources about supplement safety and regulation, discernability of supplement use through drug testing, serious adverse effects, brand names or identifiers, and reasons for supplement use.

Results: A primary concern posted by personnel who used supplements was uncertainty about the supplements that were not detectable on a drug test. Supplements to improve workout performance were the most frequently used.

Conclusions: Our pilot study suggests that military personnel may seek out peer advice about WLSP supplements on Reddit and spread misinformation about the safety and effectiveness of these products through this platform. Future directions for the monitoring of WLSP supplement use in military personnel are discussed.

(JMIR Form Res 2021;5(10):e28957) doi:10.2196/28957

KEYWORDS

dietary supplements; social media; Reddit; OPSS

Introduction

The dietary supplement market in the United States is enormous and steadily growing, with a projected $56.7 billion in annual revenue by 2024 [1]. Supplements are widely available both online and in brick-and-mortar retail outlets across the country. Recent estimates suggest that approximately 52% of the adult population [2] and 60%-70% of military personnel [3] in the United States have used dietary supplements. A meta-analysis by Santos et al [4] found that approximately 10% of the adult population in the United States has used dietary supplements specifically for weight loss [4]. Many dietary supplements are

marketed with claims that these suppress appetite, reduce fat absorption, increase the body's ability to burn fat through increased metabolism, and decrease water weight [4]. Sports performance supplements are marketed with claims that these increase energy, improve sports or workout performance, and build muscle. These performance supplements are popular among military personnel, who have reported higher use of these types of supplements than civilians [5]. In a random sample of soldiers on active duty in the army, Lieberman et al [5] found that 25% of soldiers who reported dietary supplement use did so to increase their muscle strength and 17% did so to enhance their physical performance [5].

The Dietary Supplement Health and Education Act of 1994 allows the US Food and Drug Administration (FDA) to loosely regulate supplements, but the FDA cannot mandate rigorous testing and reporting of the efficacy, safety, or purity of dietary supplements before these are placed on the market [6]. These weak regulatory conditions have led to companies marketing dietary supplements with inaccurate efficacy claims and using undisclosed proprietary formulas without rigorous premarket testing [6].

Weight loss and sports performance (WLSP) supplements have been associated with a myriad of adverse health effects, including dehydration, electrolyte imbalance, digestive dysfunction, stroke, renal failure, and death [7-11]. A study of preworkout supplement use among adults in the United States found that over half of the participants who used these supplements reported adverse effects such as lightheadedness, rapid heart rate, and nausea [12]. Despite the large market and well-documented health risks, there is no national systematic surveillance to assess the prevalence of use or adverse effects of these products in the general population.

Military personnel are particularly vulnerable to using these products and adopting other dangerous weight control behaviors because of their rigorous training regimens and long work hours and the need to reach a certain weight to be eligible for military service [13,14]. Dietary supplement use has been associated with serious adverse effects, including death, in military personnel [15,16]. For instance, Eliason and colleagues [15] reported the cases of 2 soldiers who died after ingesting dietary supplements containing 1,3-dimethylamylamine, a powerful stimulant that is a derivative of amphetamine [15]. The use of supplements containing ephedrine, which can have similar life-threatening effects, has long concerned the air force and other branches of the military [17]. Ephedrine was routinely advertised to increase energy and weight loss, but its use in dietary supplements was prohibited by the FDA in 2004 [18]. To circumvent the FDA ban on dietary supplements containing ephedrine, some consumers create an ephedrine, caffeine, and aspirin (ECA) supplement mix, colloquially referred to as an ECA or EC (ephedrine and caffeine) stack, and often combine over-the-counter drugs and supplements or food sources as ingredients for this mix [19]. ECA stacks are commonly used as preworkout supplements, with the expectation that this mix will increase performance during workouts or reduce body fat—colloquially referred to as "cutting" fat—and lead to weight loss.

In 2012, the US Department of Defense (DoD) created the Operation Supplement Safety (OPSS) campaign to increase awareness of the dangers of dietary supplement use [20]. The OPSS website was created to educate military personnel about the risk of supplement use and to address questions that personnel may have in general on dietary supplements and on supplement safety and legality. The OPSS website also has a portal for voluntary reporting of adverse effects experienced after dietary supplement use. Despite the serious risk to service members and troop readiness and the concerning ad hoc adverse effects reported through the OPSS website, the DoD does not conduct routine systematic surveillance of the use of these dietary supplements or the resulting health problems in these personnel.

Social media platforms, such as Reddit, have been employed to understand the mental health needs and experiences of military personnel [21]. Because there is no routine systematic surveillance system for dietary supplement use in military personnel, social media platforms can help researchers better understand peer-to-peer communication about the use and patterns of use of these products. As of 2019, the social media site Reddit has 430 million active monthly users, and all Reddit data are publicly available and can be used for research [22]. Reddit allows users to follow and post on specific topics in communities called subreddits. Any user can create a new subreddit or post in existing subreddits. There are subreddits specifically for military personnel broken down by military branch, such as the air force or army. In these subreddits, users can post and comment in a thread format on a variety of topics. This thread format allows users to obtain opinions and viewpoints from any Reddit user.

In several studies, Reddit has been used to assess peer-to-peer communication on topics such as mental health [23], eating disorders [24], substance use [25,26], and vaping [27-29]. Coelho et al [21] conducted a study of military personnel subreddits on Reddit to assess veterans' experiences with posttraumatic stress disorder. To our knowledge, no studies have examined discussions on supplement use by military personnel on Reddit. Therefore, we conducted a pilot study with the primary aims to assess whether military personnel subreddits have content on WLSP supplement use and to estimate intercoder agreement and a secondary aim to identify the main content of peer information on this topic. We anticipate the findings of this study will inform our understanding of the feasibility of a larger future investigation on the patterns of use of WLSP supplements among military personnel and peer-to-peer messaging around these supplements in service member spaces.

Methods

Our pilot study used data from Reddit to examine peer-to-peer communication about WLSP supplements among military personnel. The Boston Children's Hospital institutional review board determined this study to be exempt from the need to obtain informed consent given the nonidentifying nature of the data.

Data Source and Platform

Reddit hosts the following 6 publicly available military personnel subreddits for the air force, army, military, navy, and coast guard: r/AirForce, r/army, r/Military, r/navy, r/uscg, and r/USMC. To identify all potentially relevant posts between January 1, 2009, and December 31, 2019, we used the following search terms: "supplements," "dietary supplements," "ECA stack," "EC stack," and "drug test." A total of 570 potentially relevant posts were extracted for review to identify those related to WLSP supplements. Posts related to vitamins and other types of dietary supplements were excluded. A total of 64 relevant posts, to which users posted 243 relevant comments in response, were included in the analytic database for this study.

Data Capture and Coding

Posts and comments were captured as PDF files and imported into QSR International's qualitative software, NVivo 12, for coding [30]. Our study followed a data collection and reporting protocol similar to that used by Sowles and colleagues [26]. An inductive coding protocol created from commonly posted information in the Reddit posts and key terms from recent literature on WLSP dietary supplements were used to categorize posts and comments by year of posting (2009-2019) and subreddit. These posts and comments were then coded by two independent coders (KJS, JAV) for the following 6 key themes: the OPSS website, research on WLSP dietary supplements, discernability of supplement use through drug testing, brand names or identifiers, serious adverse effects, and reason for supplement use.

One of the primary aims of our pilot study was to evaluate the intercoder agreement for the Reddit posts. We assumed that feasibility would be achieved if the lower bound of the two-sided 95% exact binomial CI for the proportion of successfully coded subreddits exceeded 80%; this is equivalent to observing successful coding of at least 58/64 (91%) posts. Cohen κ [31] was used to evaluate the interrater reliability of the two ratings of the 6 key themes, and CIs for the κ coefficient were estimated using the method developed by Fleiss, Cohen, and Everitt [32]. Statistical significance was defined as a two-sided $P<.05$, and no adjustments were made for multiple comparisons. Cohen κ values in the range of 0.61-0.80 indicate substantial agreement, and values >0.80 indicate almost perfect agreement [33]. Cohen κ results indicated very strong agreement between the two coders' evaluations, with all P values <.001 (Table 1).

The word frequency analysis tool in NVivo 12 was used to identify words that were frequently used in the collected Reddit posts and comments. The tool identifies exact matches for strings of 3 or more characters. Frequent words were analyzed if they appeared in the Reddit posts 3 or more times. Common words including "the," "like," "and," and "but" were excluded from the results.

Table 1. Interrater agreement for coding of Reddit post themes appearing in military personnel subreddits from 2009 to 2019 (N=64 posts).

Themes	κ value (95% CI)	*P* value
Operation Supplement Safety	0.85 (0.56-1.0)	<.001
Other research cited[a]	0.74 (0.39-1.0)	<.001
Drug testing/positive test result	0.92 (0.75-1.0)	<.001
Products and ingredients	0.97 (0.91-1.0)	<.001
Reasons for use	0.97 (0.91-1.0)	<.001
Adverse effects	1.0[b] (1.0-1.0)	N/A[c]

[a]Research from sources other than Operation Supplement Safety (eg, the US Food and Drug Administration website).

[b]Variance was estimated to be zero as a result of perfect agreement between coders.

[c]N/A: not applicable.

Results

Peer-to-peer communication about WLSP supplements was seen in all 6 military personnel subreddits beginning in the year 2012, but no relevant posts were identified from 2009 to 2011 (Table 2). The largest number of relevant posts (23/64, 36%) was seen in the r/AirForce subreddit and in 2019 (16/64, 25%). A total of 64/64 (100%) posts in the military subreddits were successfully coded by both coders, which indicated the feasibility of our pilot study. The two-sided 95% exact binomial CI for this proportion is 94.4%-100%. Of the 64 total posts, 60 (94%) were posts asking for advice, and 4 (6%) were posts with general information.

Terms related to drug testing, such as "drug test," "illegal," "banned," and "pop," appeared frequently in posts and comments, suggesting that users were concerned with which WLSP supplements were legal and which might lead to a positive result on a drug test (Table 3). Other frequently used terms were those related to places from where supplements could be bought, such as "base," "GNC," and "online." General Nutrition Center (GNC) stores and other stores on military bases were the most frequently mentioned places from where these products could be bought. Reddit users mentioned the OPSS website 23/307 (7.5%) times in the posts and comments, and they mentioned this site in a way that conveyed it to be a source for determining if a supplement was allowed for use by military personnel. Preworkout and bodybuilding were the most frequently mentioned reasons for purchasing and using these supplements.

As shown in Table 2, ECA or EC stacks were mentioned 27/307 (8.8%) times in the Reddit posts and comments. The most commonly misused drug used in place of ephedrine in ECA

stacks was Bronkaid—an over-the-counter asthma medication easily available for purchase in most pharmacies in the United States—which was specifically mentioned 8/307 (2.6%) times in the posts and comments.

Table 2. Keywords in military personnel subreddit posts and comments by year (2009-2019) and subreddit.

Reddit post information	Frequency of words, n
Keywords searched (N=307)	
Supplements/dietary supplements	232
Drug test	84
EC[a]/ECA[b] stack	27
Number of relevant posts per subreddit (N=64)	
r/AirForce	23
r/army	13
r/Military	11
r/navy	8
r/uscg	2
r/USMC	7
Number of relevant posts per year (N=64)	
2009	0
2010	0
2011	0
2012	7
2013	5
2014	5
2015	4
2016	9
2017	8
2018	10
2019	16

[a]EC: ephedrine and caffeine.

[b]ECA: ephedrine, caffeine, and aspirin supplement mix.

Table 3. Frequently used words representing key themes in military personnel subreddit posts (N=64) and comments (N=243) from 2009 to 2019.

Key theme	Frequency of words, n
Operation Supplement Safety	
Link to Operation Supplement Safety website	23
Other research cited[a]	
Food and Drug Administration	30
Department of Defense	25
Drug testing/positive test result	
Banned	93
Illegal	30
Pop/popped/popping	53
Products and ingredients	
Animal Pak	20
Bronkaid	8
Caffeine	12
DMAA[b]	46
Ephedrine	32
Ephedra	7
Jack3d	17
Pseudoephedrine	4
Stack	16
Tea	3
Reasons for use	
Bodybuilding	12
Detox	3
Preworkout	60
Where to purchase	
Base (ie, military base)	22
GNC[c]	31
Online	14
Over the counter	12

[a]Research from sources other than Operation Supplement Safety (eg, the US Food and Drug Administration website).

[b]DMAA: 1,3-dimethylamylamine.

[c]GNC: General Nutrition Center.

Table 4 presents exemplar quotes showcasing each of the 6 key themes. Similar to the word frequency results, the quotes illustrate peer-to-peer communication by Reddit users on topics such as the safety of preworkout supplements and the likelihood of a positive result on a drug test. Adverse effects of WLSP supplement use were mentioned, including elevated heart rate and death. Many posts included misinformation about these supplements; one user stated, “if you buy [supplements] at GNC, you’re good to go,” with respect to passing a drug test, though this view was not consistent across users. Some users posted that they had failed drug tests after using supplements purchased at GNC or other stores on their base.

Reddit users asked their subreddit peers about ECA stacks, where to purchase the ingredients, and if the use of these supplements would lead to a positive result on a drug test. Several users shared similar reasons for using ECA stacks, as expressed by one user who wanted to “cut some fat and [get] below [my] max weight.” The OPSS website and other research sites, such as the FDA website, were mentioned as resources by some Reddit users.

Table 4. Exemplar quotes related to weight loss and sports performance dietary supplements for each key theme appearing in military personnel subreddits from 2009 to 2019.

Theme and subreddit	Year of posting	Quote	Coded themes
Drug testing/positive test result			
r/army	2012	"Just curious if supplements such as preworkouts or proteins are allowed in basic?"	• Regulation of supplements • Preworkout
r/army	2012	"…the Army (and pro-sports) has banned any preworkout with 1,3 Dimethylamylamine [DMAA] in them (such as Jack3d and Oxyelite Pro)."	• Regulation of supplements • DMAA[a] • Jack3d • OxyElite Pro
Operation Supplement Safety			
r/navy	2019	"https://www.opss.org/dietary-supplement-ingredients-prohibited-department-defense Looks like it's listed under the 'unapproved drugs' section."	• Operation Supplement Safety • Regulation of supplements
Other research cited[b]			
r/army	2019	"If you are ordering stuff from sketchy suppliers on the internet, who knows what's in it. https://www.fda.gov /ForConsumers/ConsumerUpdates/ucm173739.htm"	• Other research • Regulation of supplements
Products mentioned			
r/AirForce	2015	"Are we allowed to take EC stack? I've heard anecdotes of people saying you can take it and there's no problem, but just to be clear, ephedrine can cause false positives for methamphetamine. You have to buy Bronkaid or Primatene at a pharmacy in order to use the EC stack because ephedrine isn't sold as a standalone product."	• Positive test result • EC[c] stack • Ephedrine • Bronkaid
r/army	2018	"I am thinking of starting an E/C stack taking Bronkaid (OTC) and Caffeine pills in order to cut some fat and get below my max weight. I wanted to just ensure that this is ok for me to do, and that it wouldn't get me in any sort of trouble. I appreciate any and all advice, thanks!"	• Regulation of supplements • Weight loss • Bronkaid • EC stack
r/navy	2019	"E/C stack is about the only stack that's not a waste of money"	• EC stack
Reason for supplement use			
r/USMC	2018	"Want to take a preworkout with DMHA in it. DMAA is on the banned substance list, but DMHA is not. If I pop a false positive on a drug test, will I get charged? Or will they do further tests and determine it was a legal stimulant in a preworkout?"	• Regulation of supplements • Preworkout • DMAA • DMHA[d]
Adverse effects			
r/AirForce	2019	"PSA for all you PT/Mock test takers. Do not, under any circumstances, take multiple energy pills you bought at the gas station before you test. Never thought it would be a problem, but I was testing one of our officers this morning. He completed 1 lap and proceeded to vomit. His heart rate was over 210 and I was starting to worry that we would have to call an ambulance."	• Vomiting • Elevated heart rate • Energy supplements
r/army	2019	"jack3d apparently did [make you test positive on a drug test] back in like 2009-2010. I used to use it lmao but never popped hot or anything. The military banned it and the product changed to remove some sort of amphetamine it had. It also killed a [couple] dudes in the Army and at USMA though from heart attacks so it could also have been that."	• Regulation of supplements • Drug testing • Jack3d • Heart attack

[a]DMAA: 1,3-dimethylamylamine.

[b]Research from sources other than OPSS (eg, the US Food and Drug Administration website).

[c]EC: ephedrine and caffeine.

[d]DMHA: 1,5-dimethylhexylamine.

Discussion

To our knowledge, our pilot study is the first to examine peer-to-peer communication about WLSP supplements on a social media platform and document messaging around these products within all 6 publicly available military personnel subreddits from 2009 to 2019. Our results suggest that military personnel are interested in using these supplements regardless of existing DoD regulations and the potentially serious adverse effects.

The DoD has prohibited the use of some of these types of supplements because of the associated serious adverse effects [15], but the supplements that were banned for use were not clear to all Reddit users. A total of 60/64 (94%) posts were questions about supplement types, drug testing, or supplement use, while only 7/243 (2.9%) comments referenced the OPSS website. Thus, more education on this important resource is needed for service members [20].

The results of this study suggest that there is interest in using muscle-building and preworkout supplements among military personnel despite DoD prohibitions and the risk associated with the use of some types of supplements. ECA stacks were frequently mentioned in these subreddits and seemed to be favored by Reddit users over preformulated performance and preworkout supplements because of the perceived effects of the ECA ingredients on the body. Reddit users posted simple instructions for their peers explaining how users can make their own ECA stack from easily obtained ingredients. Many users suggested that the supplements had the effect of “cutting” fat, which would help them reach the weight required for service. Our results showed that the r/AirForce subreddit had the highest number of posts, but it is unclear if WLSP supplement use is higher in this branch than in the other branches of the military.

A limitation of this pilot study is that Reddit does not provide any demographic information on their users. Therefore, we could not determine the location, age, or gender of these users. There may be important differences in these demographic characteristics among military personnel who use these products. In addition, the subreddits included in our study were publicly accessible; although these communities are described as affinity groups for military personnel in different branches, it is possible that nonmilitary Reddit users join and post on these subreddits. Another limitation is that our research team could not access certain subreddits. For example, r/Marines is a private subreddit and a user has to be invited and approved by a moderator to access the subreddit. We do not know if WLSP supplements are discussed in these subreddits and—if they are—whether these conversations are similar to those in the publicly available subreddits.

Our investigation was restricted to Reddit to evaluate the feasibility of studying peer-to-peer communication about WLSP supplements among military personnel on a social media platform. Our findings suggest that such a study is not only feasible but can also provide novel insight into peer-to-peer communication about beliefs, misconceptions, concerns, and behaviors vis-à-vis WLSP supplements. Future research should focus on other social media platforms to examine similar communication among military personnel and other populations.

With the current use of dietary supplements at over 60% for military personnel [3], future studies should investigate WLSP supplement use patterns by branch of service and sources of supplements purchased, frequency of failed drug tests attributable to WLSP supplement use, chemical composition and safety of commonly used WLSP supplements, and adverse effects experienced by military personnel using WLSP supplements. The danger these products pose to the health and safety of military personnel is becoming increasingly clear even through current ad hoc channels for reporting adverse effects, which raises serious concerns about the negative impact of WLSP supplement use on the overall military readiness of the nation’s armed forces. Through the OPSS program, the DoD may consider further steps to address the use of these types of dietary supplements by military personnel. For instance, the DoD could disseminate more information for military personnel and military medical providers on the dangers of WLSP supplement use. DoD leadership could consider implementing routine, systematic surveillance to track the use of WLSP supplements within the military. Finally, the DoD could consider more stringent restrictions on the sale of these products on military bases.

Acknowledgments

Support for this work has been provided by the Ellen Feldberg Gordon Fund for Eating Disorders Prevention Research and the Strategic Training Initiative for the Prevention of Eating Disorders. SBA is supported by training grant number T76-MC00001 from the Maternal and Child Health Bureau, Health Resources and Services Administration, US Department of Health and Human Services.

Conflicts of Interest

None declared.

References

1. Total U.S. dietary supplements market size from 2016 to 2024. Statista. 2018. URL: https://www.statista.com/statistics/828481/total- dietary-supplements-market-size-in-the-us [accessed 2020-04-30]
2. Kantor ED, Rehm CD, Du M, White E, Giovannucci EL. Trends in Dietary Supplement Use Among US Adults From 1999-2012. JAMA 2016 Oct 11;316(14):1464-1474 [FREE Full text] [doi: 10.1001/jama.2016.14403] [Medline: 27727382]
3. Knapik JJ, Steelman RA, Hoedebecke SS, Farina EK, Austin KG, Lieberman HR. A systematic review and meta-analysis on the prevalence of dietary supplement use by military personnel. BMC Complement Altern Med 2014 May 02;14(1):143-119 [FREE Full text] [doi: 10.1186/1472-6882-14-143] [Medline: 24884570]
4. Santos I, Sniehotta FF, Marques MM, Carraça EV, Teixeira PJ. Prevalence of personal weight control attempts in adults: a systematic review and meta-analysis. Obes Rev 2017 Jan 21;18(1):32-50 [FREE Full text] [doi: 10.1111/obr.12466] [Medline: 27653242]
5. Lieberman H, Stavinoha TB, McGraw SM, White A, Hadden LS, Marriott BP. Use of dietary supplements among active-duty US Army soldiers. Am J Clin Nutr 2010 Oct;92(4):985-995. [doi: 10.3945/ajcn.2010.29274] [Medline: 20668050]
6. Bailey RL. Current regulatory guidelines and resources to support research of dietary supplements in the United States. Crit Rev Food Sci Nutr 2020 Nov 13;60(2):298-309 [FREE Full text] [doi: 10.1080/10408398.2018.1524364] [Medline: 30421981]
7. Cohen PA, Travis JC, Keizers PHJ, Deuster P, Venhuis BJ. Four experimental stimulants found in sports and weight loss supplements: 2-amino-6-methylheptane (octodrine), 1,4-dimethylamylamine (1,4-DMAA), 1,3-dimethylamylamine (1,3-DMAA) and 1,3-dimethylbutylamine (1,3-DMBA). Clin Toxicol (Phila) 2018 Jun 08;56(6):421-426. [doi: 10.1080/15563650.2017.1398328] [Medline: 29115866]
8. Grundlingh J, Dargan PI, El-Zanfaly M, Wood DM. 2,4-dinitrophenol (DNP): a weight loss agent with significant acute toxicity and risk of death. J Med Toxicol 2011 Sep 8;7(3):205-212 [FREE Full text] [doi: 10.1007/s13181-011-0162-6] [Medline: 21739343]
9. Or F, Kim Y, Simms J, Austin SB. Taking Stock of Dietary Supplements' Harmful Effects on Children, Adolescents, and Young Adults. J Adolesc Health 2019 Oct;65(4):455-461. [doi: 10.1016/j.jadohealth.2019.03.005] [Medline: 31176525]
10. Stickel F, Shouval D. Hepatotoxicity of herbal and dietary supplements: an update. Arch Toxicol 2015 Jun 14;89(6):851-865. [doi: 10.1007/s00204-015-1471-3] [Medline: 25680499]
11. Yen M, Ewald MB. Toxicity of weight loss agents. J Med Toxicol 2012 Jun 14;8(2):145-152 [FREE Full text] [doi: 10.1007/s13181-012-0213-7] [Medline: 22351299]
12. Jagim AR, Camic CL, Harty PS. Common Habits, Adverse Events, and Opinions Regarding Pre-Workout Supplement Use Among Regular Consumers. Nutrients 2019 Apr 16;11(4):855 [FREE Full text] [doi: 10.3390/nu11040855] [Medline: 31014016]
13. Bodell L, Forney KJ, Keel P, Gutierrez P, Joiner TE. Consequences of Making Weight: A Review of Eating Disorder Symptoms and Diagnoses in the United States Military. Clin Psychol (New York) 2014 Dec;21(4):398-409 [FREE Full text] [doi: 10.1111/cpsp.12082] [Medline: 25642105]
14. Masheb RM, Kutz AM, Marsh AG, Min KM, Ruser CB, Dorflinger LM. "Making weight" during military service is related to binge eating and eating pathology for veterans later in life. Eat Weight Disord 2019 Dec 30;24(6):1063-1070. [doi: 10.1007/s40519-019-00766-w] [Medline: 31471885]
15. Eliason MJ, Eichner A, Cancio A, Bestervelt L, Adams BD, Deuster PA. Case reports: Death of active duty soldiers following ingestion of dietary supplements containing 1,3-dimethylamylamine (DMAA). Mil Med 2012 Dec;177(12):1455-1459. [doi: 10.7205/milmed-d-12-00265] [Medline: 23397688]
16. Knapik JJ, Trone DW, Austin KG, Steelman RA, Farina EK, Lieberman HR. Prevalence, Adverse Events, and Factors Associated with Dietary Supplement and Nutritional Supplement Use by US Navy and Marine Corps Personnel. J Acad Nutr Diet 2016 Sep;116(9):1423-1442 [FREE Full text] [doi: 10.1016/j.jand.2016.02.015] [Medline: 27083989]
17. Miller SC. Safety concerns regarding ephedrine-type alkaloid-containing dietary supplements. Mil Med 2004 Feb;169(2):87-93. [doi: 10.7205/milmed.169.2.87] [Medline: 15040625]
18. Ephedra. National Institutes of Health. URL: https://ods.od.nih.gov/Health_Information/Ephedra.aspx [accessed 2020-04-30]
19. Hall M, Grogan S, Gough B. 'It is safe to use if you are healthy': A discursive analysis of men's online accounts of ephedrine use. Psychol Health 2015 Jan 02;30(7):770-782. [doi: 10.1080/08870446.2014.994632] [Medline: 25511848]
20. Department of Defense dietary supplement resource for the military community, leaders, healthcare providers and DoD civilians. Operation Supplement Safety. URL: https://www.opss.org/ [accessed 2020-04-30]
21. Coelho J. Veterans, PTSD and Social Media: Towards Identifying Trauma Text Categories using Grounded Theory. In: GROUP '20: Companion of the 2020 ACM International Conference on Supporting Group Work. 2020 Presented at: 2020 ACM International Conference on Supporting Group Work; January 6, 2020; Sanibel Island, FL. [doi: 10.1145/3323994.3369887]
22. Perez S. Reddit’s monthly active user base grew 30% to reach 430m in 2019. TechCrunch. 2019 Dec 04. URL: https://techcrunch.com/2019/12/04/reddits-monthly-active-user-base-grew-30-to-reach-430m-in-2019/?guccounter=1 [accessed 2020-04-12]

23. De Choudhury M, De S. Mental health discourse on reddit: Self-disclosure, social support, and anonymity. 2014 Presented at: Eighth International AAAI Conference on Weblogs and Social Media; June 1-4, 2014; Ann Arbor, MI p. 71-80 URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14526
24. Sowles SJ, McLeary M, Optican A, Cahn E, Krauss MJ, Fitzsimmons-Craft EE, et al. A content analysis of an online pro-eating disorder community on Reddit. Body Image 2018 Mar;24:137-144 [FREE Full text] [doi: 10.1016/j.bodyim.2018.01.001] [Medline: 29414146]
25. Pandrekar S, Chen X, Gopalkrishna G, Srivastava A, Saltz M, Saltz J, et al. Social media based analysis of opioid epidemic using Reddit. In: AMIA Annual Symposium Proceedings. 2018 Presented at: AMIA Annual Symposium; 2018; San Francisco, CA p. 867-876.
26. Sowles SJ, Krauss MJ, Gebremedhn L, Cavazos-Rehg PA. 'I feel like I've hit the bottom and have no idea what to do': Supportive social networking on Reddit for individuals with a desire to quit cannabis use. Subst Abus 2017 Jul 13;38(4):477-482 [FREE Full text] [doi: 10.1080/08897077.2017.1354956] [Medline: 28704167]
27. Brett EI, Stevens EM, Wagener TL, Leavens EL, Morgan TL, Cotton WD, et al. A content analysis of JUUL discussions on social media: Using Reddit to understand patterns and perceptions of JUUL use. Drug Alcohol Depend 2019 Jan 01;194:358-362. [doi: 10.1016/j.drugalcdep.2018.10.014] [Medline: 30472576]
28. Wang L, Zhan Y, Li Q, Zeng D, Leischow S, Okamoto J. An Examination of Electronic Cigarette Content on Social Media: Analysis of E-Cigarette Flavor Content on Reddit. Int J Environ Res Public Health 2015 Nov 20;12(11):14916-14935 [FREE Full text] [doi: 10.3390/ijerph121114916] [Medline: 26610541]
29. Zhan Y, Zhang Z, Okamoto JM, Zeng DD, Leischow SJ. Underage JUUL Use Patterns: Content Analysis of Reddit Messages. J Med Internet Res 2019 Sep 09;21(9):e13038 [FREE Full text] [doi: 10.2196/13038] [Medline: 31502542]
30. NVivo Qualitative Data Analysis Software. QSR International. URL: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home [accessed 2018-01-01]
31. Cohen J. A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement 2016 Jul 02;20(1):37-46. [doi: 10.1177/001316446002000104]
32. Fleiss JL, Cohen J, Everitt BS. Large sample standard errors of kappa and weighted kappa. Psychological Bulletin 1969;72(5):323-327. [doi: 10.1037/h0028106]
33. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med 2012;22(3):276-282. [doi: 10.11613/bm.2012.031]

Abbreviations

DoD: Department of Defense
ECA: ephedrine, caffeine, and aspirin
EC: ephedrine and aspirin
FDA: Food and Drug Administration
GNC: General Nutrition Center
OPSS: Operation Supplement Safety
WLSP: weight loss and sports performance

Edited by G Eysenbach; submitted 25.03.21; peer-reviewed by R Wierzejska, AM Auvinen; comments to author 19.04.21; revised version received 03.05.21; accepted 01.06.21; published 04.10.21.

Please cite as:
Sharp KJ, Vitagliano JA, Weitzman ER, Fitzgerald S, Dahlberg SE, Austin SB
Peer-to-Peer Social Media Communication About Dietary Supplements Used for Weight Loss and Sports Performance Among Military Personnel: Pilot Content Analysis of 11 Years of Posts on Reddit
JMIR Form Res 2021;5(10):e28957
URL: https://formative.jmir.org/2021/10/e28957
doi:10.2196/28957
PMID:34605769

Original Paper

Practical and Emotional Problems Reported by Users of a Self-guided Digital Problem-solving Intervention During the COVID-19 Pandemic: Content Analysis

Amira Hentati[1,2], MSci; Erik Forsell[1], PhD; Brjánn Ljótsson[2], PhD; Martin Kraepelien[1,2], PhD

[1]Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet & Stockholm Health Care Services, Region Stockholm, Stockholm, Sweden

[2]Division of Psychology, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

Corresponding Author:
Amira Hentati, MSci
Centre for Psychiatry Research
Department of Clinical Neuroscience
Karolinska Institutet & Stockholm Health Care Services, Region Stockholm
Norra Stationsgatan 69
Stockholm, 11364
Sweden
Phone: 46 704411425
Email: amira.hentati@ki.se

Abstract

Background: To better direct assessments and interventions toward the general population during both the ongoing COVID-19 pandemic and future crises with societal restrictions, data on the types of practical and emotional problems that people are experiencing are needed.

Objective: The aim of this study was to examine the types of practical and emotional problems that the general population is experiencing during the COVID-19 pandemic and to construct an empirically derived inventory based on the findings.

Methods: A total of 396 participants, recruited among members of the general public in Sweden who were experiencing practical and/or emotional problems during the pandemic, accessed a self-guided digital problem-solving intervention for a period of 1 week to report and solve the problems they experienced. Prior to accessing the intervention, the participants completed a short self-assessment regarding symptoms of depression and anxiety. Content analysis was used to account for the types of problems participants reported. A set of items for an inventory was later proposed based on the problem categories derived from the analysis.

Results: A majority of participants had clinically relevant symptoms of either depression or anxiety. The problems reported were categorized as 13 distinct types of problems. The most common problem was difficulty managing daily activities. Based on the categories, a 13-item inventory was proposed.

Conclusions: The 13 types of problems, and the proposed inventory, could be valuable when composing assessments and interventions for the general population during the ongoing pandemic or similar crises with societal restrictions. The most common problem was of a practical nature, indicating the importance of including examples of such problems within assessments and interventions.

Trial Registration: ClinicalTrials.gov NCT04677270; https://clinicaltrials.gov/ct2/show/NCT04677270

(JMIR Form Res 2021;5(10):e31722) doi:10.2196/31722

KEYWORDS

digital intervention; COVID-19; problem-solving; self-guided intervention; content analysis; public health; mental health; depression; anxiety; pandemic

Introduction

Background

The COVID-19 pandemic is considered a threat to the mental well-being of the general public and may increase the suicide risk for some people [1,2]. This threat consists of both emotional problems, such as anxiety, loneliness, and low mood [2,3], and practical problems, such as not being able to work remotely or to travel as before the pandemic [4,5].

Population-level efforts aiming to prevent negative mental health consequences have been called for since the first months of the COVID-19 pandemic [1]. Furthermore, remotely implementable digital interventions for treatment and prevention have been seen as critical to achieve the scalability necessary to have an impact on the health of the general public [6]. An example of a successful population-level effort during the COVID-19 pandemic is described in a previous study in which an existing intervention for extensive worry was adapted to COVID-19–related worry, transformed into a self-guided format [7], and later implemented in the Swedish regular health care system. However, because worry is only one possible problem experienced by the general population during the COVID-19 pandemic [8], there is a need to assess what types of practical and emotional problems the general population is experiencing. This could facilitate the direction of assessments and interventions toward the general public both during the ongoing pandemic and in future similar crises.

Problem-solving therapy is a well-examined intervention that was originally constructed for major depression, targeting the ability to solve problems [9]. Moreover, the ability to solve problems has been highlighted as one of several protective factors for individuals in the general population affected by the COVID-19 pandemic when considering societal suicide prevention [10]. A problem-solving intervention could thus be a suitable intervention for the general public experiencing practical or emotional problems during the ongoing pandemic. Ideally, it should be easy to gain access to such an intervention, and the intervention should be self-guided to facilitate scalability. An open access and self-guided internet-based psychological support intervention (PATH), which includes problem-solving as well as conflict management and stress management, has been examined with regard to participants' input to the program during the COVID-19 pandemic [11]. In that study, conflicts with others, worry, and difficulties concentrating stood out as the most common types of problems during the pandemic.

Aim

The aim of this study was to examine the types of practical and emotional problems that the Swedish help-seeking population is experiencing during the COVID-19 pandemic. An additional aim was to construct an empirically derived inventory to facilitate the assessment of problems and direction of interventions in the general population during the ongoing pandemic or similar crises with societal restrictions.

Methods

Setting and Study Design

This study was part of a project aiming to investigate treatment engagement with a self-guided digital problem-solving intervention between two different user interfaces. The results concerning the effect of the user interfaces on treatment engagement have been presented in a previous paper [12]. In this paper, the focus lies on the types of practical and emotional problems that the general help-seeking population in Sweden reported experiencing during the COVID-19 pandemic when using a self-guided digital problem-solving intervention. The study was approved by the Swedish national ethical review board (ID 2020-02739), and although the article does not report results of a health care intervention, the study was retrospectively registered on ClinicalTrials.gov (ID: NCT04677270, 2020-12-21).

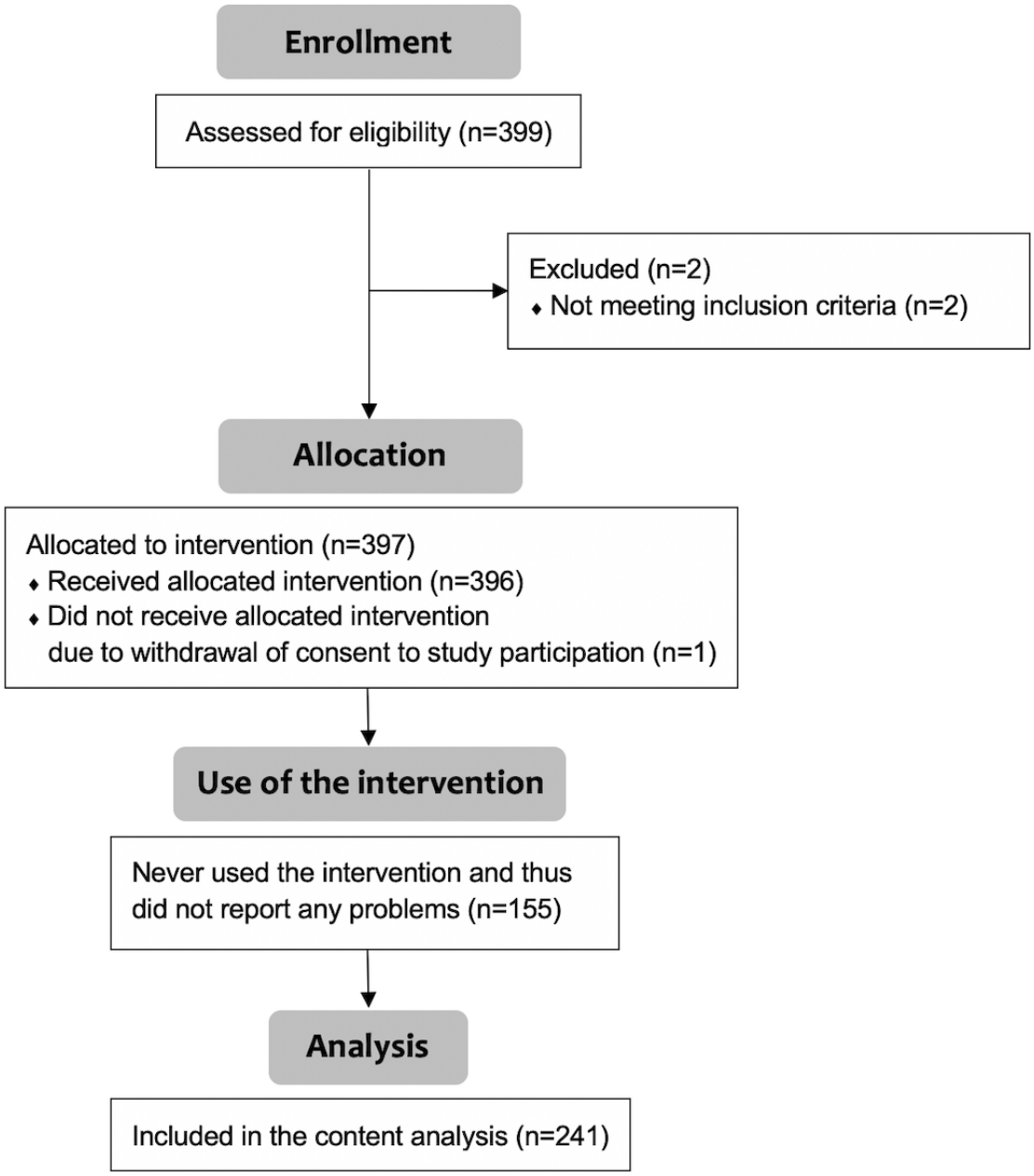
Participants and Recruitment

The target population was the general help-seeking population in Sweden who were experiencing practical and/or emotional problems during the COVID-19 pandemic. Participants were recruited nationwide in Sweden through advertisements on social media during a period of 6 weeks, between August 26 and October 6, 2020. Inclusion criteria were (1) age of 16 years or older, and (2) self-reported practical and/or emotional problems experienced during the COVID-19 pandemic. The second criterion was assessed through a question asking if participants experienced practical problems, emotional problems, practical and emotional problems, or no problems. No further definition of these problems was given at this stage of assessment.

Procedures

To register for the study, participants completed a digital self-assessment on a secure digital platform. Written informed consent was digitally provided by all participants. Of the 399 individuals who registered for the study, only 2 persons were excluded from participating, owing to not experiencing problems (ie, the second inclusion criterion was not met). Furthermore, 1 person withdrew consent for participation. A total of 396 participants were thus included, and they accessed a self-guided digital problem-solving intervention for a period of one week to report and solve problems they experienced. The study flowchart is shown in Figure 1.

Figure 1. Flowchart of the selection of study participants.



Intervention

A self-guided digital problem-solving intervention adapted to the COVID-19 pandemic was created for the study. An already existing digital problem-solving intervention, previously used as a component in a 12-week multi-component internet-delivered cognitive behavioral therapy program for individuals with major depression in Swedish regular health care [13], was used as a template and adapted for self-guidance.

The problem-solving intervention comprised psychoeducational texts and rationale, examples of problems and suggestions of solutions, pictures, instructions, and problem-solving exercises. The exercises consisted of steps where participants were able to describe and choose to work with one problem at a time; set a goal; make a list of possible solutions to the problem, including pros and cons; choose what solutions to try and plan how and when to try them out; and then evaluate the solutions and reflect on possible lessons learned. All material within the intervention was in Swedish and was adapted to the pandemic situation. Participants could read texts about four fictive users having the following problems: health-related worry, stress because of their financial situation, loneliness, and problems with household activities. In total, the intervention consisted of a single module of approximately 4800 words. The length of the intervention was 1 week, because the current study only intended to examine types of problems and platform use (data presented in [12]) and not problem-solving as a clinical intervention. The intervention could be accessed on a secure digital platform via a computer or mobile device connected to the internet.

Measurements

When registering for the study, participants completed a digital assessment comprising questions on demographics and whether they experienced problems during the COVID-19 pandemic.

Participants also completed two self-assessed short scales, the Patient Health Questionnaire-2 (PHQ-2) [14] and Generalized Anxiety Disorder-2 (GAD-2) [15], measuring symptoms of depression and anxiety, respectively. These scales were administered to assess the proportion of participants with a possible clinical symptom burden, but they were not used either for inclusion or as an outcome measure.

Data Analysis

Content analysis [16] was used to code all participants' problem-solving attempts that were reported and saved within the digital platform as well as to create categories of problems based on these. All problem-solving attempts were entered in text on the digital treatment platform. Thus, transcription of data was not needed.

The step-by-step categorization of data started with defining each problem-solving attempt as the unit of analysis [17]. A problem-solving attempt was registered every time a participant completed the first step of the problem-solving exercises, which was to define the problem. Thus, a problem-solving attempt did not necessarily constitute a complete problem-solving exercise. To create categories for the content analysis, 25 participants were selected randomly for initial coding. Three of the authors, AH, EF, and MK, coded these participants jointly and created the coding instructions as well as the categories with consensus. Then, another 25 randomly selected participants were independently coded by the three coders, using the instructions and categories created. The interrater agreement between the coders was substantial [18] (Cohen κ=0.66-0.76). After checking that the interrater reliability was acceptable, disagreements were discussed until a consensus was reached. Afterward, the remaining 346 participants were divided between the three coders, and their problem-solving attempts were coded using the identified categories. Any uncertain categorizations were discussed with all coders. Each category was defined in turn as mainly either a practical or emotional problem.

The categorization of reported problem-solving attempts was later used in a quantitative description of the content [17]: namely, the number of problem-solving attempts that belonged to each category, as well as the number of the participants who had solved at least one problem that belonged to each category. Furthermore, to ensure that the data were not skewed by a few active participants, the use of the intervention was quantified by the percentage of participants using the intervention at least once, as well as the average number of problem-solving attempts completed during the week of access. Lastly, based on the problem categories derived from the analysis, a set of items for an inventory was proposed.

Results

A majority of participants were women, were university educated, and had clinically relevant symptoms of either depression or anxiety. Table 1 shows the complete sample characteristics.

A majority of participants used the problem-solving intervention at least once during the week of access. Table 2 shows details on the use of the intervention.

The problems reported within the intervention by the participants who used the intervention at least once were categorized into 13 distinct types of problems, which can be found in Table 3. The 3 most frequent categories of problems were of a practical nature, such as difficulties initiating daily activities, problems or frustration regarding one's work and/or study situation, and problems or frustration with public health guidelines or the pandemic situation in general.

Based on the 13 distinct types of problems derived from the content analysis (see Table 3), the items shown in Textbox 1 are proposed as an inventory of practical and emotional problems during the COVID-19 pandemic.

Table 1. Sample characteristics (N=396).

Variable	Value
Female gender, n (%)	352 (88.9)
Age (years)	
Mean (SD)	40 (13)
Range	17-79
In a relationship, n (%)	246 (62.1)
Occupational status, n (%)	
Employed full-time	200 (50.5)
Employed part-time	41 (10.4)
Student	68 (17.2)
Parental leave	7 (1.8)
Unemployed	34 (8.6)
Long-term sick leave	21 (5.3)
Retired	25 (6.3)
Education, n (%)	
Primary school	7 (1.8)
Secondary school	78 (19.7)
University	311 (78.6)
Possible major depression (PHQ-2[a] score ≥3), n (%)	235 (59.3)
Possible generalized anxiety (GAD-2[b] score ≥3), n (%)	236 (59.6)
Concurrent possible depression and anxiety, n (%)	180 (44.5)
Either possible depression, anxiety, or both, n (%)	291 (73.5)

[a]PHQ-2: Patient Health Questionnaire-2.

[b]GAD-2: Generalized Anxiety Disorder-2.

Table 2. Use of the problem-solving intervention (N=396).

Variable	Value
Used the problem-solving intervention at least once, n (%)	241 (60.9)
Problem-solving attempts per participant, mean (SD)	1.13 (1.44)

Table 3. Types of problems identified from the content analysis.

Problem type	Problem category	Number of participants with problem (% of number of participants who used the intervention at least once, N=241)	Number of problems (% of total number of reported problems, N=446)	Definition	Examples
Daily activities	Practical	51 (21.2)	59 (13.2)	Difficulties initiating daily activities, staying motivated, or maintaining focus	• Not getting household activities done • Spending too much time on social media
Work and study	Practical	46 (19.1)	51 (11.4)	Problems or frustration regarding work and/or study situation	• Working at home with children • Struggling with digital work or studies
Health behaviors	Practical	40 (16.6)	44 (9.9)	Difficulties maintaining health promoting behaviors such as physical activity, satisfactory sleep patterns, or active recovery	• Getting less physical exercise than usual • Having trouble falling asleep
Family and relationship	Emotional	37 (15.4)	40 (9)	Problems related to relationships, including family, friends, or significant other	• Feeling unhappy in a relationship • Finding it difficult to establish new relationships
Health anxiety	Emotional	35 (14.5)	36 (8.1)	Affected emotionally by health fears or worry regarding self or others	• Worrying about being infected with COVID-19 • Worrying that relatives or friends will become ill with COVID-19
Pandemic guidelines	Practical	31 (12.9)	45 (10.1)	Problems or frustration with public health guidelines or the pandemic situation in general	• Being bound to home as soon as you experience the sightliest signs of symptoms of COVID-19 • Feeling frustrated at others not following the pandemic guidelines
Non–health-related anxiety or stress	Emotional	30 (12.4)	35 (7.8)	Affected emotionally by non–health-related anxiety or stress-related problems	• Experiencing social anxiety or generalized anxiety • Feeling overwhelmed
Financial issues	Practical	26 (10.8)	30 (6.7)	Financial problems or fears	• Experiencing fear of losing your job • Experiencing loss of income
Loneliness	Emotional	22 (9.1)	22 (4.9)	Affected emotionally by loneliness	• Negatively affected by having had less contact with family or friends • Experiencing social isolation
Low mood	Emotional	20 (8.3)	21 (4.7)	Low mood or feelings of meaninglessness	• Feeling sad most of the time • Experiencing apathy
Changes in emotional state apart from anxiety and low mood	Emotional	19 (7.9)	22 (4.9)	Emotional challenges other than anxiety or low mood	• Feeling angry • Experiencing low self-esteem
Health issues	Practical	19 (7.9)	23 (5.2)	Health issues related to COVID-19 or other illness regarding self or others	• Experiencing difficult symptoms of COVID-19 or other illness • Having a relative or friend with health issues
Weight and eating	Practical	18 (7.5)	18 (4)	Problems related to weight or eating	• Experiencing unintentional weight change • Struggling with binge eating

Textbox 1. Proposed inventory of practical and emotional problems during a crisis with societal restrictions.

Practical problems: do you experience…

- Difficulties initiating daily activities, staying motivated or maintaining focus? Examples of these difficulties include not getting household activities done, or spending too much time on social media.
- Problems or frustration regarding your work and/or study situation? Examples of these problems include working at home with children, or struggling with digital work or studies.
- Difficulties maintaining health promoting behaviors such as physical activity, satisfactory sleep patterns, or active recovery? Examples of these difficulties include getting less physical exercise than usual, or having trouble falling asleep.
- Problems or frustration with public health guidelines or the pandemic situation in general? Examples of these problems include being bound to home as soon as you experience the sightliest signs of symptoms of COVID-19, or feeling frustrated at others not following the pandemic guidelines.
- Financial problems or fears? Examples of these problems include experiencing fear of losing your job or of losing income.
- Health issues related to COVID-19 or other illness regarding self or others? Examples of these problems include experiencing difficult symptoms of COVID-19 or other illness, or having a relative or friend with health issues.
- Problems related to weight or eating? Examples of these problems include experiencing unintentional weight change, or struggling with binge eating.

Emotional problems: do you experience…

- Problems related to relationships, including with your family, friends, or significant other? Examples of these problems include feeling unhappy in a relationship, or finding it difficult to establish new relationships.
- Being affected emotionally by health fears or worry regarding yourself or others? Examples of these difficulties include fearing or worrying about being infected with COVID-19, or worrying that relatives or friends will become ill with COVID-19.
- Being affected emotionally by non–health-related anxiety or stress? Examples of these difficulties include experiencing social anxiety or generalized anxiety, or feeling overwhelmed.
- Being affected emotionally by loneliness? Examples of these difficulties include being negatively affected by having had less contact with family or friends, or experiencing social isolation.
- A low mood or feelings of meaninglessness? Examples of these problems include feeling sad most of the time, or experiencing apathy.
- Emotional challenges other than anxiety or low mood? Examples of these difficulties include feeling angry, or experiencing low self-esteem.

Discussion

Principal Results

In this study, COVID-19–related practical and emotional problems experienced by the Swedish help-seeking population were examined during the use of a self-guided digital problem-solving intervention. Content analysis was used to investigate the types of problems reported within the intervention.

The participants reported 13 different distinct types of problems. Practical problems, such as managing one's daily life and work situation, were most frequently reported, while loneliness, low mood, and other emotional difficulties were less common. The dominance of practical problems may have been due to participants preferring to use the digital problem-solving intervention to generate solutions to practical problems rather than emotional problems. However, it may also have to do with practical problems actually being the predominant problem type experienced by participants in this study. Additionally, it has been reported that the COVID-19 pandemic has had a vital impact on people's practical work situations [19]; societal restrictions have also led to great challenges in managing work-family balance, with sometimes minor support [20,21]. It should be further noted that the short scales PHQ-2 and GAD-2, which were used to assess the proportions of participants with possible major depression and anxiety, respectively, are probably quite sensitive to symptoms when used at the stage of screening [14]. This could help explain the high prevalence of possible clinical symptom burden among the participants, despite the fact that practical problems were more commonly reported than emotional ones.

In a previous study, COVID-19–related worry was highlighted as a target for a self-guided digital intervention [7]. Among the participants in the current study, health anxiety was reported, but it was not as common as problems of a more practical nature. This highlights the need to target a broad range of problems during a crisis involving a disease, including problems of a practical nature, as a complement to health-related worry.

The relatively low frequency of mood-related problems and feelings of loneliness may need to be interpreted within the context of the study being conducted in Sweden. Sweden has, unlike most other countries, not imposed mandatory lockdown during the COVID-19 pandemic. This may have impacted the mental well-being of the inhabitants. However, owing to voluntary restrictions recommended by the Swedish government during the pandemic, Sweden has had similar societal consequences, such as economic damage, to those of countries in lockdown [22]. As the results show, a number of people reported financial problems. These results can further be compared to the problems entered during the pandemic in the

previously mentioned PATH program, which was based in the United States [12]. In that study, the common types of problems reported were similar to those in the current study, except for a greater emphasis on interpersonal conflicts, possibly due to the separate conflict management module in the PATH program. In most cases, these types of problems would fall into the category of problems related to relationships in the proposed inventory of the current study.

Although vaccination for COVID-19 has begun worldwide, it is still not clear whether some problems, both societal and health-related, that have arisen during the pandemic will persist for some time [23,24]. We believe that the inventory proposed in this paper of practical and emotional problems during a crisis with societal restrictions could therefore be of value not only during the still ongoing pandemic but also possibly in the near future. We propose that the inventory can act as a guide when constructing both assessments and interventions related to COVID-19 problems while also providing some public health information concerning the pandemic and its consequences.

Because most items in the proposed inventory are not specific to COVID-19, the inventory may also provide helpful guidance in future similar crises. However, it is not certain that all the items will be relevant in a future crisis with societal restrictions. Health anxiety, which was reported in this study, is an example of a type of problem that is more likely to occur during crises involving a disease. However, most other items derived from the content analysis fit into a crisis with societal restrictions whether or not a disease is involved.

Based on the fact that recruitment to the study was rapid, during a relatively short period of time, we interpreted the interest in gaining access to a problem-solving intervention as high. This reflects an apparent desire for interventions of this kind. We believe that this has practical implications, as the results from this study can be used to adapt and possibly improve similar interventions, both during the still-ongoing pandemic as well as for possible persisting problems that arose due to the pandemic.

When providing digital interventions in a self-guided format, there is a risk of low adherence and low use. This was exemplified in a previous Swedish adaptation of a self-guided intervention for mental health problems [25]. We believe that one of the strengths of the current study is that a majority of participants used the digital intervention and hence contributed to the generalizability of the results.

Limitations

There are some limitations to this study that need to be acknowledged. First, the sample predominantly consisted of university-educated women, which impacts the generalizability of the results. Second, participants were recruited through advertisements on social media. We are not sure if the results or sample would have differed if additional recruitment methods would have been used. Third, we have no available data concerning problems reported by the current population before the COVID-19 pandemic, making it difficult to discern the impact of the pandemic on the problems reported. Fourth and last, the problem-solving intervention constructed for individuals with major depression that was used as a template for the development of the digital intervention used in the current study is intended to be used over several weeks [13]. In the current study, access to the intervention was limited to a period of 1 week. It is unclear whether a longer period of access to the intervention would have resulted in participants reporting additional types of problems.

Future Research

For future studies, we recommend that the inventory suggested within this paper be evaluated with regard to psychometric properties.

Conclusions

The reported problems of participants during the COVID-19 pandemic in this study fell into 13 distinct categories of problems. These can serve as targets of interventions or be of help when screening for problems in the general population during the ongoing pandemic or in future similar crises. The most frequently reported types of problems were of a practical nature, indicating the importance of giving examples of practical problems within both interventions and assessments.

Acknowledgments

This study was funded by the Government of Sweden, Ministry of Health and Social Affairs, and it used the BASS platform from the eHealth Core Facility at Karolinska Institutet, which is supported by the Strategic Research Area Healthcare Science (SFO-V). We wish to thank professor Nils Lindefors for initiating and supporting the project.

Authors' Contributions

AH and MK performed the literature search and drafted the manuscript. AH, EF, and MK performed the content analysis. All authors, AH, EF, BL, and MK, were involved in the study design and the acquisition of data; contributed to the statistical analysis and interpretation of the data; and have read, revised, and approved the manuscript. AH, EF, and MK verify the accuracy of the underlying data.

Conflicts of Interest

BL is a shareholder of DahliaQomit AB, a company specializing in online psychiatric symptom assessment, and Hedman-Lagerlöf och Ljótsson psykologi AB, a company that licenses cognitive behavior therapy manuals. AH, EF, and MK do not have any competing interests.

References

1. Galea S, Merchant RM, Lurie N. The mental health consequences of COVID-19 and physical distancing: the need for prevention and early intervention. JAMA Intern Med 2020 Jun 01;180(6):817-818. [doi: 10.1001/jamainternmed.2020.1562] [Medline: 32275292]
2. Moutier C. Suicide prevention in the COVID-19 era: transforming threat into opportunity. JAMA Psychiatry 2020 Oct 16;78(4):433. [doi: 10.1001/jamapsychiatry.2020.3746] [Medline: 33064124]
3. Canet-Juric L, Andrés ML, Del Valle M, López-Morales H, Poó F, Galli JI, et al. A longitudinal study on the emotional impact cause by the COVID-19 pandemic quarantine on general population. Front Psychol 2020 Sep 18;11:565688 [FREE Full text] [doi: 10.3389/fpsyg.2020.565688] [Medline: 33071893]
4. Bartik A, Cullen Z, Glaeser E, Luca M, Stanton C. What Jobs Are Being Done at Home During the COVID-19 Crisis? Evidence From Firm-Level Surveys. National Bureau of Economic Research. 2020 Jun. URL: https://www.nber.org/system/files/working_papers/w27422/w27422.pdf [accessed 2021-09-27]
5. Beck MJ, Hensher DA. Insights into the impact of COVID-19 on household travel and activities in Australia - the early days of easing restrictions. Transp Policy (Oxf) 2020 Dec;99:95-119 [FREE Full text] [doi: 10.1016/j.tranpol.2020.08.004] [Medline: 32836998]
6. Torous J, Jän Myrick K, Rauseo-Ricupero N, Firth J. Digital mental health and COVID-19: using technology today to accelerate the curve on access and quality tomorrow. JMIR Ment Health 2020 Mar 26;7(3):e18848 [FREE Full text] [doi: 10.2196/18848] [Medline: 32213476]
7. Wahlund T, Mataix-Cols D, Olofsdotter Lauri K, de Schipper E, Ljótsson B, Aspvall K, et al. Brief online cognitive behavioural intervention for dysfunctional worry related to the COVID-19 pandemic: a randomised controlled trial. Psychother Psychosom 2021;90(3):191-199 [FREE Full text] [doi: 10.1159/000512843] [Medline: 33212440]
8. Covid-19-Pandemins Tänkbara Konsekvenser På Folkhälsan. Folkhälsomyndigheten. 2020 Jun 24. URL: https://tinyurl.com/3mbthdxc [accessed 2021-09-27]
9. Cuijpers P, de Wit L, Kleiboer A, Karyotaki E, Ebert DD. Problem-solving therapy for adult depression: an updated meta-analysis. Eur Psychiatry 2018 Feb 01;48(1):27-37. [doi: 10.1016/j.eurpsy.2017.11.006] [Medline: 29331596]
10. Wasserman D, Iosue M, Wuestefeld A, Carli V. Adaptation of evidence-based suicide prevention strategies during and after the COVID-19 pandemic. World Psychiatry 2020 Oct;19(3):294-306 [FREE Full text] [doi: 10.1002/wps.20801] [Medline: 32931107]
11. Detweiler Guarino I, Cowan DR, Fellows AM, Buckey JC. Use of a self-guided computerized cognitive behavioral tool during COVID-19: evaluation study. JMIR Form Res 2021 May 31;5(5):e26989 [FREE Full text] [doi: 10.2196/26989] [Medline: 33973856]
12. Hentati A, Forsell E, Ljótsson B, Kaldo V, Lindefors N, Kraepelien M. The effect of user interface on treatment engagement in a self-guided digital problem-solving intervention: A randomized controlled trial. Internet Interv 2021 Dec;26:100448 [FREE Full text] [doi: 10.1016/j.invent.2021.100448] [Medline: 34471610]
13. Hedman E, Ljótsson B, Kaldo V, Hesser H, El Alaoui S, Kraepelien M, et al. Effectiveness of Internet-based cognitive behaviour therapy for depression in routine psychiatric care. J Affect Disord 2014 Feb;155:49-58. [doi: 10.1016/j.jad.2013.10.023] [Medline: 24238951]
14. Kroenke K, Spitzer RL, Williams JBW. The Patient Health Questionnaire-2: validity of a two-item depression screener. Med Care 2003 Nov;41(11):1284-1292. [doi: 10.1097/01.MLR.0000093487.78664.3C] [Medline: 14583691]
15. Kroenke K, Spitzer RL, Williams JB, Monahan PO, Löwe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med 2007 Mar 06;146(5):317-325. [doi: 10.7326/0003-4819-146-5-200703060-00004] [Medline: 17339617]
16. Krippendorff K. Content Analysis: An Introduction to its Methodology. Thousand Oaks, CA: Sage Publications; 1980.
17. Schilling J. On the pragmatics of qualitative assessment. Eur J Psychol Assess 2006 Jan;22(1):28-37. [doi: 10.1027/1015-5759.22.1.28]
18. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med 2012:276-282. [doi: 10.11613/bm.2012.031]
19. Bick A, Blandin A, Mertens K. Work From Home After the COVID-19 Outbreak. Federal Reserve Bank of Dallas. 2020 Jun. URL: https://www.dallasfed.org/-/media/documents/research/papers/2020/wp2017.pdf [accessed 2021-09-27]
20. Fisher J, Languilaire J, Lawthom R, Nieuwenhuis R, Petts RJ, Runswick-Cole K, et al. Community, work, and family in times of COVID-19. Community Work Fam 2020 May 18;23(3):247-252. [doi: 10.1080/13668803.2020.1756568]
21. Galanti T, Guidetti G, Mazzei E, Zappalà S, Toscano F. Work from home during the COVID-19 outbreak: the impact on employees' remote work productivity, engagement, and stress. J Occup Environ Med 2021 Jul 01;63(7):e426-e432 [FREE Full text] [doi: 10.1097/JOM.0000000000002236] [Medline: 33883531]

22. Sheridan A, Andersen AL, Hansen ET, Johannesen N. Social distancing laws cause only small losses of economic activity during the COVID-19 pandemic in Scandinavia. Proc Natl Acad Sci U S A 2020 Aug 25;117(34):20468-20473 [FREE Full text] [doi: 10.1073/pnas.2010068117] [Medline: 32747573]
23. Arnold D, Milne A, Samms E, Stadon L, Maskell N, Hamilton F. Are vaccines safe in patients with long COVID? A prospective observational study. medRxiv Preprint posted online on March 14, 2021. [doi: 10.1101/2021.03.11.21253225]
24. Brammer S, Branicki L, Linnenluecke MK. COVID-19, societalization, and the future of business in society. Acad Manag Perspect 2020 Nov;34(4):493-507. [doi: 10.5465/amp.2019.0053]
25. Nilsson A, Sörman K, Klingvall J, Ovelius E, Lundberg J, Hellner C. MyCompass in a Swedish context - lessons learned from the transfer of a self-guided intervention targeting mental health problems. BMC Psychiatry 2019 Jan 31;19(1):51 [FREE Full text] [doi: 10.1186/s12888-019-2039-1] [Medline: 30704424]

Abbreviations

GAD-2: Generalized Anxiety Disorder-2
PHQ-2: Patient Health Questionnaire-2

Edited by G Eysenbach; submitted 01.07.21; peer-reviewed by JC Buckey, Y Yu; comments to author 30.07.21; revised version received 19.08.21; accepted 13.09.21; published 04.10.21.

Please cite as:
Hentati A, Forsell E, Ljótsson B, Kraepelien M
Practical and Emotional Problems Reported by Users of a Self-guided Digital Problem-solving Intervention During the COVID-19 Pandemic: Content Analysis
JMIR Form Res 2021;5(10):e31722
URL: https://formative.jmir.org/2021/10/e31722
doi:10.2196/31722
PMID:34559670

Original Paper

Hospital-Based Contact Tracing of Patients With COVID-19 and Health Care Workers During the COVID-19 Pandemic in Eastern India: Cross-sectional Study

Durgesh Prasad Sahoo[1*], MBBS, MD; Arvind Kumar Singh[1*], MBBS, MD; Dinesh Prasad Sahu[1], MBBS, MD; Somen Kumar Pradhan[1], MBBS, MD; Binod Kumar Patro[1], MBBS, MD; Gitanjali Batmanabane[2], MBBS, MD; Baijayantimala Mishra[3], MBBS, MD; Bijayini Behera[3], MBBS, MD; Ambarish Das[1], MBBS; G Susmita Dora[1], MBBS, MD; L Anand[4], PhD; S M Azhar[4], MSc; Jyolsna Nair[1], MBBS; Sasmita Panigrahi[4], MSc; R Akshaya[1], MBBS; Bimal Kumar Sahoo[1], MBBS, MD; Subhakanta Sahu[1], MBBS, MD; Suchismita Sahoo[4], MSc

[1]Department of Community Medicine and Family Medicine, All India Institute of Medical Sciences, Bhubaneswar, Odisha, India

[2]All India Institute of Medical Sciences, Bhubaneswar, Odisha, India

[3]Department of Microbiology, All India Institute of Medical Sciences, Bhubaneswar, Odisha, India

[4]College of Nursing, All India Institute of Medical Sciences, Bhubaneswar, Odisha, India

[*]these authors contributed equally

Corresponding Author:
Gitanjali Batmanabane, MBBS, MD
All India Institute of Medical Sciences, Bhubaneswar
Sijua, Patrapada, Bhubaneswar
Odisha, 751019
India
Phone: 91 6742476001
Email: director@aiimsbhubaneswar.edu.in

Abstract

Background: The contact tracing and subsequent quarantining of health care workers (HCWs) are essential to minimizing the further transmission of SARS-CoV-2 infection and mitigating the shortage of HCWs during the COVID-19 pandemic situation.

Objective: This study aimed to assess the yield of contact tracing for COVID-19 cases and the risk stratification of HCWs who are exposed to these cases.

Methods: This was an analysis of routine data that were collected for the contact tracing of COVID-19 cases at the All India Institute of Medical Sciences, Bhubaneswar, in Odisha, India. Data from March 19 to August 31, 2020, were considered for this study. COVID-19 cases were admitted patients, outpatients, or HCWs in the hospital. HCWs who were exposed to COVID-19 cases were categorized, per the risk stratification guidelines, as high-risk contacts or low-risk contacts

Results: During contact tracing, 3411 HCWs were identified as those who were exposed to 360 COVID-19 cases. Of these 360 cases, 269 (74.7%) were either admitted patients or outpatients, and 91 (25.3%) were HCWs. After the risk stratification of the 3411 HCWs, 890 (26.1%) were categorized as high-risk contacts, and 2521 (73.9%) were categorized as low-risk contacts. The COVID-19 test positivity rates of high-risk contacts and low-risk contacts were 3.8% (34/890) and 1.9% (48/2521), respectively. The average number of high-risk contacts was significantly higher when the COVID-19 case was an admitted patient (number of contacts: mean 6.6) rather than when the COVID-19 case was an HCW (number of contacts: mean 4.0) or outpatient (number of contacts: mean 0.2; P=.009). Similarly, the average number of high-risk contacts was higher when the COVID-19 case was admitted in a non–COVID-19 area (number of contacts: mean 15.8) rather than when such cases were admitted in a COVID-19 area (number of contacts: mean 0.27; P<.001). There was a significant decline in the mean number of high-risk contacts over the study period (P=.003).

Conclusions: Contact tracing and risk stratification were effective and helped to reduce the number of HCWs requiring quarantine. There was also a decline in the number of high-risk contacts during the study period. This indicates the role of the implementation of hospital-based, COVID-19–related infection control strategies. The contact tracing and risk stratification approaches that were designed in this study can also be implemented in other health care settings.

(JMIR Form Res 2021;5(10):e28519) doi:10.2196/28519

KEYWORDS

COVID-19; SARS-CoV-2; risk categorization; health care personnel; virus transmission; contact tracing; pandemic; risk stratification

Introduction

With 44 million confirmed cases and over 1 million confirmed deaths affecting all countries across the world, the COVID-19 pandemic is currently the largest pandemic of the century [1]. As of August 31, 2020, 35 million COVID-19 cases and 0.06 million deaths have been reported from India [2].

By September 17, 2020, countries reported to the World Health Organization (WHO) that 14% of COVID-19 cases were health care workers (HCWs) [3]. SARS-CoV-2 infection among HCWs not only poses the risk of infection to their family members, thus contributing to community spread, but also poses this risk to other HCWs and patients. Thus, apart from stringent infection prevention and control practices for reducing the exposure to infection, the contact tracing and subsequent quarantining of HCWs are essential to minimizing further transmission. Consequently, isolation after SARS-CoV-2 infection and quarantine following exposure to a confirmed case of COVID-19 can adversely reduce the availability of human resources. To mitigate the shortage of staff in hospitals, the WHO and Centers for Disease Control and Prevention (CDC) have given recommendations for stratifying the risk following exposure into 2 categories—low-risk exposure and high-risk exposure [4,5]. The Ministry of Health and Family Welfare (MoHFW) of the Government of India has also adopted these guidelines [6].

Contact tracing is a time- and resource-intensive exercise for community settings as well as hospital settings. However, it is one of the most important methods for infectious disease prevention. In our hospital, we used different methods that were described in literature to identify people who were exposed to COVID-19 cases, like the use of closed-circuit television (CCTV) footage and duty rosters and the passive reporting of contacts by departments and via telephonic inquiry. Contact tracing by using data extracted from administrative and clinical databases, such as electronic medical records, or by using CCTV footage (a real-time locating system) has been reported previously [7,8]. Although conventional contact tracing via continuous direct observation has been considered to be the gold-standard method for accurately quantifying contact time, it requires intensive human resources and is not cost-effective [9]. Self-reporting methods can be used as alternatives to direct observation due to the lower intensity of their human resource demands; however, there is a chance of bias compromising the accuracy of the data [10].

Although the processes of contact tracing, risk stratification, and quarantine may help to reduce the transmission of infection, it is not clear whether these processes help with reducing staff shortages in an already overwhelmed health system of a resource-constrained setting. A systematic review of 22 studies concluded that an integrated strategy for contact tracing, screening, quarantine, and isolation has the potential to reduce the incidence of SARS-CoV-2 infection [11]. However, most of the studies included in this systematic review were community based. The few studies that were conducted in a health facility setting suggested using working shifts and integrating infection control practices to reduce the number of infections in health care settings [12,13]. Unlike those in community settings, very aggressive contact tracing and quarantine policies for HCWs in health care settings may be challenging to implement due to the need to balance infection control and staff shortages [14-16]. A study from India reported the beneficial effect that stratification has on minimizing staff shortages resulting from unnecessary quarantine [17]. As there is a limited amount of literature available from such settings, our study may provide information on making public health decisions in a health care setting.

In our hospital, which caters to both patients with COVID-19 and other patients, we adopted the contact tracing and risk stratification approaches described by the WHO, CDC, and MoHFW to categorize COVID-19–exposed HCWs as high-risk contacts or low-risk contacts. This study was conducted to assess the yield of hospital-based contact tracing for patients and HCWs who tested positive for COVID-19 and the risk stratification of COVID-19–exposed HCWs in the hospital—a statutory body under the aegis of the MoHFW of the Government of India. We also compared the risk categorizations of different areas (COVID-19 and non–COVID-19 areas) and different categories of index cases (outpatient department [OPD], inpatient department [IPD], and HCW cases) to assess the variations.

Methods

Study Design

This study was a process evaluation of our routine contact tracing and risk stratification mechanisms at the study site. Data from March 19 to August 31, 2020, were collected.

Study Site

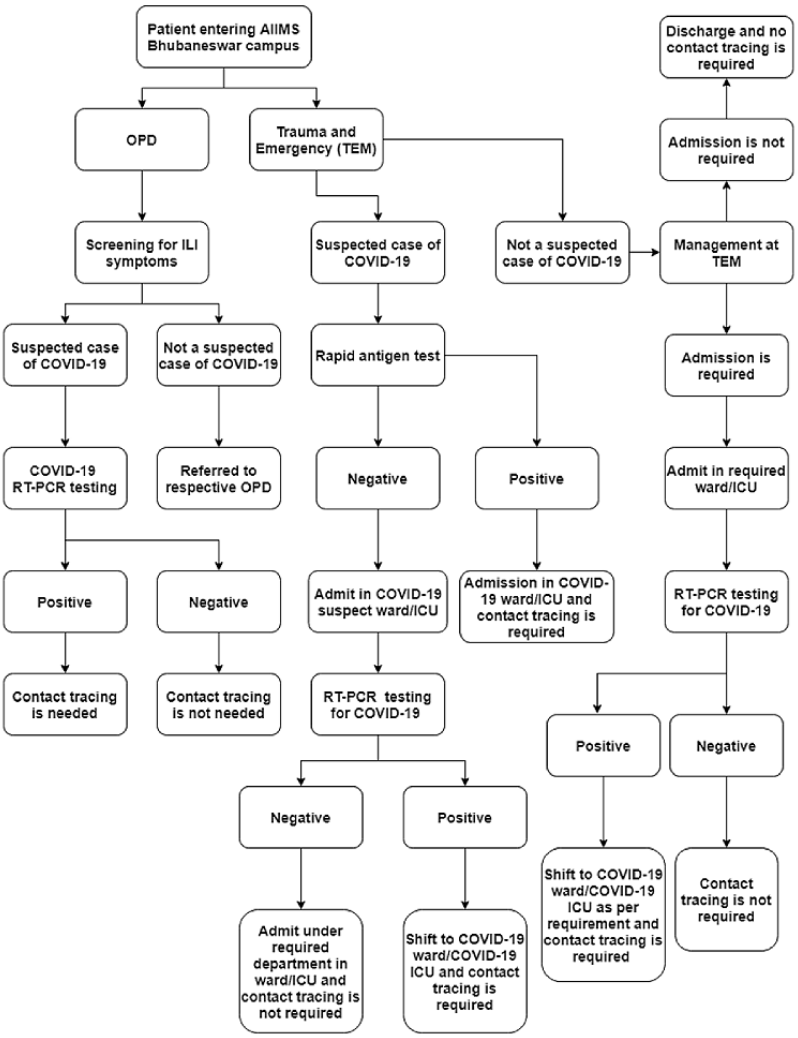
This study was conducted at the All India Institute of Medical Sciences (AIIMS), Bhubaneswar, which is a 960-bed tertiary care teaching hospital located in Bhubaneswar, the capital city of Odisha (an eastern state of India).

COVID-19–Related Clinical Services at the Study Site

Patients who were admitted to the hospital were screened for COVID-19, as per the screening algorithm depicted in Figure 1. On March 19, 2020, the first patient (the second COVID-19 case of Odisha) with COVID-19 was admitted to our hospital. The first case of COVID-19 among HCWs was reported on June 2, 2020. COVID-19 screening via reverse transcription-polymerase chain reaction (RT-PCR) tests of all newly admitted patients, irrespective of the presence of

symptoms, started on June 15, 2020. From July 10 onward, routine outpatient consultations were discontinued due to a sudden surge in the number of COVID-19 cases in the community and hospital. Hospital admission was restricted to only patients with COVID-19 and patients requiring emergency or essential intervention.

Figure 1. Algorithm of the COVID-19 testing strategy for patients admitted to the hospital. AIIMS: All India Institute of Medical Sciences; ICU: intensive care unit; ILI: influenza-like illness; OPD: outpatient department; RT-PCR: reverse transcription-polymerase chain reaction; TEM: trauma and emergency.



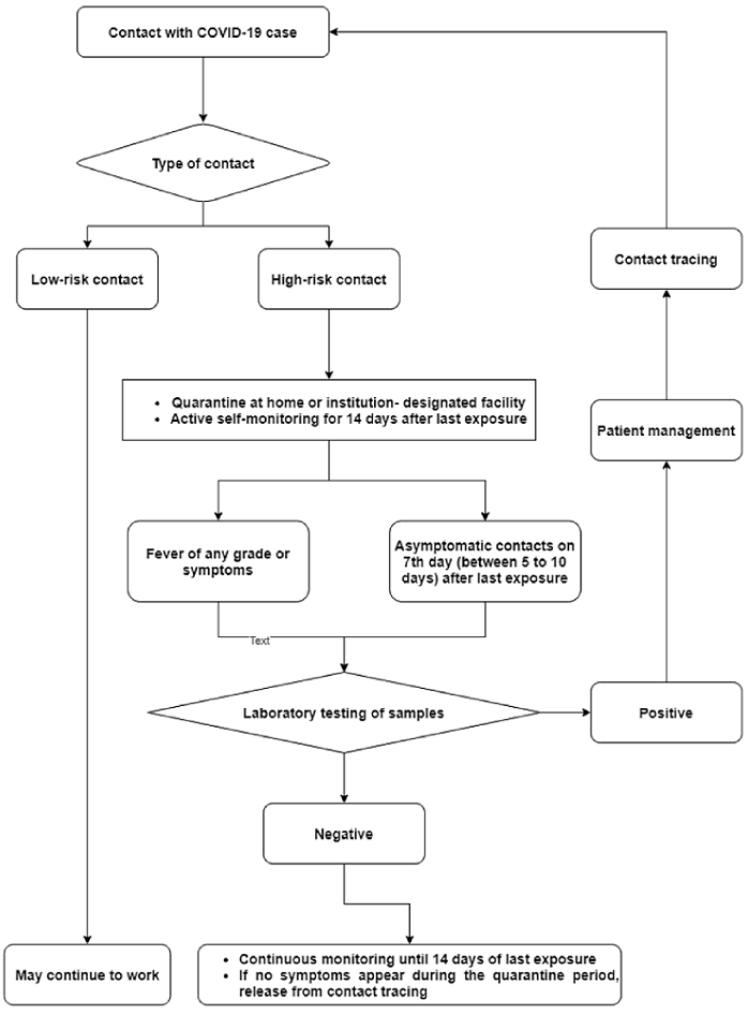
COVID-19–Related Prevention Interventions at the Study Site

Various training programs were conducted to train all cadres of HCWs in the proper use of personal protective equipment (PPE), hand hygiene measures, and other infection control practices. The use of various types of PPE in different clinical areas and hospital premises was guided by MoHFW protocols and upgraded or modified based on feedback from the contact tracing and infection control teams. Advisories were issued to all HCWs at periodic intervals for PPE compliance and infection control measures. We also introduced various behavioral and regulatory interventions to promote COVID-19–appropriate behaviors, such as a monetary penalty for not using a mask in the hospital and residential campuses.

Contact Tracing and Risk Stratification

As per the testing strategy outlined in Figure 2, testing for SARS-CoV-2 infection via the RT-PCR method was performed at the COVID-19 RT-PCR testing laboratory of the institute—an approved laboratory of the Indian Council of Medical Research of the Government of India. Patients presenting with symptoms that were consistent with COVID-19 and confirmed cases of COVID-19 were admitted in separate wards, which were referred to as *COVID-19 areas*. Patients with COVID-19 were categorized as inpatients, outpatients, or HCWs who tested positive for COVID-19. Inpatients were further categorized based on the area in which they were admitted (ie, COVID-19 areas or non–COVID-19 areas).

Figure 2. SARS-CoV-2 testing strategy for the health care workers after contact tracing.



Contact tracing was initiated by a team that was dedicated to performing contact tracing immediately after intimation from the diagnostic laboratory. Initially, contact tracing was done by physically visiting the clinical areas, personally interviewing the HCWs involved in patient care, reviewing medical records of patients and duty rosters, and viewing CCTV footage. However, this strategy was modified, due to the increase in the number of COVID-19 cases, to include a passive mechanism for contact tracing. In the later phase (from July 15 onward), the contact tracing team (CTT) directed the concerned departments to provide a list of all HCWs who had possibly come in contact with confirmed COVID-19 cases in a prescribed format. Upon obtaining the list of COVID-19–exposed HCWs, the CTT contacted each HCW telephonically to elicit histories related to the durations and types of exposures, the procedures performed on the patient, and the use of PPE during exposures. Data were collected by using a semistructured interview schedule. For cases of contact tracing related to an HCW who tested positive for COVID-19, histories related to interactions that occurred during duty break hours, during meals, and in places where HCWs are likely to be less cautious in terms of mask usage were probed during contact tracing. Exposures the occurred during the last 14 days from the date of a positive report were considered for contact tracing. The numbers of contacts were separately calculated for each positive case.

Risk categorizations (low-risk exposure and high-risk exposure) based on the criteria adopted from the WHO, CDC, and MoHFW guidelines are given in Textbox 1. A 14-day home quarantine and COVID-19 testing, which was to be conducted on the seventh day after an HCW's most recent exposure, were recommended for HCWs with high-risk exposures, whereas HCWs with low-risk exposures were recommended to continue their work. The quarantine period was considered to be a fully paid, on-duty period. Both risk categories were required to monitor symptoms and report on COVID-19 tests that were performed upon the appearance of symptoms consistent with COVID-19. In the absence of symptoms, routine testing was not recommended for low-risk contacts. However, a few HCWs with low-risk exposures were also tested upon their request. We collated data related to contact tracing and risk categorization in an Excel spreadsheet, and follow-ups of HCWs were done to inquire about symptoms and test results. The CTT regularly updated hospital authorities about their findings related to breaches in infection control practices and areas of high-risk contact and suggested specific recommendations.

Textbox 1. Risk categorization (low-risk exposure and high-risk exposure) of the health care workers who were exposed to patients who tested positive for COVID-19. The criteria were adopted from the World Health Organization, Centers for Disease Control and Prevention, and Ministry of Health and Family Welfare guidelines.

High-risk contact

- Touched body fluids of a patient (eg, touching respiratory tract secretions, blood, vomit, saliva, urine, and feces; being coughed on; touching used paper tissues with a bare hand; etc)
- Had direct physical contact with the body of a patient, including during physical examinations without personal protective equipment
- Touched or cleaned the linens, clothes, or dishes of a patient
- Lives in the same household as a patient
- Anyone who was in close proximity (within 1 meter) to a confirmed COVID-19 case and did not take precautions
- Passengers (ie, those in a vehicle) who were in close proximity (for more than 6 hours) to a symptomatic person who later tested positive for COVID-19

Low-risk contact

- Shared the same space (worked in same room or a similar situation) but did not have a high-risk exposure to a confirmed case of COVID-19
- Traveled in the same environment (bus, train, flight, or any other mode of transit) but did not have a high-risk exposure

Ethical approval to conduct this study was obtained from the Institutional Ethics Committee of AIIMS, Bhubaneswar (reference number: T/IM-NF/CMFM/20/76). Individual participant consent was not obtained, as contact tracing was a regular process for risk stratification among the HCWs. All HCWs were instructed by the hospital authorities to cooperate with the CTT.

Statistical Analysis

Statistical analyses were conducted by using Microsoft Excel 2013 and SPSS version 22.0 (IBM Corporation). Descriptive statistics were presented as means with SDs and percentages with 95% CIs. The mean number and SD of high-risk contacts and low-risk contacts among the types of patients (ie, admitted patients in a COVID-19 area, admitted patients in a non–COVID-19 area, outpatients, and HCWs) were compared. A *P* value of <.05 was considered statistically significant. We also compared the mean number of high-risk and low-risk contacts for a block of 15 days within the study period.

Results

Our analysis included data related to 360 COVID-19 cases that were reported during the study period, which included 240 (66.7%) admitted patients and IPD patients, 29 (8.1%) OPD patients, and 91 (25.3%) HCWs. Of the 269 IPD and OPD patients, 163 (60.6%) were admitted directly to a COVID-19 area, 97 (36.1%) were admitted in a non–COVID-19 area, and the rest (n=9, 3.3%) had stayed in both COVID-19 and non–COVID-19 areas (Table 1).

Table 1. Distribution of patients who tested positive for COVID-19 in the hospital from March to August 2020.

Types of patients and areas	Patients, n (%)
Type of patient (n=360)	
Inpatient department patients	240 (66.7)
Outpatient department patients	29 (8)
Health care workers	91 (25.3)
Type of area (excluding staff; n=269)	
COVID-19 area	163 (60.6)
Non–COVID-19 area	97 (36.1)
Both	9 (3.3)

The CTT identified 3411 HCWs who were exposed to any COVID-19 case in the hospital. After risk categorization, 26.1% (890/3411) of HCWs were identified as high-risk contacts, and 73.9% (2521/3411) were identified as low-risk contacts. Within 14 days of exposure to a COVID-19 case, 34 out of the 890 high-risk contacts (3.8%; 95% CI 2.7%-5.2%) and 48 out of the 2521 low-risk contacts (1.9%; 95% CI 1.4%-2.5%) tested positive for SARS-CoV-2 infection. However, among the low-risk contacts, only symptomatic HCWs were tested, and the test positivity rate among the symptomatic low-risk contacts was 48 out of 1583 (3.03%; 95% CI 2.24%-4.00%).

The mean number of high-risk contacts was 15.8 (SD 18.3) when a COVID-19 case was admitted in a non–COVID-19 area and 4.0 (SD 5.6) when the COVID-19 case was an HCW. The mean number of high-risk contacts per patient was <1 if a patient was admitted in a COVID-19 area or was provided with services on an outpatient basis. The difference in the mean numbers of

high-risk contacts among the different groups was statistically significant (*P*<.001; Table 2).

Table 2. Comparison of the average number of high-risk and low-risk contacts, with respect to the type of index case, in the hospital from March to August 2020.

Types of patients and areas	Number of cases	Number of contacts, mean (SD)	*t* test[a] (df) or ANOVA[b] test (df)	*P* value
Type of patient				
High-risk contact			4.741 (2)[c]	.009
Inpatient department patients	240	6.61 (13.895)		
Outpatient department patients	29	0.22 (0.698)		
Health care workers	91	4.02 (5.653)		
Low-risk contact			8.527 (2)[c]	.002
Inpatient department patients	240	10.81 (11.754)		
Outpatient department patients	29	3.07 (2.541)		
Health care workers	91	8.12 (6.789)		
Type of area				
High-risk contact			−10.853 (258)[d]	<.001
COVID-19 area	163	0.27 (1.207)		
Non–COVID-19 area	97	15.84 (18.268)		
Low-risk contact			−7.803 (258)[d]	<.001
COVID-19 area	163	5.93 (5.544)		
Non–COVID-19 area	97	16.19 (15.188)		

[a]A 2-tailed unpaired *t* test.

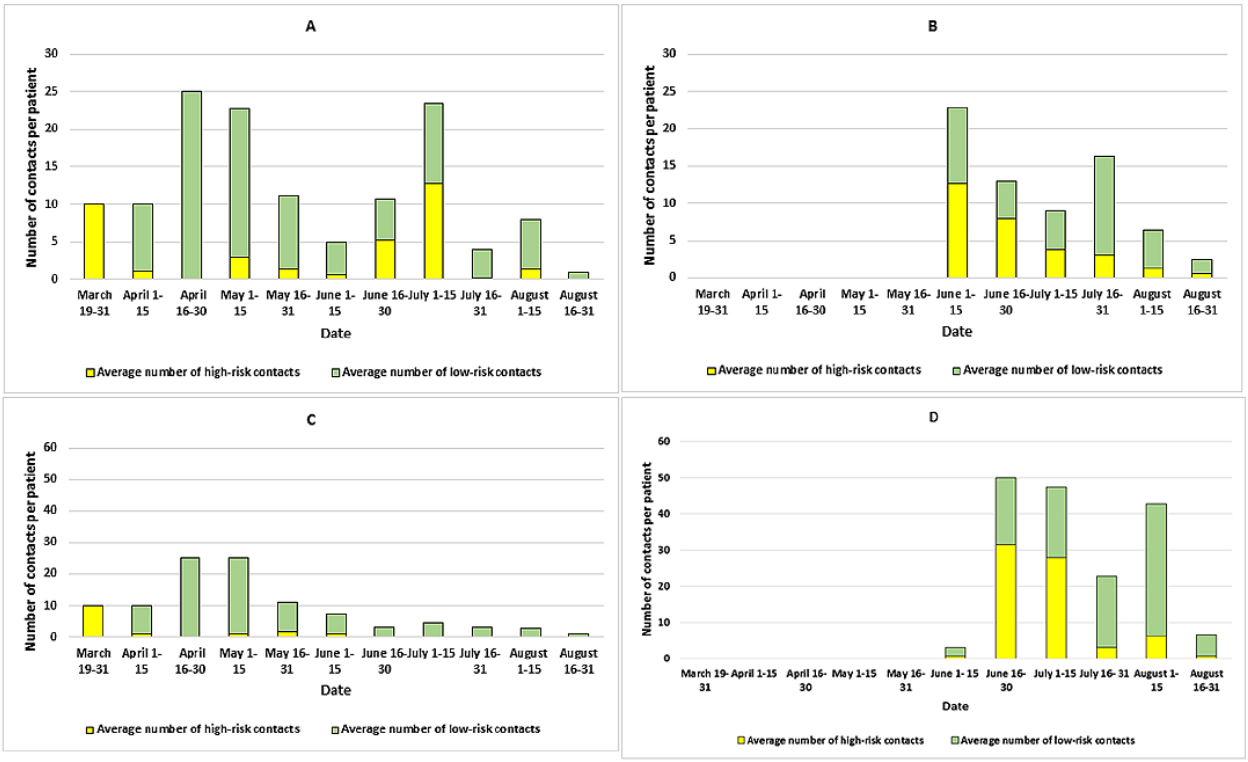
[b]ANOVA: analysis of variance.

[c]Analysis of variance test value.

[d]2-tailed unpaired *t* test value.

A significant decline in the mean number of high-risk contacts was reported over the study period (*P*=.003). In cases where an HCW was the index case, the mean number of high-risk contacts decreased from 12.7 (during June 1 to June 15, 2020) to 3.7 during July 1 to July 15, 2020, and to 0.62 during August 16 to August 31, 2020. The first case of an HCW of the hospital testing positive for COVID-19 occurred on June 2, 2020 (Figure 3). Similarly, in COVID-19 areas, the mean number of high-risk contacts was 10.0 until March 31, 2020, and these contacts involved the only patient who was admitted in a COVID-19 area at that time. Afterward, the mean number of high-risk contacts decreased to 1.0 during April 1 to June 15, 2020, and to <0.1 from June 15, 2020, onward. In the non–COVID-19 area, the mean number of high-risk contacts decreased from 31.5 (during June 16 to June 30, 2020) to 3.0 during July 16 to July 31, 2020, and to 0.7 during August 16 to August 31, 2020 (Figure 3).

Figure 3. A: Average number of contacts when the COVID-19 case was an admitted patient (March to August 2020). B: Average number of contacts when the COVID-19 case was a health care worker (March to August 2020). C: Average number of contacts when the COVID-19 case was admitted in a COVID-19 area (March to August 2020). D: Average number of contacts when the COVID-19 case was admitted in a non–COVID-19 area (March to August 2020).



Interviews with HCWs, which were conducted during contact tracing, revealed that common causes for a high-risk exposure during the provision of clinical care were the inadequate use of PPE and the nonpracticing of hand hygiene measures after having direct contact with a patient. Further, HCWs who tested positive for COVID-19 indicated that social interactions during meals and at nursing stations during duty hours, handover, travel, and the act of staying together were major contributing factors (24/91, 26%).

Discussion

Summary of Results

Our forward contact tracing of 360 COVID-19 cases, who were either patients or HCWs, resulted in the identification of 3411 exposures. After risk stratification, 26.1% (890/3411) of HCWs were categorized as high-risk contacts, and 73.9% (2521/3411) were categorized as low-risk contacts. Of the 890 high-risk contacts and 2521 low-risk contacts, 34 (3.8%) and 48 (1.9%), respectively, tested positive for SARS-COV-2 infection. We also observed a gradual decline in the average number of high-risk contacts over a period of time. HCWs were more likely to be exposed to SARS-CoV-2 infection when it was diagnosed among HCWs and patients who were admitted in a non–COVID-19 area.

Comparison With Existing Literature

A few studies from India have also reported similar proportions of high-risk contacts after risk stratification. According to a study conducted by Agarwal et al [18], in a hospital located in the eastern part of India, 7/28 (25%) of the COVID-19–exposed HCWs were high-risk contacts. However, this study had a small sample size. Another study, which was conducted by Kaur et al [17] in a hospital located in the northern part of India, reported that only 14.5% of the COVID-19–exposed HCWs were categorized as high-risk contacts. Our study reported a COVID-19 test positivity rate of 3.8% (34/890) among high-risk contacts, while the study conducted by Kaur et al [17] observed a higher test positivity rate (7.1%). In the study by Kaur et al [17], most of the COVID-19 cases (almost 85%) were HCWs, unlike in our study, where 25.3% (91/360) of COVID-19 cases were HCWs. Moreover, the higher COVID-19 test positivity rate in the Kaur et al [17] study could have been due to having more stringent criteria for stratifying HCWs into the high-risk category, as the proportion of high-risk contacts among HCWs was 14% in their study and 26.1% (890/3411) in our study. Another study conducted by Blain et al [19] reported a high COVID-19 test positivity rate among health care personnel (23.5%), but the study was only conducted for 3 COVID-19–positive cases. In our study, 1 index case had 10 high-risk contacts on average, while a study conducted by Vera et al [20] in Switzerland reported 21 high-risk contacts, which was much higher than the number reported in our study. The study from Switzerland was based on just 1 initially undiagnosed COVID-19 case.

We also observed a clear difference in the COVID-19 test positivity rates between high-risk contacts and low-risk contacts (34/890, 3.8% vs 48/2521, 1.9%), which demonstrated the

effectiveness of the risk stratification strategy. However, among the low-risk contacts who were tested, the test positivity rate was 3.03% (48/1583). Due to the very high number of contacts, all of the high-risk contacts and only the symptomatic low-risk contacts were tested. Moreover, low-risk contacts could have been exposed to SARS-CoV-2 infection outside of the hospital because, unlike high-risk contacts, they were not quarantined and continued to work. The effectiveness of contact tracing was observed in previous studies [11,14,21].

The mean number of high-risk contacts was highest when a patient was admitted in the non–COVID-19 area (number of contacts: mean 15.8) rather than when a patient was admitted in a COVID-19 area (number of contacts: mean 0.27). The mean number of high-risk contacts was higher in non–COVID-19 areas probably because the recommended level of protection in non–COVID-19 areas is different from that of COVID-19 areas. Similarly, HCWs' attitudes toward following the protocol might be better in COVID-19 areas due to the higher perceived risk. In COVID-19 areas, HCWs were completely equipped with PPE. In the non-COVID-19 area however, they were only equipped with surgical masks, N95 masks, and gloves, as per the guidelines proposed by the MoHFW, WHO, and CDC, and admitted patients were not suspected of SARS-CoV-2 infection [22-24]. Thus, stringent infection prevention and control measures also need to be adopted in areas where patients who are not suspected of SARS-CoV-2 infection are admitted. Similarly, the number of high-risk contacts was higher when the COVID-19 case was an admitted patient rather than when such cases were HCWs and outpatients. In the IPD, the HCWs perform their duties in a shift-wise manner (3 shifts in 1 day) and perform procedures. In the OPD however, HCWs only have 1 shift per day, and this can result in a fewer number of exposures. Further, the large number of high-risk contacts among COVID-19–exposed HCWs was also the result of social mixing with colleagues during duty time and in residential areas. The importance of workplace social distancing and contact tracing was mentioned in studies conducted by Ahmed et al [25] and Kretzschmar et al [26]. In a community-based study from the United States, the mean number of contacts was 2.4, whereas in our study, it was as high as 10.81 for IPD patients and as low as 0.22 for OPD patients [27]. The reason for this could be the difference in study settings. Our study was conducted in a tertiary care hospital, whereas the Miller et al [27] study was conducted among community participants.

There was also a significant reduction in the number of high-risk contacts for all categories of COVID-19 cases (ie, cases in COVID-19 areas, cases in non–COVID-19 areas, IPD cases, OPD cases, and HCW cases) over the study period (P=.003). This reflects the timely modification of infection control measures, the strict implementation of PPE protocols, and the effectiveness of providing infection control–related training to HCWs in the hospital. Similar results were observed in a study that was conducted by Hidayat et al [28] in an Indonesian COVID-19 referral hospital, where the secondary attack rate among HCWs declined from 20.1% to 3.7% over time [28]. However, at the same time, there was a decline in the total number of COVID-19–exposed contacts (low-risk and high-risk contacts combined). This might be indicative of either contact tracing activity–related fatigue in both the CTT and the departments where COVID-19 cases were detected or human resource rationing (this included the modification of duty rosters and fewer rotations of HCWs among different units), which was increasingly performed as the pandemic progressed. Thus, the implementation of strict infection prevention and control measures, PPE protocols, and continuous contact tracing can play a role in mitigating the shortage of human resources. The effectiveness of contact tracing was mentioned in a study by Stuart et al [29].

The CTT also provided regular feedback (based on inquiries from COVID-19–exposed HCWs) to the hospital administration to augment infection control measures, identified areas in which frequent breaches in protocols occurred, and suggested a mechanism for reducing the number of exposures to COVID-19. Apart from quarantine, regular feedback–based action might have helped to reduce the number of exposures to SARS-CoV-2 infection in the hospital.

Strengths

Since multiple strategies were used, such as visiting the clinical area, conducting personal interviews with the HCWs, reviewing medical records, and viewing CCTV footage, we believe that all of the possible contacts were listed, tracked, and categorized properly, as these strategies were performed by trained personnel and verified by experts. Thus, the quality of the data was expected to be satisfactory. Testing for COVID-19 was performed in an Indian Council of Medical Research–approved testing center via RT-PCR, which is considered to be the gold-standard test. All high-risk cases were continuously monitored for 14 days after their most recent exposure to SARS-CoV-2 infection, and COVID-19 testing was performed on the seventh day.

Limitations

The categorization of risk was based on the histories of the contacts, which may have increased the chances of social desirability bias affecting our results. Our data might have included misinformation, as hospital staff might have deliberately wanted to be categorized as high-risk contacts, so that they could be quarantined for 14 days and still be fully paid. There was also a chance that HCWs recalled incorrect information. Sometimes, the HCWs failed to remember patients' SARS-CoV-2 infection status and their own PPE status during patient care. Further, low-risk contacts were not routinely tested unless they were symptomatic. Therefore, we could have missed some cases, as many COVID-19 cases remain asymptomatic or paucisymptomatic.

Conclusions

Contact tracing and risk stratification were effective and helped to reduce the number of HCWs requiring quarantine. There was a decline in the number of high-risk contacts during the study period. This indicates the role of the implementation of hospital-based COVID-19–related infection control strategies. The findings obtained during contact tracing might also be beneficial for developing appropriate and strategic infection control measures. The contact tracing and risk stratification

approaches that were designed in this study can also be implemented in other health care settings.

Acknowledgments

We are grateful to the health care workers and the hospital administration for their cooperation during contact tracing.

Authors' Contributions

DP Sahoo, AKS, BKP, and GB conceptualized this study. DP Sahoo, AKS, BKP, DP Sahu, SKP, and GB developed the methodology. DP Sahoo, AKS, BKP, DP Sahu, and SKP curated the data. DP Sahoo, SKP, DP Sahu, BM, BB, AD, GSD, LA, SMA, JN, SP, RA, BKS, S Sahu, and S Sahoo performed the investigation. DP Sahoo and AKS conducted the formal analysis. AKS, BKP, and GB supervised this study. DP Sahoo and AKS wrote the original draft. BKP, DP Sahu, SKP, GB, BM, BB, AD, GSD, LA, SMA, JN, SP, RA, BKS, S Sahu, and S Sahoo wrote, reviewed, and edited this paper.

Conflicts of Interest

None declared.

References

1. WHO coronavirus (COVID-19) dashboard. World Health Organization. URL: https://covid19.who.int/ [accessed 2021-08-21]
2. #IndiaFightsCorona COVID-19 in India, vaccination, dashboard, corona virus tracker. Government of India. URL: https://www.mygov.in/covid-19 [accessed 2021-08-12]
3. Keep health workers safe to keep patients safe: WHO. World Health Organization. 2020 Sep 17. URL: https://www.who.int/news/item/17-09-2020-keep-health-workers-safe-to-keep-patients-safe-who [accessed 2021-01-12]
4. Contact tracing in the context of COVID-19. World Health Oranization. 2021 Feb 01. URL: https://www.who.int/publications/i/item/contact-tracing-in-the-context-of-covid-19 [accessed 2021-10-06]
5. Risk assessment and management of exposure of health care workers in the context of COVID-19: Interim guidance, 19 March 2020. World Health Organization. 2020 Mar 19. URL: https://apps.who.int/iris/bitstream/handle/10665/331496/WHO-2019-nCov-HCW_risk_assessment-2020.2-eng.pdf?sequence=1&isAllowed=y [accessed 2021-10-06]
6. Micro plan for containing local transmission of coronavirus disease (COVID-19). Ministry of Health and Family Welfare, Government of India. URL: https://www.mohfw.gov.in/pdf/ModelMicroplanforcontainmentoflocaltransmissionofCOVID19.pdf [accessed 2021-10-06]
7. Alamo ST, Wagner GJ, Sunday P, Wanyenze RK, Ouma J, Kamya M, et al. Electronic medical records and same day patient tracing improves clinic efficiency and adherence to appointments in a community based HIV/AIDS care program, in Uganda. AIDS Behav 2012 Feb;16(2):368-374 [FREE Full text] [doi: 10.1007/s10461-011-9996-9] [Medline: 21739285]
8. Bowen ME, Wingrave CA, Klanchar A, Craighead J. Tracking technology: lessons learned in two health care sites. Technol Health Care 2013;21(3):191-197. [doi: 10.3233/THC-130738] [Medline: 23792793]
9. Jiang L, Ng HL, Ho HJ, Leo YS, Prem K, Cook AR, et al. Contacts of healthcare workers, patients and visitors in general wards in Singapore. Epidemiol Infect 2017 Oct;145(14):3085-3095. [doi: 10.1017/S0950268817002035] [Medline: 28885136]
10. Farrahi K, Emonet R, Cebrian M. Epidemic contact tracing via communication traces. PLoS One 2014 May 01;9(5):e95133. [doi: 10.1371/journal.pone.0095133] [Medline: 24787614]
11. Girum T, Lentiro K, Geremew M, Migora B, Shewamare S. Global strategies and effectiveness for COVID-19 prevention through contact tracing, screening, quarantine, and isolation: a systematic review. Trop Med Health 2020 Nov 23;48(1):91 [FREE Full text] [doi: 10.1186/s41182-020-00285-w] [Medline: 33292755]
12. Cheng VCC, Wong SC, Chen JHK, Yip CCY, Chuang VWM, Tsang OTY, et al. Escalating infection control response to the rapidly evolving epidemiology of the coronavirus disease 2019 (COVID-19) due to SARS-CoV-2 in Hong Kong. Infect Control Hosp Epidemiol 2020 May;41(5):493-498 [FREE Full text] [doi: 10.1017/ice.2020.58] [Medline: 32131908]
13. Bai Y, Wang X, Huang Q, Gurarie D, Ndeffo-Mbah M, Fan F, et al. SARS-CoV-2 infection in health care workers: a retrospective analysis and a model study. medRxiv Preprint posted online on April 1, 2020. [FREE Full text] [doi: 10.1101/2020.03.29.20047159]
14. Keeling MJ, Hollingsworth TD, Read JM. Efficacy of contact tracing for the containment of the 2019 novel coronavirus (COVID-19). J Epidemiol Community Health 2020 Oct;74(10):861-866 [FREE Full text] [doi: 10.1136/jech-2020-214051] [Medline: 32576605]
15. Breeher L, Boon A, Hainy C, Murad MH, Wittich C, Swift M. A framework for sustainable contact tracing and exposure investigation for large health systems. Mayo Clin Proc 2020 Jul;95(7):1432-1444 [FREE Full text] [doi: 10.1016/j.mayocp.2020.05.008] [Medline: 32561146]
16. Kaito D, Matsumura K, Yamamoto R. Hospital preparedness for COVID-19: The known and the known unknown. Keio J Med 2021 Jun 25;70(2):25-34. [doi: 10.2302/kjm.2020-0011-OA] [Medline: 32830154]

17. Kaur R, Kant S, Bairwa M, Kumar A, Dhakad S, Dwarakanathan V, et al. Risk stratification as a tool to rationalize quarantine of health care workers exposed to COVID-19 Cases: Evidence from a tertiary health care center in India. Asia Pac J Public Health 2021 Jan;33(1):134-137. [doi: 10.1177/1010539520977310] [Medline: 33246367]
18. Agarwal N, Biswas B, Nair R. Ensuring safety and availability of healthcare workers in the era of COVID-19: An experience from the Eastern State of India. Indian J Med Sci 2020 Aug 21;72(2):58-64 [FREE Full text] [doi: 10.25259/ijms_128_2020]
19. Blain H, Rolland Y, Tuaillon E, Giacosa N, Albrand M, Jaussent A, et al. Efficacy of a test-retest strategy in residents and health care personnel of a nursing home facing a COVID-19 outbreak. J Am Med Dir Assoc 2020 Jul;21(7):933-936 [FREE Full text] [doi: 10.1016/j.jamda.2020.06.013] [Medline: 32674822]
20. Canova V, Schläpfer HL, Piso RJ, Droll A, Fenner L, Hoffmann T, et al. Transmission risk of SARS-CoV-2 to healthcare workers -observational results of a primary care hospital contact tracing. Swiss Med Wkly 2020 Apr 20;150:w20257 [FREE Full text] [doi: 10.4414/smw.2020.20257] [Medline: 32333603]
21. Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, Centre for Mathematical Modelling of Infectious Diseases COVID-19 working group. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. Lancet Infect Dis 2020 May;20(5):553-558 [FREE Full text] [doi: 10.1016/S1473-3099(20)30144-4] [Medline: 32171059]
22. Using personal protective equipment (PPE). Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/using-ppe.html [accessed 2020-05-12]
23. Rational use of personal protective equipment (PPE) for coronavirus disease (COVID-19). World Health Organization. URL: https://apps.who.int/iris/bitstream/handle/10665/331498/WHO-2019-nCoV-IPCPPE_use-2020.2-eng.pdf [accessed 2020-04-25]
24. Novel coronavirus disease 2019 (COVID-19): Guidelines on rational use of personal protective equipment. Ministry of Health and Family Welfare. URL: https://www.mohfw.gov.in/pdf/GuidelinesonrationaluseofPersonalProtectiveEquipment.pdf [accessed 2020-05-05]
25. Ahmed F, Zviedrite N, Uzicanin A. Effectiveness of workplace social distancing measures in reducing influenza transmission: a systematic review. BMC Public Health 2018 Apr 18;18(1):518 [FREE Full text] [doi: 10.1186/s12889-018-5446-1] [Medline: 29669545]
26. Kretzschmar ME, Rozhnova G, van Boven M. Isolation and contact tracing can tip the scale to containment of COVID-19 in populations with social distancing. Front Phys 2021 Feb 10;8:1-11 [FREE Full text] [doi: 10.3389/fphy.2020.622485]
27. Miller JS, Bonacci RA, Lash RR, Moonan PK, Houck P, Van Meter JJ, et al. COVID-19 case investigation and contact tracing in Central Washington State, June-July 2020. J Community Health 2021 Oct;46(5):918-921 [FREE Full text] [doi: 10.1007/s10900-021-00974-5] [Medline: 33689116]
28. Hidayat R, Aini N, Ilmi AFN, Azzahroh F, Giantini A. Test, trace, and treatment strategy to control COVID-19 infection among hospital staff in a COVID-19 referral hospital in Indonesia. Acta Med Indones 2020 Jul;52(3):206-213 [FREE Full text] [Medline: 33020332]
29. Stuart RL, Zhu W, Morand EF, Stripp A. Breaking the chain of transmission within a tertiary health service: An approach to contact tracing during the COVID-19 pandemic. Infect Dis Health 2021 May;26(2):118-122 [FREE Full text] [doi: 10.1016/j.idh.2020.11.003] [Medline: 33281108]

Abbreviations

AIIMS: All India Institute of Medical Sciences
CCTV: closed-circuit television
CDC: Centers for Disease Control and Prevention
CTT: contact tracing team
HCW: health care worker
IPD: inpatient department
MoHFW: Ministry of Health and Family Welfare
OPD: outpatient department
PPE: personal protective equipment
RT-PCR: reverse transcription-polymerase chain reaction
WHO: World Health Organization

Edited by G Eysenbach; submitted 05.03.21; peer-reviewed by A Lahiri, MA Bahrami, N Pawar; comments to author 02.04.21; revised version received 15.08.21; accepted 18.09.21; published 21.10.21.

Please cite as:
Sahoo DP, Singh AK, Sahu DP, Pradhan SK, Patro BK, Batmanabane G, Mishra B, Behera B, Das A, Dora GS, Anand L, Azhar SM, Nair J, Panigrahi S, Akshaya R, Sahoo BK, Sahu S, Sahoo S
Hospital-Based Contact Tracing of Patients With COVID-19 and Health Care Workers During the COVID-19 Pandemic in Eastern India: Cross-sectional Study
JMIR Form Res 2021;5(10):e28519
URL: https://formative.jmir.org/2021/10/e28519
doi:10.2196/28519
PMID:34596569

Original Paper

Perceptions and Feelings of Brazilian Health Care Professionals Regarding the Effects of COVID-19: Cross-sectional Web-Based Survey

Roberta Pires Corrêa[1], MSc; Helena Carla Castro[1,2], Prof Dr; Bruna Maria Castro Salomão Quaresma[3], PhD; Paulo Roberto Soares Stephens[1,4], Prof Dr; Tania Cremonini Araujo-Jorge[1,4], Prof Dr; Roberto Rodrigues Ferreira[1,3,4], Prof Dr

[1]Program in Education in Biosciences and Health, Oswaldo Cruz Institute, Oswaldo Cruz Foundation (Fiocruz), Rio de Janeiro, Brazil

[2]Sciences, Technology and Inclusion, Federal Fluminense University, Niterói, Brazil

[3]Laboratory of Functional Genomics and Bioinformatics, Oswaldo Cruz Institute, Oswaldo Cruz Foundation (Fiocruz), Rio de Janeiro, Brazil

[4]Laboratory of Innovations in Therapies, Education and Bioproducts, Oswaldo Cruz Institute, Oswaldo Cruz Foundation (Fiocruz), Rio de Janeiro, Brazil

Corresponding Author:
Roberto Rodrigues Ferreira, Prof Dr
Laboratory of Innovations in Therapies, Education and Bioproducts
Oswaldo Cruz Institute
Oswaldo Cruz Foundation (Fiocruz)
Av Brasil, 4365
Rio de Janeiro
Brazil
Phone: 55 21 2562 1295
Email: robertoferreira@ioc.fiocruz.br

Abstract

Background: The importance of health professionals has been recognized in COVID-19 pandemic–affected countries, especially in those such as Brazil, which is one of the top 3 countries that have been affected in the world. However, the workers' perception of the stress and the changes that the pandemic has caused in their lives vary according to the conditions offered by these affected countries, including salaries, individual protection equipment, and psychological support.

Objective: The purpose of this study was to identify the perceptions of Brazilian health workers regarding the COVID-19 pandemic impact on their lives, including possible self-contamination and mental health.

Methods: This cross-sectional web-based survey was conducted in Brazil by applying a 32-item questionnaire, including multiple-choice questions by using the Google Forms electronic assessment. This study was designed to capture spontaneous perceptions from health professionals. All questions were mandatory and divided into 2 blocks with different proposals: personal profile and COVID-19 pandemic impact.

Results: We interviewed Brazilian health professionals from all 5 Brazilian regions (N=1376). Our study revealed that 1 in 5 (23%) complained about inadequate personal protective equipment, including face shields (234/1376, 17.0%), masks (206/1376, 14.9%), and laboratory coats (138/1376, 10.0%), whereas 1 in 4 health professionals did not have enough information to protect themselves from the coronavirus disease. These professionals had anxiety due to COVID-19 (604/1376, 43.9%), difficulties in sleep (593/1376, 43.1%), and concentrating on work (453/1376, 32.9%). Almost one-third experienced traumatic situations at work (385/1376, 28.0%), which may have led to negative feelings of *fear of COVID-19* and *sadness*. Despite this situation, there was *hope* and *empathy* among their positive feelings. The survey also showed that 1 in 5 acquired COVID-19 with the most classic and minor symptoms, including headache (274/315, 87.0%), body pain (231/315, 73.3%), tiredness (228/315, 72.4%), and loss of taste and smell (208/315, 66.0%). Some of their negative feelings were higher than those of noninfected professionals (fear of COVID-19, 243/315, 77.1% vs 509/1061, 48.0%; impotence, 142/315, 45.1% vs 297/1061, 28.0%; and fault, 38/315, 12.1% vs 567/1061, 53.4%, respectively). Another worrying outcome was that 61.3% (193/315) reported acquiring an infection while working at a health facility and as expected, most of the respondents felt affected (344/1376, 25.0%) or very affected (619/1376, 45.0%) by the COVID-19.

Conclusions: In Brazil, the health professionals were exposed to a stressful situation and to the risk of self-contamination—conditions that can spell future psychological problems for these workers. Our survey findings showed that the psychological support for this group should be included in the future health planning of Brazil and of other hugely affected countries to assure a good mental health condition for the medical teams in the near future.

(JMIR Form Res 2021;5(10):e28088) doi:10.2196/28088

KEYWORDS

COVID-19; SARS-CoV-2; health professionals; Brazil; pandemic; mental health; health planning

Introduction

In December 2019, Chinese authorities notified the World Health Organization (WHO) of several cases of pneumonia of unknown etiology in the city of Wuhan [1]. In January 2020, a new coronavirus, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was identified from a patient's throat swab sample [2,3], and the WHO named the disease as "Coronavirus Disease 2019 (COVID-19)" [4,5]. On January 30, 2020, 7736 cases were confirmed in China while 82 confirmed cases were detected in 18 other countries [5,6]. On this same day, WHO declared the SARS-CoV-2 outbreak as a global health emergency [7].

The first case of COVID-19 in South America was described in Brazil in February 2020. It was a man returning from a trip to Italy, where a significant outbreak was ongoing [8]. Since then, the pandemic spread fast in Brazil, producing an emergency state. To control COVID-19, the Brazilian Ministry of Health recommended measures of social distancing, use of masks, and hand hygiene [9]. The disease spread in large capitals, followed by an increase of COVID-19 cases in smaller cities and poorer communities as well [10]. In the middle of the COVID-19 outbreak, Brazil was considered the second most affected country. On January 25, 2021, Brazil was considered as the third country with the highest number of COVID-19 cases worldwide (8.8 million) by WHO, behind India (10.6 million) and United States of America (24.7 million) [11].

Based on the report of the first 425 confirmed cases in Wuhan, the common symptoms detected included fever, dry cough, myalgia and fatigue, headache, hemoptysis, abdominal pain, and diarrhea [12]. Furthermore, studies have reported severe cases of COVID-19 with pneumonia, intestinal, liver, thrombotic, and neuronal diseases, acute respiratory distress, multiple organ failure, and death [13,14]. Because of the efforts of different countries and pharmaceutical industries, the production of more than 5 types of vaccines started and is being slowly distributed worldwide. Meanwhile, there are no specific therapies available for those already infected, who have access only to support medical assistance [15].

In the affected countries, including Brazil, a rapid increase in the demand for health services occurred, mainly for hospital beds in intensive care units [16]. The pandemic has severely affected the way of living of many people and has disrupted the already precarious health system in several countries [17]. The historic challenges regarding an insufficient number of health professionals [18] and the increase in confirmed cases led to overburdening of these individuals. COVID-19 changed not only the daily routine of business, schools, lifestyle, and economics but also profoundly changed routines inside hospitals, some of which now may not attend to diseases other than COVID-19 owing to its huge life-threatening risk [19].

Recently, researchers have described the afflictions experienced by people during the pandemic period [20,21] that goes from changing personal behavior to psychological distress, anxiety, depression, and stress. Following daily life changes, these studies have shown an important behavioral change at the beginning of the pandemic, also leading to fear of COVID-19 and insecurity. This whole process is a crucial reaction, which is mainly caused by inefficient measures to control the pandemic and the lack of psychological assistance.

During the pandemic, the world has faced shutdown, slowdown, or lockdown, and individuals have been encouraged to use masks and practice social distancing. Meanwhile, health professionals had to go in the opposite direction. These workers were directly involved in offering diagnosis and treatment care for SARS-CoV-2-infected patients with almost uninterrupted work in a life-threatening, and sometimes, frustrating perspective. Lately, besides the feasible SARS-CoV-2 contamination risk, these professionals are also at a high risk of developing psychological distress and other mental health symptoms [22]. Thus, health care professionals have been considered as one of the most vulnerable working categories to develop psychological stress and other mental health symptoms, especially in countries highly affected, such as Brazil, which now faces another wave of a new coronavirus mutant (N501Y), which is at least 50% more infective than the original strain [23]. In this work, our purpose was to identify the impact of the COVID-19 pandemic on the life and work routine of Brazilian health care professionals through the study of their self-declared perceptions and their needs during this period.

Methods

Survey Questionnaire and Validation

This study was a cross-sectional web-based survey conducted in Brazil. We prepared a 32-items questionnaire using the Google Forms electronic assessment. A combination of structured (yes/no), multiple-choice selections with 1 final open question was used. All questions were mandatory and divided into 2 blocks with different proposals: (1) personal profile (eg, age, gender, ethnicity, household income, schooling level, and professional characteristics) and (2) COVID-19 pandemic impact. Our objective in most questions was to reflect on the perceptions of health professionals about COVID-19 and allow the analysis based on the respondent's declarations during the

pandemic. It is important to note that any diagnosis pointed by the participants was not debated with them nor were they requested for any documents to assure the pathological situation or diagnosis. The questionnaire underwent an internal validation by both an expert panel of 5 and in a respondent set of 10 health professionals from a huge national health institution (Fiocruz). Experts critically reviewed the instrument and offered important feedback such as addition, deletion, and reformulation of questions and answers, and errors in the form systems used to create the questionnaire. The first approach of the survey asked for informed consent and for the autodeclaration status of "health professionals," considering all the workers in any type of unit of the National Brazilian Health System (Sistema Único de Saúde).

Recruitment and Sample

The invitation to answer the questionnaire was distributed nationally to health professionals in different health institutions through emails, WhatsApp groups, and social media (Facebook). This study was designed to capture spontaneous perceptions from health professionals and had no epidemiological purpose. Their motivation to access and answer the forms relied on the altruistic feeling of the participant to collaborate with the research. Although the survey was not designed to follow strict representative numbers of health professionals in all Brazilian regions, a study of the last available census of Brazilian health professionals [24] was previously prepared to ascertain that all the geographical regions were covered with a sufficient representativity to be considered a national assessment. To reduce the bias of the result at a specific point, the survey was kept open around 3 weeks, from September 12 to October 5, 2020, collecting 1476 answers in the closure of the investigation. The final set of data was obtained after excluding duplicate answers through email confirmation (n=88) and answers in which the participant presented a contradictory statement related to his/her status of the health professional (n=12, retired, I am not working yet, I am not in the health area, salesman, primary school teacher), achieving 1376 answers that were finally analyzed.

Data Analysis

Data exploration, analysis, and cleaning were performed using the Python programming language (version 3.6) with the Jupyter interface. During the analysis, the percentage of participants who selected each response was computed, and the Pandas and NumPy libraries were used together with Matplotlib library for the table generation. Chi-square analysis was performed whenever necessary to statistically confirm differences between any specific group of interest. The participants indicated using a 5-point scale how much COVID-19 affected their lives (1=not affected and 5=very affected) and since the beginning of the pandemic, how much they thought about COVID-19 (1=not at all and 5=very much). The level of anxiety was measured by averaging the participants' scores (ranging from 1 to 5) so that the higher the average, the greater the anxiety of the individual was expected regarding COVID-19. Word clouds were prepared in the WordArt program. This approach was previously validated by other studies [25].

Ethical Committee Approval

The ethical approval for this research was obtained from the Research Ethics Committee of the Oswaldo Cruz Institute-CEP FIOCRUZ/IOC under the number CAAE: 34985420.0.0000.5248. All respondents gave informed consent before their entry into the study.

Results

After the web-based questionnaire distribution, 1376 answers came from all 5 Brazilian regions in a regional percentage distribution, following the same trends observed in the data from the last census of the health care professionals available at the Brazilian Health Ministry. The analysis of the demographic part of the questionnaire section showed that most of the respondents were females (1159/1376, 84.2%) in the age range of 31-50 years (830/1376, 60.3%) (Table 1). The female proportion in the survey was higher in the general population, but in the health profession, this is common owing to the influence of nursing and auxiliary nursing staffs that are ~85% females [26]. Accordingly, we found that the nursing staff (graduate/postgraduate nurses, nursing technicians, and nursing auxiliary) was the largest group answering this survey (669/1376, 48.6%). Since health staffs have a wide variety of professionals—partly legally regulated and others dealing with new professions that are under legislation [27,28]—the survey proposed 10 professional categories but registered 33 types of professions. In the descending order, the survey registered answers from nursing technicians/auxiliary (447/1376, 32.5%), nurses (228/1376, 16.6%), medical doctors (129/1376, 9.3%), physiotherapists/physical educators (128/1376, 9.2%), laboratory, radiology, and other technicians and technologists (75/1376, 5.5%), pharmacy professionals (50/1376, 3.6%), health community agents (17/1376, 1.2%), dentists (14/1376, 1.0%), administrative staff (12/1376, 1.0%), other types of health agents (n=6), and other 13 types of professions including mental therapy workers, social assistants, speech therapists, nutrition professionals, biologists, biomedical scientists, and others (270/1376, 19.6%). The wide reach of our survey corresponds to the general profile of the health professionals produced by the Brazilian Health Ministry at 1 month before the study [28], thus confirming that the survey participants represent this category of workers for the study of their perceptions.

Table 1. Profile of the Brazilian health professionals enrolled in this study (N=1376).

Demographic characteristics	Values, n (%)
Brazilian regions	
Southeast	929 (67.5)
South	149 (10.8)
Central West	140 (10.2)
Northeast	92 (6.7)
North	66 (4.8)
Gender	
Female	1159 (84.2)
Male	215 (15.6)
Nonidentified	2 (0.1)
Age (years)	
18-30	287 (20.9)
31-40	467 (33.9)
41-50	363 (26.4)
51-60	201 (14.6)
>60	58 (4.2)
Ethnicity	
European-derived	724 (53.6)
African-derived	601 (43.7)
Asiatic	23 (0.0)
Indigenous	2 (0.0)
Nonidentified	26 (0.0)
Educational level	
University grade/postgraduate	903 (65.6)
Complete technical/high school level	438 (31.8)
Incomplete university grade	30 (2.2)
Incomplete technical/high school level	5 (0.4)
Household monthly income (USD)	
<52 USD	10 (0.7)
>52-260 USD	33 (2.4)
>260-500 USD	403 (29.3)
>500-1500 USD	598 (43.5)
≥1500 USD	332 (24.1)
Sharing the house with family/friends	
No	44 (3.2)
1-3 persons	1033 (75.1)
≥4 persons	299 (21.7)

More than 50% of the respondents declared themselves as European-derived people (724/1376, 53.6%), with African-derived people constituting 43.7% (601/1376) of the participants (Table 1)—a proportion lower than that in the general composition of the Brazilian population, formed in 2010 majorly by African-derived people (50.9%). Concerning the education level, 31.8% (438/1376) completed the technical level and 65.6% (903/1376) had university grades (Table 1), as expected for the health working force [29,30]. Approximately 67.6% (930/1376) of the respondents had a family income higher than 500 USD and lived with 1-3 persons at home (Table 1). Among them, 39.7% (546/1376) were frontline health care

workers during the COVID-19 outbreak, and only 19% were not working due to unemployment, retirement, or temporary leave from work due to risk factors for COVID-19 infection (Table 2). We also analyzed the amount of distress in relation to economic income and educational qualification, but no correlation was identified between these factors in this group of participants. In this survey of health professionals, 55.0% (757/1376) of them worked in the public sector and 76.8% (1057/1376) of them reported that all personal protective equipment (PPE) was available (Table 2). According to this, 23.2% (319/1376) who complained about inadequate PPE said that the scarcest items were face shields (234/1376, 17.0%), masks (206/1376, 14.9%), and laboratory coats (138/1376, 10.0%). One in 4 health professionals who answered the survey reported that they had not enough information to protect themselves from the coronavirus disease (360/1376, 26.2%). Regarding their personal information source, 40.0% (551/1376) reported the data published by the Brazilian Ministry of Health or the WHO, 26.7% (368/1376) on television, and 18.5% (254/1376) on social networks (Facebook/Instagram/WhatsApp/internet).

When asked about being infected by SARS-CoV-2, almost 22.9% (315/1376) reported positiveness (Table 3), confirming recent data showing rates of infection from 17.8% to 25% depending on the specific type of health profession [28]. However, a major proportion of the respondents did not know if they got infected (289/1376, 21.0%) and 56% (771/1376) reported that did not get COVID-19 (Table 3). Those who were infected by SARS-CoV-2 (315/1376, 22.9%) described majorly (263/315, 83.5%) 3 or more symptoms, with only 5.4% (17/315) being asymptomatic. The most recurrent symptoms were headache (274/315, 87.0%), body pain (231/315, 73.3%), tiredness (228/315, 72.4%), and loss of taste and smell (208/315, 66.0%). Regarding the worst outcome and the severe form of COVID-19, 6.0% (19/315) of the health professionals responding to the survey reported experiences of hospitalization and 0.3% (1/315) reported receiving intubation and invasive ventilation in the intensive care units (Table 3).

Critically, 61.3% (193/315) answered that they were infected with SARS-CoV-2 while working at a health facility, whereas 15.2% (48/315) did not know where they were infected, and 13.3% (42/315) assumed that they got infected from their own family members that had COVID-19. An important result is that 48.3% (152/315) of those health professionals who acquired COVID-19 reported that family members or friends living in the same house also got infected, and 27.0% (85/315) think that they probably were the source of their infection (Table 3).

We also asked that the participants that had COVID-19 to report their feelings during the pandemic period by using not only closed options but also allowing an additional open choice. Figure 1 shows the word cloud images of positive (Figure 1A) and negative (Figure 1B) feelings reported by all the health professionals responding to the COVID-19 perception survey, highlighting the hope and fear of COVID-19 as the predominant feelings, respectively. In this question, 110 participants who did not get COVID-19 chose to answer, thus allowing a quantitative analysis comparing both groups of respondents—those who got COVID-19 and those who did not—confirming that the 3 most recurrently stressful/negative feelings described by those who got COVID-19 were fear of COVID-19 (243/315, 77.0%), insecurity (158/315, 50.0%), and sadness (142/315, 45.0%), as shown in Figure 2. Positive feelings were also reported, including hope, empathy, compassion, relief, and tranquility (Figure 1A). The only significant difference between the 2 groups was found in the feeling of compassion, which was frequently more reported in the group that did not have COVID-19 (Figure 2). The group that got infected expressed some negative feelings at a higher frequency than those that did not get COVID-19, including fear (243/315, 77.1% vs 509/1061, 48.0%), impotence (142/315, 45.1% vs 297/1061 28.0%), and fault (38/315, 12.1% vs 567/1061, 53.4%), respectively. Insecurity, sadness, frustration, rage, shame, and concern were similarly reported. To assess the effect of the pandemic on stress, regardless of whether infected or not with SARS-CoV-2, we elaborated a question with affirmative sentences in which they could mark more than one option (Table 4). Approximately 43.9% (604/1376) of the health professionals pointed to "*I had an* anxiety due to COVID-19," whereas 43.1% (593/1376) selected "*I experienced difficulties in falling asleep*." Furthermore, 32.9% (453/1376) reported "*I had difficulty in concentrating*" and 28.0% (385/1376) reported "*I experienced traumatic situations at work*" (Table 4). From those who pointed *difficulties in falling asleep, in concentrating and lost interest in activities*, 60.0% (826/1376) also reported having a regular or bad institutional support. In addition, during the pandemic, 15.8% (217/1376) developed depression, 33.6% (463/1376) developed general anxiety, and 8.2% (113/1376) developed panic disorder (Table 4). According to our survey, to face the pandemic challenges and to deal with difficulties in this period, Brazilian health professionals received emotional support from family or friends, or from religion, spirituality, or faith (918/1376, 66.7%), and only 8.6% (119/1376) accessed professional psychological and teletherapy services (Table 4). A large percentage (1170/1376, 85.0%) reported receiving support from their immediate bosses at work, half of whom were considered as good/excellent and the other half as regular/bad (Table 4).

Table 2. Labor characteristics of the working places and personal protection equipment and COVID-19 information acquired by health professionals who participated in the national survey (N=1376).

Labor information	Respondents, n (%)
Work during the pandemic	
Not working (unemployed)	164 (11.9)
Working on the front line of COVID-19	546 (39.7)
Working, not on the front line of COVID-19	566 (41.1)
Retired/temporarily away owing to comorbidities	100 (7.3)
Health system working place	
Public sector	636 (46.2)
Private sector/philanthropy hospitals	323 (23.5)
Both public and private sectors	127 (9.2)
Family residences	71 (5.2)
Health education institute	18 (1.3)
Web-based surveillance	11 (0.8)
Not working (retired/unemployed, others)	164 (11.9)
Receive sufficient information to prevent infection	
Yes	996 (72.4)
No	360 (26.2)
Did not answer	20 (1.5)
Have access to adequate safety equipment at work	
Yes	1057 (76.8)
No	319 (23.2)
Considered as a person from the risk groups	
No	882 (64.1)
Yes	494 (35.9)
Perceived alterations in daily routines	
Yes	970 (70.5)
No	406 (29.5)
Sources of information about COVID-19	
Ministry of Health/World Health Organization websites	551 (40.0)
Television	368 (26.7)
Internet sites/Facebook/Instagram	254 (18.5)
Newspapers and journals	109 (7.9)
At work	10 (0.7)
Radio	12 (0,9)
Refuse to get more information	6 (0.4)
Friends and family members	20 (1.5)
Science journals	23 (1.7)
Other media/all the sources	23 (1.7)

Table 3. COVID-19 pandemic impact on Brazilian health professionals who participated in the national survey.

COVID-19 self-reported information	Respondents, n (%)
SARS-CoV-2 infection	
Survey participants	1376 (100.0)
Got infected	315 (22.9)
Did not get infected/do not know	1061 (77.1)
Symptoms developed (n=315, among only responders with COVID-19)	
One or two symptoms	33 (10.5)
Three or more symptoms	263 (83.5)
Headache	274 (87.0)
Body pain	231 (73.3)
Tiredness	228 (72.4)
Loss of taste and smell	208 (66.0)
Dry cough	171 (54.3)
Fever	152 (48.3)
Diarrhea	144 (45.7)
Breath difficulty	128 (40.6)
Minor symptoms (chest pressure, skin eruptions, conjunctivitis, vomiting)	149 (47.3)
Asymptomatic	17 (5.4)
Worsening of the clinical symptoms (n=315, among only responders with COVID-19)	
No worsening	287 (91.1)
Hospitalized in the infirmary	19 (6.0)
Hospitalized in the intensive care unit without intubation	8 (2.5)
Hospitalized in the intensive care unit with intubation	1 (0.3)
Where he/she presumes to have got infected (among only responders with COVID-19)	
Positive history of COVID-19	315 (100.0)
Working in a health facility	193 (61.3)
Do not know	48 (15.2)
From family or friends	42 (13.3)
Public transportation	16 (5.1)
Supermarket/others	16 (5.1)
Persons living in the same place got COVID-19 (n=315, among only responders with COVID-19)	
Yes	152 (48.3)
No	134 (42.5)
Do not know	29 (9.2)
Persons living in the same place got COVID-19 (n=1061, among responders with negative history of COVID-19)	
Yes	91 (8.6)
No	816 (76.9)
Do not know	154 (14.5)
Think have transmitted it to family/friends (n=315, among only responders with COVID-19)	
Yes	85 (27.0)
No	181 (57.5)
Do not know	49 (15.6)

Figure 1. Word cloud images showing the qualitative frequencies of positive (A) and negative (B) feelings reported by health professionals in the COVID-19 perception survey conducted in Brazil (September-October 2020).

Figure 2. Feelings reported by the health professionals in the survey, showing frequencies of answers in the group reporting experience of acquiring COVID-19 (black bars) in comparison with those that did not acquire COVID-19 (white bars). Asterisks indicate significant differences (P<.05) between the two groups, as indicated by chi-square analysis.

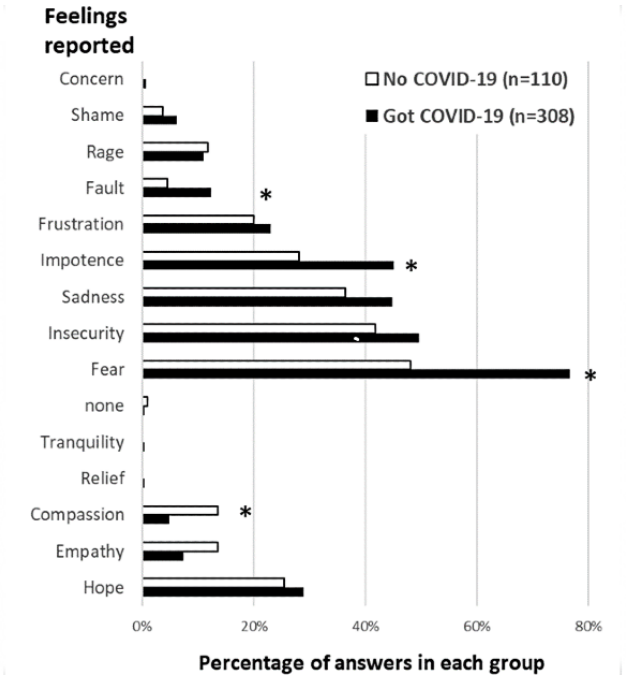
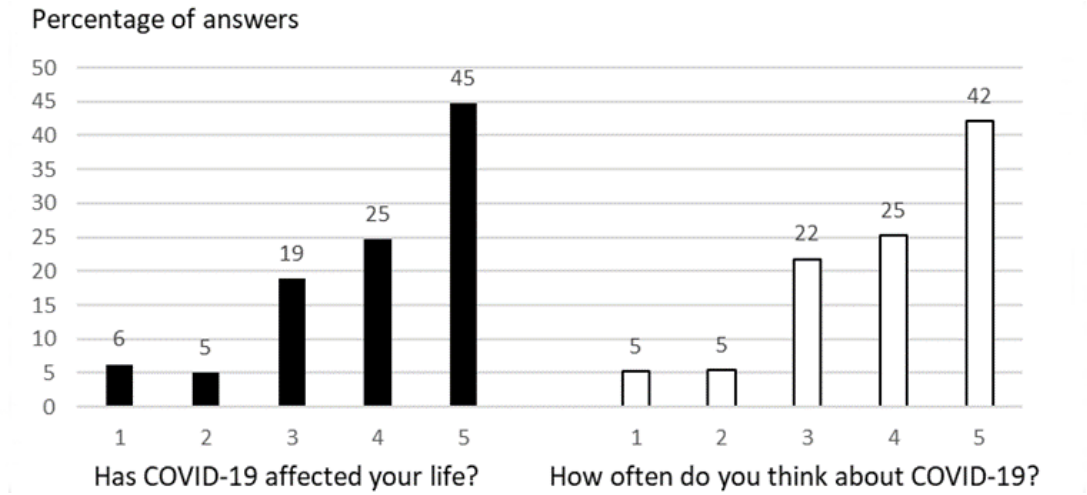


Table 4. Effect of the COVID-19 pandemic on mental health/stress situations of Brazilian health professionals who participated in the survey (N=1376).

Answers concerning stress at work	Respondents, n (%)
Agreement with this statement	
I had an anxiety due to COVID-19	604 (43.9)
I experienced difficulties falling asleep	593 (43.1)
I had difficulty concentrating	453 (32.9)
I lost interest in activities I used to do	447 (32.5)
I experienced traumatic situations at work	385 (28.0)
I do not feel safe leaving home	372 (27.2)
I did not go through these issues	274 (19.9)
I had the need to seek psychological help	182 (13.2)
Have family people depending on special care	
Yes	662 (48.1)
No	714 (51.9)
Diagnosis of adjustment disorder during pandemic	
No	612 (44.5)
Yes, general anxiety	463 (33.6)
Yes, depression	217 (15.8)
Yes, panic	113 (8.2)
Received emotional support from others	
Yes, from friends/family/religion/social networking	918 (66.7)
Yes, from professional support	119 (8.6)
No	339 (24.6)
Received support from immediate boss at work	
Excellent	202 (14.7)
Good	387 (28.1)
Regular	306 (22.2)
Bad	269 (29.5)
Do not have bosses	212 (15.4)

The survey ended with 2 questions asking for a general opinion based on a 5-point scale and related to the general impact of COVID-19 in their lives (Figure 3). Question A: *Has COVID-19 affected your life?* (1=did not affect and 5=affected very much) and question B: *How often, since the beginning of the pandemic, do you think about COVID-19?* (1=almost never, 5=very often). Most of the respondents (963/1376, 70.0%) felt affected (344/1376, 25.0%) or very affected (619/1376, 45.0%) by the COVID-19 pandemic. This was in accordance with their answer to the second question, where a high proportion of health sector workers thought a lot (344/1376, 25.0%) or very much (578/1376, 42.0%) about the disease. Considering that, the higher the proportion of health sector workers overanalyzed about the disease, the greater the anxiety of the individual regarding COVID-19 was expected—both questions indicate this scenario of anxiety due to COVID-19 among most of the health care professionals.

Figure 3. Scores attributed by the health professionals participating in the survey to 2 questions regarding general perception of the impact of the COVID-19 pandemic in their life. Score 1 represents none/few and score 5 represents very much/very often/a lot. The questions were "Has COVID-19 affected your life?" (answers in black bars) and "How often, since the beginning of the pandemic, do you think about COVID-19?" (answers in white bars).



Discussion

Principal Findings

At the end of 2019, COVID-19 was described as a disease that could be easily transmitted and rapidly spread by the SARS-CoV-2 virus [1,31]. Therefore, health professionals are one of the most exposed groups to this disease and to its psychosocial consequences as they are responsible for caring and dealing with patients infected by the virus on a daily basis [32,33]. The nation-by-nation number of deaths and infections of health professionals is still increasing [34,35]; in September 2020, about 570,000 health professionals became infected with the SARS-CoV-2 virus and 2500 died due to the disease in the Americas [28].

Our purpose in this study was to identify the impact of the COVID-19 pandemic on the life and work routine of Brazilian health professionals through the analysis of their perceptions and feelings during this period. Most health professionals who responded to the survey were females, European-derived, aged 31-40 years, and located in the southeast of Brazil, the most populous region. The data from our research showed a higher percentage of female health professionals' participation in relation to males (our data: 1159/1376, 84.2% women and 215/1376, 15.6% men) confirming the literature, which shows that a greater number of health professionals in Brazil are females [32,33] and females are the most effective at work in the face of the COVID-19 pandemic [36]. Interestingly, we verified through our results similar impressions reported by individuals of both genders. Among the health professionals who worked during the pandemic, women represented the largest proportion—they also being the ones that perform the major care functions at home. Even in different studies carried out in Latin American countries, women constituted the highest proportion: Chile (72.6%) [37], Ecuador (68.3%) [38], Argentina (71%) [39], Bolivia (72.9%) [40], and Peru (71%) [41]. According to the United Nations Entity for Gender Equality and the Empowerment of Women [36], 70% of the global health workers are women (eg, nurses, midwives, community health workers) also working as cleaners, caterers, and launderers in health facilities; they have few leadership positions (30%) and lower salaries. In the pandemic scenario, women have been in a huge demand and have been professionally highly affected as children and the older adults depend on them even more, without schools or helpers to support them [42].

According to our survey, despite the highly stressful scenario, these female health workers did not respond differently from men in this pandemic situation, especially in terms of their feelings. They also felt the fear of COVID-19, sadness, hope, empathy, and insecurity while they cared for their family's demands and social and economic problems. Literature shows that women are dealing with these health and stressful issues and social and economic problems totally by themselves, thus highlighting the need for creating gender-specific programs to help these women in the near future [36,42]. Some authors such as Campos et al [43] reported that even though most health care professionals are females, the death risk is higher (52.8 times higher) among younger men than among older women. They justified the higher death rates among younger men to be caused by the highly patriarchal nature of the Brazilian society with a very strong masculine pride and that men do not acknowledge their fragility or seek for assistance.

During the pandemic, the problems in the health care environment included the use of PPE that was intended for other individuals (eg, size for bigger men used by women or smaller persons) and even the absence of these materials as well as life support to use with the patients (eg, respiratory equipment). As the pandemic spread across the globe, the adequate provision of PPE for health professionals was a constant concern [44]. A cross-sectional study conducted in Latin America (Brazil, Colombia, and Ecuador) showed that at least 70% of the health professionals reported a lack of PPE [45]. This concern was reinforced by the answers of our participants, in which 1 in 5 complained about missing PPE such as face shields, masks, and laboratory coats. It is important to notice that the distribution of PPE to health institutions should be a government policy, especially in Brazil that has a huge public system called the

Sistema Único de Saúde [46] that is always in massive demand and that requires mobilization of the national health industry to respond to the challenge of facing the pandemic. Unfortunately, this has not been done and the costs of PPE have been increased [47]. The scarcity of PPEs has also been reported in other Brazilian studies with smaller groups and in other countries [41,43,44,48], especially in those needed to protect frontline health professionals. In Italy, PPE shortages might be among the relevant factors contributing to the high burden of infection and hospital staff deaths, similar to what our survey indicated for Brazilian health workers [49].

Based on the fact that the recent vaccines developed against coronavirus are still not available for everyone in all countries, including Brazil [50], and that the number of new infections is growing at an alarming rate, especially those caused by new mutant strains [51], the knowledge about preventive steps is still essential to disrupt the chain of virus transmission among health professionals. In our study, the astonishing evidence was that 1 in 4 health professionals (26%) indicated a lack of enough information to protect themselves from COVID-19. It means that these health professionals work with insecurities and worry about being infected during their journey times—many of them who work as frontline health care professionals. Some studies described different aspects of health professionals from Brazil during the pandemic with lesser numbers of participants from specific states or regions and different evaluation aspects, sometimes including examining the psychological impact of the COVID-19 pandemic such as those reported by Campos et al [43], Duarte et al [52], and Cotrin et al [53]. Some of these studies showed that Brazil had the largest preponderance of death records caused by COVID-19, especially among nursing professionals, because of several factors such as direct contact with patients, the frequency in performing different procedures, and the lack or inadequate use of PPEs, among others. Our work added to these factors that misinformation (61%) contributed to the lack of precise knowledge about COVID-19 since almost 1 in 5 workers choose social networks as their source of information. The profusion of news on social networks, most of them without any validation on their authenticity, is becoming a huge social problem that compromises the ability to distinguish between facts, opinions, or fake news [54]. The problem is so serious that Dr Tedros Adhanom Ghebreyesus, the WHO Director-General called this news situation as an infodemic that should be fought against, leading to some efforts to create strategies to help on this issue [22,55]. The need for further awareness campaigns and knowledge of safe interventions to combat the spread of COVID-19 still remain, requiring that the health sectors increase the access to precise information about this disease [56]. These data also reinforced the identification of these workplaces as high-risk environments in Brazil as well as in other countries [57-59]. Although some studies with smaller groups pointed that economic income and educational qualification had some correlation with COVID-19, we did not observe them as a direct factor to be considered in this group. It is important to notice that our group showed a professional distribution similar to that described by national and international reports of the Brazilian medical team, which may suggest that these factors as more related to specific groups or regions in our country [53].

From the beginning of the SARS-CoV-2 outbreak, concerns have been raised about its effect on mental health [60,61]. According to WHO, mental health is defined as "*a state of well-being in which each individual realizes their own potential and can cope with the normal stress of life, can work productively and is able to contribute to their community,*" and it is more important than physical health, especially when it comes to stressful situations such as the COVID-19 pandemic [62]. Several studies have been published describing the mental profile of patients with COVID-19 who developed symptoms of anxiety, depression, psychological distress, and insomnia [63-65]. An American survey included 1651 respondents from all 50 states and reported that 60% of the health professionals had a higher risk of emotional distress/burnout during the COVID-19 pandemic [66]. Hair cortisol evaluation is a suitable biomarker for an individual's exposure to stressful events. A study conducted in Argentina on 234 health professionals showed that 40% of the sample population presented hair cortisol values outside of the healthy reference range in the course of the COVID-19 pandemic, thereby showing a direct correlation with the perceived stress and the emotional exhaustion component of burnout [39]. In Canada, by surveying health professionals, Wilbiks et al [67] described that there was an elevated level of depressive symptomatology in that population. The prevalence of stress, anxiety, and depression in frontline health care professionals caring for patients with COVID-19 was already described for some small groups such as those described with a convenience sample of 364 health workers, including physicians, nurses, pharmacists, and laboratory technicians [58]. Like our study that identified positive feelings from the participants, they described positive attitudes from all participants with mostly moderate COVID-19 psychological stress levels.

The literature also described a systematic review that evaluated 29 studies, with a total sample size of around 22,000 health professionals. Similarly to our survey in which several health professionals experienced anxiety due to COVID-19 (N=9680, 44%), depression (N=7480, 34%), and insomnia (N=7260, 33%), the review showed that 21 papers described the prevalence of depression, 23 reported the prevalence of anxiety, and 9 studies have reported the prevalence of stress [68]. COVID-19 changed the lives of everybody worldwide [69], and our survey reinforced that the Brazilian health professionals were also affected at a high level at 70% (N=963) and this is apparently directly associated with higher levels of psychological and physical stress.

Limitations

Our study has some limitations, which need to be considered. The findings are not generalizable to all categories of health care professionals, as it is a compilation of all respondents' impressions. Another important piece of information that should be deemed is the total period of the COVID-19 pandemic. We evaluated the perceptions and feelings of these professionals in a specific time frame. Therefore, longitudinal studies are recommended in Brazil. Despite the self-report questionnaire being one of the most widely used assessments, its use rather than a clinical assessment reduced the power of our findings. Another limitation of this study was that most of the respondents

were those who used or operated the internet, which only constitutes a partial section of the society. However, this study could suggest a general overview of the perceptions and feelings present in the health professionals in Brazil.

Conclusions

In every country, during the COVID-19 pandemic, health care professionals had to work under pressure with risks of affecting their physical and mental health, by being on the front line and assisting to save lives. Our data showed that the COVID-19 pandemic affected overall 70% of the Brazilian health professionals, according to their answers to our survey. However, most of the feelings did not change when comparing those who did get infected to those who did not, men or women, suggesting that to be exposed to this work environment and the pandemic situation are enough to develop negative feelings, such as fear of COVID-19, sadness, and insecurity, despite keeping "alive" their hope. These negative feelings are probably maintained by knowing situations such as (1) absence of Brazilian strategies at the national level for mass testing of the population, (2) absence of effective public policies that reduce the cases of COVID-19, and (3) absence of sanitary measures carried out in a centralized manner by states and municipalities (not only guaranteed by calling the judiciary), especially in states where the epidemic is most severe. Altogether, these feelings and perceptions reported in this work are alarming and must be well addressed with interventions that enhance the quality of life of the health professionals. There is an urgent need for regular monitoring of potential stress disorders, aiming to reduce the associated side effects in the longer run. Therefore, health policymakers should plan actions to control and prevent mental disorders in this category of professionals as soon as possible. One of the actions that should be implemented in each hospital, clinic, and asylum is the creation of multidisciplinary groups that may attend and monitor the medical staff, including all involved, not only for training but also to dialogue and identify burnout situations before they deeply and irreversibly affect this group that is so stressed out in this pandemic. This also includes the assurance of vaccination (2 doses taken) for all of them.

Acknowledgments

We thank Coordenação de Aperfeiçoamento de Pessoal de Nível Superio (CAPES), Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) (E-26/210.844/2019; 201.983/2020), and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (311361/2015-3) for the fellowships and financial support.

Conflicts of Interest

None declared.

References

1. Chen J. Pathogenicity and transmissibility of 2019-nCoV-A quick overview and comparison with other emerging viruses. Microbes Infect 2020 Mar;22(2):69-71 [FREE Full text] [doi: 10.1016/j.micinf.2020.01.004] [Medline: 32032682]
2. Hui DS, I Azhar E, Madani TA, Ntoumi F, Kock R, Dar O, et al. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health - The latest 2019 novel coronavirus outbreak in Wuhan, China. Int J Infect Dis 2020 Feb;91:264-266 [FREE Full text] [doi: 10.1016/j.ijid.2020.01.009] [Medline: 31953166]
3. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol 2020 Apr;5(4):536-544 [FREE Full text] [doi: 10.1038/s41564-020-0695-z] [Medline: 32123347]
4. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA 2020 Apr 07;323(13):1239-1242. [doi: 10.1001/jama.2020.2648] [Medline: 32091533]
5. Burki TK. Coronavirus in China. Lancet Respir Med 2020 Mar;8(3):238 [FREE Full text] [doi: 10.1016/S2213-2600(20)30056-4] [Medline: 32027848]
6. Pan Y, Guan H. Imaging changes in patients with 2019-nCov. Eur Radiol 2020 Jul;30(7):3612-3613 [FREE Full text] [doi: 10.1007/s00330-020-06713-z] [Medline: 32025790]
7. Yi Y, Lagniton PNP, Ye S, Li E, Xu R. COVID-19: what has been learned and to be learned about the novel coronavirus disease. Int J Biol Sci 2020;16(10):1753-1766 [FREE Full text] [doi: 10.7150/ijbs.45134] [Medline: 32226295]
8. Rodriguez-Morales AJ, Gallego V, Escalera-Antezana JP, Méndez CA, Zambrano LI, Franco-Paredes C, et al. COVID-19 in Latin America: The implications of the first confirmed case in Brazil. Travel Med Infect Dis 2020;35:101613 [FREE Full text] [doi: 10.1016/j.tmaid.2020.101613] [Medline: 32126292]
9. Lima-Costa MF, Mambrini JVDM, Andrade FBD, Peixoto SWV, Macinko J. Social distancing, use of face masks and hand washing among participants in the Brazilian Longitudinal Study of Aging: the ELSI-COVID-19 initiative. Cad Saude Publica 2020;36Suppl 3(Suppl 3):e00193920 [FREE Full text] [doi: 10.1590/0102-311X00193920] [Medline: 33053062]
10. Dantas RCC, de Campos PA, Rossi I, Ribas RM. Implications of social distancing in Brazil in the COVID-19 pandemic. Infect Control Hosp Epidemiol 2020 May 08:1-2 [FREE Full text] [doi: 10.1017/ice.2020.210] [Medline: 32383418]
11. World Health Organization. Europe COVID-19 Response Fund in Eastern Mediterranean, Africa, Western Pacific. 2021 Feb 12. URL: https://covid19.who.int [accessed 2021-02-12]

12. Yufika A, Wagner AL, Nawawi Y, Wahyuniati N, Anwar S, Yusri F, et al. Parents' hesitancy towards vaccination in Indonesia: A cross-sectional study in Indonesia. Vaccine 2020 Mar 04;38(11):2592-2599. [doi: 10.1016/j.vaccine.2020.01.072] [Medline: 32019704]
13. Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. Lancet 2020 Feb 15;395(10223):507-513 [FREE Full text] [doi: 10.1016/S0140-6736(20)30211-7] [Medline: 32007143]
14. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet 2020 Feb;395(10223):497-506. [doi: 10.1016/s0140-6736(20)30183-5]
15. Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, et al. Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. Int J Infect Dis 2020 May;94:91-95 [FREE Full text] [doi: 10.1016/j.ijid.2020.03.017] [Medline: 32173574]
16. Velavan TP, Meyer CG. The COVID-19 epidemic. Trop Med Int Health 2020 Mar;25(3):278-280 [FREE Full text] [doi: 10.1111/tmi.13383] [Medline: 32052514]
17. Beatriz R, Rudi R, Leticia N, Paula S, Ana Maria M, Adriano M. Necessidades de Infraestrutura do SUS em Preparo à COVID-19: leitos de UTI, respiradores e ocupação hospitalar. IEPS. 2020 Mar 01. URL: https://ieps.org.br/wp-content/uploads/2020/04/IEPS-NT3.pdf [accessed 2021-09-24]
18. Cimerman S, Chebabo A, Cunha CAD, Rodríguez-Morales AJ. Deep impact of COVID-19 in the healthcare of Latin America: the case of Brazil. Braz J Infect Dis 2020;24(2):93-95 [FREE Full text] [doi: 10.1016/j.bjid.2020.04.005] [Medline: 32335078]
19. Maffioli EM, Hernandes Rocha TA, Vivas G, Rosales C, Staton C, Nickenig Vissoci JR. Addressing inequalities in medical workforce distribution: evidence from a quasi-experimental study in Brazil. BMJ Glob Health 2019;4(6):e001827 [FREE Full text] [doi: 10.1136/bmjgh-2019-001827] [Medline: 31798991]
20. Tan W, Hao F, McIntyre RS, Jiang L, Jiang X, Zhang L, et al. Is returning to work during the COVID-19 pandemic stressful? A study on immediate mental health status and psychoneuroimmunity prevention measures of Chinese workforce. Brain Behav Immun 2020 Jul;87:84-92 [FREE Full text] [doi: 10.1016/j.bbi.2020.04.055] [Medline: 32335200]
21. Rajkumar RP. COVID-19 and mental health: A review of the existing literature. Asian J Psychiatr 2020 Aug;52:102066 [FREE Full text] [doi: 10.1016/j.ajp.2020.102066] [Medline: 32302935]
22. Huang Y, Zhao N. Chinese mental health burden during the COVID-19 pandemic. Asian J Psychiatr 2020 Jun;51:102052 [FREE Full text] [doi: 10.1016/j.ajp.2020.102052] [Medline: 32361387]
23. Le Page M. Threats from new variants. New Scientist 2021 Jan;249(3316):8-9. [doi: 10.1016/s0262-4079(21)00003-8]
24. DATASUS. Número de profissionais de saúde por habitante Períodos Disponíveis Seleções Disponíveis. Government of Brazil. URL: https://datasus.saude.gov.br/ [accessed 2021-02-12]
25. Vasconcellos-Silva PR, Carvalho D, Lucena C. Word frequency and content analysis approach to identify demand patterns in a virtual community of carriers of hepatitis C. Interact J Med Res 2013 Jul 04;2(2):e12 [FREE Full text] [doi: 10.2196/ijmr.2384] [Medline: 23827772]
26. Machado MH, Filho WA, De Lacerda WF, De Oliveira E, Lemos W, Wermelinger M, et al. CARACTERISTICAS GERAIS DA ENFERMAGEM: O PERFIL SÓCIO DEMOGRÁFICO. revenf 2016 Jan 27;7(ESP):9. [doi: 10.21675/2357-707x.2016.v7.nesp.686]
27. Bem ID, Gomes T, Santos MD, Delduque M, Alves S. Legislativo e Saúde: a regulamentação de profissões para o Sistema Único de Saúde. Biblioteca Virtual em Saúde MS. URL: http://bvsms.saude.gov.br/bvs/periodicos/ccs_artigos/legislativo_saude_regulamentacao_profissoes.pdf [accessed 2021-09-24]
28. Alerta epidemiológica COVID-19 en personal de salud. Organización Panamericana de la Salud. URL: https://fi-admin.bvsalud.org/document/view/vunfw [accessed 2021-09-24]
29. Carvalho MND, Gil CRR, Costa EMOD, Sakai MH, Leite SN. Needs and dynamics of the Primary Healthcare workforce in Brazil. Cien Saude Colet 2018 Jan;23(1):295-302 [FREE Full text] [doi: 10.1590/1413-81232018231.08702015] [Medline: 29267833]
30. Rizzotto MLF, Gil CRR, Carvalho DC, Fonseca ALN, Santos MF. Força de trabalho e gestão do trabalho em saúde: revelações da avaliação externa do Programa de Melhoria do Acesso e da Qualidade na Atenção Básica no Paraná. Workforce and work management in health/primary care: revelations of External Evaluation of the National Program for Access and Quality Improvement in Primary Care in Paraná 2014;38:237-251. [doi: 10.5935/0103-1104.2014S018]
31. Editorial. How epidemiology has shaped the COVID pandemic. Nature 2021 Jan;589(7843):491-492. [doi: 10.1038/d41586-021-00183-z] [Medline: 33505041]
32. Lai J, Ma S, Wang Y, Cai Z, Hu J, Wei N, et al. Factors Associated With Mental Health Outcomes Among Health Care Workers Exposed to Coronavirus Disease 2019. JAMA Netw Open 2020 Mar 02;3(3):e203976 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.3976] [Medline: 32202646]
33. Jahan I, Ullah I, Griffiths MD, Mamun MA. COVID-19 suicide and its causative factors among the healthcare professionals: Case study evidence from press reports. Perspect Psychiatr Care 2021 Feb 05 [FREE Full text] [doi: 10.1111/ppc.12739] [Medline: 33547666]

34. Bandyopadhyay S, Baticulon RE, Kadhum M, Alser M, Ojuka DK, Badereddin Y, et al. Infection and mortality of healthcare workers worldwide from COVID-19: a systematic review. BMJ Glob Health 2020 Dec;5:12 [FREE Full text] [doi: 10.1136/bmjgh-2020-003097] [Medline: 33277297]
35. Erdem H, Lucey DR. Healthcare worker infections and deaths due to COVID-19: A survey from 37 nations and a call for WHO to post national data on their website. Int J Infect Dis 2021 Jan;102:239-241 [FREE Full text] [doi: 10.1016/j.ijid.2020.10.064] [Medline: 33130210]
36. In focus: gender equality matters in COVID-19 response, Featured: The shadow pandemic-violence against women. UN Women. URL: https://www.unwomen.org/en/news/in-focus/in-focus-gender-equality-in-covid-19-response [accessed 2021-02-12]
37. Iruretagoyena M, Vial MR, Spencer-Sandino M, Gaete P, Peters A, Delgado I, et al. Longitudinal assessment of SARS-CoV-2 IgG seroconversionamong front-line healthcare workers during the first wave of the Covid-19 pandemic at a tertiary-care hospital in Chile. BMC Infect Dis 2021 May 26;21(1):478 [FREE Full text] [doi: 10.1186/s12879-021-06208-2] [Medline: 34039287]
38. Pazmiño Erazo EE, Alvear Velásquez MJ, Saltos Chávez IG, Pazmiño Pullas DE. Factors Associated With Psychiatric Adverse Effects in Healthcare Personnel During the COVID-19 Pandemic in Ecuador. Rev Colomb Psiquiatr 2021;50(3):166-175 [FREE Full text] [doi: 10.1016/j.rcp.2020.12.007] [Medline: 33735001]
39. Ibar C, Fortuna F, Gonzalez D, Jamardo J, Jacobsen D, Pugliese L, et al. Evaluation of stress, burnout and hair cortisol levels in health workers at a University Hospital during COVID-19 pandemic. Psychoneuroendocrinology 2021 Jun;128:105213 [FREE Full text] [doi: 10.1016/j.psyneuen.2021.105213] [Medline: 33845387]
40. Zhang SX, Chen J, Afshar JA, Alvarez-Risco A, Dai H, Li J, et al. Succumbing to the COVID-19 Pandemic?Healthcare Workers Not Satisfied and Intend to Leave Their Jobs. Int J Ment Health Addict International Journal of Mental Health and Addiction; 2021 . [doi: 10.1007/s11469-020-00418-6]
41. Raraz Vidal JG, Allpas Gomez HL, Torres Salome FK, Cabrera Patiño WM, Ramos Gómez RP, Alcántara Leyva LM, et al. Work conditions and personal protective equipment against COVID-19 in health personnel, Lima-Peru. RFMH 2021 Mar 15;21(2):335-345. [doi: 10.25176/rfmh.v21i2.3608]
42. Castro MC, Massuda A, Almeida G, Menezes-Filho NA, Andrade MV, de Souza Noronha KVM, et al. Brazil's unified health system: the first 30 years and prospects for the future. The Lancet 2019 Jul 01;394(10195):345-356 [FREE Full text] [doi: 10.1016/S0140-6736(19)31243-7]
43. Viana Campos AC, Pereira Colares Leitão L. Letalidade da COVID-19 entre profissionais de saúde no Pará, Brasil. J Health NPEPS 2021;6(1):22-34. [doi: 10.30681/252610105190]
44. The Lancet. COVID-19: protecting health-care workers. Lancet 2020 Mar 21;395(10228):922 [FREE Full text] [doi: 10.1016/S0140-6736(20)30644-9] [Medline: 32199474]
45. Martin-Delgado J, Viteri E, Mula A, Serpa P, Pacheco G, Prada D, et al. Availability of personal protective equipment and diagnostic and treatment facilities for healthcare workers involved in COVID-19 care: A cross-sectional study in Brazil, Colombia, and Ecuador. PLoS One 2020;15(11):e0242185 [FREE Full text] [doi: 10.1371/journal.pone.0242185] [Medline: 33175877]
46. Castro MC, Massuda A, Almeida G, Menezes-Filho NA, Andrade MV, de Souza Noronha KVM, et al. Brazil's unified health system: the first 30 years and prospects for the future. Lancet 2019 Jul 27;394(10195):345-356. [doi: 10.1016/S0140-6736(19)31243-7] [Medline: 31303318]
47. Burki T. Global shortage of personal protective equipment. Lancet Infect Dis 2020 Jul;20(7):785-786 [FREE Full text] [doi: 10.1016/S1473-3099(20)30501-6] [Medline: 32592673]
48. Shortage of personal protective equipment endangering health workers worldwide. WHO. URL: https://www.who.int/news/item/03-03-2020-shortage-of-personal-protective-equipment-endangering-health-workers-worldwide [accessed 2021-02-12]
49. Lapolla P, Mingoli A, Lee R. Deaths from COVID-19 in healthcare workers in Italy-What can we learn? Infect Control Hosp Epidemiol 2021 Mar;42(3):364-365 [FREE Full text] [doi: 10.1017/ice.2020.241] [Medline: 32408922]
50. Forni G, Mantovani A, COVID-19 Commission of Accademia Nazionale dei Lincei, Rome. COVID-19 vaccines: where we stand and challenges ahead. Cell Death Differ 2021 Feb;28(2):626-639 [FREE Full text] [doi: 10.1038/s41418-020-00720-9] [Medline: 33479399]
51. Zeng L, Li D, Tong W, Shi T, Ning B. Biochemical features and mutations of key proteins in SARS-CoV-2 and their impacts on RNA therapeutics. Biochem Pharmacol 2021 Jul;189:114424 [FREE Full text] [doi: 10.1016/j.bcp.2021.114424] [Medline: 33482149]
52. Duarte H, Daros Vieira R, Cardozo Rocon P, Andrade ACDS, Wittmer VL, Capellini VK, et al. Factors associated with Brazilian physical therapists' perception of stress during the COVID-19 pandemic: a cross-sectional survey. Psychol Health Med 2021 Jan 25:1-12. [doi: 10.1080/13548506.2021.1875133] [Medline: 33487038]
53. Cotrin P, Moura W, Gambardela-Tkacz CM, Pelloso FC, Santos LD, Carvalho MDDB, et al. Healthcare Workers in Brazil during the COVID-19 Pandemic: A Cross-Sectional Online Survey. Inquiry 2020;57:46958020963711 [FREE Full text] [doi: 10.1177/0046958020963711] [Medline: 33034257]
54. Hartley K, Vu MK. Fighting fake news in the COVID-19 era: policy insights from an equilibrium model. Policy Sci 2020 Sep 09:1-24 [FREE Full text] [doi: 10.1007/s11077-020-09405-z] [Medline: 32921821]

55. Sear RF, Velasquez N, Leahy R, Restrepo NJ, Oud SE, Gabriel N, et al. Quantifying COVID-19 Content in the Online Health Opinion War Using Machine Learning. IEEE Access 2020;8:91886-91893. [doi: 10.1109/access.2020.2993967]
56. Melesie Taye G, Bose L, Beressa TB, Tefera GM, Mosisa B, Dinsa H, et al. COVID-19 Knowledge, Attitudes, and Prevention Practices Among People with Hypertension and Diabetes Mellitus Attending Public Health Facilities in Ambo, Ethiopia. Infect Drug Resist 2020;13:4203-4214 [FREE Full text] [doi: 10.2147/IDR.S283999] [Medline: 33262615]
57. Passos RG, Silveira MB, Abrahão JS. Exploratory assessment of the occurrence of SARS-CoV-2 in aerosols in hospital facilities and public spaces of a metropolitan center in Brazil. Environ Res 2021 Apr;195:110808 [FREE Full text] [doi: 10.1016/j.envres.2021.110808] [Medline: 33513382]
58. Elbqry MG, Elmansy FM, Elsayed AE, Mansour B, Tantawy A, Eldin MB, et al. Effect of COVID-19 stressors on healthcare workers' performance and attitude at Suez Canal university hospitals. Middle East Curr Psychiatry 2021 Jan 26;28(1):4. [doi: 10.1186/s43045-021-00084-x]
59. Bonano JC, Huddleston JI. Perioperative Medical and Surgical Coronavirus Disease 2019 Issues: Keeping Surgeons, Operating Room Teams, and Patients Safe. J Arthroplasty 2021 Jul;36(7S):S46-S48 [FREE Full text] [doi: 10.1016/j.arth.2021.01.047] [Medline: 33618956]
60. Giacalone A, Rocco G, Ruberti E. Physical Health and Psychosocial Considerations During the Coronavirus Disease 2019 Outbreak. Psychosomatics 2020;61(6):851-852 [FREE Full text] [doi: 10.1016/j.psym.2020.07.005] [Medline: 32861431]
61. Xiang Y, Yang Y, Li W, Zhang L, Zhang Q, Cheung T, et al. Timely mental health care for the 2019 novel coronavirus outbreak is urgently needed. Lancet Psychiatry 2020 Mar;7(3):228-229 [FREE Full text] [doi: 10.1016/S2215-0366(20)30046-8] [Medline: 32032543]
62. Mental health: strengthening our response. WHO. URL: https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response [accessed 2021-09-24]
63. Khanagar SB, Al-Ehaideb A, Vishwanathaiah S, Maganur PC, Varadarajan S, Patil S. Depression, Anxiety, and Psychological Distress among Healthcare Providers during the Outbreak of the Life-threatening Coronavirus Disease (COVID-19). J Contemp Dent Pract 2020 May 01;21(5):471-472. [Medline: 32690824]
64. Kibbey MM, Fedorenko EJ, Farris SG. Anxiety, depression, and health anxiety in undergraduate students living in initial US outbreak "hotspot" during COVID-19 pandemic. Cogn Behav Ther 2021 Sep;50(5):409-421. [doi: 10.1080/16506073.2020.1853805] [Medline: 33433271]
65. Mazza MG, De Lorenzo R, Conte C, Poletti S, Vai B, Bollettini I, COVID-19 BioB Outpatient Clinic Study group, et al. Anxiety and depression in COVID-19 survivors: Role of inflammatory and clinical predictors. Brain Behav Immun 2020 Oct;89:594-600 [FREE Full text] [doi: 10.1016/j.bbi.2020.07.037] [Medline: 32738287]
66. Sharma M, Creutzfeldt C, Lewis A, Patel PV, Hartog C, Jannotta G, et al. Health-care Professionals' Perceptions of Critical Care Resource Availability and Factors Associated With Mental Well-being During Coronavirus Disease 2019 (COVID-19): Results from a US Survey. Clin Infect Dis 2021 May 18;72(10):e566-e576 [FREE Full text] [doi: 10.1093/cid/ciaa1311] [Medline: 32877508]
67. Wilbiks JMP, Best LA, Law MA, Roach SP. Evaluating the mental health and well-being of Canadian healthcare workers during the COVID-19 outbreak. Healthc Manage Forum 2021 Jul;34(4):205-210 [FREE Full text] [doi: 10.1177/08404704211021109] [Medline: 34098760]
68. Salari N, Khazaie H, Hosseinian-Far A, Khaledi-Paveh B, Kazeminia M, Mohammadi M, et al. The prevalence of stress, anxiety and depression within front-line healthcare workers caring for COVID-19 patients: a systematic review and meta-regression. Hum Resour Health 2020 Dec 17;18(1):100 [FREE Full text] [doi: 10.1186/s12960-020-00544-1] [Medline: 33334335]
69. Szylovec A, Umbelino-Walker I, Cain BN, Ng HT, Flahault A, Rozanova L. Brazil's Actions and Reactions in the Fight Against COVID-19 from January to March 2020. Int J Environ Res Public Health 2021 Jan 11;18(2):1-16 [FREE Full text] [doi: 10.3390/ijerph18020555] [Medline: 33440812]

Abbreviations

PPE: personal protective equipment
WHO: World Health Organization

Edited by G Eysenbach; submitted 20.02.21; peer-reviewed by E Ruberti, RP Rajkumar; comments to author 10.05.21; revised version received 19.06.21; accepted 01.08.21; published 22.10.21.

Please cite as:
Corrêa RP, Castro HC, Quaresma BMCS, Stephens PRS, Araujo-Jorge TC, Ferreira RR
Perceptions and Feelings of Brazilian Health Care Professionals Regarding the Effects of COVID-19: Cross-sectional Web-Based Survey
JMIR Form Res 2021;5(10):e28088
URL: https://formative.jmir.org/2021/10/e28088
doi:10.2196/28088
PMID:34519656

Corrigenda and Addenda

Correction: Precision Public Health Campaign: Delivering Persuasive Messages to Relevant Segments Through Targeted Advertisements on Social Media

Jisun An[1], PhD; Haewoon Kwak[1], PhD; Hanya M Qureshi[2], BA; Ingmar Weber[3], PhD

[1]School of Computing and Information Systems, Singapore Management University, Singapore, Singapore

[2]Yale School of Medicine, Yale University, New Haven, CT, United States

[3]Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar

Corresponding Author:
Jisun An, PhD
School of Computing and Information Systems
Singapore Management University
80 Stamford Road
Singapore, 178902
Singapore
Phone: 65 6826 4809
Email: an.jisun.221@gmail.com

Related Article:

Correction of: http://formative.jmir.org/2021/9/e22313/

(JMIR Form Res 2021;5(10):e33922) doi:10.2196/33922

In "Precision Public Health Campaign: Delivering Persuasive Messages to Relevant Segments Through Targeted Advertisements on Social Media" (JMIR Form Res 2021;5(9):e22313) the authors noted one error.

In the originally published manuscript, the order of authors Ingmar Weber and Hanya M Qureshi was reversed. This has been corrected to reflect that Hanya M Qureshi is the paper's third author and Ingmar Weber is the paper's fourth author. The author affiliations have been renumbered accordingly.

The full list of authorship and affiliations in the originally published version appeared as follows:

> *Jisun An[1], PhD; Haewoon Kwak[1], PhD; Ingmar Weber[2], PhD; Hanya M Qureshi[3], BA*
>
> *[1]School of Computing and Information Systems, Singapore Management University, Singapore, Singapore*
>
> *[2]Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar*
>
> *[3]Yale School of Medicine, Yale University, New Haven, CT, United States*

The full list of authorship and affiliations has been updated as follows in the corrected version:

> *Jisun An[1], PhD; Haewoon Kwak[1], PhD; Hanya M Qureshi[2], BA; Ingmar Weber[3], PhD*
>
> *[1]School of Computing and Information Systems, Singapore Management University, Singapore, Singapore*
>
> *[2]Yale School of Medicine, Yale University, New Haven, CT, United States*
>
> *[3]Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar*

The correction will appear in the online version of the paper on the JMIR Publications website on October 5, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 29.09.21; this is a non–peer-reviewed article;accepted 29.09.21; published 05.10.21.

Please cite as:
An J, Kwak H, Qureshi HM, Weber I
Correction: Precision Public Health Campaign: Delivering Persuasive Messages to Relevant Segments Through Targeted Advertisements on Social Media
JMIR Form Res 2021;5(10):e33922
URL: https://formative.jmir.org/2021/10/e33922
doi:10.2196/33922
PMID:34609948

Corrigenda and Addenda

Correction: Electronic Video Consent to Power Precision Health Research: A Pilot Cohort Study

Arash Naeim[1*], MD, PhD; Sarah Dry[2*], MD; David Elashoff[2], PhD; Zhuoer Xie[3], MD; Antonia Petruse[4], MBA; Clara Magyar[2], PhD; Liliana Johansen[4], MPH; Gabriela Werre[4], MSHSD; Clara Lajonchere[5], PhD; Neil Wenger[2], MD, MPH

[1]UCLA Center for SMART Health, Clinical and Translational Science Institute, David Geffen School of Medicine at UCLA, Los Angeles, CA, United States

[2]David Geffen School of Medicine at UCLA, Los Angeles, CA, United States

[3]Mayo Clinic, Rochester, MN, United States

[4]Embedded Clinical Research and Innovation Unit, CTSI Office of Clinical Research, Los Angeles, CA, United States

[5]Institute for Precision Health, David Geffen School of Medicine at UCLA, Los Angeles, CA, United States

[*]these authors contributed equally

Corresponding Author:
Arash Naeim, MD, PhD
UCLA Center for SMART Health
Clinical and Translational Science Institute
David Geffen School of Medicine at UCLA
10911 Weyburn Ave
Los Angeles, CA, 90095
United States
Phone: 1 3107948118
Email: anaeim@mednet.ucla.edu

Related Article:

Correction of: https://formative.jmir.org/2021/9/e29123

(JMIR Form Res 2021;5(10):e33891) doi:10.2196/33891

In “Electronic Video Consent to Power Precision Health Research: A Pilot Cohort Study” (JMIR Form Res 2021;5(9):e29123) the authors noted six errors.

The title of the originally published article appeared as follows:

> *Electronic Video Consent to Power Precision Research: A Pilot Cohort Study*

In the corrected version, the title has been revised to:

> *Electronic Video Consent to Power Precision Health Research: A Pilot Cohort Study*

In the original version, author Liliana Johansen's name was displayed incorrectly as follows:

> *Lilliana Johansen*

In the revised version, it has been corrected as follows:

> *Liliana Johansen*

In the originally published paper, Affiliation 1 appeared as follows:

> *UCLA Center for SMART Health, Clinical Translational Science Institute, David Geffen School of Medicine at UCLA, Los Angeles, CA, United States*

In the revised version, it has been revised to:

> *UCLA Center for SMART Health, Clinical and Translational Science Institute, David Geffen School of Medicine at UCLA, Los Angeles, CA, United States*

In the originally published paper, Affiliation 2 appeared as follows:

> *David Geffen UCLA School of Medicine, Los Angeles, CA, United States*

In the revised version, it has been revised to:

> *David Geffen School of Medicine at UCLA, Los Angeles, CA, United States*

In the originally published paper, Affiliation 4 appeared as follows:

> *Office of Clinical Research, Clinical and Translational Science Institute, David Geffen School of Medicine at UCLA, Los Angeles, CA, United States*

In the revised version, it has been revised to:

> *Embedded Clinical Research and Innovation Unit, CTSI Office of Clinical Research, Los Angeles, CA, United States*

In the originally published paper, the following email address of the corresponding address was provided:

arashnaeim@gmail.com

In the revised version, it has been changed to:

anaeim@mednet.ucla.edu

The corrections will appear in the online version of the paper on the JMIR Publications website on October 21, 2021, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 28.09.21; this is a non–peer-reviewed article;accepted 05.10.21; published 21.10.21.

Please cite as:
Naeim A, Dry S, Elashoff D, Xie Z, Petruse A, Magyar C, Johansen L, Werre G, Lajonchere C, Wenger N
Correction: Electronic Video Consent to Power Precision Health Research: A Pilot Cohort Study
JMIR Form Res 2021;5(10):e33891
URL: https://formative.jmir.org/2021/10/e33891
doi:10.2196/33891
PMID:34673529

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org