JMIR Formative Research

Impact Factor (2023): 2.0
Volume 4 (2020), Issue 6 ISSN 2561-326X Editor in Chief: Amaryllis Mavragani, PhDc

Contents

Original Papers

Corrigenda and Addenda

Original Paper

A Mobile Health Intervention System for Women With Coronary Heart Disease: Usability Study

Avijit Sengupta[1*], MSc; Theresa Beckie[2*], PhD; Kaushik Dutta[1*], PhD; Arup Dey[3*], MSc; Sriram Chellappan[3*], PhD

[1]Information Systems and Decision Sciences, University of South Florida, Tampa, FL, United States

[2]College of Nursing, University of South Florida, Tampa, FL, United States

[3]College of Engineering, University of South Florida, Tampa, FL, United States

[*]all authors contributed equally

Corresponding Author:
Kaushik Dutta, PhD
Information Systems and Decision Sciences
University of South Florida
4202 E Fowler Avenue
Tampa, FL, 33620
United States
Phone: 1 8139746338
Email: duttak@usf.edu

Abstract

Background: Coronary heart disease (CHD) is the leading cause of death and disability among American women. The prevalence of CHD is expected to increase by more than 40% by 2035. In 2015, the estimated cost of caring for patients with CHD was US $182 billion in the United States; hospitalizations accounted for more than half of the costs. Compared with men, women with CHD or those who have undergone coronary revascularization have up to 30% more rehospitalizations within 30 days and up to 1 year. Center-based cardiac rehabilitation is the gold standard of care after an acute coronary event, but few women attend these valuable programs. Effective home-based interventions for improving cardiovascular health among women with CHD are vital for addressing this gap in care.

Objective: The ubiquity of mobile phones has made mobile health (mHealth) behavioral interventions a viable option to improve healthy behaviors of both women and men with CHD. First, this study aimed to examine the usability of a prototypic mHealth intervention designed specifically for women with CHD (herein referred to as HerBeat). Second, we examined the influence of HerBeat on selected health behaviors (self-efficacy for diet, exercise, and managing chronic illness) and psychological (perceived stress and depressive symptoms) characteristics of the participants.

Methods: Using a single-group, pretest, posttest design, 10 women participated in the 12-week usability study. Participants were provided a smartphone and a smartwatch on which the HerBeat app was installed. Using a web portal dashboard, a health coach monitored participants' ecological momentary assessment data, their behavioral data, and their heart rate and step count. Participants then completed a 12-week follow-up assessment.

Results: All 10 women (age: mean 64.4 years, SD 6.3 years) completed the study. The usability and acceptability of HerBeat were good, with a mean system usability score of 83.60 (SD 16.3). The participants demonstrated statistically significant improvements in waist circumference (P=.048), weight (P=.02), and BMI (P=.01). Furthermore, depressive symptoms, measured with the Patient Health Questionnaire-9, significantly improved from baseline (P=.04).

Conclusions: The mHealth prototype was feasible and usable for women with CHD. Participants provided data that were useful for further development of HerBeat. The mHealth intervention is expected to help women with CHD self-manage their health behaviors. A randomized controlled trial is needed to further verify the findings.

(JMIR Form Res 2020;4(6):e16420) doi:10.2196/16420

KEYWORDS

coronary heart disease; mobile health technology; behavior change interventions; women; mobile phone

Introduction

Center-based cardiac rehabilitation (CBCR) is a multidisciplinary, comprehensive, evidence-based intervention with proven morbidity and mortality benefits [1-4]. Outpatient CBCR in the United States generally takes place three times per week over 12 weeks [1,2]. Cardiac rehabilitation is the gold standard of care for the secondary prevention of cardiovascular (CV) disease and focuses on healthy behaviors, including physical activity (PA), healthy eating, psychosocial counseling for stress management, medication adherence, and smoking cessation [1,2]. Although CBCR provides irrefutable health benefits compared with usual care, significant underutilization and lack of access make CBCR programs beneficial only to the few who have health insurance and transportation to the facility [3,4].

CBCR referral is a health care quality performance metric [5,6], yet for three decades, only 10% to 20% of eligible women have attended CBCR, with up to a 56% dropout rate [7-18]. CBCR underutilization stems from numerous intrapersonal, interpersonal, logistical, programmatic, and health system barriers [19,20]. Inadequate health insurance and copayments of up to US $250 per session deter women from CBCR participation [21]. Socioeconomically deprived women who face transportation challenges, family or work obligations, depression, anxiety, or low social support are especially unable to use CBCR [22-27]. These limitations have prompted a call to redesign CBCR for women [7,28,29].

Home-based cardiac rehabilitation (HBCR) offers a potential solution as it avoids conflicts with competing demands of daily life; however, limited evidence exists that HBCR is effective and will reach more women [30-32]. Our study is a direct response to the call to action to expand the reach of secondary prevention to women unable to attend CBCR [7]. We explore the feasibility of delivering technology supported behavior change interventions to women with coronary heart disease (CHD). On the basis of our previous proof-of-concept research [33-35], we translated our gender-specific, motivationally enhanced CBCR program to a prototype of a mobile health (mHealth) home-based behavioral intervention (referred to here as HerBeat). HerBeat has the potential to improve health behaviors and CV risk factors in women with CHD by overcoming barriers inherent in CBCR, expanding reach to the majority of women without access to CBCR, and integrating a home-based program seamlessly into their lives.

Up to 80% of CHD events are attributed to unhealthy behaviors [36]; adherence to health behaviors unquestionably improves CV health [8,36,37]. Fortunately, CBCR practice standards are widely disseminated and readily adaptable for a gender-specific HBCR, based on the results from HBCR studies [38,39]. CBCR-eligible patients given the choice between HBCR and CBCR are up to four times more likely to participate in HBCR [40-42]. Compared with CBCR, HBCR overcomes logistical barriers to access, the need for expensive facilities, specialized exercise equipment, and high personnel costs and provides education, coaching, and monitoring by a health coach through, when available, wearable sensors and smartphones that are potentially operational 24 hours a day, 7 days a week [20,43]. Moreover, HBCR assesses daily PA, whereas CBCR only measures supervised exercise sessions [44]. Most CHD patients spend over 5000 waking hours yearly, independent of medical providers [45], and thus, arming them with behavior change techniques (BCTs) that can be implemented anytime is crucial.

A BCT is defined as an observable and replicable intervention component designed to redirect causal processes that regulate behavior, a technique proposed to be an active ingredient [46]. Unlike most mHealth interventions that deliver text messages at preset times, largely unrelated to patient behavior [47], HerBeat delivers personalized, just-in-time adaptive interventions comprising gender-specific, behavior theory–based BCTs in response to proximal behaviors and moods. Theoretically derived BCTs delivered anytime and anywhere are essential to forming and maintaining health behaviors into lifelong habits. We used four specific BCTs as we designed interventions to be deployed through HerBeat: (1) goals and planning, (2) feedback and monitoring, (3) shaping knowledge, and (4) repetition and substitution [46]. For the BCT goals and planning, we used the subtechniques *goal setting* and *review behavior goal* for creating instantiations of the intervention. For the BCT feedback and monitoring, we used subtechniques such as *feedback on behavior*, *self-monitoring of behavior*, *monitoring of outcomes*, and *feedback on outcomes*. We used subtechniques such as *instructions on how to perform the behavior* and *information about antecedents* for the BCT shaping knowledge, and for the BCT repetition and substitution, we used subtechniques such as *graded tasks* and *habit formation and habit reversal*. Higher levels of self-monitoring/management and unsupervised exercise inherent in HBCR vs CBCR can aid transition from active intervention to lifelong self-management seamlessly.

First, the purpose of this study was to examine the usability of a prototype of HerBeat for women with CHD. Second, we sought to examine the potential influence of the prototype on health behaviors (eating habits, PA, and goal setting) and psychosocial characteristics (self-efficacy [SE], depressive symptoms, and perceived stress) at the 12-week follow-up visit.

Methods

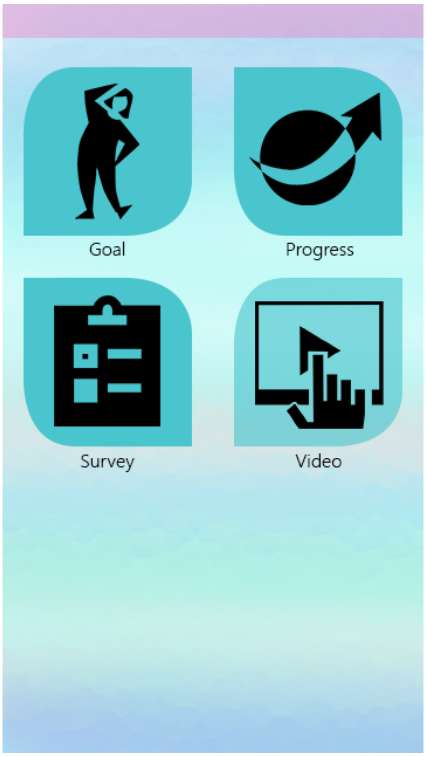
Design Overview

We used a smartwatch app and a smartphone app to collect data on a patient's daily PA, heart rate, eating episodes, and mood. Data from the sensors embedded in the smartwatch are interpreted as step counts and heart rate and are sent to the smartphone via Bluetooth and then to a cloud drive via Wi-Fi or 4G. All data uploaded to the cloud are then downloaded immediately and uploaded to a server over a secured virtual private network connection through the public internet. Data in the server are analyzed and projected on the dashboard for the health coach to view. The old data are archived and then refreshed by the most recent data on the dashboard every 10 min.

Intervention

The HerBeat prototype included a wrist-worn smartwatch (Moto 360 2nd Gen, Android Wear OS 2.0) and a smartphone (Samsung Galaxy S6, Android 7.0), with the app installed on both devices and a web-based dashboard for monitoring participant data. The 4 features of the prototype included (1) goal setting, (2) progress, (3) ecological momentary assessment (EMA) surveys, and (4) videos (see Figure 1). The goal setting feature allowed participants to set multiple walking goals for up to 60 min each. Study participants were tasked with setting their own PA goals in terms of the number of minutes walked. Participants were also prompted to report their readiness to begin PA and their current level of energy on a scale of 1 to 10. After setting a PA goal, each participant was sent a motivational message that encouraged exercise. Data about the participant goal setting and subsequent PA performance were monitored through a web-based dashboard in real time by a trained professional.

Figure 1. Main menu screen of HerBeat application.



The progress function permitted participants to review the number of minutes walked, number of steps taken, and distance covered in miles. If participants had not completed their goal when seeking progress, they were presented with the number of minutes remaining to goal completion. If a goal was completed, the participant was sent a gender-specific graphic user interface (GUI) with a congratulatory message for achieving their goal. The EMAs are described in the Measurement section. The final feature provided participants access to 9 customized short videos, developed by the principal investigator (PI) with expertise in behavioral medicine and women's CV health, on healthy eating behavior and on guidelines for safe PA. The app also sent two types of behavior change intervention messages. If the participant had not set a PA goal by 4 PM daily, a message prompting them to exercise was sent. If participants were proactive in setting and achieving walking goals, they were sent a positive reinforcing message. Figures 2-5 show some of the examples of GUIs of intervention screens. The dashboard was used by the health coach to remotely monitor participants' PA (step count), heart rate, goal setting behavior, responses to the EMA surveys, and frequency of accessing the health videos. The health coach, via the dashboard, also monitored episodes of Wi-Fi and Bluetooth disconnections. The health coach sent a personalized, encouraging message to engage with HerBeat to the participant's smartphone about once a week.

Figure 2. Example of graphic user interfaces of interventions (congratulatory message for achieving goal).

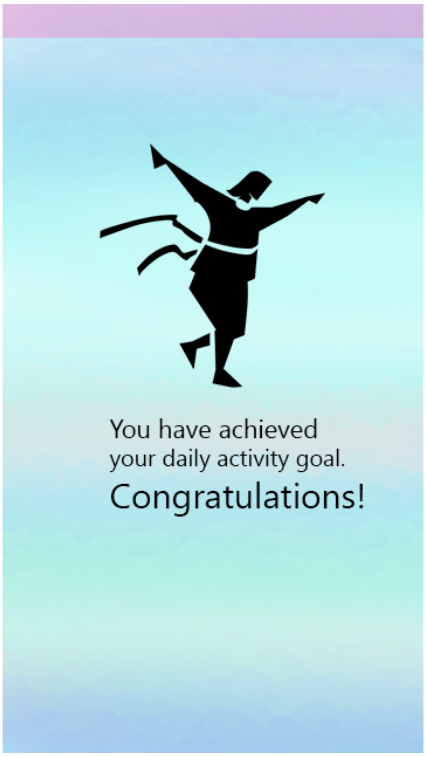


Figure 3. Example of graphic user interfaces of interventions (physical activity schedule).

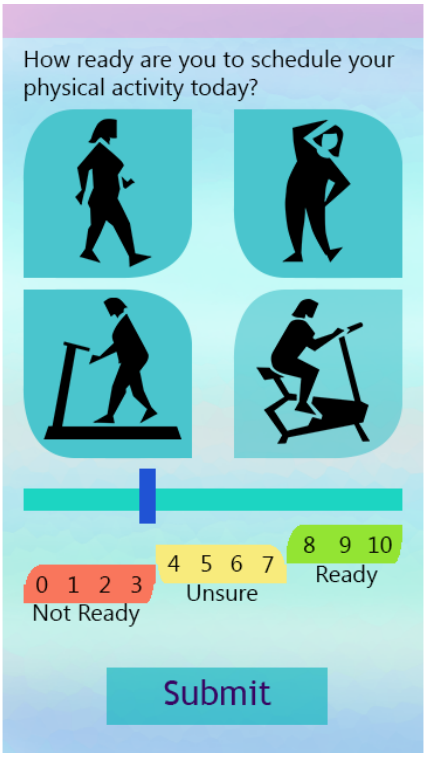


Figure 4. Example of graphic user interfaces of interventions (prompting to exercise).

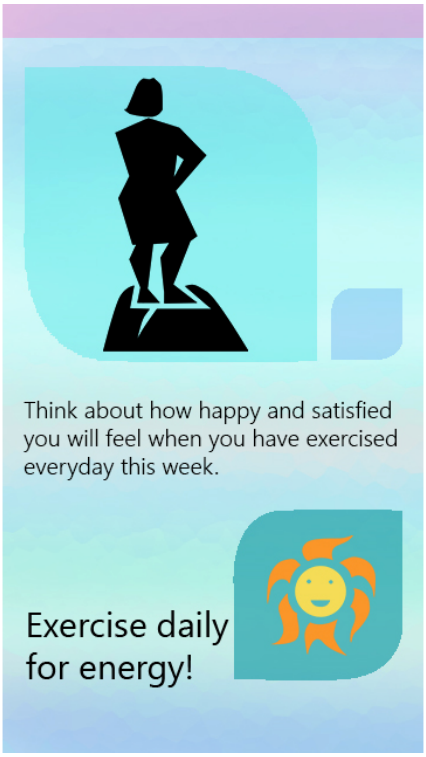
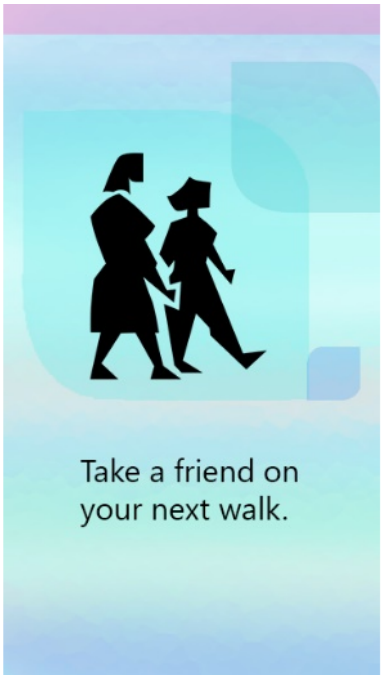


Figure 5. Example of graphic user interfaces of interventions (positive reinforcing message).



Recruitment

After obtaining approval from the university institutional review board and using a single-group, pretest, posttest design, we recruited participants from a university-affiliated outpatient cardiology clinic between May 2018 and August 2018. The cardiology clinic is part of an academic medical center and is staffed by faculty members who are physicians in the Division of Cardiovascular Sciences. The clinic provides state-of-the-art services and treatment options. Participants were recruited by the PI (health coach), who had access to the clinic's electronic health record system. Potential participants who were scheduled to see their health care provider were approached for inclusion in the study after they had completed their clinic visit. Women were eligible for the study if they were aged 50 years or older; diagnosed with an acute coronary syndrome or coronary revascularization in the last 10 years; able to read, speak, and understand English; and able to participate in a PA, such as walking, unaided. We also sought verbal clearance from their cardiologist to participate in the study. Study exclusion criteria included residing outside a 50-mile radius of the study site; a psychiatric condition including dementia, delirium, or

schizophrenia or actively undergoing acute psychiatric treatment; prior neurological brain disorders; current use of illicit drugs and/or chronic alcohol use at the discretion of the PI; or life-limiting comorbid conditions (eg, metastatic cancer).

Study Procedures

The informed consent form clearly explained that study participation was voluntary and participants could withdraw from the study at any time without jeopardizing their health care. Their decision to withdraw had no impact on their relationship with their cardiologist. If they wished to withdraw study participation, they needed only to inform the PI, and no further data would be collected from that time onward.

After baseline assessment was completed, the participants were trained by a graduate student with technical expertise to use the smartwatch and the smartphone that were provided for the duration of the study. Technical questions were answered by one of the study personnel via telephone or in person. The participants were then asked to use the prototype for 12 weeks and return for a follow-up visit when data collection was completed and the hardware was returned. We did not explicitly request that participants improve their health behaviors because our primary focus was the usability of the HerBeat prototype.

Data collected from self-report questionnaires and physical assessments were maintained in a database using Research Electronic Data Capture software. The smartwatch streamed step count and heart rate data continuously between 6 AM and 10 PM daily every 3 min via Bluetooth and Wi-Fi to a Health Insurance Portability and Accountability Act–compliant server. We archived participants' data by a study identification number to protect their identity. Daily backend jobs processing all collected information were automated on the server side. The resulting information was stored in a Structured Query Language (SQL) format for easy retrieval. The web portal dashboard was created from the SQL data to present the data to the health coach.

Measures

Usability

We evaluated the participants' perceptions of usefulness, ease of use, and satisfaction with HerBeat with the System Usability Scale (SUS) [48]. The SUS was first introduced in 1986 and consists of 10 items. The study by Lewis and Sauro [48] suggested that SUS has two different factors. The first factor consists of 8 items on a 5-point scale that measures how usable the system is, and the second factor measures how easy it is to learn the system. These correlated factors have reasonable reliability (coefficient α of .91 and .70, respectively) and correlated highly with the overall SUS [48]. A sensitivity analysis conducted by Lewis and Sauro [48] suggested that using data from 19 tests had a significant *test* by *scale* interaction, providing additional evidence of the differential utility of the scale. The SUS is frequently used by both researchers and practitioners, given the adequate reliability data and ease of implementation. Scoring guidelines of the SUS recommend transforming the scale to a 0 to 100 range. The SUS yields a single number representing a composite measure of overall usability, and scores for individual items have very limited meaning on their own. SUS follows a specific rubric and reverse scoring of certain items to calculate the final usability composite score from the scores against individual items.

Sociodemographic and Health History

At baseline, we collected data on cardiac history, comorbidities, medications, and CV risk factors as well as sociodemographic attributes, including age, marital status, work status, education, occupation, living arrangements, insurance status, and income.

Psychosocial

Dietary SE was measured using the 20-item Eating Habits Confidence Survey consisting of a 5-point scale ranging from 1 to 5, with higher scores reflecting higher SE [49]. This instrument has shown strong internal consistency reliability in overweight postmenopausal women [50]. Exercise SE was measured with the 12-item Exercise Confidence Survey asking participants to rate their confidence in maintaining an exercise routine when facing various barriers. Scores range from 12 to 60; higher scores reflect higher SE [49]. Participants' perceptions of their SE for managing chronic illness were assessed with a 6-item instrument, with scores ranging from 6 to 60 [51]. The scale measures the perceived adaptability of survey participants to manage different aspects of chronic diseases, such as pain and fatigue, and the scores demonstrate good reliability (Cronbach α coefficient .89) [51]. The Perceived Stress Scale (PSS) [52] consists of 14 items that are measured on a 5-point scale, with higher scores reflecting greater perceived stress. PSS scores are obtained by first reversing the scores on the 7 positive items and then summing across all 14 items. The coefficient α reliability for the PSS was .85, and the validity of PSS was established by showing substantial correlations between the scale and standard symptomatology measures [52]. Depressive symptoms were measured with the Patient Health Questionnaire-9 (PHQ-9), with scores ranging from 0 to 17, with higher scores reflecting more depressive symptoms [53]. Scores from the PHQ-9 questionnaire items showed strong reliability and validity when used by researchers to measure major depressive disorder [54], depression [55] (Cronbach α .85), and depression in patients with CHD (Cronbach α .90) [56].

Behavioral

Eating behavior was assessed using the 13-item Rapid Eating Assessment for Participants-Short Form (REAP-S) [57]. Possible scores range from 13 to 39, with a higher score indicating better diet quality [58]. Self-reported PA was assessed using the 7-day recall International Physical Activity Questionnaire-Short Form (IPAQ-SF), which measures PA intensity, frequency, and duration [59]. Items in the IPAQ-SF were structured to provide separate scores on walking and moderate- and vigorous-intensity activities. The IPAQ-SF questionnaire showed moderate to strong reliability (intraclass correlation coefficient [ICC]) in prior research conducted with college students (ICC=0.71-0.89) [60], Chinese youth (ICC=0.43-0.83) [61], pregnant women (ICC=0.81-0.84) [62], and individuals with schizophrenia [63]. Step count and distance walked were measured objectively with the Moto 360 smartwatch over 12 weeks. The smartwatch

determines step count via processing readings from the accelerometer and gyroscope sensor, from which we estimate the distance walked. The heart rate was collected continuously when the watch was worn. The smartwatch was worn every day, except while bathing, sleeping, or swimming.

Physiological

With participants in light clothing without shoes, weight was measured to the nearest 0.1 kg using research precision–grade, calibrated, digital scales, and height was measured to the nearest 0.1 cm using a freestanding stadiometer. BMI was calculated as weight (kg)/height (m^2). Waist circumference, assessed just above the uppermost lateral border of the right ilium using a Gulick tape measure, was calculated to the nearest 0.1 cm as the mean of the second and third measures [64,65]. Blood pressure (BP) was obtained with a calibrated automated monitor according to the standard protocol [66].

Ecological Momentary Assessments

Participants completed brief 1- to 2-min surveys sent to their smartphone at 8 random times throughout the day. These surveys asked about their current activity, location, mood, eating episodes, and who they were with.

Data Analysis

Data analysis techniques are applied to gain insights into the patient's activity, heart rate, and EMA survey response data. Patients' physical activities and EMA responses are analyzed through a decision rule-based expert system as well as by the health coach. These data are used to send standard preprogrammed intervention messages to the patients by the system and to help the health coach to customize intervention messages to send to patients through the dashboard at the right time to maximize their impact.

Descriptive statistics (eg, univariate graphical and numerical statistics, bivariate distributions, scatterplots, and counts/percentages) were generated and summarized for all study data. Paired *t* tests were used to compare continuously measured variables from baseline to the 12-week posttest measures. Given the small sample size, we were generally underpowered to perform parametric statistics. The α level was set at $P \leq .05$. Qualitative field notes were summarized across all participants for themes.

Results

Participants

A total of 11 participants signed the informed consent form, and 10 participants completed data collection. Of the 10 participants, 2, both recently experiencing traumatic life events, engaged very little with HerBeat. Most of the participants were white (8/10, 80%), married, or partnered (6/10, 60%) women with a mean age of 64 years (range 53-75 years; SD 6 years; see Table 1). The majority of participants had health insurance and an income of at least US $40,000 annually; 5 participants worked full time. All participants had CHD, with 2 participants diagnosed with a myocardial infarction and one with heart failure. Moreover, 50% (5/10) of participants had undergone a percutaneous coronary intervention, and 20% (2/10) of participants had undergone a coronary artery bypass graft surgery. None of the participants had ever attended a CBCR program.

The participants had multiple comorbidities including diabetes mellitus (3/10, 30%), osteoarthritis (4/10, 40%), and orthopedic disorders (2/10, 20%), and one participant was being treated for skin cancer (1/10, 10%; see Table 2). The participants exhibited traditional CV disease risk factors, including dyslipidemia, hypertension, physical inactivity, familial heart disease, and being overweight. Most participants had never used tobacco; former smokers had a mean of 23.75 (SD 19.3) pack-years of smoking. Participants were prescribed numerous evidence-based CV medications to treat their chronic conditions.

Table 3 summarizes the baseline and 12-week follow-up physiological and psychosocial participant characteristics. Although we observed no changes in BP, the participants had statistically significant improvements in waist circumference (P=.048), weight (P=.02), and BMI (P=.01). Furthermore, participants' depressive symptoms significantly improved from baseline (P=.04).

SE for exercise, diet, and managing chronic illness was not statistically significantly different from baseline, although it trended in the desired direction. Participants also demonstrated nonsignificant improvements in REAP-S scores and perceived stress.

Table 1. Participants' sociodemographic data (n=10).

Characteristics	Values
Age (years)	
Mean (SD)	64.4 (6.3)
Range	53-75
Race or ethnicity, n (%)	
White	8 (80)
Black, African American	1 (10)
Asian/Pacific Islander	1 (10)
Education, n (%)	
Community college	5 (50)
4-year college incomplete	1 (10)
4-year degree	1 (10)
Master's degree	2 (20)
Doctoral degree	1 (10)
Employment status, n (%)	
Employed full time	5 (50)
Not employed or retired	5 (50)
Marital status, n (%)	
Married/partnered	6 (60)
Divorced	2 (20)
Widowed	2 (20)
Primary insurance status, n (%)	
Private insurance	6 (60)
TriCare (military/veterans)	1 (10)
Medicaid	2 (20)
Medicare	1 (10)
Annual household income (US $), n (%)	
20,000 to <40,000	3 (30)
40,000 to <80,000	2 (20)
80,000 to <100,000	2 (20)
≥100,000	3 (30)

Table 2. Clinical characteristics of participants (n=10).

Characteristics	Value, n (%)
Cardiovascular disease diagnosis	
Coronary heart disease	8 (80)
Myocardial infarction	1 (10)
Congestive heart failure	1 (10)
Comorbidities	
Diabetes	3 (30)
Arthritis	4 (40)
Orthopedic disorder	2 (20)
Skin cancer	1 (10)
Cardiovascular risk factors	
Overweight (BMI 25.0-29.9 kg/m^2) or obese (BMI >30 kg/m^2)	6 (60)
Familial heart disease (onset before 60 years and 50 years in mother and father, respectively)	2 (20)
Dyslipidemia	10 (100)
Hypertension	6 (60)
Physical inactivity (<30 min 5 times weekly)	8 (80)
Tobacco use	
Never	4 (40)
Former	6 (60)
Medication classes prescribed	
Beta blocker	8 (80)
Calcium channel blocker	2 (20)
Angiotensin-converting enzyme inhibitor	4 (40)
Angiotensin receptor blocker	3 (30)
Statin	10 (100)
Insulin	2 (20)
Metformin	2 (20)
Aspirin	9 (90)
Clopidogrel	5 (50)
Other antiplatelet	3 (30)

Table 3. Physiological and psychosocial characteristics (n=10).

Characteristics[a]	Baseline, mean (SD)	12-week follow-up, mean (SD)	*P* value
Systolic blood pressure (mm Hg)	129.2 (12.3)	141.5 (18.9)	NS[b]
Diastolic blood pressure (mm Hg)	76.7 (8.7)	73.6 (9.2)	NS
Waist (cm)	97.7 (14.7)	95.4 (12.6)	.048
Weight (kg)	80.5 (19.7)	79.1 (18.6)	.02
BMI (kg/m^2)	29.2 (6.0)	28.7 (5.8)	.01
Self-Efficacy Scale for Managing Chronic Disease	45.4 (12.5)	48.2 (7.6)	NS
Self-efficacy for exercise behavior	52.5 (7.6)	54.4 (6.2)	NS
Self-efficacy for diet	88.8 (6.0)	89.6 (6.8)	NS
Perceived Stress Scale	13.3 (6.7)	9.9 (6.9)	NS
Patient Health Questionnaire-9	5.5 (5.4)	2.9 (3.8)	.04
Rapid Eating Assessment for Participants-Short Form	32.7 (3.5)	33.7 (2.7)	NS
International Physical Activity Questionnaire (last 7 days)			
Days of moderate physical activity	3.0 (2.4)	3.4 (2.3)	NS
Minutes per day of moderate physical activity	35.7 (35.3)	63.1 (52.8)	NS
Minutes sitting on 1 week day	330.0 (124.1)	331.0 (212.6)	NS
Days walked at least 10 min per day	5.4 (2.3)	5.5 (1.7)	NS
System Usability Scale	N/A[c]	83.6 (16.4)	N/A

[a]$\alpha \leq .05$.

[b]NS: not significant.

[c]N/A: not applicable.

Engagement With the Prototype

Over the course of the study, participants (n=8) collectively set 132 goals, with a mean of 16.5 (SD 17.3) goals per participant for a collective total of 3335 min of walking, with a mean of 34.72 (SD 41.68) min per participant (see Table 4) per week. Most of the walking goals were set between 9 AM and 11 AM and 5 PM and 6 PM. Over the course of the study, smartwatches allocated to the participants collectively recorded 4933 min of walking, with a mean of 22.02 (SD 35.32) min per participant per day. That is, the participants walked more than they intended when setting a goal. Over 12 weeks, each participant walked a mean of 28 days (out of a possible 90 days) and took a mean of 3718.8 (SD 3826.0) steps per day. The group responded to 830 EMA surveys and accessed 8 health educational videos 165 times during the study. The participants accessed more videos related to healthy eating behavior (137/165, 83%) than those related to PA (28/165, 17%). The participants received a total of 265 automated intervention messages based on their progress toward their goals.

Table 4. Participants' engagement (N=8).

HerBeat features	Value, range	Value, mean (SD)
Goals		
Number of goals set per participant	3-52	16.5 (17.3)
Walking goal set (minutes) per participant per week	1-60	34.72 (41.68)
Progress		
Daily walking (minutes)	1-132	22.02 (35.32)
Daily steps per participant	3-21,179	3718.8 (3826.0)
Daily miles per participant	0.1-10.6	1.86 (1.9)
Videos		
Number of times health videos were accessed per participant per week	0-17	1.96 (1.76)
Number of ecological momentary assessment survey responses per participant per week	0-36	8.64 (9.45)
Behavior change messages acknowledged per participant per week	0-7	2.75 (2.65)

Usability

The mean score on the SUS was 83.60 (SD 16.4). Participants generally found HerBeat to be easy to learn and use. They also found the functionalities to be well integrated, and they felt confident in using HerBeat. The participants did not find it unnecessarily complex or cumbersome to use (see Table 5).

Table 5. Descriptive statistics of the System Usability Scale items.

No	Item	Value, mean (SD)
1	I think I would like to use this system frequently.	79.5 (2.13)
2	I found the product unnecessarily complex. (R)[a]	83.5 (1.15)
3	I thought the product was easy to use.	86.6 (1.29)
4	I think that I would need the support of a technical person to be able to use this product. (R)	91.0 (2.25)
5	I found that the various functions in this product were well integrated.	83.1 (1.79)
6	I thought that there was too much inconsistency in this product. (R)	81.1 (1.26)
7	I would imagine that most people would learn to use this product very quickly.	82.2 (1.28)
8	I found the product very cumbersome to use. (R)	78.7 (2.14)
9	I felt very confident using the product.	87.7 (2.17)
10	I needed to learn a lot of things before I could get going with this product.	82.2 (1.02)

[a](R)=reversed scored item.

They also reported not requiring the support of a technical person to use HerBeat. Only one patient required a home visit to address a technical issue. Participants' themes derived from field notes mostly involved technical issues. The most frequent complaint was the short battery life of the smartwatch. We rectified this problem after valuable participants' input. Some working participants found it difficult to carry both a personal phone and a study phone and respond to EMA surveys during the day. A participant who worked in a library sought permission from her supervisor to carry the study phone and respond to the EMA surveys. One participant requested taking HerBeat with her to Europe to allow her to track her activity while on vacation.

Participants' feedback also led to the redesign of some of the GUIs of the EMA survey. Although there was minimal contact between the health coach and the participants during the 12 weeks and participants went on vacation during the study, they voiced reassurance that their progress was being monitored by the health coach via the dashboard. Participants had no adverse events during the study, and there were no issues raised about privacy concerns.

Data captured during our study suggest that at least one of the participants set a walking goal of 1 min and at least one of the participants watched no health-related videos during the study. We probed the corresponding participants during the final debriefing session about these data. For the first observation, the participant suggested that the walking goal of 1 min was mistakenly set while exploring the goal setting function at the very beginning of the study. The participant's intention was to navigate further inside the goal setting function. Regarding the second observation, the participant chose not to watch any health-related videos because she felt well informed about these health behaviors.

Discussion

Principal Findings

The primary aim of this study was to determine the usability of our mHealth system, HerBeat, with a cohort of women with CHD before proceeding with the development of a comprehensive home-based secondary prevention intervention. Our secondary aim was to evaluate the influence of HerBeat on various psychosocial and health behaviors of the participants. To our knowledge, this is the first study to evaluate the usability of a gender-specific mHealth app for secondary prevention of CHD in women. The main finding of the study was that the system was acceptable and usable in its prototypic form. The level of engagement of participants with HerBeat was greater than anticipated, given the relatively primitive features. We developed HerBeat to avoid high data entry burden and designed gender-specific GUIs to foster engagement. Given that 80% of health-related apps are abandoned after only 2 weeks [67], the engagement of the participants with our prototype was good, particularly when they were given little prodding for using the technology. We viewed this as an encouragement to proceed with the expanded version of HerBeat, with increased involvement of the health coach.

Additional Findings

Comparisons of user engagement with mobile apps of participants with characteristics similar to the participants in our study are difficult to make because usability was defined differently in these studies [68-71]. Some described metrics such as app usage frequency, duration, data registration, or responsiveness of the user to daily tasks. In addition to the often low participant numbers, dropouts, and short study duration, conclusions about engagement are difficult to draw. Completion of tasks within the app, such as completion of an education module, was a typical measure of use in studies with a focus on healthy lifestyle. Forman et al [68] gauged engagement by

patient completion of at least one prescribed daily task. In other studies, emphasis was placed on logging medication intake or physical measurements [69,70]. The authors did not report the acceptability of a data entry requirement. We made the decision early during development, based on numerous interviews with patients, to avoid the requirement of data entry to reduce respondent burden. In an uncontrolled single-group, pretest, posttest design [72], participants were required to log daily BP measurements for 55 days; however, it was unclear whether all patients logged BP on each of the 55 days. Patients in one small study of both heart failure and CHD participants [73] appreciated medication reminders and PA information. However, they felt that daily requirements for data entry or other responses were inconvenient. Clearly, high data entry burden is a usability issue [74].

We did not see evidence of the message fatigue reported by others [75]. The fact that the participants responded to 830 EMA surveys over 12 weeks was, in our opinion, quite remarkable. Although the number of EMA survey responses was greater than expected, the responses declined over time. Educational videos on healthy eating behavior were viewed more often than videos related to physical activities, presumably because eating a healthy diet is often a daily or hourly struggle between reflex and self-control. Participants may have viewed the videos to seek assistance with making healthy eating decisions. Eating and body weight regulation is a complex process that involves both metabolic and hormonal control mechanisms and neurocognitive processes involved with memories, expectations, and evaluation of food and the consequences of eating [76]. The decisions about what and when to eat are a balance between reflexive behavior and higher-level cognitive processes. Eating can be reflexive and automatic by the mere smell of a favored food [77]. This reflexive eating can be opposed by dietary restraint of choosing a healthy food that involves higher-level cognitive processes to counter the power of tempting environmental stimuli [78].

On the basis of decision rules related to participants' responses employed in HerBeat, some intervention messages were deployed more frequently than others. Most participants exceeded the walking goals they set. In other words, most of the time participants did not abruptly stop their walk after achieving their PA goal but rather exercised beyond the goal. We hypothesize that this may reflect low SE when setting the goal, followed by greater confidence when they surpassed the goal. Although we did not set a target for time spent walking or for step count, the participants' daily step count was relatively modest. A common goal of 10,000 steps per day has been perpetuated by the lay press and is often used as the default by software programs on wearables and smartphones [79]. In the United States, the average number of steps accrued daily (measured by smartphones) is approximately 4800; worldwide, it is approximately 5000 [80]. There is sparse data on the number of daily steps needed for health [81,82] or clinical outcomes and mortality [83]. In the Women's Health Study, a cohort of 16,741 women with a mean age of 72 years wore accelerometers to measure their steps per day over 7 days [84]. Women who averaged 4400 steps per day had significantly lower mortality rates during a follow-up of 4.3 years compared with the least active women who took approximately 2700 steps per day. As more steps per day were accrued, mortality rates progressively decreased before leveling at approximately 7500 steps per day [84].

Although we did not expect participant health behaviors, SE, perceived stress, or depressive symptoms to improve with a limited functionality prototype, we nonetheless observed significant reductions in waist circumference, weight, and BMI as well as reduced depressive symptoms after study participation. These improvements were unexpected because the research team had minimal contact with the participants during the 12 weeks, and we did not prompt them to set goals for walking. Participants reported minimal positive changes in their SE for exercise, diet, or managing chronic illness, but scores nonetheless trended in the expected direction. From baseline to the 12-week follow-up, there was a modest increase in the mean minutes of moderate-intensity exercise. There were no reductions in participants' time spent sitting. The primary purpose of this study was to examine the usability of the system, and secondarily, to examine behavior change after the 12-week study. With a more robust version of the system, we will examine the effectiveness of the system in a randomized clinical trial.

Limitations

Our findings must be balanced with the limitations of the study. First, this was a small convenience sample from a single study site. With multiple statistical testing, we may have capitalized on chance findings. The generalizability of the findings is limited to women with CHD. Furthermore, we used a nonexperimental design without a control group. Second, this was a usability test of a minimal viable product with minimal contact from the research team. Third, our study was not long enough to evaluate any sustained behavior change. A randomized controlled trial with a larger sample is needed to better understand the optimal way of providing secondary prevention through digital health interventions. However, the aim of our study was to examine the usability, viability, and user requirements for developing a more comprehensive mHealth intervention for women with CHD.

Future Directions

This usability study has encouraged us to develop a comprehensive mHealth behavior change intervention that targets PA, healthy eating, stress management, medication adherence, and smoking cessation. Such a home-based system is not intended to replace CBCR but rather to offer behavior change theory–based interventions in real time to individuals as they live their lives, particularly for those who cannot access CBCR. Evidence for the effectiveness of self-management of multiple health behaviors for improved outcomes will require a larger, randomized controlled trial of a longer duration. A pilot randomized study of the next version of HerBeat is currently underway.

Our formative evaluation of HerBeat helped us to refine our design strategy for the next trial. We plan to incorporate more provision for the user to communicate with the health coach as a group as well as individually. We have expanded the EMA

surveys to target more behaviors of relevance to CV health. With feedback from the participants, we have developed many more meaningful BCTs that are deployed using decision rules in response to the participants' responses to the EMA surveys. We have enhanced the dashboard to be more visually usable by the health coach. Finally, we have resolved some of the problems with the wearable sensor by implementing the use of a different smartwatch that has a long battery life.

Conclusions

CV disease remains the leading cause of death worldwide. Healthy lifestyle behaviors are critical to CV health. We designed a mHealth prototype specifically for women with CHD to assist them with behavioral self-management. The participants found the prototype easy to use over 12 weeks and were receptive to setting walking goals and responding to EMA surveys. Mobile technology is an innovative and scalable approach to reducing the risk factors of CV disease, but evidence related to acceptability remains limited. Our study has contributed to the limited data on the usability of mobile apps for CV disease self-management.

Acknowledgments

This work was supported by the National Science Foundation (National I-Corps Program), the University of South Florida Bull Ring Accelerator Grant, and the University of South Florida Women in Leadership and Philanthropy Heart Health Faculty Research Award to the principal investigator (TB).

Conflicts of Interest

None declared.

References

1. Balady GJ, Williams MA, Ades PA, Bittner V, Comoss P, Foody JM, American Heart Association Exercise, Cardiac Rehabilitation,Prevention Committee, the Council on Clinical Cardiology, American Heart Association Council on Cardiovascular Nursing, American Heart Association Council on Epidemiology and Prevention, American Heart Association Council on Nutrition, Physical Activity,Metabolism, American Association of Cardiovascular and Pulmonary Rehabilitation. Core components of cardiac rehabilitation/secondary prevention programs: 2007 update: a scientific statement from the American Heart Association Exercise, Cardiac Rehabilitation, and Prevention Committee, the Council on Clinical Cardiology; the Councils on Cardiovascular Nursing, Epidemiology and Prevention, and Nutrition, Physical Activity, and Metabolism; and the American Association of Cardiovascular and Pulmonary Rehabilitation. Circulation 2007 May 22;115(20):2675-2682. [doi: 10.1161/CIRCULATIONAHA.106.180945] [Medline: 17513578]
2. Smith Jr SC, Benjamin EJ, Bonow RO, Braun LT, Creager MA, Franklin BA, World Heart Federation and the Preventive Cardiovascular Nurses Association. AHA/ACCF secondary prevention and risk reduction therapy for patients with coronary and other atherosclerotic vascular disease: 2011 update: a guideline from the American Heart Association and American College of Cardiology Foundation. Circulation 2011 Nov 29;124(22):2458-2473. [doi: 10.1161/CIR.0b013e318235eb4d] [Medline: 22052934]
3. Anderson L, Taylor R. Cardiac rehabilitation for people with heart disease: an overview of Cochrane systematic reviews. Cochrane Database Syst Rev 2014 Dec 12;177(12):CD011273. [doi: 10.1002/14651858.CD011273.pub2] [Medline: 25503364]
4. Anderson L, Oldridge N, Thompson D, Zwisler AD, Rees K, Martin N, et al. Exercise-based cardiac rehabilitation for coronary heart disease: Cochrane systematic review and meta-analysis. J Am Coll Cardiol 2016 Jan 5;67(1):1-12 [FREE Full text] [doi: 10.1016/j.jacc.2015.10.044] [Medline: 26764059]
5. Drozda Jr JJ, Messer J, Spertus J, Abramowitz B, Alexander K, Beam CT, et al. ACCF/AHA/AMA-PCPI 2011 performance measures for adults with coronary artery disease and hypertension: a report of the American College of Cardiology Foundation/American Heart Association Task Force on Performance Measures and the American Medical Association-Physician Consortium for Performance Improvement. Circulation 2011 Jul 12;124(2):248-270. [doi: 10.1161/CIR.0b013e31821d9ef2] [Medline: 21670226]
6. Thomas RJ, Balady G, Banka G, Beckie TM, Chiu J, Gokak S, et al. 2018 ACC/AHA clinical performance and quality measures for cardiac rehabilitation: a report of the American College of Cardiology/American Heart Association Task Force on Performance Measures. J Am Coll Cardiol 2018 Apr 24;71(16):1814-1837 [FREE Full text] [doi: 10.1016/j.jacc.2018.01.004] [Medline: 29606402]
7. Balady GJ, Ades PA, Bittner VA, Franklin BA, Gordon NF, Thomas RJ, American Heart Association Science Advisory and Coordinating Committee. Referral, enrollment, and delivery of cardiac rehabilitation/secondary prevention programs at clinical centers and beyond: a presidential advisory from the American Heart Association. Circulation 2011 Dec 20;124(25):2951-2960. [doi: 10.1161/CIR.0b013e31823b21e2] [Medline: 22082676]

8. Mosca L, Benjamin E, Berra K, Bezanson JL, Dolor RJ, Lloyd-Jones DM, et al. Effectiveness-based guidelines for the prevention of cardiovascular disease in women-2011 update: a guideline from the American Heart Association. Circulation 2011 Mar 22;123(11):1243-1262 [FREE Full text] [doi: 10.1161/CIR.0b013e31820faaf8] [Medline: 21325087]
9. Fihn S, Gardin J, Abrams J, Berra K, Blankenship JC, Dallas AP, American College of Cardiology Foundation, American Heart Association Task Force on Practice Guidelines, American College of Physicians, American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, Society of Thoracic Surgeons. 2012 ACCF/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease: a report of the American College of Cardiology Foundation/American Heart Association Task Force on practice guidelines, and the American College of Physicians, American Association for Thoracic Surgery, Preventive Cardiovascular Nurses Association, Society for Cardiovascular Angiography and Interventions, and Society of Thoracic Surgeons. J Am Coll Cardiol 2012 Dec 18;60(24):e44-164 [FREE Full text] [doi: 10.1016/j.jacc.2012.07.013] [Medline: 23182125]
10. Samayoa L, Grace SL, Gravely S, Scott LB, Marzolini S, Colella TJ. Sex differences in cardiac rehabilitation enrollment: a meta-analysis. Can J Cardiol 2014 Jul;30(7):793-800 [FREE Full text] [doi: 10.1016/j.cjca.2013.11.007] [Medline: 24726052]
11. Colella TJ, Gravely S, Marzolini S, Grace SL, Francis JA, Oh P, et al. Sex bias in referral of women to outpatient cardiac rehabilitation? A meta-analysis. Eur J Prev Cardiol 2015 Apr;22(4):423-441. [doi: 10.1177/2047487314520783] [Medline: 24474091]
12. Supervía M, Medina-Inojosa JR, Yeung C, Lopez-Jimenez F, Squires RW, Pérez-Terzic CM, et al. Cardiac rehabilitation for women: a systematic review of barriers and solutions. Mayo Clin Proc 2017 Mar 13 epub ahead of print [FREE Full text] [doi: 10.1016/j.mayocp.2017.01.002] [Medline: 28365100]
13. Dreyer RP, Xu X, Zhang W, Du X, Strait KM, Bierlein M, et al. Return to work after acute myocardial infarction: comparison between young women and men. Circ Cardiovasc Qual Outcomes 2016 Feb;9(2 Suppl 1):S45-S52 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.115.002611] [Medline: 26908859]
14. Lavie CJ, Bennett A, Arena R. Enhancing cardiac rehabilitation in women. J Womens Health (Larchmt) 2017 Aug;26(8):817-819. [doi: 10.1089/jwh.2017.6476] [Medline: 28613967]
15. Turk-Adawi KI, Grace SL. Narrative review comparing the benefits of and participation in cardiac rehabilitation in high-, middle- and low-income countries. Heart Lung Circ 2015 May;24(5):510-520 [FREE Full text] [doi: 10.1016/j.hlc.2014.11.013] [Medline: 25534902]
16. Peters AE, Keeley EC. Trends and predictors of participation in cardiac rehabilitation following acute myocardial infarction: data from the behavioral risk factor surveillance system. J Am Heart Assoc 2017 Dec 29;7(1):e007664 [FREE Full text] [doi: 10.1161/JAHA.117.007664] [Medline: 29288154]
17. Li S, Fonarow GC, Mukamal K, Xu H, Matsouaka RA, Devore AD, et al. Sex and racial disparities in cardiac rehabilitation referral at hospital discharge and gaps in long-term mortality. J Am Heart Assoc 2018 Apr 6;7(8):e008088 [FREE Full text] [doi: 10.1161/JAHA.117.008088] [Medline: 29626153]
18. Kotseva K, Wood D, de Bacquer D, EUROASPIRE investigators. Determinants of participation and risk factor control according to attendance in cardiac rehabilitation programmes in coronary patients in Europe: EUROASPIRE IV survey. Eur J Prev Cardiol 2018 Aug;25(12):1242-1251. [doi: 10.1177/2047487318781359] [Medline: 29873511]
19. Resurrección DM, Motrico E, Rigabert A, Rubio-Valera M, Conejo-Cerón S, Pastor L, et al. Barriers for nonparticipation and dropout of women in cardiac rehabilitation programs: a systematic review. J Womens Health (Larchmt) 2017 Aug;26(8):849-859. [doi: 10.1089/jwh.2016.6249] [Medline: 28388314]
20. Moghei M, Turk-Adawi K, Isaranuwatchai W, Sarrafzadegan N, Oh P, Chessex C, et al. Cardiac rehabilitation costs. Int J Cardiol 2017 Oct 1;244:322-328. [doi: 10.1016/j.ijcard.2017.06.030] [Medline: 28622943]
21. Beckman AL, Bucholz EM, Zhang W, Xu X, Dreyer RP, Strait KM, et al. Sex differences in financial barriers and the relationship to recovery after acute myocardial infarction. J Am Heart Assoc 2016 Oct 14;5(10):003923 [FREE Full text] [doi: 10.1161/JAHA.116.003923] [Medline: 27742618]
22. Grace SL, Gravely-Witte S, Kayaniyil S, Brual J, Suskin N, Stewart DE. A multisite examination of sex differences in cardiac rehabilitation barriers by participation status. J Womens Health (Larchmt) 2009 Feb;18(2):209-216 [FREE Full text] [doi: 10.1089/jwh.2007.0753] [Medline: 19183092]
23. Marzolini S, Brooks D, Oh PI. Sex differences in completion of a 12-month cardiac rehabilitation programme: an analysis of 5922 women and men. Eur J Cardiovasc Prev Rehabil 2008 Dec;15(6):698-703. [doi: 10.1097/HJR.0b013e32830c1ce3] [Medline: 18981927]
24. Mochari H, Lee JR, Kligfield P, Mosca L. Ethnic differences in barriers and referral to cardiac rehabilitation among women hospitalized with coronary heart disease. Prev Cardiol 2006;9(1):8-13 [FREE Full text] [doi: 10.1111/j.1520-037x.2005.3703.x] [Medline: 16407697]
25. Sanderson BK, Bittner V. Women in cardiac rehabilitation: outcomes and identifying risk for dropout. Am Heart J 2005 Nov;150(5):1052-1058. [doi: 10.1016/j.ahj.2004.12.027] [Medline: 16290995]
26. Beckie TM, Beckstead JW. Predicting cardiac rehabilitation attendance in a gender-tailored randomized clinical trial. J Cardiopulm Rehabil Prev 2010;30(3):147-156 [FREE Full text] [doi: 10.1097/HCR.0b013e3181d0c2ce] [Medline: 20216324]

27. Pedersen M, Overgaard D, Andersen I, Baastrup M, Egerod I. Mechanisms and drivers of social inequality in phase II cardiac rehabilitation attendance: a convergent mixed methods study. J Adv Nurs 2018 May 17 epub ahead of print. [doi: 10.1111/jan.13715] [Medline: 29772607]
28. Sandesara PB, Lambert CT, Gordon NF, Fletcher GF, Franklin BA, Wenger NK, et al. Cardiac rehabilitation and risk reduction: time to 'rebrand and reinvigorate'. J Am Coll Cardiol 2015 Feb 3;65(4):389-395 [FREE Full text] [doi: 10.1016/j.jacc.2014.10.059] [Medline: 25634839]
29. Lavie CJ, Arena R, Franklin BA. Cardiac rehabilitation and healthy life-style interventions: rectifying program deficiencies to improve patient outcomes. J Am Coll Cardiol 2016 Jan 5;67(1):13-15 [FREE Full text] [doi: 10.1016/j.jacc.2015.09.103] [Medline: 26764060]
30. Antypas K, Wangberg SC. An internet- and mobile-based tailored intervention to enhance maintenance of physical activity after cardiac rehabilitation: short-term results of a randomized controlled trial. J Med Internet Res 2014 Mar 11;16(3):e77 [FREE Full text] [doi: 10.2196/jmir.3132] [Medline: 24618349]
31. Bjarnason-Wehrens B, Bott D, Benesch L, Bischoff KO, Buran-Kilian B, Gysan D, et al. Long-term results of a three-week intensive cardiac out-patient rehabilitation program in motivated patients with low social status. Clin Res Cardiol 2007 Feb;96(2):77-85. [doi: 10.1007/s00392-007-0461-0] [Medline: 17160566]
32. Neubeck L, Redfern JU, Fernandez R, Briffa T, Bauman A, Freedman SB. Telehealth interventions for the secondary prevention of coronary heart disease: a systematic review. Eur J Cardiovasc Prev Rehabil 2009 Jun;16(3):281-289. [doi: 10.1097/HJR.0b013e32832a4e7a] [Medline: 19407659]
33. Beckie TM, Beckstead JW, Schocken DD, Evans ME, Fletcher GF. The effects of a tailored cardiac rehabilitation program on depressive symptoms in women: a randomized clinical trial. Int J Nurs Stud 2011 Jan;48(1):3-12 [FREE Full text] [doi: 10.1016/j.ijnurstu.2010.06.005] [Medline: 20615504]
34. Beckie TM, Beckstead JW. The effects of a cardiac rehabilitation program tailored for women on their perceptions of health: a randomized clinical trial. J Cardiopulm Rehabil Prev 2011;31(1):25-34 [FREE Full text] [doi: 10.1097/HCR.0b013e3181f68acc] [Medline: 21037482]
35. Beckie TM, Beckstead JW. The effects of a cardiac rehabilitation program tailored for women on global quality of life: a randomized clinical trial. J Womens Health (Larchmt) 2010 Nov;19(11):1977-1985 [FREE Full text] [doi: 10.1089/jwh.2010.1937] [Medline: 20846018]
36. Spring B, Moller AC, Colangelo LA, Siddique J, Roehrig M, Daviglus ML, et al. Healthy lifestyle change and subclinical atherosclerosis in young adults: coronary artery risk development in young adults (CARDIA) study. Circulation 2014 Jul 1;130(1):10-17 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.113.005445] [Medline: 24982115]
37. Arnett D, Blumenthal R, Albert M, Buroker AB, Goldberger ZD, Hahn EJ, et al. 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation 2019 Sep 10;140(11):e596-e646. [doi: 10.1161/CIR.0000000000000678] [Medline: 30879355]
38. Buckingham SA, Taylor RS, Jolly K, Zawada A, Dean SG, Cowie A, et al. Home-based versus centre-based cardiac rehabilitation: abridged Cochrane systematic review and meta-analysis. Open Heart 2016 Sep 14;3(2):e000463 [FREE Full text] [doi: 10.1136/openhrt-2016-000463] [Medline: 27738516]
39. Anderson L, Sharp G, Norton R, Dalal H, Dean SG, Jolly K, et al. Home-based versus centre-based cardiac rehabilitation. Cochrane Database Syst Rev 2017 Jun 30;6:CD007130 [FREE Full text] [doi: 10.1002/14651858.CD007130.pub4] [Medline: 28665511]
40. Tang LH, Berg SK, Christensen J, Lawaetz J, Doherty P, Taylor RS, et al. Patients' preference for exercise setting and its influence on the health benefits gained from exercise-based cardiac rehabilitation. Int J Cardiol 2017 Apr 1;232:33-39. [doi: 10.1016/j.ijcard.2017.01.126] [Medline: 28159358]
41. Beatty AL, Truong M, Schopfer DW, Shen H, Bachmann JM, Whooley MA. Geographic variation in cardiac rehabilitation participation in medicare and veterans affairs populations: opportunity for improvement. Circulation 2018 May 1;137(18):1899-1908 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.117.029471] [Medline: 29305529]
42. Schopfer DW, Krishnamurthi N, Shen H, Duvernoy CS, Forman DE, Whooley MA. Association of veterans health administration home-based programs with access to and participation in cardiac rehabilitation. JAMA Intern Med 2018 May 1;178(5):715-717 [FREE Full text] [doi: 10.1001/jamainternmed.2017.8039] [Medline: 29356813]
43. Thomas RJ, Beatty AL, Beckie TM, Brewer LC, Brown TM, Forman DE, et al. Home-based cardiac rehabilitation: a scientific statement from the American Association of Cardiovascular and Pulmonary Rehabilitation, the American Heart Association, and the American College of Cardiology. J Am Coll Cardiol 2019 Jul 9;74(1):133-153. [doi: 10.1016/j.jacc.2019.03.008] [Medline: 31097258]
44. Kaminsky LA, Brubaker PH, Guazzi M, Lavie CJ, Montoye AH, Sanderson BK, et al. Assessing physical activity as a core component in cardiac rehabilitation: a position statement of the American Association of Cardiovascular and Pulmonary Rehabilitation. J Cardiopulm Rehabil Prev 2016;36(4):217-229. [doi: 10.1097/HCR.0000000000000191] [Medline: 27307067]
45. Asch DA, Muller RW, Volpp KG. Automated hovering in health care-watching over the 5000 hours. N Engl J Med 2012 Jul 5;367(1):1-3. [doi: 10.1056/NEJMp1203869] [Medline: 22716935]

46. 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]
47. Hutchesson MJ, Rollo ME, Krukowski R, Ells L, Harvey J, Morgan PJ, et al. eHealth interventions for the prevention and treatment of overweight and obesity in adults: a systematic review with meta-analysis. Obes Rev 2015 May;16(5):376-392. [doi: 10.1111/obr.12268] [Medline: 25753009]
48. Lewis JR, Sauro J. The Factor Structure of the System Usability Scale. In: Proceedings of the International Conference on Human Centered Design. 2009 Presented at: HCD'09; July 19-24, 2009; San Diego, CA, USA p. 94-103. [doi: 10.1007/978-3-642-02806-9_12]
49. Sallis JF, Pinski RB, Grossman RM, Patterson TL, Nader PR. The development of self-efficacy scales for healthrelated diet and exercise behaviors. Health Educ Res 1988;3(3):283-292. [doi: 10.1093/her/3.3.283]
50. Decker JW, Dennis KE. The eating habits confidence survey: reliability and validity in overweight and obese postmenopausal women. J Nurs Meas 2013;21(1):110-119. [doi: 10.1891/1061-3749.21.1.110] [Medline: 23786138]
51. Lorig KR, Ritter P, Stewart AL, Sobel DS, Brown BW, Bandura A, et al. Chronic disease self-management program: 2-year health status and health care utilization outcomes. Med Care 2001 Nov;39(11):1217-1223. [doi: 10.1097/00005650-200111000-00008] [Medline: 11606875]
52. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385-396. [doi: 10.2307/2136404] [Medline: 6668417]
53. 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]
54. Fann JR, Bombardier CH, Dikmen S, Esselman P, Warms CA, Pelzer E, et al. Validity of the patient health questionnaire-9 in assessing depression following traumatic brain injury. J Head Trauma Rehabil 2005;20(6):501-511. [doi: 10.1097/00001199-200511000-00003] [Medline: 16304487]
55. Adewuya AO, Ola BA, Afolabi OO. Validity of the patient health questionnaire (PHQ-9) as a screening tool for depression amongst Nigerian university students. J Affect Disord 2006 Nov;96(1-2):89-93. [doi: 10.1016/j.jad.2006.05.021] [Medline: 16857265]
56. Stafford L, Berk M, Jackson HJ. Validity of the hospital anxiety and depression scale and patient health questionnaire-9 to screen for depression in patients with coronary artery disease. Gen Hosp Psychiatry 2007;29(5):417-424. [doi: 10.1016/j.genhosppsych.2007.06.005] [Medline: 17888808]
57. Segal-Isaacson CJ, Wylie-Rosett J, Gans KM. Validation of a short dietary assessment questionnaire: the Rapid Eating and Activity Assessment for Participants short version (REAP-S). Diabetes Educ 2004;30(5):774, 776, 778 passim. [doi: 10.1177/014572170403000512] [Medline: 15510530]
58. Johnston CS, Bliss C, Knurick JR, Scholtz C. Rapid eating assessment for participants [shortened version] scores are associated with healthy eating index-2010 scores and other indices of diet quality in healthy adult omnivores and vegetarians. Nutr J 2018 Sep 28;17(1):89 [FREE Full text] [doi: 10.1186/s12937-018-0399-x] [Medline: 30266095]
59. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc 2003 Aug;35(8):1381-1395. [doi: 10.1249/01.MSS.0000078924.61453.FB] [Medline: 12900694]
60. Dinger MK, Behrens TK, Han JL. Validity and reliability of the international physical activity questionnaire in college students. Am J Health Educ 2006;37(6):337-343. [doi: 10.1080/19325037.2006.10598924]
61. Wang C, Chen P, Zhuang J. Validity and reliability of international physical activity questionnaire-short form in Chinese youth. Res Q Exerc Sport 2013 Dec;84(Suppl 2):S80-S86. [doi: 10.1080/02701367.2013.850991] [Medline: 24527570]
62. Sanda B, Vistad I, Haakstad LA, Berntsen S, Sagedal LR, Lohne-Seiler H, et al. Reliability and concurrent validity of the international physical activity questionnaire short form among pregnant women. BMC Sports Sci Med Rehabil 2017;9:7 [FREE Full text] [doi: 10.1186/s13102-017-0070-4] [Medline: 28316789]
63. Duncan MJ, Arbour-Nicitopoulos K, Subramanieapillai M, Remington G, Faulkner G. Revisiting the international physical activity questionnaire (IPAQ): assessing physical activity among individuals with schizophrenia. Schizophr Res 2017 Jan;179:2-7. [doi: 10.1016/j.schres.2016.09.010] [Medline: 27623360]
64. American College of Sports Medicine. ACSM's Guidelines For Exercise Testing And Prescription. Eighth Edition. New York: lippincott Williams & Wilkins; 2010.
65. Cornier M, Després JP, Davis N, Grossniklaus DA, Klein S, Lamarche B, American Heart Association Obesity Committee of the Council on Nutrition, Physical Activity and Metabolism, Council on Arteriosclerosis, Thrombosis and Vascular Biology, Council on Cardiovascular Disease in the Young, Council on Cardiovascular Radiology and Intervention, Council on Cardiovascular Nursing, Council on Epidemiology and Prevention, Council on the Kidney in Cardiovascular Disease, Stroke Council. Assessing adiposity: a scientific statement from the American Heart Association. Circulation 2011 Nov 1;124(18):1996-2019. [doi: 10.1161/CIR.0b013e318233bc6a] [Medline: 21947291]
66. Pickering TG, Hall JE, Appel LJ, Falkner BE, Graves J, Hill MN, et al. Recommendations for blood pressure measurement in humans and experimental animals: part 1: blood pressure measurement in humans: a statement for professionals from

the Subcommittee of Professional and Public Education of the American Heart Association Council on High Blood Pressure Research. Circulation 2005 Feb 8;111(5):697-716. [doi: 10.1161/01.CIR.0000154900.76284.F6] [Medline: 15699287]
67. Baldwin JL, Singh H, Sittig DF, Giardina TD. Patient portals and health apps: pitfalls, promises, and what one might learn from the other. Healthc (Amst) 2017 Sep;5(3):81-85 [FREE Full text] [doi: 10.1016/j.hjdsi.2016.08.004] [Medline: 27720139]
68. Forman DE, LaFond K, Panch T, Allsup K, Manning K, Sattelmair J. Utility and efficacy of a smartphone application to enhance the learning and behavior goals of traditional cardiac rehabilitation: a feasibility study. J Cardiopulm Rehabil Prev 2014;34(5):327-334. [doi: 10.1097/HCR.0000000000000058] [Medline: 24866355]
69. Johnston N, Bodegard J, Jerström S, Åkesson J, Brorsson H, Alfredsson J, et al. Effects of interactive patient smartphone support app on drug adherence and lifestyle changes in myocardial infarction patients: a randomized study. Am Heart J 2016 Aug;178:85-94 [FREE Full text] [doi: 10.1016/j.ahj.2016.05.005] [Medline: 27502855]
70. Widmer RJ, Allison TG, Lerman LO, Lerman A. Digital health intervention as an adjunct to cardiac rehabilitation reduces cardiovascular risk factors and rehospitalizations. J Cardiovasc Transl Res 2015 Jul;8(5):283-292 [FREE Full text] [doi: 10.1007/s12265-015-9629-1] [Medline: 25946990]
71. Hägglund E, Lyngå P, Frie F, Ullman B, Persson H, Melin M, et al. Patient-centred home-based management of heart failure. Findings from a randomised clinical trial evaluating a tablet computer for self-care, quality of life and effects on knowledge. Scand Cardiovasc J 2015 Aug;49(4):193-199. [doi: 10.3109/14017431.2015.1035319] [Medline: 25968968]
72. Bengtsson U, Kjellgren K, Hallberg I, Lindwall M, Taft C. Improved blood pressure control using an interactive mobile phone support system. J Clin Hypertens (Greenwich) 2016 Feb;18(2):101-108 [FREE Full text] [doi: 10.1111/jch.12682] [Medline: 26456490]
73. Layton AM, Whitworth J, Peacock J, Bartels MN, Jellen PA, Thomashow BM. Feasibility and acceptability of utilizing a smartphone based application to monitor outpatient discharge instruction compliance in cardiac disease patients around discharge from hospitalization. Int J Telemed Appl 2014;2014:415868 [FREE Full text] [doi: 10.1155/2014/415868] [Medline: 25574165]
74. 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]
75. Kim S, So J. How message fatigue toward health messages leads to ineffective persuasive outcomes: examining the mediating roles of reactance and inattention. J Health Commun 2018;23(1):109-116. [doi: 10.1080/10810730.2017.1414900] [Medline: 29272208]
76. Davidson TL, Jones S, Roy M, Stevenson RJ. The cognitive control of eating and body weight: it's more than what you 'think'. Front Psychol 2019;10:62 [FREE Full text] [doi: 10.3389/fpsyg.2019.00062] [Medline: 30814963]
77. Jones A, Robinson E, Duckworth J, Kersbergen I, Clarke N, Field M. The effects of exposure to appetitive cues on inhibitory control: a meta-analytic investigation. Appetite 2018 Sep 1;128:271-282 [FREE Full text] [doi: 10.1016/j.appet.2018.06.024] [Medline: 29935289]
78. Appelhans B. Neurobehavioral inhibition of reward-driven feeding: implications for dieting and obesity. Obesity (Silver Spring) 2009 Apr;17(4):640-647 [FREE Full text] [doi: 10.1038/oby.2008.638] [Medline: 19165160]
79. Torjesen I. Sixty seconds on . . . exercise. Br Med J 2018 Jul 10;362:k3006. [doi: 10.1136/bmj.k3006] [Medline: 29991492]
80. Althoff T, Sosič R, Hicks JL, King AC, Delp SL, Leskovec J. Large-scale physical activity data reveal worldwide activity inequality. Nature 2017 Jul 20;547(7663):336-339 [FREE Full text] [doi: 10.1038/nature23018] [Medline: 28693034]
81. Bassett Jr DR, Toth LP, LaMunion SR, Crouter SE. Step counting: a review of measurement considerations and health-related applications. Sports Med 2017 Jul;47(7):1303-1315 [FREE Full text] [doi: 10.1007/s40279-016-0663-1] [Medline: 28005190]
82. Kraus WE, Yates T, Tuomilehto J, Sun J, Thomas L, McMurray JJ, et al. Relationship between baseline physical activity assessed by pedometer count and new-onset diabetes in the NAVIGATOR trial. BMJ Open Diabetes Res Care 2018;6(1):e000523 [FREE Full text] [doi: 10.1136/bmjdrc-2018-000523] [Medline: 30073088]
83. Dwyer T, Pezic A, Sun C, Cochrane J, Venn A, Srikanth V, et al. Objectively measured daily steps and subsequent long term all-cause mortality: the tasped prospective cohort study. PLoS One 2015;10(11):e0141274 [FREE Full text] [doi: 10.1371/journal.pone.0141274] [Medline: 26536618]
84. Lee I, Shiroma EJ, Kamada M, Bassett DR, Matthews CE, Buring JE. Association of step volume and intensity with all-cause mortality in older women. JAMA Intern Med 2019 May 29 epub ahead of print. [doi: 10.1001/jamainternmed.2019.0899] [Medline: 31141585]

Abbreviations

BCT: behavior change technique
BP: blood pressure
CBCR: center-based cardiac rehabilitation
CHD: coronary heart disease
CV: cardiovascular
EMA: ecological momentary assessment

GUI: graphic user interface
HBCR: home-based cardiac rehabilitation
ICC: intraclass correlation coefficient
IPAQ-SF: International Physical Activity Questionnaire-Short Form
mHealth: mobile health
PA: physical activity
PHQ: Patient Health Questionnaire
PI: principal investigator
PSS: Perceived Stress Scale
REAP-S: Rapid Eating Assessment for Participants-Short Form
SE: self-efficacy
SQL: Structured Query Language
SUS: System Usability Scale

Edited by G Eysenbach; submitted 26.09.19; peer-reviewed by J King, J Whiteley; comments to author 19.11.19; revised version received 05.03.20; accepted 11.03.20; published 03.06.20.

Please cite as:
Sengupta A, Beckie T, Dutta K, Dey A, Chellappan S
A Mobile Health Intervention System for Women With Coronary Heart Disease: Usability Study
JMIR Form Res 2020;4(6):e16420
URL: https://formative.jmir.org/2020/6/e16420
doi:10.2196/16420
PMID:32348270

Original Paper

Patients' User Experience of a Blended Face-to-Face and Web-Based Smoking Cessation Treatment: Qualitative Study

Lutz Siemer[1,2]; Somaya Ben Allouch[1,3], PhD; Marcel E Pieterse[2], PhD; Marjolein Brusse-Keizer[4], PhD; Robbert Sanderman[2,5], PhD; Marloes G Postel[2,6], PhD

[1]Technology, Health & Care Research Group, Saxion University of Applied Sciences, Enschede, Netherlands

[2]Centre for eHealth and Well-Being Research, University of Twente, Enschede, Netherlands

[3]Digital Life Research Group, Amsterdam University of Applied Science, Amsterdam, Netherlands

[4]Medical School Twente, Medisch Spectrum Twente, Enschede, Netherlands

[5]Department of Health Psychology, University Medical Center Groningen, University of Groningen, Groningen, Netherlands

[6]Tactus, Enschede, Netherlands

Corresponding Author:
Lutz Siemer
Technology, Health & Care Research Group
Saxion University of Applied Sciences
Postbus 70000
Enschede, 7500 KB
Netherlands
Phone: 49 17678025906
Email: l.siemer@utwente.nl

Abstract

Background: Blended web-based and face-to-face (F2F) treatment is a promising electronic health service because the strengths of one mode of delivery should compensate for the weaknesses of the other.

Objective: The aim of this study was to explore this compensation by examining patients' user experience (UX) in a blended smoking cessation treatment (BSCT) in routine care.

Methods: Data on patients' UX were collected through in-depth interviews (n=10) at an outpatient smoking cessation clinic in the Netherlands. A content analysis of the semantic domains was used to analyze patients' UX. To describe the UX, the Hassenzahl UX model was applied, examining 4 of the 5 key elements of UX from a user's perspective: (1) patients' standards and expectations, (2) apparent character (pragmatic and hedonic attributes), (3) usage situation, and (4) consequences (appeal, emotions, and behavior).

Results: BSCT appeared to be a mostly positively experienced service. Patients had a positive-pragmatic standard and neutral-open expectation toward BSCT at the treatment start. The pragmatic attributes of the F2F sessions were mostly perceived as positive, whereas the pragmatic attributes of the web sessions were perceived as both positive and negative. For the hedonic attributes, there seemed to be a difference between the F2F and web sessions. Specifically, the hedonic attributes of the web sessions were experienced as mostly negative, whereas those of the F2F sessions were experienced as mostly positive. For the usage situation, the physical and social contexts were experienced positively, whereas the task and technical contexts were experienced negatively. Nevertheless, the consequential appeal of BSCT was positive. However, the consequential emotions and behavior varied, ultimately resulting in diverse combinations of consequential appeal, emotions, and behavior (positive, negative, and mixed).

Conclusions: This study provided insights into the UX of a blended treatment, and the results support the expectation that in a blended treatment, the strengths of one mode of delivery may compensate for the weaknesses of the other. However, in this certain setting, this is mainly achieved in only one way: F2F sessions compensated for the weaknesses of the web sessions. As a practical conclusion, this may mean that the web sessions, supported by the strengths of the F2F sessions, offer an interesting approach for further improving the blended treatment. Our theoretical findings reflect the relevance of the aspects of hedonism, such as fun, joy, or happiness in the UX, which were not mentioned in relation to the web sessions and were only scarcely mentioned in relation to the F2F sessions. Future research should further investigate the role of hedonistic aspects in a blended treatment and whether increased enjoyment of a blended treatment could increase treatment adherence and, ultimately, effectiveness.

(JMIR Form Res 2020;4(6):e14550) doi:10.2196/14550

KEYWORDS

smoking cessation; cognitive therapy; blended treatment; smoking; user experience; tobacco; patient perspective

Introduction

Blended Treatment

Health care is undergoing a sea change driven by the progress in digital technology [1]. One of the interesting innovations is blended treatment—a combination of the Web-based and face-to-face (F2F) therapy [2,3]. Blended treatment is a promising electronic health (eHealth) service because it is expected that the strengths of one mode of delivery will compensate for the weaknesses of the other [3-9]. For example, it is the strength of F2F treatment to be able to provide the personal attention of a professional that could compensate for the lack of F2F contact in Web-based treatment. In turn, one of the unique features of Web-based care is the accessibility, anytime and anywhere, which could compensate for the time in between F2F sessions when patients need support. Until now, there has been no final definition for blended treatment [3,6], and it is offered in various formats. The literature on blended treatment mentions different modes of delivery (eg, mainly Web-based [10,11], mainly F2F [12,13], 50-50 blend of Web-based and F2F [14]), different orders of F2F- and Web-based treatment (eg, sequential [10] or integrated [8,15]), and different tools for its use (such as platforms, emails, short message service, text messaging, and apps [5,16]). The intervention in this study is an integrated 50-50 blend of F2F treatment and treatment via a Web platform.

User Experience and Blended Treatment

One of the main elements clarifying the individual's use of services in general [17] and eHealth services, such as blended treatment, in particular [18], is the user experience (UX). UX refers to what people personally encounter, undergo, or live through while using, interacting with, or being confronted passively with systems [19]. Systems can denote products, services, and artifacts—separately or combined in one form or another—that a person can interact with [20].

Usually, the term UX refers to products, services, and objects that a person interacts with through a user interface [21]. However, for this study, we widened the scope of this term to explore the UX of a service (ie, blended treatment) that alternately uses computer-mediated communication *via* a user interface and F2F communication in counselling sessions.

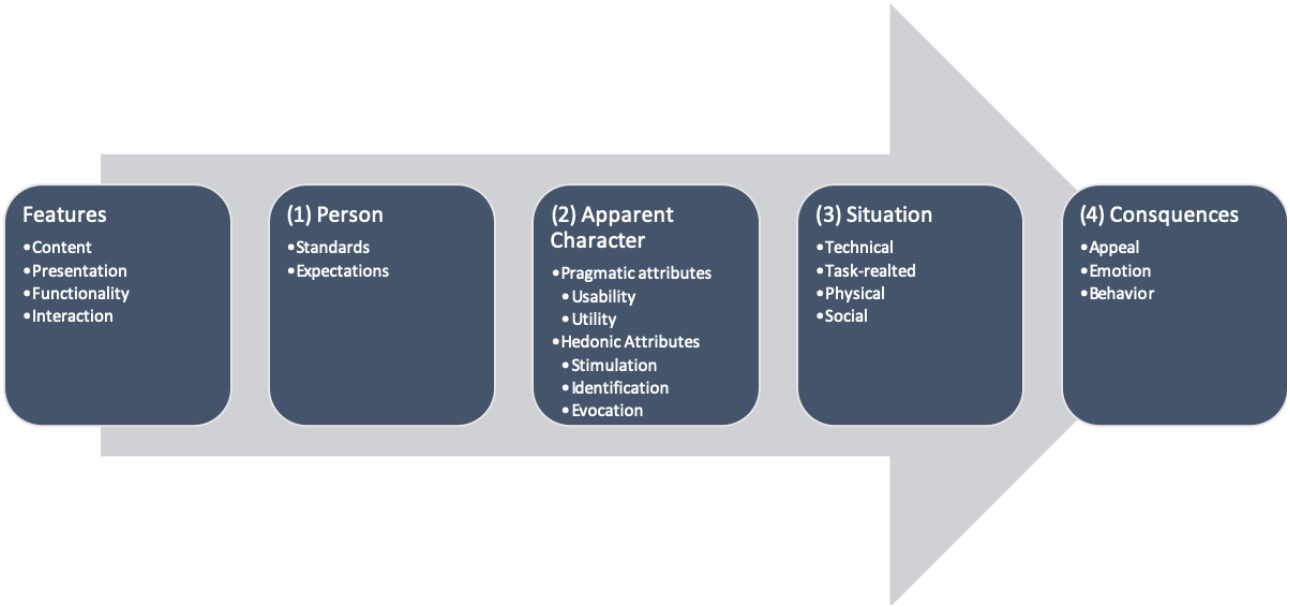
Although a number of studies have examined the blended treatment [15], little is known about the patients' UX specifically with blended treatments. An evaluation study (n=7) of a blended cognitive behavioral treatment for major depression [14] showed that while the patients' pretreatment expectations were mainly neutral and some skeptical patients found it hard to start with the Web-based sessions, most patients appeared to have positive attitudes toward the blended treatment afterward. Another study [22] (n=14) on internet-based cognitive behavioral therapy for depression supported by short F2F consultations found that a sense of relatedness in terms of feeling connected to the therapist and being able to identify with the Web-based treatment may increase patients' adherence to the blended treatment. Both the studies suggest that the elements of patients' UX, such as expectations, usability, and identification, play a role in adherence to a blended treatment and should further be explored.

Patients' User Experience

For the patients' perspective on the blended care treatment, Hassenzahl's model of UX from a user's perspective was adapted [21,23-25]. This process-oriented constructivist model defines five key elements and their functional relations (Figure 1). Basically, the model states that while getting in contact with the *features* of a product or service, a process is triggered, in which the user constructs the UX (this is illustrated by the grey arrow in Figure 1). In the beginning, the user constructs—moderated by the person's *standards and expectations*—an *apparent character* of the product or service. Moderated by the specific usage *situation*, the apparent character will then finally mediate a number of *consequences.*

The features of the service (in this case the blended treatment) are selected and combined by the treatment developers independently of the patients that ultimately follow the treatment. Since the features are not constructed by the users, the product features only play a minor role in this study. In turn, the focus is placed on the patients' response to the treatment's features to explore the UX from the user's perspective in a narrower sense. This means that the UX from a user's perspective is built based on only four of these five key elements: (1) the patient's standards and expectations, (2) the apparent character, (3) the usage situation, and (4) the consequences. In the following paragraphs, each key element is described and illustrated by examples of how it applies to the blended treatment in this study.

Figure 1. Key elements of the user experience model.



Features

The *features* of a product or service refer to its *content*, *presentation*, *functionality*, and *interaction* [23,24]. The content of the treatment of this study—Blended Smoking Cessation Treatment (BSCT)—refers, for example, to the behavioral change techniques for smoking cessation [26] that comprises BSCT. The presentation refers to the clinical surrounding as BSCT is part of the routine care setting of a hospital. Functionality and interaction refer to the F2F and Web-based sessions, which offer synchronous interactions with the counselor (eg, functions, such as providing feedback on behavior and building rapport) and asynchronous counselor-independent interactions with the Web-based system (eg, functions, such as self-recording of smoking behavior *via* a Web-based smoking diary). More details about the study intervention are provided below in the Methods section.

Person

The patients' *standards* and *expectations* are based on their experiences with the other services [23,24] with which the patient can compare BSCT. If a patient compares BSCT with, for example, earlier experiences in health care, smoking cessation support, F2F treatment, or use of computers and internet, the patient may start BSCT with a subjective standard, such as "using the computer for treatment is too difficult for me," or with an expectation, such as "blended treatment will be more comfortable because I can partly do treatment at home."

Apparent Character

When confronted with a service, an apparent character is constructed by the user. The apparent character is a cognitive structure representing *pragmatic* and *hedonic* attributes [23,24]. Pragmatic attributes refer to the *utility* (eg, "supporting," "useful") and *usability* (eg, "clear" and "easy to use") of a service, such as BSCT. Hedonic attributes of BSCT refer to *stimulation* (eg, "novel and interesting" and "makes me think"), *identification* (eg, "my style"), and *evocation* (eg, "reminds me of filling in tax forms").

Situation

The usage situation moderates the consequences of the apparent character [23,24] and refers to the *technical*, *task-related*, *physical*, and *social* contexts. These situations are different between patients and over the course of the treatment, especially for the Web-based sessions. For example, filling in a smoking diary while being on your own in a silent surrounding may result in different consequences than doing this in the living room with a partner and children around you.

Consequences

The fit of the apparent character and the usage situation leads to three consequences: *appeal*, *emotions*, and *behavior* [23,24]. For patients, BSCT, for example, may appeal as "fine" while feeling "satisfied" and "adhering to the treatment."

Aims of This Study

As UX has been shown as an important factor in explaining the behavior of a user in general [27], and patients' use of health care services in particular [28], the aim of this study is—from a UX point of view – to explore whether in blended treatment the strength of one mode of delivery may compensate for the weaknesses of the other. By applying Hassenzahl's model of UX to qualitatively describe the patients' UX of BSCT in routine care, the question what positive and negative experiences patients have with BSCT in general and with the F2F sessions and the Web sessions in particular will be addressed. This research will contribute to a deeper understanding of the facilitators and barriers to blended treatment, which will provide new insights for both scientific research on blended treatment and its improvement in clinical practice. It is expected that the application of the findings on UX elements in furthering the development of blended treatment will lead to better treatment outcomes.

Methods

Study Intervention

BSCT is a clinician-led intervention [1] which combines F2F and Web-based treatment delivered in routine care settings at the Outpatient Smoking Cessation Clinic (Stoppen met Roken Poli [SRP]) of the Department of Pulmonary Medicine at Medisch Spectrum Twente Hospital in Enschede, The Netherlands. BSCT is derived from the Dutch Guideline Tobacco Addiction [29], fulfilling the requirements of the Dutch care module for smoking cessation [30]. The treatment is based on both the F2F treatment as usual at SRP [31,32] and Web-based treatment at Tactus Addiction Treatment (www.rokendebaas.nl). A team of clinical experts from both the organizations developed BSCT, striving for a 50-50 mix with constant alternating of F2F and Web-based treatments by replacing five of the usual ten F2F sessions with appropriate Web-based sessions. This treatment design decision was made based on the randomized controlled trial (LiveSmokefree study [8]), which compared the effectiveness of BSCT with F2F treatment. The order, planning, mode of delivery, and main content of the BSCT sessions is shown in Table 1. The details of BSCT have been described in earlier papers [8,15]). To provide an impression of the look and feel of the Web interventions, Multimedia Appendix 1 shows screenshots of the Web sessions of BSCT.

Table 1. Order, planning, mode of delivery, and main content of the blended smoking cessation treatment sessions.

Session	Week	Mode of delivery	Content
1	1	Face-to-face	Goal setting
2	3	Web-based	Measures for self-control
3	5	Face-to-face	Dealing with withdrawal
4	7	Web-based	Breaking habits
5	9	Face-to-face	Dealing with triggers
6	11	Web-based	Food for thought
7	14	Face-to-face	Think differently
8	18	Web-based	Do differently
9	22	Face-to-face	Action plan
10	26	Web-based	Closure

Setting and Participants

The current study is a substudy of the LiveSmokefree study—a single-center randomized controlled noninferiority trial with parallel group design, which examines the effectiveness of BSCT as compared with F2F treatment. The inclusion criteria for the LiveSmokefree study were (1) aged 18 years or older, (2) willing to quit smoking, (3) current daily smoker (at least one cigarette a day), and (4) speaking/reading/writing Dutch.

A purposive sample (n=10) of the participants from the blended arm of the LiveSmokefree study [8] that had already ended the treatment was selected, striving for a heterogeneous mix of patients regarding the characteristics (Table 2) that were expected to influence the patients' UX (ie, age, sex, educational level, adherence, counselling, and quitting success). For recruitment, the patients were called and invited by the research assistants to participate in the UX study. The participation was voluntary; patients had to sign an informed consent form and received no incentives.

Table 2. Purposive sample.

Characteristics	Randomization number									
	10	12	14	25	27	34	53	75	106	509
Age (years)	77	54	68	71	37	45	60	65	37	58
Sex (m: male; f: female)	m	m	m	f	f	m	m	f	m	m
Education level	Low[a]	Mid/high[b]	Mid/high	Low	Low	Mid/high	Low	Mid/high	Low	Mid/high
Internet skills[c]	28	34	37	38	38	46	36	40	40	39
Nicotine dependence[d]	5	7	4	6	7	4	3	2	7	6
#Adherence F2F[e]	3	4	6	5	2	3	5	2	8	2
#Adherence Web[f]	3	2	8	2	0	3	6	9	7	2
#Adherence BSCT[g]	6	6	14	7	2	6	11	11	15	4
Adherence F2F[h]	N	N	Y	Y	N	N	Y	N	Y	N
Adherence Web[i]	N	N	Y	N	N	N	Y	Y	Y	N
Adherence BSCT[j]	N	N	Y	N	N	N	Y	N	Y	N
Counselor[k]	A	B	B	B	B	A	A	C	C	B
Stopped smoking[l]	Yes	No	Yes	No	No	No	Yes	No	Yes	No

[a]Low: lower than vocational education and training.

[b]Mid/high: vocational education and training or higher.

[c]Internet skills: range 10-60; higher number indicates better skills.

[d]Nicotine dependence: Fagerström, range 0-10; higher numbers indicate higher nicotine dependency.

[e]#Adherence F2F: adherence to face-to-face (F2F) sessions, range 0-8, based on the 8 activities belonging to F2F sessions; higher number indicates higher adherence.

[f]#Adherence Web: adherence to Web sessions, range 0-10, based on the 10 activities belonging to Web sessions; higher number indicates higher adherence.

[g]Adherence BSCT: adherence to blended smoking cessation treatment (BSCT) in general, sum of #Adherence F2F and #Adherence Web, range 0-18; higher number indicates higher adherence.

[h]Adherence F2F: categorical classification of adherence to the F2F sessions based on a 60% threshold (Y= adherent; N=nonadherent).

[i]Adherence Web: categorical classification of adherence to the Web sessions based on a 60% threshold (Y= adherent; N=nonadherent).

[j]Adherence BSCT: categorical classification of adherence to BSCT in general based on a 60% threshold (Y= adherent; N=nonadherent).

[k]Counselor: who carried out the treatment.

[l]Stopped smoking: self-reported abstinence.

Ethics

Both the LiveSmokefree study and this substudy on patients' UX were approved by the accredited Medical Research Ethics Committee Twente (P14-37/NL50944.044.14). The LiveSmokefree study was registered in the Dutch Trial Registration (NTR5113).

Data Collection

Qualitative data about the patients' UX was collected by in-depth semi-structured interviews. The interview guide (Multimedia Appendix 2) was developed following the key elements of the UX [23,24] to elicit both the patients' standards and expectations toward BSCT, the apparent character of BSCT (usability, utility, stimulation, identification, and evocation), the usage situation (technical, tasks, physical, and social), and the consequences (appeal, emotions, and behavior). Additional interview questions were created from a clinical perspective addressing practicalities (eg, intake procedure, treatment procedure, and adherence) and ideas for the improvement of current BSCT.

The interviews were conducted by the first author (LS) between October 2016 and March 2017. Because LS is not a Dutch native speaker, he was supported by trained Dutch research assistants to avoid possible ambiguities and linguistic misunderstandings. On the date of the interview, the interviewees were picked up from the waiting area of the SRP and led to a neutral meeting room. After receiving permission for audio recording, the interviewer read a written introduction, which emphasized that the patient was invited to recall and describe ("tell stories") their UX. After this briefing, a general stimulus ("Can you, first of all, tell us what your experiences with the blended treatment are? We would like to hear all the events and experiences that were important to you.") was used to start. Interviews followed a detailed written interview guide (Multimedia Appendix 2), but were open-ended in nature, allowing the interviewers to ask

probing questions and to follow up on interesting topics and experiences related to BSCT.

The audio-recordings were transcribed verbatim by trained research assistants following the guidelines for data preparation and transcription, as described by McLellan et al [33], and were subsequently analyzed using the qualitative data analysis software ATLAS.ti Version 8.3.1 (ATLAS.ti Scientific Software Development GmbH).

Auxiliary Data

The data regarding the patients' age, sex, education level, internet skills, nicotine dependence, and counselor (Table 2) were acquired from the LiveSmokefree study database, for which the data were collected using a Web-based questionnaire that the patients completed at the beginning of the treatment. A detailed description of the variables and their measurements can be found in the protocol article of the LiveSmokefree study [8]. The patients' characteristics were reported as medians with IQRs or as numbers using SPSS version 24.

The data about adherence and smoking status (Table 2) were acquired from a dataset build in 2018 for a paper on adherence to BSCT [15]. Based on 18 patient activities that reflect the course of the treatment (eg, attending a F2F session or completing a Web-based task), an adherence score ranging from 0 (nonadherent to any activity after the first treatment session) to 18 (adherent to all activities) was calculated for each patient. The patients' adherence rates were reported as medians with IQR using SPSS version 24. Based on a 60% threshold for both the F2F sessions and the Web sessions [15], the patients were classified as adherent or nonadherent to BSCT. Questions regarding adherence were also asked in the interviews (see above), which might have led to different assessments (eg, patient #25). To examine the self-reported smoking status (stopped smoking: Yes/No), data from both the in-depth interviews and the follow-up Web-based-questionnaires of the LiveSmokefree study 6-month after the treatment start were used. In case the interview and questionnaire data contradicted each other, the interview data were considered superior.

Codebook Development

Based on the semi-structured interview guide, content analysis was used to analyze all the interviews. The codebook was developed by two research team members (LS, SA), building on the interview guide and the research goals related to the clinical setting (eg, ideas for improvement of BSCT) [34]). The codes were grouped in semantic domains and intercoder agreement was analyzed per semantic domain using the intercoder analysis feature of Atlas.ti 8.2.4. The disagreements were discussed, and the codebook was revised until acceptable agreement (Krippendorff c-α-binary 0.650-0.928) for each semantic domain was achieved. The codes, their description, and the intercoder agreement per semantic domain are displayed in Multimedia Appendix 3.

Paraphrasing and Regrouping

After coding, all coded Dutch quotes were paraphrased in English by LS and collected in a table (Multimedia Appendix 4). Applying Hassenzahl's model of UX from a user's perspective [23,24], the semantic domains of the codes were revised by linking the codes to the four of the five key elements, which form the UX from a user's perspective: (1) patients' standards and expectations; (2) apparent character (pragmatic attributes: usability, utility; hedonic attributes: stimulation, identification, and evocation); (3) usage situation (physical, social, technical, task); and (4) consequences (appeal, emotion, behavior). Finally, the UX was described for each key element distinguishing as far as possible between BSCT in general (ie, the experience of BSCT as a whole) from the two modes of delivery (ie, the F2F sessions and the Web sessions). Furthermore, in describing the UX, an attempt was made to make a distinction between the positive and negative UX, which is based on the idea that UX is a "primarily evaluative feeling (good/bad) while interacting with a product or service" [25]. Ultimately, we summarized the variety of consequences in three kinds of combinations of consequential appeals, emotions, and behavior.

Results

Overview

In the following, the patient characteristics are presented first. Then, the positive and negative statements for each key element are described. As far as possible, this is done first for BSCT in general and then for the F2F and Web sessions. It is to be noted that the analysis and presentation methods were clarified after the interview phase, and the statements were not always available in every area.

Participants

Patients' characteristics are shown in Table 2. The median age of the patients was 59.0 years (IQR 43.0-68.8), and the majority were males (7/10). Half (5/10) of the patients' educational level was lower than vocational education and training. The median internet skill level (range 10-50, higher numbers indicate higher skills [8]) was 38.0 (IQR 35.5-40.0), and the median nicotine dependence (Fagerström range: 0-10, higher numbers indicate higher dependency [35]) was 5.5 (IQR 3.8-7.0).

Patients' Standards and Expectations

In general, the patients approached BSCT mostly with a positive-pragmatic standard and a neutral-open expectation. None of the patients had followed a blended treatment or a Web-based treatment before. Therefore, their standards and expectations were based mainly on earlier experiences with F2F sessions, with earlier stop smoking attempts, and with ICT use in general. Only one patient (#34) used health-apps (Mindfulness, Stoptober). However, most of the patients (7/10) had received F2F counseling before, participated in a group therapy (#34), or were familiar with mindfulness (#34, #53).

For *F2F sessions*, positive standards predominated. Patients said, for example, that "Human touch is important" (#75), quitting is easier with F2F support ("with help stopping will be easier" [#53, #14]), F2F treatment is "ideal," and it "adapts to your competencies" (#12). One patient, however, considered that it "can be hard if you dislike the counselor" (#14).

Building amongst others on *earlier stop smoking attempts*, patients had the standard that the quitting success may depend on themselves (reporting, eg, "Stopping you have to do for yourself." [#12]; "Treatment only makes sense if you have the will to stop" [#14]; quitting is "more a mental than a physical problem" [#27]; "You have to be strong" [#27]; or "You just have to do the things" [#53]), on missing support ("With help stopping will be easier" [#53, #14], and on stress ("Relapses are caused by stress" [#10, #34, #75]).

For *ICT-use in general*, while being familiar with using ICT (eg, searching the Web, using email/WhatsApp), the majority of patients showed a pragmatic standard "Computer is a tool" (#509, #75); "I am not a computer freak" (#12); "Computer is not my way" (#10); or "I am neither a forerunner nor a left behind" (#25). Only one patient (#34) reported that he "personalizes his mobile." Most patients also emphasized that they do not prefer computer-mediated communication over F2F communication because it "leads to misunderstandings" (#53), "it is easier to cheat online" (#34, #12), "it is easier to do sloppy" (34), "online information is not as important as written on paper" (#25), or "I do not trust internet information" (#509). Three of the patients (#106, #27, #25) reported that they use mobile devices (smartphone, tablet) more often, for example, "I use the laptop less since I have a tablet" (#25) or "I prefer mobile over PC" (#27).

Referring to *BSCT in general*, most patients (#106, #34, #27, #75, #25) described their *expectations* as "neutral" or "not clear," while some (#53. #34, #10) emphasized to expect support from BSCT, saying, for example, that they want the counselor to be "a driving force" (#10) or that they expect "to get more grip on smoking cessation" (#34). One patient (#14) remarked that BSCT "is new and sounds interesting."

Apparent Character of Blended Smoking Cessation Treatment

While being confronted with BSCT and moderated by their standards and expectations, the apparent character of BSCT that the patients constructed, seemed to be both positive and negative. The *pragmatic attributes* (*usability* and *utility*) were experienced mostly positive while the *hedonic attributes* (*stimulation*, *identification*, and *evocation*), especially for the Web sessions, tended to be negative.

Pragmatic Attributes of Blended Smoking Cessation Treatment

BSCT's pragmatic attributes (*usability* and *utility*) were experienced as good. However, some patients also criticized pragmatic aspects of BSCT, especially of the Web sessions, which indicated possibilities for further improvements.

Usability

Most patients experienced the *usability of BSCT in general* as positive, reporting, for example, that the "intake was good" (#75, #10 ,#14), "there have been no problems" (#509), "everyone was kind" (#34) "everything was clear and easy to use" (#53), "all was quite logical" (#14), the "treatment was picked up well" (#53, #27, #10), "BSCT parts connected to each other" (#106, #75, #14), and that "the intervals between sessions were fine" (#53). One patient (#14) reported "less travelling" (Note: BSCT patients only had to attend 5 F2F sessions at the clinic, compared with 10 F2F sessions in the F2F treatment as usual) as an advantage of BSCT, while another patient (#10) found that "still having to travel to the hospital at all" is a disadvantage. Further negative aspects of usability reported were the "long waiting list" before treatment start (#14, #12) (Note: regular waiting time before treatment start was around two months), "the long waiting times" in the waiting area before start of a F2F session (#14), that it was "not clear where to turn to outside the office hours" (#14), that "intervals between sessions were too long" (#10, #25), and that "not everything was explained in detail" (#25) and that the patient was "surprised about the order of the sessions" (#25).

The *usability of the F2F sessions* was experienced as "easy" (#25) or as "easier than web" (#27). Yet some patients criticized "that the counselor did not have enough time" (#12, #10, #14) or that the sessions were "slow and time consuming" (#27, #14).

Six patients (#509, #106, #53, #27, #75, #14), experienced the *usability ofWeb sessions* as "easy to use," while three patients (#34, #10, #25) reported the opposite ("not easy to use"). The patients criticized that the Web sessions were "too time demanding" (#509, #106), there was "a lot of repetition" (#10, #53, #27, #14), they "did not get immediate response" (#14), they "did not receive online assignments" (#27), and the "login would have been easier if you do not have to remember your password" (#27). Furthermore, two patients (#509, #106) reported that they did the smoking registration on paper before doing it on the Web because "it was simpler" (#106). However, this was "double work" (#106). Yet the patients liked "to be notified about new Web content automatically" (#75), that "emails and phone calls raised awareness" (#34), that "filling in forms online was handy" (#34), and "online saved time" (#25).

Utility

With regard to the *utility of BSCT in general*, the patients experienced the utility as positive, finding that "all BSCT parts were helpful—some more, some less" (#53), BSCT "matched my quitting process" (#53), "all has been discussed" (#106), "there was progress" (#106), BSCT "offered support" (#27), or "Web only would not have offered what I needed" (#75).

The *utility of the F2F session* was experienced as positive by most patients (7/10) also. Patients reported that F2F "offered flexibility" (#75) as "I could talk to the counselors about all of my problems" (#509), "all has been discussed" (#14), that with F2F "it was easier to ask questions" (#14), F2F "you got direct answers" (#14), F2F "stimulated more than web" (#10), and F2F "with medication was better than medication only" (#106). The counselors "reinforced" (#53, #25), "stimulated" (#53, #14, #10), "offered support" (#53, #25, #14), "shared good metaphors" (#53), and "explained everything very well" (14). Three patients experienced the F2F session as not useful, saying that the counselors "did not offer enough support" (#34, #27, 12), "did not reinforce" (#12), "did not motivate" (#12), "did not discuss all alternatives" (#34), and "asked too much questions" (#27).

For the *utility of the Web sessions, there were both positive and negative experiences*. Some patients had a predominantly negative experience saying that "reporting *via* Web was too time demanding" (#509), that Web "offered too much information" (#106), that Web "did not match my quitting process" (#27, #14), that "a computer does not answer" (#14) and that Web "does not work for me" (#75). Furthermore, ideas for improvement were reported, such as "an App would be better than web" (#34) and other services should be included, such as "short reinforcements *via* WhatsApp, emails, in-between sessions, video instructions, helpdesk, chat support, short instructions" (#34) and "audio information" (text to speech) (#27). However, patients also reported positive experiences saying that the Web "offered support in difficult moments" (#53), Web "offered tips" (#53), and "it was good to have information available online" (#27, #14, #34).

Hedonic Attributes of Blended Smoking Cessation Treatment

For the hedonic attributes (*stimulation*, *identification*, and *evocation*), BSCT was experienced both positively and negatively. While some patients felt *stimulated* by BSCT, others reported being demotivated. Especially for the Web sessions, most patients reported low *identification*. Also, the Web sessions *evoked* mostly negative comparisons and induced several ideas for improvements.

Stimulation

Patients reported both positive and negative stimulation by *BSCT in general* and rather low stimulation referring to the *F2F sessions* and *Web sessions*.

For *BSCT in general*, patients—on the one hand—felt stimulated to "quit smoking" (#14), to "discuss costs of smoking" (#12), to "think" (#106, #34), to "dig deeper" (#509), or to "look back" (#75). Patients also reported that the carbon monoxide measurements during the F2F sessions stimulated quitting (#53, #12). On the other hand, patients reported that "BSCT did not offer new things" (#34) or was "not interesting" (#14), and that certain interventions (ie, dealing with tempters) were "not new" (#25). Furthermore, patients were demotivated by "always the same questions" (#27), by "digging too deep" (#27), and by contradictory goals (quitting smoking vs weight reduction) (#27).

For the *F2F sessions*, patients said that the "counselor had no impact" (#27, #12, #14, #25). However, some patients (#12, #509, #34) reported that they were reinforced by the counselors to use the Web sessions.

For the *Web sessions*, one patient said, that Web "broadened your awareness" (#75), whereas the majority of patients reported no or low stimulation saying that "online won't get through to me" (#53, #34, #14, #25), "online exchange with the counselor did not affect extraordinary" (#25, #509), and to be demotivated by the Web sessions (#10, #509) or computer use (#106).

Identification

For *BSCT in general* patients could identify linking to individual features, such as "perseverance" or "self-control." However, for the Web sessions, most patients reported low identification. The ones showing higher identification with the Web sessions did this by referring to personal contact with the counselor. Patients found it easier with the *F2F sessions* than the *Web sessions* of BSCT.

Related to *BSCT in general*, patients reported that BSCT linked to individual features, such as "perseverance" (#75), "self-control" (#75), "the ability to work based on reading and writing" (#75), "IT-skills" (#10), and "age" (#10). However, one patient (#27) reported that she "felt treated like a child" and that she "lost her rhythms."

For the *F2F sessions* patients reported that these "felt more familiar" (#106), that patients liked "the F2F sessions the most" (#53) and "talking to the ladies" (#10) (Note: by this the male patient (#10) refers to the female counselors).

For the *Web sessions*, most patients reported low identification, saying, "I don't feel like it much" (#106) or "not to like online" (#106, #10), that "online is not my style" (#12, #75, #10, #25), to "prefer on paper" (#25), or being "too stupid for IT" (#10). One patient (#75) showed a higher identification with the Web sessions, emphasizing "Web I did for myself," "I know why I did Web," and "I understood the process." In turn, she criticized saying that "online did not give the opportunity to make it more personal" (#75). Three patients reported that the Web parts supported their personal contact with the counselor, mentioning that *via* Web parts "I had contact with her" and "they knew something about me" (#509), that "during the F2F sessions it became clear that the counselor reads the Web content" (#25), that "I had the idea that it is used on the other side" (#53), and that "you knew there is someone behind it" (#34). In turn, three patients reported that "you didn't know who has written the content" (#15), that "computer did not talk to you" (#12, #14), and that "you did not get the feeling that there is a human being on the other side" (#12).

Evocation

For the *Web sessions*, the patients reported several negative comparisons, such as "Web was like handling a machine, because you are not sitting opposite to each other" (#106), Web sessions were like "bookkeeping" (#53, #34, #14), like "a manual" (#53), like "filling in tax forms" (#10), and like "paper" (#27).

Situation

For the usage situation, mostly the *technical* context had a negative impact on the UX. Especially the *Web sessions* depended on the technical factors, which were criticized. Furthermore, referring to the *task* context of *BSCT in general*, some patients reported not having enough time for the treatment. Both the *physical* and the *social* context were described as mostly positive.

Technical

For the technical situation, the patients referred to the *Web sessions*, criticizing by saying that Web "did not work on iPad" (#34, #10, #25, #75). Although the patients had been informed at start of the treatment that the software for the Web session could not be used on tablet computers, they would have preferred to use tablets because the "Tablet is always on, Laptop

not" (#34, #75, #14, #25) and tablet "is more comfortable" (#10), or because they (#10, #25) moved from laptop to tablet during BSCT. Furthermore, for the use of computers for the Web sessions, the patients criticized by saying that they "had to start up the laptop, which takes time" (#106, #34, #14).

Task

Referring to tasks, patients reported not to have enough time for the BSCT "because of other tasks" (#509) or "because of family tasks" (#106), or to feel "sometimes stressed—sometimes relaxed" (#27).

Physical

For the *F2F sessions*, the patients reported little about the physical usage situation, mentioning only "that I live close to the hospital" (#25) and "that the treatment took place in the old building, which was not a nice place" (#34, #27) (Note: Between the patients treatment and the interviews the department moved to a new building).

For the *Web sessions*, the patients shared more information about the physical usage situation reporting that they did the Web sessions at "my own home office" (#25, #509), in a "hobby room upstairs, which is a nice place" (#10), "upstairs, where it is quite hot in the summer" (#14), "with the laptop at the dining table with wife and children around me" (#53), "in the kitchen" (#106), and "with laptop lying on the bed in the sleeping room" (#509).

Social

For the social situation during *BSCT in general*, most patients reported feeling supported by the family, saying that everyone "supported" (#53, #25) and "complimented" (#53), that "family motivated stopping" (#106) and "nearly no one in our family smokes" (#14), that "my partner stimulated" (#509, #10), "offered incentives" (#14, #53), "accompanied" (#14), "gave feedback on better health conditions" (#53) and "does not smoke" (#509), and that "children supported" (#27), "children were positive about quitting" (#53) and "my son also quit" (#14). One patient said he (#25") "lives alone" and "did not tell much about BSCT"; she reported that "everyone was sceptic of the quitting success." One patient (#27) reported that "her partner did not support," "questioned the Web sessions," and broke "the agreement to smoke outside only."

For friends and colleagues, the patients reported that "none of my friends smoke" (#10), "no one smokes inside" (#14), and that "colleagues also have positive experiences with cessation treatment" (#53). Furthermore, one patient emphasized that he "stimulates others to quit smoking" (#10).

Consequences

Overall, *BSCT in general* had a positive *appeal*, while *emotions* (eg, "satisfaction") varied. Again, there was clear distinction between the *F2F sessions* and the *Web sessions*. Similar to the emotional consequences, the behavioral consequences (*adherence, quitting*) also varied, ultimately resulting in diverse combinations of consequential appeal, emotions, and behavior.

Appeal

For six patients (#106, #53, #27, #75, #14, #25), *BSCT in general*, appealed to be "good." The patients reported that BSCT was a "mix of talking and reading" (#14) and it "offered variety" (#75). The "shared information both F2F and Web was fine" (#106) and "Web only would not have been so easy" (#53). F2F sessions and Web sessions were "quite different" (#34); "sometime F2F was better—sometimes Web was better" (#14) and "Web was an extension of F2F" (#53). One patient (#27) emphasized the medical treatment saying "Champix was good."

The *F2F sessions* mostly appealed to be "good." The patients reported that the F2F sessions were "fine" (#509) or "finer than web" (#106) and that the F2F sessions were "most important" (#53) or "most important at treatment start" (#34). One patient (#12) emphasized "that only F2F touches your heart" and that he would go for F2F "100% in all facets." However, one patient (#27) said that the F2F sessions were "whiny." For the counselors, one patient (#27) described her counselor as "nice," while another patient (#34) said that his counselor had a "stiff posture" and that she was "annoying," "pedantic" and "cumbersome."

For the *Web sessions*, the majority of patients reported a negative appeal, saying that the Web sessions "yielded nothing" (#509, #75, #14), were "a lot" (#509, #27), "cumbersome" (#106), "boring" (#34, #27), "tiring" (#27), "nonsense" (#12, #10), and "dead" (#10). However, one patient (#75) said that "Web was nice" while others—also referring to positive appeal—reported that the Web sessions could be done "comfortable at home" (#34) and that Web was "a serious matter" (#25), although she would not go for "Web only."

Emotion

Emotional consequences varied—some patients were satisfied with *BSCT in general*, some not. Again, there was a distinction between *F2F sessions* and *Web sessions*, but not as clear as for the appeal.

While two patients (#34, #25) said that they were not satisfied with *BSCT in general*, three patients reported to be satisfied (#27, #10) or "thankful" (#106). Furthermore, referring to negative emotions about BSCT in general, patients reported "feeling abandoned, left alone" (#12), "tension and the need to relax physically" (#75), and "contradictions between quitting smoking and weight reduction" (#27). One patient said that the F2F sessions and Web sessions stimulated "the same moods" (#25). The mood during the *F2F sessions* was "good" (#53, #27), while *Web sessions* were experienced as "unpleasant" (#27) and "making me nervous" (#34). One patient reported to feel "guilty because I did not stick to appointments" (#27).

Behavior

During the interviews, three patients (#14, #53, #25) reported that they *adhered* to *BSCT in general*, doing both the F2F sessions and the Web sessions. One of them (#14) said he "could have stopped after four sessions" because he was "sure not to need it in the future." However, he continued BSCT "to do the counselors and researchers a favor." Five patients (#106, #34, #27, #10, #25) reported that they found the *Web sessions*

"sloppy." Furthermore, one patient (#27) mentioned that she "forgot about some of her sessions."

Based on the auxiliary data (Table 2), medium adherence to *BSCT in general* (range 0-18, higher number indicate higher adherence) was 6.5 (IQR 5.50-11.75). Based on a 60% threshold for both the *F2F sessions* and the *Web sessions* [15], three patients (#14, #53, #106) were classified as adherent to BSCT in general. One patient (#75) was classified as adherent to the Web sessions but not to the F2F sessions, while another patient (#25)—one of the patients who reported to be adherent to BSCT in general during the interview—was classified as adherent to the F2F sessions but not to the Web sessions. Five patients (#509, #34, #27. #12, #10) were classified as nonadherent, because they neither adhered to the F2F sessions nor to the Web sessions.

Based on the interviews and the auxiliary data, four patients (#10, #14, #106, #53) reported successful *quitting*. One (#106) mentioned that "I had no problems because I had medication (Champix)" and "I threw away my last shags." The other one (#53) mentioned that he told himself "Never again!" and "Enough!" (Basta!), and that "he saved money for the holidays with his family." Two patients (#75, #509) reported that they reduced smoking during BSCT.

Combinations of Consequential Appeal, Emotions, and Behavior

The variety of consequential appeals, emotions, and behavior could be summarized in three types of combinations: "positive," "negative," and "mixed" consequences.

Three patients (#14, #53, #106) experienced "positive" consequences. BSCT appealed to be good and they felt "satisfied"/"thankful," adhered to the treatment and quit smoking.

On the contrary, another three patients (#12, #34, #509) experienced "negative" consequences: The Web sessions appealed negative ("nonsense," "boring," and "yielded nothing") and BSCT in general resulted in negative emotions (abandoned/not satisfied). Ultimately, they did not adhere to the treatment and did not quit smoking.

Mixed consequences: Three (#25, #27, #75) of the four remaining patients did not quit smoking, while one (#10) did. Interestingly, BSCT in general appealed "good" to the nonquitters (#25, #27, #75) while—for the quitter (#10) at least, the Web sessions appealed to be "nonsense." Although two of the nonquitters (#25, #75) reported negative emotions ("tension"/"not satisfied"), these two patients at least partly adhered to BSCT (#25 adherent to F2F sessions; #75 adherent to Web sessions). In turn, the third nonquitter (#27) reported positive emotions ("satisfied") but did not adhere at all.

To the remaining quitter (#10), although the Web sessions appealed to be "nonsense" and he did not adhere to BSCT in general, he reported positive emotions ("satisfied") and ultimately quit smoking.

Discussion

Principal Findings

This study aimed to provide insight in the UX of a blended treatment. In the light of this study, the expectation that the strength of one mode of delivery can compensate for the weaknesses of the other in blended treatment, can be partially supported because the F2F sessions compensated for the weaknesses of the Web sessions so that BSCT in general was mostly experienced positively.

Our study described the UX of a BSCT using Hassenzahl's key elements of UX from a user's perspective [23,24]. Overall, BSCT in general appeared to be a mostly positively experienced service. Patients had a positive-pragmatic standard and neutral-open expectation toward BSCT in general at treatment start, and the pragmatic attributes of the F2F session were mostly perceived as positive while the pragmatic attributes of the Web sessions were perceived as both positive and negative. For the hedonic attributes, there seems to be a difference between the F2F and Web sessions. Specifically, the hedonic attributes of the Web sessions were experienced mostly negative while the hedonic attributes of the F2F sessions were mostly positive. For the usage situation, the physical and social context was experienced positively while the task and technical context was experienced negatively. Nevertheless, the consequential appeal of BSCT in general was positive. However, the consequential emotions and behavior varied, ultimately resulting in diverse combinations of consequential appeal, emotions, and behavior (positive, negative, and mixed).

Although patients' pretreatment expectations toward BSCT were neutral and the Web sessions appealed negative, overall BSCT in general appeared to be positively experienced afterwards. This is in line with an evaluation study (n=7) by Kooistra et al [14] of a blended cognitive behavioral treatment for major depression. However, our study provides a more differentiated insight in why the Web sessions were appraised negatively. Applying Hassenzahl's distinction between pragmatic and hedonic attributes [23], our findings suggest that while patients experienced the pragmatic attributes (usability, utility) of the Web sessions in general as more positive, the negative hedonic attributes (stimulation, identification, and evocation) of the Web sessions led to a combination of negative consequences, such as negative appeal, negative emotions, and low adherence.

Interestingly, although the hedonistic gap made the Web sessions appeal negatively, the overall BSCT was experienced positively. This could support the assumption that in blended treatment the strength of one mode (ie, F2F) may compensate for the weaknesses of the other (ie, Web) [4,5]. This is further supported by our findings about relatedness and identification, which are in line with a qualitative study (n=14) by Wilhelmsen et al (2013) [22] on the internet-based cognitive behavioral therapy for depression supported by short F2F consultations. Three of our patients that adhered to BSCT and ultimately quit smoking showed rather low identification with the Web sessions but had positive appeal and emotions toward BSCT in general, especially

toward the F2F sessions. This positive overall appraisal may have cancelled out the negative appeal of the Web sessions.

We mainly found that the F2F sessions compensated for the weaknesses of the Web sessions. Yet three patients reported that the Web sessions influenced their personal contact with the counselor positively. Although the Web sessions mostly had a low identification and a negative appeal, the Web sessions supported the F2F sessions because these patients felt more related to the counselor. However, even though the Web sessions may have supported the F2F sessions, it should be noted that none of the patients indicated that the Web sessions compensated for the F2F sessions. It remains undecided if this is because there was no need for compensation as the F2F sessions were overall positive, or that the Web sessions were not able to compensate. It should also be taken into account that the routine care in the hospital was not Web based, the patients were of older age, and did not have an affinity for the internet (although they reported to have sufficient internet skills), and the patients' preferences for modes of delivery were not taken into account as they were not free to choose for BSCT because they were included in a randomized controlled trial. These factors may additionally explain the low positive impact of the Web sessions.

The emotional and behavioral consequences varied, ultimately resulting in three types of combinations of appeal, emotions (eg, satisfaction), and behavior (adherence, quitting): "positive," "negative," and "mixed." These types can be used to work on UX profiles that can support further development of blended care and improve the matching between the treatment and patient [3].

Implication for Future Research and Clinical Practice

Further work needs to be done to investigate how the integration of F2F and Web treatments can be carried out to ultimately increase the effectiveness and efficiency of a blended treatment. This study provides a hint to explore this question by emphasizing the relevance of hedonic attributes in the UX. Even if the UX was predominantly positive because the hedonistic gap in the area of the Web sessions was compensated relatively easily by the F2F sessions, this does not mean that BSCT cannot be further improved to increase adherence and long-term abstinence. Hedonism could be a starting point for this. Further research on the following questions could be useful:

Could the hedonistic gap in the Web sessions be not only due to the mode of delivery, but also the concrete content of the Web sessions? Perhaps, it was precisely the interventions that the patients experienced as nonhedonistic, which were the part of the Web sessions. This was neither explicitly considered in the treatment design nor asked for in detail in the interviews. However, this might have been the case because more standard exercises and messages could be offered on the Web more easily. A stronger involvement of patients in the early design stages of the Web sessions may help to prevent the hedonistic gap.

May hedonism play a less prominent role in the health care context than in the other domains? Patients tend to approach a health problem with a pragmatic-neutral expectation, such as "What's important is that it works. As long as it helps, I can also accept that it is unpleasant." Consequently, hedonistic aspects, such as fun, enjoyment, pleasure, and aesthetics may not be expected in the first place, and therefore, may not be missed. Moreover, this may be compensated relatively easily by positive experiences with the counselors. However, if hedonism was less important in health care, it would contradict our conclusion that it should receive more attention.

Could both scientific research and clinical practice use insights from persuasive systems design [36,37], "nudging" [38] and "funology" [23,24] to address the hedonic gap, which may negatively influence smoking cessation patients who are usually a highly motivated target group [39]? Persuasive design features, such as primary task support (eg, tailoring, personalization), dialogue support (eg, rewards, liking), credibility support (eg, real-world feel), social support (eg, normative influence, competition), and hedonic aspects (eg, fun, enjoyment, pleasure, and aesthetics) may play a role in sustaining patients' motivation to adhere to the treatment and quit smoking.

How do the apparent character and the consequential appeal and emotions relate to the quitting behavior? On the one hand, apparently a negative appeal (ie, missing hedonic attributes) may lead to consequential combination of negative appeal, emotions, and behavior (ie, neither adhere nor quit). On the other hand, it is also possible to distinguish between diverse episodic UXs ultimately leading to a cumulative UX [20], for example, a motivated patient may start with a positive UX but after failing to quit or relapsing, the patient's standards and expectations may change during the treatment, which can then lead to a negative appeal and ultimately to a cumulative negative UX. The cumulative UX would then not be the result of a linear process as in the model of Hassenzahl (Figure 1). Rather, in a circular process, consequences (ie, quitting), apparent character, expectations, and standards would influence each other.

Strengths and Limitations

The data and model used in this study provided a rich insight into the UX of a blended treatment for smoking cessation in an ambulant clinical setting. Though this study yielded valuable knowledge for the understanding and improvement of BSCT and the matching of patients and treatment, limitations should be noted when interpreting the findings. First, the sample of patients used in this study was a purposive sample that was intended to represent the heterogeneity of the patients of an outpatient cessation clinic. Hence, it is uncertain, if the rather small sample (n=10) is representative of the population referring to characteristics, such as sex, age, internet skills, or educational level, and if the thematic saturation was reached with this sample size. It should also be considered that patients did not choose BSCT on their own but were randomly assigned to it because they participated in a randomized controlled trial. However, the high degree of consensus in the findings may indicate generalization of our main conclusions. Second, the interviews were conducted retrospectively. Conducting additional interviews at treatment start and during treatment could have offered a more valid insight in the process of patients' UX construction (eg, for the standards, expectations, and apparent character). Third, the software that was used for the Web sessions was developed around 2005, which may have led to

technical incommodities (eg, Web software is “Flash”-based and nonresponsive; not mobile device–compatible), which may have negatively impacted the UX. This assumption is based on the fact that patients often stated that they would have liked to do the Web sessions on their mobile device. We assume that a newer mobile device–compatible software with similarly good pragmatic attributes as the previous Flash-software could have also improved the hedonistic attributes and thus have led to more positive consequences. Fourth, the interviews were conducted with the first patients that followed the new blended version of the smoking cessation treatment. At that time, the treatment still had some teething problems, such as being new for the originally F2F counselors. We did not integrate the counselors’ views on the uptake of BSCT, and therefore, we cannot compensate for bias through inadequate treatment fidelity. Fifth, as long-term abstinence is the goal of a smoking cessation treatment, prolonged follow-up analysis of patients’ UX could reveal a different picture. For example, some patients may continue using the Web-based modality and benefit from this at a later stage, resulting in a UX that would be more in favor of the Web-based treatment. Conversely, relapse to smoking at a later stage may lead to a negative adjustment of the UX of the blended treatment. Sixth, we could not elaborate further on which specific parts of the Web sessions were experienced positively or negatively as we did not ask for these in detail in the interviews. Seventh, the study interventions are selected and combined by the researchers and treatment developers without considering the individual patients who ultimately followed the treatment. This resulted in a rather inflexible approach of blending (five Web-based sessions and five F2F sessions in a fixed sequence and with equivalent content) to allow for comparisons with the F2F treatment as usual in the randomized controlled trial (LiveSmokefree study) [8]. This inflexible approach is due to the research design and may limit the potential of blending. In daily practice, the blending of Web-based and F2F interventions may lead to a flexible exchangeability of all intervention components, which would foster a treatment that is highly tailored to the patient’s needs and abilities and could lead to a different UX.

Conclusions

This study provides insight into the key elements of the UX of a blended treatment for smoking cessation and supports the expectation that in a blended treatment, one mode of delivery may compensate for the weaknesses of the other. However, in this certain setting, this could be mainly achieved in only one way: F2F sessions compensated for the weaknesses of the Web sessions. As a practical conclusion, this may mean that the Web sessions supported by the strength of the F2F sessions, offer an interesting approach for further improving the blended treatment in this specific context. Our theoretical findings reflect the relevance of the aspects of hedonism, such as fun, joy, or happiness in UX [23], which were not mentioned in relation to the Web sessions and only scarcely mentioned in relation to the F2F sessions. Future research should further investigate the role of hedonistic aspects in the blended treatment and if increased enjoyment of the blended treatment could increase the treatment adherence and ultimately its effectiveness.

Acknowledgments

This research was funded by the Technology, Health, and Care research group at the Saxion University of Applied Sciences and the Centre for eHealth and Well-Being Research at the University of Twente. The authors would like to thank the directors, personnel, and patients of the Medisch Spectrum Twente and Tactus Addiction Treatment.

Authors' Contributions

LS, SA, MP, MK, and MEP identified the study questions and designed the study. LS is the principal investigator and prepared the first draft of this manuscript. LS, SA, MP, MK, MEP, and RS edited this manuscript. LS, SA, and MP revised the manuscript. All authors approved the final version of this manuscript for publication.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Screenshots of the Web sessions of Blended Smoking Cessation Treatment.
[PDF File (Adobe PDF File), 1447 KB - formative_v4i6e14550_app1.pdf]

Multimedia Appendix 2
Blended Smoking Cessation Treatment user experience Interview Guide (Dutch).
[PDF File (Adobe PDF File), 171 KB - formative_v4i6e14550_app2.pdf]

Multimedia Appendix 3
Codes, code description, and intercoder agreement per semantic domain.
[PDF File (Adobe PDF File), 81 KB - formative_v4i6e14550_app3.pdf]

Multimedia Appendix 4
[PDF File (Adobe PDF File), 159 KB - formative_v4i6e14550_app4.pdf]

References

1. 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]
2. Kooistra LC, Wiersma JE, Ruwaard J, van Oppen P, Smit F, Lokkerbol J, et al. Blended vs face-to-face cognitive behavioural treatment for major depression in specialized mental health care: study protocol of a randomized controlled cost-effectiveness trial. BMC Psychiatry 2014 Oct 18;14:290 [FREE Full text] [doi: 10.1186/s12888-014-0290-z] [Medline: 25326035]
3. Wentzel J, van der Vaart R, Bohlmeijer ET, van Gemert-Pijnen JE. Mixing online and face-to-face therapy: how to benefit from blended care in mental health care. JMIR Ment Health 2016 Feb 9;3(1):e9 [FREE Full text] [doi: 10.2196/mental.4534] [Medline: 26860537]
4. Barak A, Hen L, Boniel-Nissim M, Shapira N. A comprehensive review and a meta-analysis of the effectiveness of internet-based psychotherapeutic interventions. J Technol Hum Serv 2008;26(2-4):109-160. [doi: 10.1080/15228830802094429]
5. Kemmeren LL, van Schaik DJ, Riper H, Kleiboer AM, Bosmans JE, Smit JH. Effectiveness of blended depression treatment for adults in specialised mental healthcare: study protocol for a randomised controlled trial. BMC Psychiatry 2016 Apr 21;16:113 [FREE Full text] [doi: 10.1186/s12888-016-0818-5] [Medline: 27102812]
6. Erbe D, Eichert H, Riper H, Ebert DD. Blending face-to-face and internet-based interventions for the treatment of mental disorders in adults: systematic review. J Med Internet Res 2017 Sep 15;19(9):e306 [FREE Full text] [doi: 10.2196/jmir.6588] [Medline: 28916506]
7. Postel MG, Witting M, van Gemert-Pijnen JE. Blended behandeling in de geestelijke gezondheidszorg [Blended treatment in mental health]. Dth : kwartaaltijdschrift voor directieve therapie en hypnose 2013;33(3):210-221 [FREE Full text]
8. Siemer L, Pieterse ME, Brusse-Keizer MG, Postel MG, Allouch SB, Sanderman R. Study protocol for a non-inferiority trial of a blended smoking cessation treatment versus face-to-face treatment (LiveSmokefree-Study). BMC Public Health 2016 Nov 24;16(1):1187 [FREE Full text] [doi: 10.1186/s12889-016-3851-x] [Medline: 27881108]
9. van der Vaart R, Witting M, Riper H, Kooistra L, Bohlmeijer ET, van Gemert-Pijnen LJ. Blending online therapy into regular face-to-face therapy for depression: content, ratio and preconditions according to patients and therapists using a Delphi study. BMC Psychiatry 2014 Dec 14;14:355 [FREE Full text] [doi: 10.1186/s12888-014-0355-z] [Medline: 25496393]
10. Harrington KF, McDougal JA, Pisu M, Zhang B, Sadasivam RS, Houston TK, CHART Collaborative Group. Web-based smoking cessation intervention that transitions from inpatient to outpatient: study protocol for a randomized controlled trial. Trials 2012 Aug 1;13:123 [FREE Full text] [doi: 10.1186/1745-6215-13-123] [Medline: 22852802]
11. Massoudi B, Blanker MH, van Valen E, Wouters H, Bockting CL, Burger H. Blended care vs usual care in the treatment of depressive symptoms and disorders in general practice [BLENDING]: study protocol of a non-inferiority randomized trial. BMC Psychiatry 2017 Jun 13;17(1):218 [FREE Full text] [doi: 10.1186/s12888-017-1376-1] [Medline: 28610561]
12. Bruinsma A, Kampman M, Exterkate CC, Hendriks GJ. [An exploratory study of 'blended' cognitive behavioural therapy (CBT) for patients with a panic disorder: results and patients' experiences]. Tijdschr Psychiatr 2016;58(5):361-370 [FREE Full text] [Medline: 27213635]
13. Månsson KN, Ruiz ES, Gervind E, Dahlin M, Andersson G. Development and initial evaluation of an Internet-based support system for face-to-face cognitive behavior therapy: a proof of concept study. J Med Internet Res 2013 Dec 10;15(12):e280 [FREE Full text] [doi: 10.2196/jmir.3031] [Medline: 24326065]
14. Kooistra LC, Ruwaard J, Wiersma JE, van Oppen P, van der Vaart R, van Gemert-Pijnen JE, et al. Development and initial evaluation of blended cognitive behavioural treatment for major depression in routine specialized mental health care. Internet Interv 2016 May;4:61-71 [FREE Full text] [doi: 10.1016/j.invent.2016.01.003] [Medline: 30135791]
15. Siemer L, Brusse-Keizer MG, Postel MG, Allouch SB, Bougioukas AP, Sanderman R, et al. Blended smoking cessation treatment: exploring measurement, levels, and predictors of adherence. J Med Internet Res 2018 Aug 1;20(8):e246 [FREE Full text] [doi: 10.2196/jmir.9969] [Medline: 30068503]
16. Kleiboer A, Smit J, Bosmans J, Ruwaard J, Andersson G, Topooco N, et al. European COMPARative Effectiveness research on blended Depression treatment versus treatment-as-usual (E-COMPARED): study protocol for a randomized controlled, non-inferiority trial in eight European countries. Trials 2016 Aug 3;17(1):387 [FREE Full text] [doi: 10.1186/s13063-016-1511-1] [Medline: 27488181]
17. Liébana-Cabanillas F, Muñoz-Leiva F, Sánchez-Fernández J, Jesús MI. The moderating effect of user experience on satisfaction with electronic banking: empirical evidence from the Spanish case. Inf Syst E-Bus Manag 2015;14(1):141-165. [doi: 10.1007/s10257-015-0277-4]
18. Ramtohul I. The adoption of e-health services: Comprehensive analysis of the adoption setting from the user s perspective. Health Policy Technol 2015 Sep;4(3):286-293. [doi: 10.1016/j.hlpt.2015.04.007]
19. Roto V, Lee M, Pihkala K, Castro B, Vermeeren A, Law E, et al. All About UX. URL: http://www.allaboutux.org [accessed 2020-02-05]

20. Roto V, Law E, Vermeeren A, Hoonhout J. All About UX. 2011 Feb 11. User Experience White Paper. Bringing clarity to the concept of user experience. Result from Dagstuhl Seminar on Demarcating User Experience, September 15-18, 2010 URL: http://allaboutux.org/files/UX-WhitePaper.pdf [accessed 2020-02-05]
21. Law EL, Roto V, Hassenzahl M, Vermeeren A, Kort J. Understanding, Scoping and Defining User Experience: A Survey Approach. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2009 Presented at: CHI'09; April 4-9, 2009; MA, Boston, USA p. 719-728. [doi: 10.1145/1518701.1518813]
22. Wilhelmsen M, Lillevoll K, Risør MB, Høifødt R, Johansen M, Waterloo K, et al. Motivation to persist with internet-based cognitive behavioural treatment using blended care: a qualitative study. BMC Psychiatry 2013 Nov 7;13:296 [FREE Full text] [doi: 10.1186/1471-244X-13-296] [Medline: 24199672]
23. Hassenzahl M. The Thing and I (Summer of '17 Remix). In: Blythe M, Monk A, editors. Funology 2. Cham: Springer; 2018:17-31.
24. Hassenzahl M. The thing and I: understanding the relationship between user and product. In: Blythe MA, Overbeeke K, Monk AF, Wright PC, editors. Funology. Dordrecht: Springer Netherlands; 2003:31-42.
25. Hassenzahl M. User Experience (UX): Towards an Experiential Perspective on Product Quality. In: Proceedings of the 20th Conference on l'Interaction Homme-Machine. User experience (UX): towards an experiential perspective on product quality. Proceedings of the 20th Conference on l'Interaction Homme-Machine; 2008 Presented at: IHM'08; September 2-5, 2008; Metz, France p. 11-15. [doi: 10.1145/1512714.1512717]
26. Michie S, Hyder N, Walia A, West R. Development of a taxonomy of behaviour change techniques used in individual behavioural support for smoking cessation. Addict Behav 2011 Apr;36(4):315-319. [doi: 10.1016/j.addbeh.2010.11.016] [Medline: 21215528]
27. Castañeda JA, Muñoz-Leiva F, Luque T. Web Acceptance Model (WAM): Moderating effects of user experience. Inform Manag 2007 Jun;44(4):384-396. [doi: 10.1016/j.im.2007.02.003]
28. Haun JN, Lind JD, Shimada SL, Martin TL, Gosline RM, Antinori N, et al. Evaluating user experiences of the secure messaging tool on the Veterans Affairs' patient portal system. J Med Internet Res 2014 Mar 6;16(3):e75 [FREE Full text] [doi: 10.2196/jmir.2976] [Medline: 24610454]
29. Kwaliteitsinstituut voor de Gezondheidszorg CBO. Nederlands Huisartsen Genootschap. 2009. Richtlijn Behandeling van Tabaksverslaving herziening 2009 [Guideline on Tobacco Addiction revision 2009] URL: https://www.nhg.org/sites/default/files/content/nhg_org/uploads/tabaksverslaving-herziening.pdf [accessed 2020-03-11]
30. Mijnregister. Den Haag: Partnership Stop met Roken; 2009. Zorgmodule Stoppen met Roken URL: https://kwaliteitsregisterstopmetroken.nl/userfiles/Nieuws/Zorgmodule_Stoppen_met_Roken.pdf [accessed 2020-03-11]
31. Christenhusz L, Prenger R, Pieterse M, Seydel E, van der Palen J. Cost-effectiveness of an intensive smoking cessation intervention for COPD outpatients. Nicotine Tob Res 2012 Jun;14(6):657-663. [doi: 10.1093/ntr/ntr263] [Medline: 22180589]
32. Pieterse M, van der Palen J, Hagens P. A randomised controlled trial with 18-months follow-up in COPD patients comparing a smoking reduction programme with a placebo self-help intervention, on reduction followed by sustained cessation. Dutch Asthma Foundation grant no. 3408036, NTR trial ID: NTR2227. In press 2009.
33. McLellan E, MacQueen KM, Neidig JL. Beyond the Qualitative Interview: Data Preparation and Transcription. Field Methods 2003;15(1):63-84. [doi: 10.1177/1525822x02239573]
34. DeCuir-Gunby JT, Marshall PL, McCulloch AW. Developing and using a codebook for the analysis of interview data: An example from a professional development research project. Field Methods 2011;23(2):136-155. [doi: 10.1177/1525822x10388468]
35. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström Test for Nicotine Dependence: a revision of the Fagerström Tolerance Questionnaire. Br J Addict 1991 Sep;86(9):1119-1127. [doi: 10.1111/j.1360-0443.1991.tb01879.x] [Medline: 1932883]
36. 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 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
37. 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]
38. Vlaev I, King D, Dolan P, Darzi A. The theory and practice of 'Nudging': changing health behaviors. Public Admin Rev 2016;76(4):550-561. [doi: 10.1111/puar.12564]
39. Perski O, Herd N, West R, Brown J. Perceived addiction to smoking and associations with motivation to stop, quit attempts and quitting success: A prospective study of English smokers. Addict Behav 2019 Mar;90:306-311 [FREE Full text] [doi: 10.1016/j.addbeh.2018.11.030] [Medline: 30476896]

Abbreviations

BSCT: Blended Smoking Cessation Treatment
eHealth: electronic health
F2F: face-to-face

SRP: Outpatient Smoking Cessation Clinic (Dutch: Stoppen met Roken Poli)
UX: user experience

Edited by G Eysenbach; submitted 01.05.19; peer-reviewed by M Schulte, G Khalil; comments to author 24.06.19; revised version received 10.10.19; accepted 27.01.20; published 03.06.20.

Please cite as:
Siemer L, Ben Allouch S, Pieterse ME, Brusse-Keizer M, Sanderman R, Postel MG
Patients' User Experience of a Blended Face-to-Face and Web-Based Smoking Cessation Treatment: Qualitative Study
JMIR Form Res 2020;4(6):e14550
URL: https://formative.jmir.org/2020/6/e14550
doi:10.2196/14550
PMID:32343245

Original Paper

Feasibility of In-Home Sensor Monitoring to Detect Mild Cognitive Impairment in Aging Military Veterans: Prospective Observational Study

Adriana Seelye[1,2,3,4], PhD; Mira Isabelle Leese[1], BA; Katherine Dorociak[1], PhD; Nicole Bouranis[3,4], MA; Nora Mattek[3,4], MPH; Nicole Sharma[3,4], BA; Zachary Beattie[3,4], PhD; Thomas Riley[3,4], BSc; Jonathan Lee[2,4], BA; Kevin Cosgrove[3,4], MSEE, PE; Nicole Fleming[3,4], BSc; Jessica Klinger[1,2], MSW; John Ferguson[1,5], PhD; Greg John Lamberty[1,2], PhD; Jeffrey Kaye[3,4], MD

[1]Minneapolis Veterans Affairs Health Care System, Minneapolis, MN, United States

[2]Department of Psychiatry, University of Minnesota, Minneapolis, MN, United States

[3]Department of Neurology, Oregon Health and Science University, Portland, OR, United States

[4]Oregon Center for Aging and Technology, Portland, OR, United States

[5]Division of Rehabilitation Science, Department of Rehabilitation Medicine, University of Minnesota, Minneapolis, MN, United States

Corresponding Author:
Adriana Seelye, PhD
Minneapolis Veterans Affairs Health Care System
1 Veterans Dr
Minneapolis, MN,
United States
Phone: 1 6127252000
Email: adriana.seelye@va.gov

Abstract

Background: Aging military veterans are an important and growing population who are at an elevated risk for developing mild cognitive impairment (MCI) and Alzheimer dementia, which emerge insidiously and progress gradually. Traditional clinic-based assessments are administered infrequently, making these visits less ideal to capture the earliest signals of cognitive and daily functioning decline in older adults.

Objective: This study aimed to evaluate the feasibility of a novel ecologically valid assessment approach that integrates passive in-home and mobile technologies to assess instrumental activities of daily living (IADLs) that are not well captured by clinic-based assessment methods in an aging military veteran sample.

Methods: Participants included 30 community-dwelling military veterans, classified as healthy controls (mean age 72.8, SD 4.9 years; n=15) or MCI (mean age 74.3, SD 6.0 years; n=15) using the Clinical Dementia Rating Scale. Participants were in relatively good health (mean modified Cumulative Illness Rating Scale score 23.1, SD 2.9) without evidence of depression (mean Geriatrics Depression Scale score 1.3, SD 1.6) or anxiety (mean generalized anxiety disorder questionnaire 1.3, SD 1.3) on self-report measures. Participants were clinically assessed at baseline and 12 months later with health and daily function questionnaires and neuropsychological testing. Daily computer use, medication taking, and physical activity and sleep data were collected via passive computer monitoring software, an instrumented pillbox, and a fitness tracker watch in participants' environments for 12 months between clinical study visits.

Results: Enrollment began in October 2018 and continued until the study groups were filled in January 2019. A total of 201 people called to participate following public posting and focused mailings. Most common exclusionary criteria included nonveteran status 11.4% (23/201), living too far from the study site 9.4% (19/201), and having exclusionary health concerns 17.9% (36/201). Five people have withdrawn from the study: 2 with unanticipated health conditions, 2 living in a vacation home for more than half of the year, and 1 who saw no direct benefit from the research study. At baseline, MCI participants had lower Montreal Cognitive Assessment (P<.001) and higher Functional Activities Questionnaire (P=.04) scores than healthy controls. Over seven months, research personnel visited participants' homes a total of 73 times for technology maintenance. Technology maintenance visits were more prevalent for MCI participants (P=.04) than healthy controls.

Conclusions: Installation and longitudinal deployment of a passive in-home IADL monitoring platform with an older adult military veteran sample was feasible. Knowledge gained from this pilot study will be used to help develop acceptable and effective home-based assessment tools that can be used to passively monitor cognition and daily functioning in older adult samples.

(JMIR Form Res 2020;4(6):e16371) doi:10.2196/16371

KEYWORDS

aging; mild cognitive impairment; activities of daily living; technology

Introduction

Background

By year 2050, the number of people living with dementia is projected to triple to 115 million as the world's aging population continues to grow rapidly [1,2]. There will be an increased demand for health care institutions and researchers to respond and develop preventative strategies to address the growing needs of the aging population. An important subgroup of the larger aging population is aging military veterans. Aging military veterans are at an elevated risk for developing dementia because of their unique military histories (eg, traumatic brain injury and posttraumatic stress disorder) [3-6] and their increased incidence of vascular risk factors (eg, diabetes, hypertension, and hyperlipidemia) [7-9]. In 2017, the prevalence of Alzheimer disease (AD) among US military veterans was 750,000, an increase of 166% from 2014 [10]. On average, the expense related to lifetime dementia care is US $350,174 (in 2018) per person, which is US $150,303 more than the expense for those without dementia [11]. As neurodegenerative disorders such as AD progress slowly and over a long period of time [12], early detection of dementia is crucial and has the potential to reduce the number of individuals and caregivers affected by the disease [10]. Identifying individuals with mild cognitive impairment (MCI), which often represents the prodromal stage of several neurodegenerative diseases (including AD) [13], could lead to targeted interventions that ultimately improve daily function and independence.

Although basic activities of daily living such as bathing, grooming, and eating are affected later in the course of neurodegenerative diseases after cognitive impairment has progressed and significantly impacted daily functions, subtle difficulties performing instrumental activities of daily living (IADLs) are present earlier in the course of disease. These IADLs are behavioral signs that may signal the progression from normal aging to MCI providing the opportunity for early detection and intervention [14-19]. IADLs are cognitively complex daily activities that require multiple constituent cognitive processes to perform accurately and efficiently and that are crucial for independent living. Examples of IADLs include managing medications and finances, driving a motor vehicle, and using every day technology such as computers and mobile phones [20]. However, early detection measures for MCI currently face many challenges and fail to identify change in real time because of infrequent episodic clinic visits [21]. The sensitivity and specificity of cognitive screeners used to detect impaired cognition has an unacceptable proportion of false positives and negatives and fails to approximate a patient's real-world difficulties. In addition, patient self-report and collateral information can be biased or unreliable because of worry, stress, or forgetfulness [21,22]. Therefore, a comprehensive neuropsychological battery has been considered the standard for clarifying the nature and extent of someone's cognitive deficits [23]. Unfortunately, these neuropsychological evaluations are time-consuming, expensive, do not provide a clear appraisal of one's functional performance, and are not widely accessible to older adults who face socioeconomic and geographic barriers to specialty care services.

An alternative assessment approach to capture changes in cognitive status is the unobtrusive collection of continuous activity data over long periods of time [21,22]. Real-world assessment technologies have allowed researchers to continuously monitor cognitively demanding functional activities in one's home environment to identify abnormal activity patterns predictive of MCI [22,24,25]. Everyday consumer devices (eg, medication pillbox, and home computer) are being used to continuously observe cognitively challenging IADLs. This offers a practical, low cost, and noninvasive approach to assessing changes in one's daily functioning [19,26].

This Study and Objectives

Despite a growing interest in the use of real-world assessment approaches to detect early signs of MCI, there are still many gaps in the current literature. This paper has described a 12-month pilot study, Promote Independent Aging (PRIA) is among the first to deploy unobtrusive sensor-based assessment technologies in the homes of aging military veterans in the community. PRIA partners with the Oregon Center for Aging and Technology (ORCATECH) [21] and the Collaborative Aging Research using Technology (CART) initiative [27], utilizing components of the ORCATECH-CART in-home and mobile sensor assessment platform and infrastructure to identify and monitor meaningful changes in routine daily activities that are affected by MCI, such as computer use, medication taking, physical activity, and sleep.

In this paper, we have described the sample of participants recruited into the study, the clinical assessment procedures, and the in-home sensor-based assessment platform and monitoring of the activity measures (outcome variables) in the study. We have discussed our experiences with feasibility (eg, recruitment, retention, installation, and in-home technology maintenance visits) of the in-home monitoring technologies. Finally, we have concluded with a discussion about future directions, limitations, and clinical implications of this research.

Methods

Study Participants

Participants were 30 community-dwelling older adult military veterans. Of this group, 15 were classified as healthy controls and 15 were classified as MCI using established clinical and research measures. All participants were recruited from the Minneapolis-Saint Paul, Minnesota metropolitan area, and gave written informed consent before participating in study activities. The protocol was approved by the Minneapolis Veterans Affairs Health Care System's (MVAHCSs) institutional review board (IRB #4748-A). Enrollment began in October 2018 and continued on a rolling basis until January 2019. Participants were compensated US $40 per month over the course of 12 months. See Textbox 1 for participant inclusion criteria.

Participants were recruited through recruitment letters and advertising (eg, flyers) targeted toward patients seen at the MVAHCS in primary care and specialty clinics serving older adult military veterans. Initial screening data were also pulled from the Veterans Affairs (VA) Informatics and Computing Infrastructure and the computerized patient record system to screen military veterans for eligibility based on age, location, and health history. Recruitment letters were sent out to patients who met basic eligibility criteria from these datasets, and follow-up calls were made within 2 weeks to interested parties. Potential candidates were also pulled from a clinical database that includes neuropsychological, clinical, and demographic information from over 2000 military veterans referred for outpatient neuropsychological evaluations since 2014 at the MVAHCS (VA IRB#: 4637-B).

Textbox 1. Promote Independent Aging inclusion criteria.

- Aged 65 years or older
- Live within 30 miles of the Minneapolis Veterans Affairs Health Care System
- Live independently in their home (living with a companion or spouse was allowed but not as a caregiver)
- Take at least one medication daily and willing to use the instrumented study pillbox
- Have a home internet connection
- Own a computer and use it at least once per week
- Relatively healthy for age (no poorly controlled or unstable medical conditions or major neurological disorders)
- Absence of moderate to severe depression (Geriatric Depression Scale score-15≥7) [28]
- Absence of moderate to severe anxiety (Generalized anxiety disorder-7 questionnaire score>5) [29]
- No impaired global cognition (Montreal Cognitive Assessment sex, age, and education adjusted z-scores<−2) [30]
- Do not meet criteria for dementia (having a global Clinical Dementia Rating Scale score of less than or equal to 0.5 indicating no major impairment in daily functioning) [31]

Clinical Assessment Procedures

Participants were assessed at the MVAHCS Geriatric Research, Education and Clinical Center at baseline and 12 months. Research staff met with participants as well as their study informant (usually a spouse, close family member, or friend) during the baseline and final study visit. A battery of standardized neuropsychological tests, health assessments, and daily function questionnaires (eg, Functional Activities Questionnaire [32]) was administered (see Table 1), with a subset of this assessment battery from the Uniform Data Set of the National Alzheimer's Disease Coordinating Center [33] in addition to other well-validated measures used in prospective National Institute on Aging (NIA)–funded longitudinal aging cohort studies. Health assessments consisted of a review of medical histories, medication lists, and completion of the Modified Cumulative Illness Rating Scale (MCIRS) [34,35]. The neuropsychological examination included the following battery of well-established and validated tests assessing multiple cognitive domains: attention and processing speed (Number Span Forward, Trail Making Part A, Stroop Color Naming, Stroop Word Reading) [33,36,37], working memory (Number Span Backward) [33], Memory (Craft Story Recall; Consortium to Establish a Registry for Alzheimer's Disease delayed recall and recognition; and Benson Complex Figure Delayed Recall) [38-40], language (Multilingual Naming Test and Category Fluency) [39,41,42], executive functioning (Stroop Color-Word; Verbal Fluency; Trail Making Part B) [33,36,37], and visuospatial construction (Benson Complex Figure Copy) [40].

Table 1. Promote Independent Aging study visit information.

Assessment	Baseline visit	Final study visit (month 12)
Consent and authorization forms	+[a]	N/A[b]
Demographics form	+	N/A
Socioeconomic and employment form	+	+
Physical assessment form	+	+
Mobility form	+	+
Personal and family health history	+	+
Modified Cumulative Illness Rating Scale [33,34]	+	+
Physical Activity Scale for the Elderly [43]	+	+
Pittsburg Sleep Quality Index [44]	+	+
Older Americans Resources and Services activities of daily living and IADL[c] [45]	+	+
Everyday Cognition Questionnaire self and informant [46,47]	+	+
Functional Activities Questionnaire [31]	+	+
Clinical Dementia Rating self and informant [30]	+	+
Habits form	+	+
Cognitive status form	+	+
Montreal Cognitive Assessment [29]	+	+
Neuropsychological examination (see text)	+	+
Generalized anxiety disorder 7-item [28]	+	+
Geriatric Depression Scale-15 item [27]	+	+
University of California, Los Angeles Loneliness Scale [48]	+	+
Lubben Social Network Scale [49,50]	+	+
RAND 36-Item Health Survey [51]	+	+

[a]+: measure is administered.

[b]N/A: not applicable.

[c]IADL: instrumental activities of daily living.

In-Home Activity Monitoring Platform and Installation

Daily activity data were collected using a well-established unobtrusive in-home activity assessment system installed in the home of each study participant. The in-home assessment platform is developed and managed by the ORCATECH [21]. ORCATECH is a National Institutes of Health (NIH)/NIA–funded research center that develops, implements, and supports leading-edge technologies for clinical research. The specific devices used in the study were chosen because they are included in the NIA-funded ORCATECH in-home technology assessment platform and are compatible with the ORCATECH technology infrastructure. This research platform is currently widely deployed across the United States as part of the VA and NIH CART initiative. In this study, the ORCATECH platform installation occurred within 4 weeks of the participant's baseline study visit and took place at the participant's home. Aside from the brief Web-based surveys, the devices used in this study gathered information from participants passively and did not require training or new learning. The pillbox used in the study was a standard 7-day pillbox familiar to participants and did not require formal instruction. Participants were current computer users at study entry, and no training was required for using their personal computers or completing web-based surveys. The activity tracker watch required no training, as it was worn like a regular watch. The passive nature of the data collection is critical to feasibility and long-term retention because of low participant effort and burden. Study devices (such as, pillbox, watch, and computer software, described in detail in the following sections) were purchased and maintained by ORCATECH and installed by VA research personnel. VA and ORCATECH research personnel monitored technology through the ORCATECH Management Console interface on a weekly basis to ensure that all devices were working properly. In the event of technical difficulty, research personnel repaired or replaced technology within 1 to 2 weeks. At the end of the 12 months, ORCATECH study devices were removed from participants' homes.

Hub Computer (Raspberry Pi 3 Model B, Pencoad, Wales)

The hub computer [52] received and transferred all deidentified sensor data collected at the participant's home (medication taking and fitness tracker) via a secure virtual private network

(VPN) connection to the secure ORCATECH research server. The hub computer, which is placed unobtrusively in the home, broadcasts a wireless network in the participant's home, acts as a client to a wireless or wired router, and checks in with the ORCATECH server to ensure that the in-home monitoring devices are up to date and properly identified. The in-home activity data were sent from the hub computer to the ORCATECH server on a continuous basis and was deleted from the device afterwards.

Medication Tracking Pillbox (TimerCap iSort, Moorpark, CA)

The pillbox [53] recorded timestamps of when the lids of the 7-day pillbox (one lid for each day of the week) were opened or closed and transmitted the information to the hub computer via Bluetooth Low Energy (BLE). The pillbox recorded whether or not the compartment was opened (and closed) and the time or times of day that it was opened. If the pillbox did not have a connection to the hub computer, data were cached locally to the device until the next successful connection. The pillbox caches 2 to 3 weeks of data. Data were transmitted securely to servers at ORCATECH via a VPN, and pillboxes were linked to participant ID numbers.

Wrist-Worn Fitness Tracker (Nokia Steel, Issy-les-Moulineaux, France)

The wrist-worn wearable device [54] collected physical activity data (ie, steps taken and time spent sleeping) and transmitted information to the hub computer via BLE. This device communicates data acquired several times a day to the hub computer. If the wearable device was not able to connect to the hub computer, data were cached locally to the device until the next successful connection. The wearable device caches 3 weeks of data. Data were transmitted securely to servers at ORCATECH via a VPN.

Computer Use Monitoring Software (Worktime Corporate, Woodbridge, Ontario)

Computer use monitoring software was installed on participants' own computers by VA research personnel. This commercially available software collected information about number and duration of computer sessions such as log-in/log-off times, active/idle times, and time spent on types of applications (internet and documents). Advanced Encryption Standard encrypted data (FIPS 140-2 compliant) was transmitted to ORCATECH servers via Transmission Control Protocol connection. Document, names, and Web URLs were excluded from monitoring. Worktime Corporate [55] does not record keystrokes, passwords, emails, chats, document content, or screen content. This computer software is only compatible with Windows 7, 8, 8.1, and 10. The computer software is not compatible with Mac operating system; 8 participants with Mac computers did not have Worktime Corporate installed on their personal computers. Thus, 8 participants did not have computer metrics measured in this study.

Web-Based Health Update Questionnaire (5-10 Min Per Week)

Participants received a brief, weekly Web-based 13-item health questionnaire (see Multimedia Appendix 1) that asked questions about events and behaviors that could affect in-home monitoring activity patterns (eg, medication changes, falls, injuries, health changes, emergency room visits, depression, changes to living space, vacations, and visitors) [21,56]. This survey was administered via the Qualtrics Survey Platform [57] and sent through email every Monday at 9 AM (Central Time [CT]). If a participant failed to complete the survey by Wednesday of each week, another survey was sent automatically on Wednesday at 9 AM (CT). If the participant failed to complete the follow-up survey by Thursday or Friday of each week, phone calls were made to each participant to ensure data capture and quality.

Survey for Memory Attention and Reaction Time (5-10 Min Per Month)

Participants received a monthly Web-based memory test (see Multimedia Appendix 2) called the Survey for Memory Attention and Reaction Time (SMART) [58], which included four short cognitive tasks (including versions of the Trail Making Test B and the Stroop Color-Word Interference tasks). The SMART survey was administered via the Qualtrics Survey Platform and was sent on the last Monday of each month at 9 AM (CT). If a participant failed to complete the survey by Wednesday of each week, another survey was sent automatically on Wednesday at 9 AM (CT). If the participant failed to complete the follow-up survey by Thursday or Friday of each week, phone calls were made to each participant to ensure data capture and quality.

Statistical Analyses

Recruitment and retention numbers, baseline demographic and clinical characteristics, and common technological difficulties are presented for the overall cohort as well as by study group (MCI vs intact cognition). Differences between the two study groups were assessed using 2-tailed *t* tests or the Wilcoxon rank sum test for continuous variables (depending on the distribution) or by using the Pearson chi-square test or Fisher exact test for categorical variables (depending on cell size). All summaries and analyses were performed using SPSS version 24.

Results

Participant, Recruitment, and Retention

PRIA research personnel received 201 calls to participate. A total of 150 people were screened by research personnel. People were not screened either because they did not return research personnel phone calls or because people called after study slots were filled. Of the 150 screened, the most common exclusion criteria included nonveteran status (17/150, 11.3%), living too far from the study site (14/150, 9.3%), and having exclusionary physical or mental health concerns (27/150, 18.0%). Of those enrolled, 35 participants had the full clinical assessment, and 32 participants had the research technology platform installed in their home. On average, the MCI group had 185.3 (SD 27.1; n=15) days of follow-up (range 113-213) and the cognitively

intact group had 150.9 (SD 33.8; n=15) days of follow-up in a 7-month monitoring period (range 63-205).

Following baseline evaluation, 5 participants withdrew from the study. Specifically, 2 participants withdrew before installation of the technology because they were away from the metropolitan area for half of the year without an internet connection, and 1 participant withdrew because they saw no direct benefit from this research study. Two participants withdrew following technology installation because of unanticipated acute medical events. A summary of demographic and clinical characteristics of the final sample (N=30) is presented in Table 2.

Table 2. Participant baseline demographics and clinical characteristics (N=30).

Variable	Total	Mild cognitive impairment (n=15)	Healthy controls (n=15)	*P* value
Age at baseline (years), mean (SD)	73.5 (5.4)	74.3 (6.0)	72.8 (4.9)	.46
Sex (male), n (%)	28 (93)	14 (93)	14 (93)	N/A[a]
Race (white), n (%)	30 (100)	15 (100)	15 (100)	N/A
Education (years), mean (SD)	14.9 (2.0)	14.9 (2.3)	15.0 (1.9)	.86
Montreal Cognitive Assessment [29], mean (SD)	24.5 (2.3)	22.9 (1.8)	26.1 (1.5)	<.001
Geriatric Depression Scale [27], mean (SD)	1.3 (1.3)	1.5 (1.5)	1.1 (1.1)	.42
Generalized anxiety disorder [28], mean (SD)	1.3 (1.6)	1.7 (1.7)	0.9 (1.4)	.16
Functional Activities Questionnaire [31], mean (SD)	1.0 (1.6)	1.6 (1.8)	0.4 (1.1)	.04
Modified Cumulative Illness Rating Scale [33,34], mean (SD)	23.1 (2.9)	24.0 (3.3)	22.1 (2.2)	.08
Pittsburg Sleep Quality Index [44], mean (SD)	5.7 (3.2)	6.5 (3.7)	5.0 (2.4)	.21
Everyday Cognition Questionnaire (ECog) informant [18,46,47], mean (SD)	1.3 (0.3)	1.3 (0.4)	1.2 (0.3)	.39
ECog participant [18,46,47], mean (SD)	1.4 (0.5)	1.5 (0.7)	1.3 (0.2)	.17
Physical Activity Scale for the Elderly [43], mean (SD)	157.8 (57.9)	172.3 (66.5)	143.3 (45.5)	.18
Health comorbidities (positive for the condition), n (%)				
Atrial fibrillation	4 (13)	4 (27)	0 (0)	.03
Diabetes	5 (17)	2 (13)	3 (20)	.62
Hypertension	19 (63)	12 (80)	7 (47)	.06
Hypercholesterolemia	24 (80)	13 (87)	11 (73)	.36
Sleep apnea	17 (57)	9 (60)	8 (53)	.71

[a]N/A: not applicable.

Common Technical Difficulties

Hub Computer (Raspberry Pi 3 Model B)

The most common technical difficulty associated with the hub computer was the loss of connection to the hub computer, requiring research personnel to update the hub computer remotely. However, sometimes these updates needed to be done manually. For example, an unexpected license expiration lead to a temporary VPN handshake failure, which affected all 30 participants and delayed data capture for some participants up to 1 month. This VPN handshake failure could not be fixed remotely, and research personnel were required to update the hub computer manually across all 30 homes. Over a 7-month monitoring period, 13% (4/30) of participants in the entire sample required two or more in-home technology maintenance visits to repair the hub computer; all 4 were MCI participants, *P*=.10 (see Table 3).

Table 3. Technology device repairs and reminder phone calls over a 7-month monitoring period (N=30).

Variable	Total sample	Mild cognitive impairment (n=15)	Healthy controls (n=15)	*P* value
Total device repair visits, mean (SD)	2.4 (1.7)	3.1 (1.9)	1.8 (1.2)	.04[a]
Participants requiring >2 hub computer visit, n (%)	4 (13)	4 (27)	0 (0)	.10
Participants requiring >1 pillbox visit, n (%)	7 (23)	6 (40)	1 (7)	.08
Participants requiring >1 watch visit, n (%)	8 (27)	5 (33)	3 (20)	.68
Participants requiring >1 worktime visit, n (%)	11 (37)	6 (40)	5 (33)	>.99
Total reminder phone calls, mean (SD)	2.7 (2.9)	2.9 (3.2)	2.5 (2.8)	.67
Participants requiring >1 Web-based health questionnaire reminder call, n (%)	19 (63)	12 (80)	7 (47)	.13
Participants requiring >1 Survey for Memory Attention and Reaction Time survey reminder call, n (%)	18 (60)	8 (53)	10 (67)	.71

[a]Group comparisons were made using independent *t* tests for continuous variables or Fisher exact tests (2-tailed) for categorical variables.

Medication Tracking Pillbox (TimerCap iSort)

The most common technical difficulties associated with the pillbox included battery issues (which required a battery change) and batteries falling out of the pillbox (which could be prevented by taping the battery door shut with a piece of masking tape). Furthermore, broken compartment lids and broken contact pieces within the instrumented pillbox required research personnel to replace the lids or the pillbox altogether. Finally, there were issues associated with syncing the pillbox to the hub computer, which required research personnel to reset the Pi's BLE connection. Over a 7-month monitoring period, 23% (7/30) of participants in the entire sample required one or more in-home technology maintenance visits to repair the instrumented pillbox; 6 were MCI participants, and 1 was a healthy control participant; *P*=.08 (see Table 3).

Wrist-Worn Fitness Tracker Watch (Nokia Steel)

The most common technical difficulties associated with the watch included low batteries within the watch, which required a battery change (batteries should last at least six months) as well as broken sensors within the watch, which required research personnel to replace the watch altogether. Furthermore, three watch faces were broken, which required research personnel to replace the watch. Finally, there were issues associated with synchronizing the watch to the hub computer, which required research personnel to reset the Pi's BLE connection. Over a 7-month monitoring period, 27% (8/30) of participants in the entire sample required one or more in-home technology maintenance visits to repair the fitness tracker watch: 5 were MCI participants and 3 were healthy control participants; *P*=.68 (see Table 3).

Computer Use Monitoring Software (Worktime Corporate)

The most common technical difficulties associated with the computer use monitoring software was the removal of Worktime Corporate by malware detection and prevention programs installed on the participant's computer requiring research personnel to redownload the software on participants' computers. Furthermore, it came to our attention that Worktime Corporate was not compatible with certain versions of Mac computers as well as tablets or mobile phones. Thus, Worktime Corporate was only installed in 22 homes because 8 participants owned Mac devices. Two participants were married and shared the same computer with separate log-on accounts. Over a 7-month monitoring period, 36% (11/30) participants in the entire sample required one or more in-home technology maintenance visits to repair Worktime software: 6 were MCI participants and 5 were healthy control participants; *P*>.99 (see Table 3).

Web-Based Health Update Questionnaire

The most common technical difficulties associated with the Web-based health update questionnaire was that the email containing a link to the questionnaire would sometimes go to a participant's Spam folder. This issue was mitigated by asking participants to save the email address in their contacts (or showing them how to do so). Furthermore, some research participants would complete the survey all the way through and then receive a message that they had failed to complete the survey. This would require participants to fill out the same survey twice. If participants failed to fill out their questionnaire by Thursday of each week, they would receive a reminder phone call from research personnel. Over a 7-month monitoring period, 63% (19/30) of participants in the entire sample required one or more reminder phone calls to complete the Web-based health questionnaire: 12 were MCI participants and 9 were healthy control participants; *P*=.13 (see Table 3).

Survey for Memory Attention and Reaction Time

The most common technical difficulties associated with the SMART survey was screen freezing. The freezing during a participant's session was most often related to a portion of code that transferred mouse activity data back to the ORCATECH servers. Research staff discovered that the transfers were being called multiple times on each task and eventually overloaded the browser with unnecessary transfers. This issue was mitigated by having a programmer apply logic to check if there were transfers in progress, which ultimately reduced the frequency of transfers. Furthermore, although the SMART survey was compatible with every internet browser, operating system, and device, participants had trouble completing the SMART survey on their mobile phone. A total of 24 people completed the

SMART survey on a computer, 4 participants used their mobile phone to complete the survey, and 2 participants used their tablet. Of note, the 2 tablet users were in the MCI group and 3 of the 4 mobile phone users were in the MCI group. Over a 7-month monitoring period, 60% (18/30) of participants in the entire sample required one or more reminder phone calls to complete the Web-based SMART survey: 8 were MCI participants and 10 were healthy control participants; *P*=.71 (see Table 3).

Discussion

Principal Findings

The results in this study demonstrate the feasibility of engaging older adult military veterans in an observational research study using in-home IADL monitoring technology. Recruitment was successful as evidenced by a high number of phone calls received from older adults to participate and study enrollment completed within 6 months. Retention was high; 86% (26/30) of individuals who were initially consented were retained 7 months into the study follow-up period. Five military veterans dropped out from the study; only one dropped out because of study-specific concerns. The 2 individuals who withdrew because of unanticipated medical events expressed interest in remaining in the study and continue contributing. Furthermore, frequency of in-home technology maintenance visits was relatively low across the sample given the duration of monitoring follow-up, as shown in Table 3. Overall, technology maintenance visits were more prevalent for MCI participants (*P*=.04) than healthy controls, although reminder telephone calls to complete Web-based surveys were not (*P*=.67). One possible explanation for this finding is that when the devices (eg, watch and pillbox) in the study lost internet connection with the hub computer because of software or firmware updates, healthy control participants were better able to successfully troubleshoot and problem solve technological difficulties remotely over the phone with research staff compared with MCI participants, preventing research staff from making additional trips to repair devices. In contrast, because of their MCIs, MCI participants may have been less successful in understanding and retaining complex instructions over the phone and organizing and executing device repairs or resets requiring multiple steps without in-person assistance. Troubleshooting device and internet connection issues is a cognitively complex task that requires executive functions, memory, language, visual spatial abilities, and processing speed. Compared with the devices used in the study, the Web-based surveys required less troubleshooting by participants. In general, once participants initiated taking the Web-based surveys by clicking a link in their email, the survey software program worked well. Our results indicate/suggest that the sample of MCI participants remembered to take the Web-based surveys as reliably as healthy controls, but they had more difficulty carrying out cognitively complex tasks with higher executive demands such as troubleshooting device resets and repairs as well as healthy controls.

Various factors could be related to the high rates of retention and participants' motivation to participate in this kind of research. First, real-world assessment research allows for rapport and relationship-building between research staff and study participants from the onset of the study. Real-world assessment research requires the installation of technology in the participants' home, as compared with scenario-based assessment research where the interaction occurs in a clinic-based setting [24]. Research staff are welcomed into the homes of aging military veterans and able to develop fulfilling relationships with research participants. Real-world assessment research also facilitates a deeper understanding about a person's needs and daily functioning compared with conducting a clinical interview in a clinic-based setting. Developing this unique rapport/relationship with research participants started at the baseline study visit where research staff met with study participants and their study informant (usually a spouse, close family member, or friend). The information acquired from the participant and informant during the study visit created a holistic picture of each participant, allowing research personnel to generate a multifaceted view of each person. Measures such as the CDR [31] and PASE [43] include questions about the participant's hobbies, physical activities, family, and about recent events that happen in their lives. These questions allowed research personnel to connect to the study informant and the study participant and develop trust and familiarity before they entered the participant's home.

Strengths and Limitations

The overall experience with the ORCATECH platform was positive. Research technical staff were able to learn quickly how to install and trouble shoot technology, and research participants were generally receptive to the technology. However, as is common when working with diverse technologies, there were some limitations particularly with commercially available devices and software integrated into the ORCATECH platform. For example, 8 older adults in our study have macOS-based computers so Worktime Corporate [55] could not be installed, thus preventing usage data capture. Furthermore, as Worktime Corporate can only be installed on traditional computers, we failed to capture information on participant's mobile phones or tablets. In future studies, monitoring software should be developed that is compatible across all platforms, operating systems, and devices so that all data are captured. Other limitations include the durability of the commercially available devices used in the study. The pillboxes had limited durability, which required research personnel to replace pillbox lids and pillboxes frequently. Furthermore, watches had to be replaced because of broken watch faces (3/30, 10% of participants over a span of 7 months). In contrast, across other ORCATECH studies (including CART) involving over 200 primarily nonveteran, female aging participants, 2.0% (5/250) participants have broken the face of their watch over a span of multiple years. Other limitations included the homogeneity of our sample (30/30, 100% white; 28/30, 93% male), non-inclusion of military veterans who lived outside of the metropolitan area (eg, rural populations), and noninclusion of military veterans who lived in another home during the winter (*snow birds)*. This study was funded by the Veterans Health Administration, which required recruitment of military veterans only. The older adult military veteran population in Minnesota,

where the study was conducted, is largely represented by white men [59]. Future studies will expand recruitment to nonveteran samples that are more heterogeneous and balanced regarding gender and race. Despite the homogeneity of our cohort, there was diversity within this sample regarding the cohort's varying health comorbidities (mean MCIRS score 23.1, SD 2.9) and educational attainment. Duration of education ranged from 9 to 18 years (mean education 14.9, SD 2.0). Other in-home technology and aging studies report higher educational levels and lower MCIRS levels in their cohorts; thus, this research helps to increase our knowledge about the feasibility of in-home technologies in a wider array of older adults [16,19,21].

Future Directions and Conclusions

Future analyses will establish which sensor monitored IADL variables best discriminate between MCI and healthy controls and will explore trajectories of IADL functioning in MCI and healthy controls. Few studies [22,60] discuss the acceptability of the technology used over time, especially from the participant's perspective. Future directions will include deployment of an in-home monitoring technology perception survey to the military veterans in this cohort. This survey will allow us to capture participants' perceptions about the in-home technology they have been using in the study as well as their perceptions about how they would want to use in-home monitoring technology in the future. Furthermore, very few studies [24] have explored ethics related to in-home monitoring which is important to discuss to implement these technologies in people's homes. This survey will explore ethical issues associated with real-world assessment research such as privacy, security concerns, and who should have access to the data. This will give us a better idea of how comfortable older adult military veterans feel with the usage of their activity data in the future. Other future directions include using this technology to extend beyond IADL and cognitive assessment to help identify people at risk for acute medical events (stroke and infections). This research study also was able to install technology and monitor a participant's activity in a rehabilitation center following an unanticipated medical event (data not shown), demonstrating the flexibility of the methodology. Future studies should investigate the feasibility of incorporating these methodologies in a variety of health care settings outside of the home, including inpatient medical facilities when higher levels of care for individuals are temporarily required. Future studies using activity monitoring technologies to detect and monitor cognitive decline should aim to increase inclusion of individuals who are typically underrepresented in aging and dementia research such as younger older adults, people of color, adults living in rural communities, and individuals with low socioeconomic status.

Knowledge gained from this pilot study will be used to help develop acceptable and effective home-based assessment tools that can be used within the VA system to monitor cognition and daily functioning in aging military veterans. The results and lessons provided by this study have importantly been incorporated into improving the national CART initiative platform, which has now been deployed to diverse cohorts of older adults including rural-residing military veterans.

Acknowledgments

The authors would like to thank their colleagues, research assistants, technicians, programmers, analysts, and volunteers who invited them into their homes and lives. This work was supported by pilot project CX001669-01 from the US Department of Veterans Affairs Clinical Science Research and Development Service The contents do not represent the views of the US Department of Veterans Affairs or the United States Government. This research was also supported in part by the Collaborative Aging-in-place Research Using Technology initiative (NIH grant U2C AG0543701; Veteran's Administration grant IIR 17-144), the Oregon Roybal Center for Translational Research on Aging (NIA P30 AG024978), and the Oregon Clinical Translational Research Institute Clinical and Translational Science Award (NCATS UL1 TR002369). Some data were collected and managed using REDCap electronic data capture tools hosted at Oregon Health and Science University [61]. The Collaborative Aging-in-place Research Using Technology initiative is funded by the Office of the Director, NIH, National Center for Advancing Translational Sciences, National Institute of Biomedical Imaging and Bioengineering, National Institute of Nursing Research, NIA, National Institute of Neurological Disorders and Stroke, National Cancer Institute, and the Veteran's Administration.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sample questions for the web-based health update questionnaire.
[DOCX File , 42 KB - formative_v4i6e16371_app1.docx]

Multimedia Appendix 2
Opening page of the Survey for Memory, Attention, Responding, and Thinking.
[DOCX File , 67 KB - formative_v4i6e16371_app2.docx]

References

1. Alzheimer's Disease International. London: Alzheimer's Disease International; 2013. Policy Brief for Heads of Government: The Global Impact of Dementia 2013-2050 URL: https://www.alz.co.uk/research/GlobalImpactDementia2013.pdf [accessed 2020-03-06]
2. Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, Ferri CP. The global prevalence of dementia: a systematic review and metaanalysis. Alzheimers Dement 2013 Jan;9(1):63-75.e2. [doi: 10.1016/j.jalz.2012.11.007] [Medline: 23305823]
3. Weiner MW, Friedl KE, Pacifico A, Chapman JC, Jaffee MS, Little DM, et al. Military risk factors for Alzheimer's disease. Alzheimers Dement 2013 Jul;9(4):445-451 [FREE Full text] [doi: 10.1016/j.jalz.2013.03.005] [Medline: 23809365]
4. Sibener L, Zaganjor I, Snyder HM, Bain LJ, Egge R, Carrillo MC. Alzheimer's Disease prevalence, costs, and prevention for military personnel and veterans. Alzheimers Dement 2014 Jun;10(3 Suppl):S105-S110 [FREE Full text] [doi: 10.1016/j.jalz.2014.04.011] [Medline: 24924663]
5. Vincent AS, Roebuck-Spencer TM, Cernich A. Cognitive changes and dementia risk after traumatic brain injury: implications for aging military personnel. Alzheimers Dement 2014 Jun;10(3 Suppl):S174-S187 [FREE Full text] [doi: 10.1016/j.jalz.2014.04.006] [Medline: 24924669]
6. Mohlenhoff BS, O'Donovan A, Weiner MW, Neylan TC. Dementia risk in posttraumatic stress disorder: the relevance of sleep-related abnormalities in brain structure, amyloid, and inflammation. Curr Psychiatry Rep 2017 Oct 16;19(11):89 [FREE Full text] [doi: 10.1007/s11920-017-0835-1] [Medline: 29035423]
7. Yu W, Ravelo A, Wagner TH, Phibbs CS, Bhandari A, Chen S, et al. Prevalence and costs of chronic conditions in the VA health care system. Med Care Res Rev 2003 Sep;60(3 Suppl):146S-167S. [doi: 10.1177/1077558703257000] [Medline: 15095551]
8. Yaffe K, Hoang TD, Byers AL, Barnes DE, Friedl KE. Lifestyle and health-related risk factors and risk of cognitive aging among older veterans. Alzheimers Dement 2014 Jun;10(3 Suppl):S111-S121 [FREE Full text] [doi: 10.1016/j.jalz.2014.04.010] [Medline: 24924664]
9. Keane J, Meier JL, Noth RH, Swislocki AL. Computer-based screening of veterans for metabolic syndrome. Metab Syndr Relat Disord 2009 Dec;7(6):557-561. [doi: 10.1089/met.2009.0021] [Medline: 19558266]
10. Alzheimer's Association. 2019 Alzheimer's disease facts and figures. Alzheimers Dement 2019;15(3):321-387. [doi: 10.1016/j.jalz.2019.01.010]
11. Jutkowitz E, Kane RL, Gaugler JE, MacLehose RF, Dowd B, Kuntz KM. Societal and family lifetime cost of dementia: implications for policy. J Am Geriatr Soc 2017 Oct;65(10):2169-2175 [FREE Full text] [doi: 10.1111/jgs.15043] [Medline: 28815557]
12. Villemagne VL, Burnham S, Bourgeat P, Brown B, Ellis KA, Salvado O, Australian Imaging Biomarkers and Lifestyle (AIBL) Research Group. Amyloid β deposition, neurodegeneration, and cognitive decline in sporadic Alzheimer's disease: a prospective cohort study. Lancet Neurol 2013 Apr;12(4):357-367. [doi: 10.1016/S1474-4422(13)70044-9] [Medline: 23477989]
13. Petersen RC. Mild cognitive impairment as a diagnostic entity. J Intern Med 2004 Sep;256(3):183-194 [FREE Full text] [doi: 10.1111/j.1365-2796.2004.01388.x] [Medline: 15324362]
14. Jekel K, Damian M, Wattmo C, Hausner L, Bullock R, Connelly PJ, et al. Mild cognitive impairment and deficits in instrumental activities of daily living: a systematic review. Alzheimers Res Ther 2015;7(1):17 [FREE Full text] [doi: 10.1186/s13195-015-0099-0] [Medline: 25815063]
15. Seelye A, Hagler S, Mattek N, Howieson DB, Wild K, Dodge HH, et al. Computer mouse movement patterns: A potential marker of mild cognitive impairment. Alzheimers Dement (Amst) 2015 Dec 1;1(4):472-480 [FREE Full text] [doi: 10.1016/j.dadm.2015.09.006] [Medline: 26878035]
16. Seelye A, Mattek N, Howieson DB, Austin D, Wild K, Dodge HH, et al. Embedded online questionnaire measures are sensitive to identifying mild cognitive impairment. Alzheimer Dis Assoc Disord 2016;30(2):152-159 [FREE Full text] [doi: 10.1097/WAD.0000000000000100] [Medline: 26191967]
17. Chen Y, Denny KG, Harvey D, Farias ST, Mungas D, DeCarli C, et al. Progression from normal cognition to mild cognitive impairment in a diverse clinic-based and community-based elderly cohort. Alzheimers Dement 2017 Apr;13(4):399-405 [FREE Full text] [doi: 10.1016/j.jalz.2016.07.151] [Medline: 27590706]
18. Farias ST, Lau K, Harvey D, Denny KG, Barba C, Mefford AN. Early functional limitations in cognitively normal older adults predict diagnostic conversion to mild cognitive impairment. J Am Geriatr Soc 2017 Jun;65(6):1152-1158 [FREE Full text] [doi: 10.1111/jgs.14835] [Medline: 28306147]
19. Seelye A, Mattek N, Sharma N, Witter P, Brenner A, Wild K, et al. Passive assessment of routine driving with unobtrusive sensors: a new approach for identifying and monitoring functional level in normal aging and mild cognitive impairment. J Alzheimers Dis 2017;59(4):1427-1437 [FREE Full text] [doi: 10.3233/JAD-170116] [Medline: 28731434]
20. Gold DA. An examination of instrumental activities of daily living assessment in older adults and mild cognitive impairment. J Clin Exp Neuropsychol 2012;34(1):11-34. [doi: 10.1080/13803395.2011.614598] [Medline: 22053873]
21. Kaye JA, Maxwell SA, Mattek N, Hayes TL, Dodge H, Pavel M, et al. Intelligent Systems For Assessing Aging Changes: home-based, unobtrusive, and continuous assessment of aging. J Gerontol B Psychol Sci Soc Sci 2011 Jul;66(Suppl 1):i180-i190 [FREE Full text] [doi: 10.1093/geronb/gbq095] [Medline: 21743050]

22. Lyons BE, Austin D, Seelye A, Petersen J, Yeargers J, Riley T, et al. Pervasive computing technologies to continuously assess Alzheimer's disease progression and intervention efficacy. Front Aging Neurosci 2015;7:102 [FREE Full text] [doi: 10.3389/fnagi.2015.00102] [Medline: 26113819]
23. Bondi MW, Smith GE. Mild cognitive impairment: a concept and diagnostic entity in need of input from neuropsychology. J Int Neuropsychol Soc 2014 Feb;20(2):129-134 [FREE Full text] [doi: 10.1017/S1355617714000010] [Medline: 24490866]
24. Lussier M, Lavoie M, Giroux S, Consel C, Guay M, Macoir J, et al. Early detection of mild cognitive impairment with in-home monitoring sensor technologies using functional measures: a systematic review. IEEE J Biomed Health Inform 2019;23(2):838-847. [doi: 10.1109/jbhi.2018.2834317]
25. Lussier M, Adam S, Chikhaoui B, Consel C, Gagnon M, Gilbert B, et al. Smart home technology: a new approach for performance measurements of activities of daily living and prediction of mild cognitive impairment in older adults. J Alzheimers Dis 2019;68(1):85-96. [doi: 10.3233/JAD-180652] [Medline: 30775978]
26. Austin J, Klein K, Mattek N, Kaye J. Variability in medication taking is associated with cognitive performance in nondemented older adults. Alzheimers Dement (Amst) 2017;6:210-213 [FREE Full text] [doi: 10.1016/j.dadm.2017.02.003] [Medline: 28349120]
27. Oregon Health & Science University. Collaborative Aging Research Using Technology (CART) URL: https://www.ohsu.edu/collaborative-aging-research-using-technology [accessed 2020-03-04]
28. Yesavage JA, Brink T, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: a preliminary report. J Psychiatr Res 1982;17(1):37-49. [doi: 10.1016/0022-3956(82)90033-4] [Medline: 7183759]
29. 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]
30. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc 2005 Apr;53(4):695-699. [doi: 10.1111/j.1532-5415.2005.53221.x] [Medline: 15817019]
31. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 1993 Nov;43(11):2412-2414. [doi: 10.1212/wnl.43.11.2412-a] [Medline: 8232972]
32. Pfeffer RI, Kurosaki TT, Harrah CH, Chance JM, Filos S. Measurement of functional activities in older adults in the community. J Gerontol 1982 May;37(3):323-329. [doi: 10.1093/geronj/37.3.323] [Medline: 7069156]
33. Morris JC, Weintraub S, Chui HC, Cummings J, Decarli C, Ferris S, et al. The Uniform Data Set (UDS): clinical and cognitive variables and descriptive data from Alzheimer Disease Centers. Alzheimer Dis Assoc Disord 2006;20(4):210-216. [doi: 10.1097/01.wad.0000213865.09806.92] [Medline: 17132964]
34. Linn BS, Linn MW, Gurel L. Cumulative illness rating scale. J Am Geriatr Soc 1968 May;16(5):622-626. [doi: 10.1111/j.1532-5415.1968.tb02103.x] [Medline: 5646906]
35. Miller MD, Paradis CF, Houck PR, Mazumdar S, Stack JA, Rifai A, et al. Rating chronic medical illness burden in geropsychiatric practice and research: application of the Cumulative Illness Rating Scale. Psychiatry Res 1992 Mar;41(3):237-248. [doi: 10.1016/0165-1781(92)90005-n] [Medline: 1594710]
36. Reitan RM. The relation of the trail making test to organic brain damage. J Consult Psychol 1955 Oct;19(5):393-394. [doi: 10.1037/h0044509] [Medline: 13263471]
37. Stroop JR. Studies of interference in serial verbal reactions. J Exp Psychol 1935;18(6):643-662. [doi: 10.1037/h0054651]
38. Craft S, Newcomer J, Kanne S, Dagogo-Jack S, Cryer P, Sheline Y, et al. Memory improvement following induced hyperinsulinemia in Alzheimer's disease. Neurobiol Aging 1996;17(1):123-130. [doi: 10.1016/0197-4580(95)02002-0] [Medline: 8786794]
39. Morris JC, Heyman A, Mohs RC, Hughes JP, van Belle G, Fillenbaum G, et al. The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part I. Clinical and neuropsychological assessment of Alzheimer's disease. Neurology 1989 Sep;39(9):1159-1165. [doi: 10.1212/wnl.39.9.1159] [Medline: 2771064]
40. Possin KL, Laluz VR, Alcantar OZ, Miller BL, Kramer JH. Distinct neuroanatomical substrates and cognitive mechanisms of figure copy performance in Alzheimer's disease and behavioral variant frontotemporal dementia. Neuropsychologia 2011 Jan;49(1):43-48 [FREE Full text] [doi: 10.1016/j.neuropsychologia.2010.10.026] [Medline: 21029744]
41. Ivanova I, Salmon DP, Gollan TH. The multilingual naming test in Alzheimer's disease: clues to the origin of naming impairments. J Int Neuropsychol Soc 2013 Mar;19(3):272-283 [FREE Full text] [doi: 10.1017/S1355617712001282] [Medline: 23298442]
42. Gollan TH, Weissberger GH, Runnqvist E, Montoya RI, Cera CM. Self-ratings of Spoken Language Dominance: A Multi-Lingual Naming Test (MINT) and Preliminary Norms for Young and Aging Spanish-English Bilinguals. Biling (Camb Engl) 2012 Jul;15(3):594-615 [FREE Full text] [doi: 10.1017/S1366728911000332] [Medline: 25364296]
43. Washburn RA, Smith KW, Jette AM, Janney CA. The Physical Activity Scale for the Elderly (PASE): development and evaluation. J Clin Epidemiol 1993 Feb;46(2):153-162. [doi: 10.1016/0895-4356(93)90053-4] [Medline: 8437031]
44. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res 1989 May;28(2):193-213. [doi: 10.1016/0165-1781(89)90047-4] [Medline: 2748771]

45. Doble S, Fisher A. The dimensionality and validity of the Older Americans Resources and Services (OARS) Activities of Daily Living (ADL) Scale. J Outcome Meas 1998;2(1):4-24. [Medline: 9661728]
46. Farias ST, Mungas D, Reed BR, Cahn-Weiner D, Jagust W, Baynes K, et al. The measurement of everyday cognition (ECog): scale development and psychometric properties. Neuropsychology 2008 Jul;22(4):531-544 [FREE Full text] [doi: 10.1037/0894-4105.22.4.531] [Medline: 18590364]
47. Farias ST, Mungas D, Harvey DJ, Simmons A, Reed BR, Decarli C. The measurement of everyday cognition: development and validation of a short form of the Everyday Cognition scales. Alzheimers Dement 2011 Nov;7(6):593-601 [FREE Full text] [doi: 10.1016/j.jalz.2011.02.007] [Medline: 22055976]
48. Russell D, Peplau LA, Ferguson ML. Developing a measure of loneliness. J Pers Assess 1978 Jun;42(3):290-294. [doi: 10.1207/s15327752jpa4203_11] [Medline: 660402]
49. Lubben JE. Assessing social networks among elderly populations. Fam Commun Health 1988;11(3):42-52. [doi: 10.1097/00003727-198811000-00008]
50. Lubben J, Blozik E, Gillmann G, Iliffe S, von Renteln Kruse W, Beck JC, et al. Performance of an abbreviated version of the Lubben Social Network Scale among three European community-dwelling older adult populations. Gerontologist 2006 Aug;46(4):503-513. [doi: 10.1093/geront/46.4.503] [Medline: 16921004]
51. Hays RD, Sherbourne CD, Mazel RM. The RAND 36-Item Health Survey 1.0. Health Econ 1993 Oct;2(3):217-227. [doi: 10.1002/hec.4730020305] [Medline: 8275167]
52. Raspberry Pi. 2019. Raspberry Pi 3 Model B URL: https://www.raspberrypi.org/products/raspberry-pi-3-model-b/
53. TimerCap LLC. 2019. URL: https://www.timercap.com/ [accessed 2020-03-03]
54. Withings. 2019. URL: https://www.withings.com/ [accessed 2020-03-03]
55. WorkTime. 2019. URL: https://www.worktime.com/ [accessed 2020-03-03]
56. Seelye A, Mattek N, Sharma N, Riley T, Austin J, Wild K, et al. Weekly observations of online survey metadata obtained through home computer use allow for detection of changes in everyday cognition before transition to mild cognitive impairment. Alzheimers Dement 2018 Feb;14(2):187-194 [FREE Full text] [doi: 10.1016/j.jalz.2017.07.756] [Medline: 29107052]
57. Qualtrics. 2019. URL: https://www.qualtrics.com/education/ [accessed 2020-03-03]
58. Seelye A, Mattek N, Reynolds C, Sharma N, Wild K, Kaye J. The Survey for Memory, Attention, and Reaction Time (SMART): A Brief Online Personal Computing-Based Cognitive Assessment for Healthy Aging and Mild Cognitive Impairment. 2016 Presented at: Alzheimer's Association International Conference; 2017; London, England URL: https://alz-journals.onlinelibrary.wiley.com/doi/10.1016/j.jalz.2017.07.235 [doi: 10.1016/j.jalz.2017.07.235]
59. Minneapolis VA. 2018. Minnesota Department of Veterans Affairs 2018 Annual Report URL: https://mn.gov/mdva/assets/2018-mdva-annual-report_tcm1066-365087.pdf [accessed 2020-03-03]
60. Holthe T, Halvorsrud L, Karterud D, Hoel K, Lund A. Usability and acceptability of technology for community-dwelling older adults with mild cognitive impairment and dementia: a systematic literature review. Clin Interv Aging 2018;13:863-886 [FREE Full text] [doi: 10.2147/CIA.S154717] [Medline: 29765211]
61. 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. [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]

Abbreviations

AD: Alzheimer disease
BLE: Bluetooth Low Energy
CART: Collaborative Aging Research using Technology
CT: Central Time
IADL: instrumental activities of daily living
IRB: institutional review board
MCI: mild cognitive impairment
MCIRS: Modified Cumulative Illness Rating Scale
MVAHCS: Minneapolis Veterans Affairs Health Care System
NIA: National Institute on Aging
NIH: National Institutes of Health
ORCATECH: Oregon Center for Aging and Technology
PRIA: Promote Independent Aging
SMART: Survey for Memory Attention and Reaction Time
VA: Veterans Affairs
VPN: virtual private network

Edited by G Eysenbach; submitted 27.09.19; peer-reviewed by P Urwyler, M Gately, J Vidal-Alaball; comments to author 28.11.19; revised version received 10.01.20; accepted 04.02.20; published 08.06.20.

Please cite as:
Seelye A, Leese MI, Dorociak K, Bouranis N, Mattek N, Sharma N, Beattie Z, Riley T, Lee J, Cosgrove K, Fleming N, Klinger J, Ferguson J, Lamberty GJ, Kaye J
Feasibility of In-Home Sensor Monitoring to Detect Mild Cognitive Impairment in Aging Military Veterans: Prospective Observational Study
JMIR Form Res 2020;4(6):e16371
URL: https://formative.jmir.org/2020/6/e16371
doi:10.2196/16371
PMID:32310138

Original Paper

A Digital Smoking Cessation Program for Heavy Drinkers: Pilot Randomized Controlled Trial

Christopher W Kahler[1], PhD; Amy M Cohn[2], PhD; Catherine Costantino[3], MA; Benjamin A Toll[4], PhD; Nichea S Spillane[5], PhD; Amanda L Graham[6,7], PhD

[1]Department of Behavioral and Social Sciences, Center for Alcohol and Addiction Studies, Brown University School of Public Health, Providence, RI, United States

[2]University of Oklahoma Health Sciences Center, Oklahoma City, OK, United States

[3]Brown University School of Public Health, Providence, RI, United States

[4]Medical University of South Carolina, Charleston, SC, United States

[5]Department of Psychology, University of Rhode Island, South Kingston, RI, United States

[6]Innovations Center, Truth Initiative, Washington, DC, United States

[7]Georgetown University Medical Center, Lombardi Comprehensive Cancer Center, Washington, DC, United States

Corresponding Author:
Christopher W Kahler, PhD
Department of Behavioral and Social Sciences
Center for Alcohol and Addiction Studies
Brown University School of Public Health
Box G-S121-4
Providence, RI
United States
Phone: 1 4018636651
Email: Christopher_Kahler@brown.edu

Abstract

Background: Heavy drinking (HD) is far more common among smokers compared with nonsmokers and interferes with successful smoking cessation. Alcohol-focused smoking cessation interventions delivered by counselors have shown promise, but digital versions of these interventions—which could have far greater population reach—have not yet been tested.

Objective: This pilot randomized controlled trial aimed to examine the feasibility, acceptability, and effect sizes of an automated digital smoking cessation program that specifically addresses HD using an interactive web-based intervention with an optional text messaging component.

Methods: Participants (83/119, 69.7% female; 98/119, 82.4% white; mean age 38.0 years) were daily smokers recruited on the web from a free automated digital smoking cessation program (BecomeAnEX.org, *EX*) who met the criteria for HD: women drinking 8+ drinks/week or 4+ drinks on any day and men drinking 15+ drinks/week or 5+ drinks on any day. Participants were randomized to receive EX with standard content (EX-S) or an EX with additional content specific to HD (EX-HD). Outcomes were assessed by web-based surveys at 1 and 6 months.

Results: Participants reported high satisfaction with the website and the optional text messaging component. Total engagement with both EX-S and EX-HD was modest, with participants visiting the website a median of 2 times, and 52.9% of the participants enrolled to receive text messages. Participants in both the conditions showed substantial, significant reductions in drinking across 6 months of follow-up, with no condition effects observed. Although smoking outcomes tended to favor EX-HD, the condition effects were small and nonsignificant. A significantly smaller proportion of participants in EX-HD reported having a lapse back to smoking when drinking alcohol (7/58, 16%) compared with those in EX-S (18/61, 41%; χ^2_1=6.2; *P*=.01).

Conclusions: This is the first trial to examine a digital smoking cessation program tailored to HD smokers. The results provide some initial evidence that delivering such a program is feasible and may reduce the risk of alcohol-involved smoking lapses. However, increasing engagement in this and other web-based interventions is a crucial challenge to address in future work.

Trial Registration: ClinicalTrials.gov NCT03068611; https://clinicaltrials.gov/ct2/show/NCT03068611

(JMIR Form Res 2020;4(6):e7570) doi:10.2196/formative.7570

KEYWORDS

smoking cessation; alcohol drinking; internet; text messaging; therapy

Introduction

Background

Cigarette smokers have substantially higher rates of alcohol consumption than nonsmokers [1,2], and both observational [3,4] and clinical studies [5-8] find that greater alcohol use predicts a reduced odds of smoking cessation. Alcohol use is a common smoking relapse precipitant [9,10], with over one-third of heavy drinkers reporting alcohol use at the time of a smoking lapse, often in combination with being around other people who are smoking and with experiencing positive affect [11,12]. Laboratory-based experiments show that alcohol has a direct pharmacologic effect on increasing the urge to smoke, which is beyond the effects of alcohol-related cues (ie, beyond the learned association between drinking and smoking) [13]. This increase in the urge to smoke accounts, in part, for alcohol's effect on smoking. Heavy drinking (HD) is associated with a substantially greater risk of smoking lapse compared with moderate drinking [11]. Experimental research indicates that this effect of HD is not due to dose-dependent effects of alcohol on the urge to smoke and may instead reflect changes in regulatory control due to intoxication [13].

Given the deleterious impact of alcohol use on smoking cessation, a number of behavioral interventions have been developed specifically to address HD among HD smokers who are seeking smoking cessation treatment. In the first trial of its kind, Kahler et al [14] found that a smoking cessation treatment designed to motivate reduction of alcohol consumption and mitigate the risk of alcohol-involved smoking lapses resulted in significantly reduced alcohol consumption over 6 months, with small positive effects on smoking cessation compared with standard cessation treatment. Toll et al [15] incorporated a brief intervention to reduce alcohol use in HD smokers calling a state quitline, which resulted in significantly higher rates of smoking abstinence at 7 months compared with standard quitline counseling, with a trend toward reduced HD. In a pilot efficacy trial, Ames et al [16] found that an integrated smoking and alcohol intervention, compared with a smoking-only intervention, resulted in somewhat greater cigarette smoking abstinence and significantly reduced alcohol use among treatment completers. Finally, Correa-Fernandez et al [17] found that a combined alcohol and smoking intervention did not significantly improve smoking cessation outcomes compared with a smoking-only intervention but significantly reduced drinking in those who successfully quit smoking. Together, these studies highlight the potential value of interventions that simultaneously address smoking and HD and the positive public health impact such combined interventions may have if they have sufficient population reach.

Internet-based smoking cessation interventions have great potential to reduce smoking rates at the population level, given their demonstrated reach and effectiveness. In 2017, 36% of US smokers searched the web for information about quitting smoking [18], and hundreds of thousands register on web-based cessation programs each year [18,19]. Evidence-based internet interventions deliver the core components of cessation treatment through engaging, multimodal formats, oftentimes free of charge; are scalable and cost-efficient; and yield quit rates comparable with in-person and phone-based interventions [20,21]. However, across smoking cessation websites, alcohol is only minimally addressed [22,23]. Given that one-fourth to one-third of smokers seeking cessation services are heavy drinkers [5,24,25], enhancing internet-based smoking cessation interventions with alcohol-related content could have a substantial population impact.

Objectives

Despite the potential promise and importance of addressing HD in smoking cessation treatment, models for incorporating an HD intervention into digital smoking cessation programs have not yet been developed. The objective of this study was to examine the feasibility and acceptability of a digital smoking cessation program that was developed to address HD specifically. HD smokers were randomly assigned to either a standard publicly available program or a modified version of the program that addressed HD in depth. Feasibility was assessed by recruitment rates, and acceptability was assessed by examining engagement and satisfaction with the elements of the program. Preliminary effect sizes were obtained on smoking and alcohol use outcomes at 1 and 6 months postrandomization with the hypothesis that addressing HD in depth would lead to less alcohol use and more smoking abstinence than the standard program. Reduced odds of alcohol-involved lapses were examined as a putative mechanism of action for the HD intervention.

Methods

Participants and Procedure

This pilot trial used a 2-group randomized control design. All procedures were approved by the Brown University Institutional Review Board, and the study was registered on ClinicalTrials.gov (NCT03068611). Participants were 119 newly registered users on BecomeAnEX.org (EX), a digital smoking cessation program developed and managed by Truth Initiative—a nonprofit public health organization in the United States dedicated to ending cigarette smoking (formerly the American Legacy Foundation). EX was developed in collaboration with the Mayo Clinic Nicotine Dependence Center aligning with the tobacco clinical practice guidelines [26] and first went on the web in 2008. Study recruitment for this project was conducted over 7 weeks (from May 10, 2017, to July 3, 2017).

New registered users on EX were eligible for the study if they met the following inclusion criteria: (1) current daily smoker, (2) met the National Institute on Alcohol Abuse and Alcoholism (NIAAA) criteria for the past month (8+ drinks/week or 4+ drinks on ≥1 occasion in the past month for women and 15+ drinks/week or 5+ drinks on ≥1 occasion in the past month for

men), (3) ≥18 years old, (4) willing to provide contact information, (5) no prior use of the EX website, and (6) US residence based on Internet Protocol (IP) address. Individuals were ineligible if they reported a history of severe alcohol withdrawal symptoms because we did not wish to recommend abstaining from alcohol to participants for whom unsupervised alcohol withdrawal could be dangerous.

All procedures were conducted on the web. Recruitment information asked new registered users about their interest in participating in a research study *to develop and test a version of BecomeAnEX that is designed specifically for smokers who drink alcohol*. Interested individuals, then, completed the study eligibility screening questions. Those screened as eligible then read and electronically signed a web-based informed consent document that described study procedures, including the saliva collection procedure at follow-ups for those reporting smoking abstinence. Next, they verified their email address and completed the baseline survey. Individuals who were ineligible, who did not complete the consent, or who did not verify their email were directed to the standard EX (EX-S) program but were not enrolled in the study. Participants were compensated US $20 via an Amazon gift card code for their time completing the baseline survey.

Randomization

Eligible participants were assigned on a 1:1 basis by a computer algorithm to either EX with standard content (EX-S) or EX with additional content specific to HD (EX-HD). To ensure a balance of potentially key background factors, block randomization was conducted within 8 blocks formed by a 2 × 2 × 2 matrix of the following factors: sex (male and female), age (≤30 years old and >30 years old), and frequency of HD (<weekly and ≥weekly). The age cutoff of 30 years was chosen to ensure that both conditions had a very similar number of participants who would be classified as young adults, for whom the use of mobile technology and the context of alcohol consumption may be different when compared with older adults. Those randomized had access only to their respective treatment condition.

Follow-Ups

At 1 and 6 months postrandomization, all participants were sent an email to complete a follow-up survey. Those who did not complete the survey within 10 days of the initial email were called by study staff. Participants were compensated US $30 for the 1-month and US $50 for the 6-month follow-up via an Amazon gift card code.

Intervention Content

Website

Primary elements of the EX-S site are summarized in Table 1. A *My Quit Plan* page displayed a checklist for users that showed if each of the site's core components had been completed and provided the recommended next steps. The EX-HD website included all elements of EX-S and additional pages addressing alcohol use (Table 1), aligning with our previous counselor-delivered interventions [14,15] and including content mirroring and linking to NIAAA's *Rethinking Drinking* website [27]. A *Managing Alcohol Use* tab was added to *My Quit Plan* in EX-HD so that participants could view and print their alcohol change plan. Participants were given no specific guidance about how often to use the website, consistent with how EX is used by the public. Participants had access to the website throughout their participation in the study and retained that access as registered users after study participation concluded.

Table 1. Website elements.

Element[a]	Description
Elements in EX-Standard	
(1) Quit Date	A tool for participants to select a quit smoking date, which is recommended to occur within 2 weeks from registration.
(2) Cigarette Tracker	An interactive exercise to help participants identify the time of day and triggers for smoking by tracking cigarette use.
(3) Beat Your Smoking Triggers	An interactive exercise to help participants formulate strategies to dissociate cigarettes from smoking triggers.
(4) Build Your Support System	An interactive exercise for participants to identify helpful supporters to improve their efforts to quit as well as unsupportive people in their life they may need to avoid while quitting.
(5) Choose a Quit Smoking Aid	A strategic planning exercise where participants indicate their plan for pharmacotherapy use.
(6) EX Community	A large web-based social network of current and former smokers.
(7) Educational Content	Didactic, multimodal content that addresses strategies to prepare for quit day, cope with slips, and prevent relapse. It includes videos about tobacco addiction and pharmacotherapy.
Additional elements in EX-Heavy Drinker	
(1) Alcohol and Quitting Smoking	A content page on the risk of smoking relapse associated with heavy alcohol use.
(2) What's Your Drinking Pattern	An interactive exercise that provides personalized normative feedback on drinking using gender-based national US drinking norms.
(3) Risks of Heavy Drinking	Content page on NIAAA[b] guidelines for low-risk drinking and interactive effects of smoking and HD[c] on head and neck cancer risk.
(4) Managing Drinking	An interactive exercise for participants to evaluate the importance of changing drinking while quitting smoking and articulate their own reasons for change.
(5) Benefits of Changing Drinking	An interactive exercise where participants endorse which benefits they might experience if they stopped or cut down on drinking.
(6) Make a Plan to Manage Drinking	An interactive exercise for setting goals regarding drinking during the quit smoking attempt (eg, stop drinking for a specified time after quitting and set a maximum daily alcohol consumption limit) and making plans to minimize the risk of smoking when drinking (eg, drinking only with people who do not smoke or in places where smoking is not allowed).
(7) Strategies for Limiting Drinking	An interactive exercise where participants select strategies they could use for limiting drinking such as practicing drink refusal skills and self-monitoring drinking.
(8) Alcohol and Staying Quit	A content page highlighting the ongoing risk of smoking relapse associated with drinking alcohol.

[a]Participants were randomized to either EX with standard content (EX-S) or EX with additional content specific to HD (EX-HD). Those who were randomized to EX-HD received all EX-S content as well as the EX-HD content. Individuals received only the intervention web pages to which they were randomized each time they logged into the website.

[b]NIAAA: National Institute on Alcohol Abuse and Alcoholism.

[c]HD: heavy drinking.

Text Messaging

The EX text message program used was consistent with empirically supported text messaging smoking cessation programs in both frequency of texting and duration of the program [28]. Users signed up for text messaging during registration and had to reply *OK* to an initial message to complete text message enrollment; they could unenroll at any time by texting *STOP*. Messages encouraged interaction with the program through yes/no, true/false, and multiple-choice questions and provided quitting advice, positive reinforcement, reminders to avoid smoking triggers, information about nicotine and nicotine withdrawal, and tailored support around a set quit date.

Text messaging was the same intensity in both conditions. Participants received 2 messages/day before their selected quit date, 3-5 messages/day for 2 weeks starting on their quit date, and then 1-2 messages/day through 6 weeks past the quit date; the maximum number of messages they could receive was 128. Only 1 EX-S text message directly addressed alcohol use. In the EX-HD program, 24 of the EX-S messages (18.8%) were replaced with an alcohol-focused message. These texts were developed through an iterative process that involved extracting key content from previous alcohol-focused smoking interventions [12,14,15] and adapting it for a short messaging format consistent with the messages used in the EX-S text messaging program. Messages were piloted with a sample of 12 participants who gave feedback on content and tone. These texts provided information about the effect of HD on health and quitting smoking (eg, "For smokers who drink regularly, more than 35% of slips back to smoking happen when they are drinking alcohol"), encouraged and reinforced reductions in

drinking (eg, "Great! You're showing your commitment to your health; consider the benefits of drinking less and record them here"), provided links to alcohol-focused content on EX-HD, and reminded participants to anticipate situations in which they might drink (eg, "Any events coming up this week where you might be around alcohol? Reply with Y or N").

Measures

Baseline Data

At baseline, demographic information was collected about sex, age, race/ethnicity, education, employment, income, and marital status. Cigarette dependence was assessed using the Fagerström Test for Cigarette Dependence [29], and the perceived importance of quitting smoking was assessed with a single-item scale (0=*not at all important* to 10=*extremely important*) [14]. Severity of alcohol problem was assessed at baseline using the short inventory of problems [30]. Participants were asked about the importance of cutting down or stopping drinking while quitting smoking using a single-item scale (0=*not at all important* to 10=*extremely important*) and if they planned to cut down or stop drinking while quitting smoking (0=*no*, 1=*possibly*, 2=*probably*, and 3=*definitely*) [14].

Website Utilization

Utilization data from the 6 months following randomization were extracted from the EX database, including number of website visits and pages viewed. Utilization of EX-S content was assessed by tracking if participants completed each of the 7 core EX activities, and utilization of EX-HD content was assessed by tracking completion of the 8 HD activities.

Satisfaction With and Perceptions of the Program

At 1 month, satisfaction with the overall program was assessed using the 8-item client satisfaction questionnaire (CSQ-8) [31], which assesses satisfaction with provided services on a 1 to 4 scale, where higher scores indicate higher levels of satisfaction (alpha=.92). In addition, all participants provided single-item ratings of specific aspects of the program, including rating their *overall experience* with the website (1=*poor* to 5=*excellent*) and the extent to which the website made them consider how alcohol might affect their quitting (1=*Did not use*, 2=*Not at all*, 3=*A little*, 4=*Moderately*, and 5=*Very much/a lot*). Those who enrolled in text messaging were asked about their *overall experience* with the text messaging program (1=*poor* to 5=*excellent*) and if it helped them quit smoking and manage their drinking, respectively, (from 1=*completely disagree* to 5=*strongly agree*). They also rated if the number of text messages on alcohol was 1=*too few*, 2=*just right*, or 3=*too many*.

Smoking Outcomes

At baseline and follow-up, cigarette use was assessed using a daily report of smoking behavior over the past 7 days [32]. The primary smoking outcome was a self-reported 7-day abstinence from smoking cigarettes and other combustible tobacco products at 1 and 6 months [33]. Continuous abstinence was a secondary smoking outcome, defined as having made a quit attempt with no reported slips at both the 1- and 6-month follow-ups [33]. Participants reporting smoking abstinence at 6 months were invited to provide a saliva sample by mail via a Salimetrics collection kit to biochemically verify smoking status [34,35]. Participants were paid US $25 for returning the kit within 48 hours of receipt. Abstinence was confirmed with a cotinine concentration of <15 ng/mL.

Alcohol Outcomes

At baseline and follow-up, participants reported past 30-day drinking, including the number of days they drank 4+ drinks (for women)/5+ drinks (for men), and provided a daily report of drinking over the past 7 days [32]. The primary alcohol outcome was the number of HD days in the past 30 days at 1 and 6 months, with the 30-day window chosen to capture variability in this more uncommon behavior. The secondary outcome was the total number of drinks consumed in the past 7 days at those follow-ups. Although this 7-day assessment window is narrower than that used for inclusion in the study, this secondary outcome may provide a better estimate of total drinking as it is based on day-level reporting rather than using a quantity-frequency estimate.

Smoking Lapses

A series of questions were used to assess the perceptions of the causes of initial smoking lapses following previous work in the area [11,12]. At follow-up, participants were asked if they had made a quit attempt since enrolling. Those reporting a quit attempt were asked if they had smoked since that attempt, and if so, "What was the major cause of your smoking?," where *drinking alcohol* was 1 of the 10 response options, including *don't know* and *other*. Those selecting *drinking alcohol* were classified as having an alcohol-involved lapse.

Analysis Plan

All analyses were conducted using SAS 9.4 [36]. We first examined the number of participants passing study milestones (eg, screening, consent, baseline, and follow-up) and the baseline characteristics of the sample. The acceptability of EX-HD was examined by comparing EX-S and EX-HD on website and text message utilization and program satisfaction using *t* tests and nonparametric tests. In exploratory analyses, we also examined if baseline characteristics predicted engagement metrics. The preliminary efficacy of EX-HD was examined by comparing smoking and alcohol use outcomes using *t* tests and chi-square tests. We used chi-square analyses to test if EX-HD reduced the odds of alcohol-involved smoking lapses compared with EX-S. In exploratory analyses, we examined if engagement metrics were associated with perceptions of the program. We also conducted generalized estimating equations (GEE) covarying age, sex, and baseline frequency of HD (ie, the variables included in the randomization scheme) to examine if treatment effects across 1 and 6 months were moderated by gender, motivation to change (ie, perceived importance of quitting smoking and reducing drinking, respectively), level of website engagement, and text messaging enrollment. An alpha level of .05 was used for all analyses, without adjustment for multiple testing, because the primary purpose of the analyses was to identify patterns of results that might inform future modifications and improvement of the EX-HD program.

Results

Feasibility

Figure 1 shows the Consolidated Standards of Reporting Trials diagram of participant flow through study milestones. During the recruitment period, there were a total of 816 newly registered EX users with US IP addresses, who each received an invitation to participate in the study. Of these, 57.9% (473/816) completed screening, and 38.3% of those were screened as eligible. A third (34.3%) of participants who were screened as eligible were not enrolled in the study because they did not complete the consent process, did not confirm their contact information via email, or did not complete the baseline survey, leaving 119 enrolled participants, with 61 randomized to EX-S and 58 to EX-HD. Overall, 14.6% of all new US users of EX during the recruitment window participated in the study.

Table 2 shows the demographic and clinical characteristics of the enrolled participants overall and by condition. Follow-up rates were 72.3% at 1 month and 52.9% at 6 months. There were no significant differences in demographics, baseline clinical variables, or intervention condition between follow-up completers and noncompleters. At the 6-month follow-up, 14% (17/119) of participants reported 7-day abstinence from smoking and nicotine and were sent a saliva collection kit. Of these, 42% (7/17) returned the kit, with 4 having abstinence confirmed by testing. Three participants reported smoking in the past week at the time they returned the sample, of whom 2 were over the cutoff of 15 ng/mL.

Figure 1. A Consolidated Standards of Reporting Trials diagram showing participant flow. EX-HD: BecomeAnEX with additional content specific to heavy drinking; EX-S: BecomeAnEX with standard content.

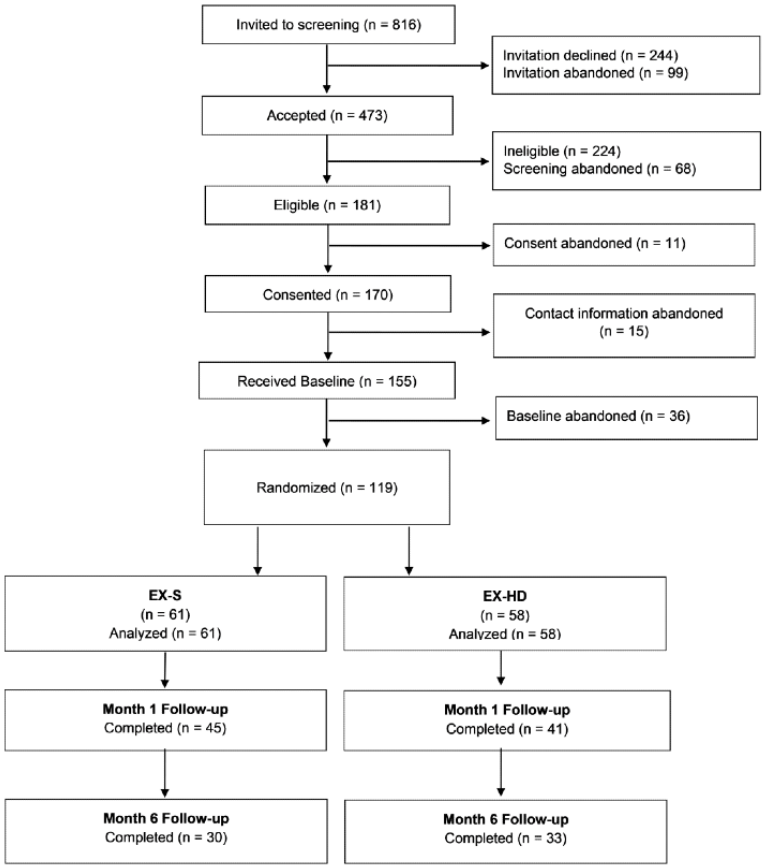


Table 2. Baseline demographic and clinical characteristics for the entire sample and by treatment condition.

Variable	Overall (N=119)	EX-S[a] (n=61)	EX-HD[b] (n=58)
Female gender, n (%)	83 (69.7)	42 (69)	41 (71)
Age (years), mean (SD)	38.0 (11.8)	38.6 (12.9)	37.1 (10.6)
Race, n (%)			
White	98 (82.4)	50 (82)	48 (83)
Black/African American	13 (10.9)	6 (10)	7 (12)
Native American	1 (0.8)	0 (0)	1 (2)
Asian	0 (0.0)	0 (0)	0 (0)
Pacific Islander	0 (0.0)	0 (0)	0 (0)
Multiple races	6 (5.0)	5 (8)	1 (2)
Other	1 (0.8)	0 (0)	1 2)
Hispanic/Latino	4 (3.4)	1 (2)	3 (5)
Education, n (%)			
Less than high school	7 (5.9)	4 (7)	3 (5)
High school	20 (16.8)	8 (13)	12 (21)
Some college/tech/business	66 (55.5)	37 (61)	29 (50)
College graduate or higher	26 (21.8)	12 (20)	14 (24)
Employment status, n (%)			
Unemployed	21 (17.6)	9 (15)	12 (21)
Employed part/full time	84 (70.6)	48 (79)	36 (62)
Other[c]	14 (11.8)	4 (7)	10 (17)
Income, n (%)			
Less than US $30,000	52 (43.7)	28 (46)	24 (41)
US $30,000-US $59,999	32 (26.9)	18 (29)	14 (24)
US $60,000-US $89,999	18 (15.1)	6 (10)	12 (21)
US $90,000 or more	17 (14.3)	9 (15)	8 (14)
Marital status, n (%)			
Married or cohabiting	47 (39.5)	26 (43)	21 (36)
Single (never married)	36 (30.3)	17 (28)	19 (33)
Widow/separated/divorced	36 (30.3)	18 (29)	18 (31)
Clinical characteristics, mean (SD)			
Cigarettes per day in past 7 days	17.1 (9.1)	15.9 (74)	18.4 (10.5)
FTCD[d] total score	5.1 (2.3)	4.9 (2.3)	5.3 (2.3)
SIP[e] total score	9.8 (10.1)	10.1 (9.7)	9.5 (10.6)
Number of HD[f] days[g]	6.9 (6.8)	7.3 (7.2)	6.6 (6.4)
Number of drinks in the past 7 days	16.0 (12.8)	14.2 (11.7)	17.8 (13.7)
Importance of quitting smoking[h]	9.1 (1.6)	9.2 (1.5)	9.1 (1.6)
Importance of reducing drinking[h]	6.5 (3.6)	7.1 (3.5)	5.9 (3.6)
Intent to change drinking,[i] n (%)			
No	9 (7.6)	3 (5)	6 (10)
Possibly	29 (24.4)	13 (21)	16 (28)

Variable	Overall (N=119)	EX-S[a] (n=61)	EX-HD[b] (n=58)
Probably	33 (27.7)	16 (26)	17 (29)
Definitely	48 (40.3)	29 (47)	19 (33)

[a]EX-S: EX-standard.

[b]EX-HD: EX-heavy drinker.

[c]Includes students, homemakers, and retired individuals.

[d]FTCD: Fagerström Test for Cigarette Dependence.

[e]SIP: short inventory of problems.

[f]HD: heavy drinking.

[g]In the 30 days before study enrollment.

[h]Single-item rating from 0=not important to 10=extremely important.

[i]Participants were asked if they planned to cut down or stop drinking while quitting smoking.

Website Engagement and Acceptability

Website engagement, as operationalized by site visits, total page views, pages viewed per visit, total minutes on the site, and completion of EX-S activities are shown in Table 3 and were similar by condition. Just under half of the participants (48.7%) completed <1 EX activity. Just over two-thirds (42/61, 69%) of the participants set a quit date at least once through the website, with 13% (8/61) setting a quit date more than once. The median time spent on the site was 8 min in both conditions, with just over a quarter of participants spending more than 25 min on the site.

Among those assigned to EX-HD, 74% (43/58) of participants did not complete any EX-HD activities and 12% (7/58) completed more than half (Table 3). Exploratory analyses of the clinical factors in Table 2 that might be related to engagement with EX-HD indicated that 35% (6/17) of those who reported they would *probably* change their drinking completed more than half of EX-HD exercises compared with 0% (0/6), 6% (1/16), and 0% (0/19) of those reporting *no*, *possibly*, or *definitely*, respectively (χ^2_3=12.6; *P*=.006). Other smoking and drinking characteristics were not significantly associated with high EX-HD website engagement.

With respect to acceptability, participants in both the conditions rated the overall experience with the website very similarly, with means between *good* and *very good* (Table 3). Contrary to expectations, participants in the both conditions provided similar ratings about how much the website made them consider the impact of alcohol on their quit efforts, with means near the anchor of *moderately*. In an exploratory analysis within the EX-HD condition, ratings on this item were significantly higher among those who had completed at least one EX-HD activity (mean 4.67, SD 0.71) compared with those who had not (mean 3.58, SD 1.39; t_{38}=2.25; *P*=.03).

Table 3. Website engagement and acceptability.

Variable	Experimental group		*P* value[c]
	EX-S[a] (control)	EX-HD[b]	
Sample size, n	61	58	N/A[d]
Number of visits, median (IQR)	2 (1-3)	2 (1-3)	.47
Number of pages viewed, median (IQR)	17 (6-31)	13 (4-33)	.48
Pages viewed per visit, median (IQR)	6.7 (3.3-11.3)	5.6 (3.0-11.3)	.49
Minutes on website, median (IQR)	8.1 (3.3-25.9)	8.1 (1.3-27.7)	.86
EX-S activities completed, n (%)			
Set a quit date	42 (69)	39 (67)	.85
Number of activities completed, n (%)			.80
No exercises	13 (21)	15 (26)	
1 exercise	17 (28)	13 (22)	
2 exercises	15 (25)	15 (26)	
3 exercises	8 (13)	6 (10)	
4+ exercises	8 (13)	9 (15)	
EX-HD activities completed, n (%)			
No exercises	N/A	43 (74)	N/A
1 exercise	N/A	5 (9)	N/A
2 exercises	N/A	1 (2)	N/A
3 exercises	N/A	2 (3)	N/A
4+ exercises	N/A	7 (12)	N/A
Self-reported website satisfaction			
Total respondents, n (%)	45 (74)	40 (69)	N/A
Overall experience, mean (SD)[e]	3.87 (0.97)	3.75 (0.95)	.58
Caused reflection on alcohol use[f], mean (SD)	3.91 (1.14)	3.82 (1.33)	.75

[a]EX-S: EX-standard.

[b]EX-HD: EX-heavy drinker.

[c]*P* values were determined using *t* tests for continuous variables, chi-square tests for categorical variables, and Wilcoxon rank sum tests for non-normal variables.

[d]N/A: not applicable.

[e]Response scale is 1-5 (1=Poor, 2=Fair, 3=Good, 4=Very Good, and 5=Excellent).

[f]Response scale 1-5 (1=Did not use, 2=Not at all, 3=A little, 4=Moderately, and 5=Very much/a lot).

Text Messaging Engagement and Acceptability

Table 4 also shows text messaging engagement and acceptability. With respect to text messaging engagement, participants in EX-HD were significantly more likely to unenroll from text messaging compared with those in EX-S. However, the number of messages sent or received did not differ significantly by condition. Participants in EX-S and EX-HD rated the overall experience with the texting program very similarly, with means between *good* and *very good*. Likewise, they rated the helpfulness of the program very similarly. With regard to whether the text program helped them manage drinking while quitting smoking, those in EX-HD rated this item numerically higher (closer to the *somewhat agree* anchor), although this difference was not statistically significant. Regarding the number of alcohol-focused texts, participants in EX-HD rated these as slightly over *just right*, whereas those in EX-S rated them as slightly under *just right*, a difference which was significant.

Table 4. Text messaging engagement and acceptability.

Variable	Experimental group		*P* value[c]
	EX-S[a] (control)	EX-HD[b]	
Text message enrollment, n (%)			.46
Never signed up	16 (26)	16 (28)	
Signed up/no initiation	15 (25)	9 (16)	
Fully enrolled	30 (49)	33 (57)	
Unenrollment rate, n (%)	3 (13)[d]	14 (42)[e]	.01
Messaging frequency, median (IQR)[f]			
Messages received by participant	116 (36-142)	86 (27-153)	.95
Messages sent from participant	13 (3-26)	9 (3-14)	.47
Text messaging acceptability			
Total respondents, n (%)	25 (41)	22 (38)	N/A[g]
Overall experience, mean (SD)[h]	3.64 (1.00)	3.86 (1.25)	.90
Helpful with quitting smoking,[i] mean (SD)	3.84 (1.10)	3.86 (1.13)	.94
Helpful with managing drinking,[i] mean (SD)	2.96 (1.34)	3.64 (1.14)	.07
Number of alcohol-related messages,[j] mean (SD)	1.89 (0.67)	2.32 (0.57)	.02

[a]EX-S: EX-standard.

[b]EX-HD: EX-heavy drinker.

[c]*P* values were determined using t tests for continuous variables, chi-square tests for categorical variables, and Wilcoxon rank sum tests for non-normal variables.

[d]N=30.

[e]N=33.

[f]Of participants who were fully enrolled in the text messaging program, including those who were unenrolled.

[g]N/A: not applicable.

[h]Response scale is 1-5 (1=Poor, 2=Fair, 3=Good, 4=Very Good, 5=Excellent).

[i]Response scale is 1-5 (1=Completely disagree, 2=Somewhat disagree, 3=Neither agree nor disagree, 4=Somewhat agree, 5=Strongly agree).

[j]Response scale is 1-3 (1=Too few, 2=Just right, 3=Too many).

Overall Program Satisfaction

CSQ-8 ratings were very similar between EX-S (mean 3.14, SD 0.59) and EX-HD (mean 3.07, SD 0.63; t_{83}=0.57; *P*=.57).

Alcohol Use Outcomes

Primary drinking outcomes (means and SDs) and between-group effect sizes are shown in Table 5. The number of drinks per week and percent HD days showed significant reductions from baseline to both 1-month and 6-month follow-up, regardless of condition (*P*s<.05). However, there was no clear pattern of differential reductions in drinking by condition, and between-group effect sizes were small.

Table 5. Summary of alcohol use and smoking outcomes at baseline and at each follow-up.

Variables	EX-S[a] (n=61)	EX-HD[b] (n=58)	Effect size[c]	*P* value
Drinking outcomes[d], mean (SD)				
Number of HD[e] days in the past 30 days				
Baseline (EX-S=61, EX-HD=58)	7.3 (7.2)	6.6 (6.4)	−0.01	.94
1 month (EX-S=45, EX-HD=40)	2.9 (4.7)	4.4 (6.3)	0.14	.36
6 months (EX-S=30, EX-HD=33)	2.6 (3.5)	3.8 (6.7)	0.03	.92
Total number of drinks in the past 7 days[f]				
Baseline (EX-S=61, EX-HD=58)	14.2 (11.7)	17.8 (13.7)	0.26	.16
1 month (EX-S=45, EX-HD=41)	10.6 (12.1)	8.8 (10.1)	−0.23	.29
6 months (EX-S=30, EX-HD=33)	8.1 (8.2)	7.7 (10.1)	−0.14	.57
Smoking outcomes[g]				
Smoking abstinence, n (%)				
1 month (past 7 days)	10 (16)	10 (17)	0.02	.90
6 months (past 7 days)	9 (15)	11 (19)	0.11	.54
6 months—biochemically verified	1 (2)	3 (5)	0.20	.36[h]
Continuous abstinence[i]	2 (3)	5 (9)	0.23	.22[h]
Smoking lapse behavior, n (%)				
Reporting any lapse[j]	77 (89)	73 (84)	−0.14	.51
Reporting alcohol-involved lapse[j]	18 (41)	7 (16)	−0.56	.01

[a]EX-S: EX-standard.

[b]EX-HD: EX-heavy drinker.

[c]Effect size is expressed as Cohen *d* for continuous data and Cohen *h* for dichotomous outcomes.

[d]*P* values and effect sizes are calculated using *t* tests on log-transformed data due to positive skewness and kurtosis.

[e]HD: heavy drinking.

[f]Obtained from a 7-day timeline follow-up.

[g]Denominator is all 119 participants with those participants with missing data on smoking status considered as smoking. Abstinence was defined as self-reported abstinence from all combustible tobacco products.

[h]*P* values were calculated using Fisher exact tests rather than chi-square tests due to small cell sizes.

[i]Continuous abstinence was defined as those who made a quit attempt and reported no slips for 6 months.

[j]Denominator was 87 participants who reported having made a quit attempt at either the 1-month or 6-month follow-up.

Smoking Outcomes

Smoking outcomes are shown in Table 5. The effect sizes were small and not statistically significant.

Alcohol-Involved Lapses

A total of 90 participants (EX-S: n=46 and EX-HD: n=44) completed a 1-month or 6-month smoking lapse survey. Of those 90, 87 (97%) made a quit smoking attempt since enrollment. Although the percentage of participants reporting a lapse was similar across conditions (Table 4), EX-HD participants were significantly less likely to report having an alcohol-involved lapse.

Moderators, Engagement, and Outcomes: Exploratory Analyses

Moderators

For all smoking and alcohol use outcomes reported above, exploratory analyses indicated no significant interactions between gender and treatment condition, *P*s>.45. Likewise, neither perceived importance of quitting smoking nor perceived importance of cutting down on drinking while quitting smoking interacted with treatment condition significantly, *P*s>.45.

Engagement

Due to the non-normal distribution of website utilization metrics (ie, site visits, pages viewed, and EX-S activities completed), we dichotomized each using a median split and summed them to create a 0-3 ordinal scale, where higher scores reflected greater engagement. Website engagement was not significantly correlated with overall program satisfaction as assessed by the

CSQ-8 (r_{85}=0.16; *P*=.15). In GEE models, greater website engagement was associated with significantly higher odds of smoking abstinence (odds ratio 1.72, 95% CI 1.20-2.49; *P*=.003) but was not significantly associated with either drinking outcome. Text messaging enrollment was not significantly correlated with website engagement (r_{119}=0.10; *P*=.29), or with overall program satisfaction (r_{85}=0.07; *P*=.55). Text messaging enrollment was not significantly associated with smoking or drinking outcomes, and the effect of treatment conditions on smoking and drinking outcomes did not differ based on the level of website engagement or text messaging enrollment, *P*s>.40.

Discussion

Principal Findings

This pilot randomized controlled trial was built on a well-established evidence-based smoking cessation website to address HD in the context of smoking cessation, incorporating key elements of our previously tested person-delivered approaches for addressing HD in smoking cessation. The findings provided strong support for the feasibility of recruiting participants who are HD smokers enrolled in a publicly available digital smoking cessation program. Almost 60% of invited individuals agreed to be screened, a rate higher than that in a previous study from this website [37]. In general, registrants on EX who elect to be screened for research tend to be heavier smokers and more motivated to quit compared with those who do not screen [37]. Of those who screened, almost 40% were eligible, with the vast majority subsequently enrolled in the trial; such screening and enrollment rates compare favorably with other studies on web-based smoking cessation programs [38,39]. Participants recruited from EX tended to be female, although HD is more common in men [4]. This gender imbalance likely reflects that women are more likely to look for health information on the web [40]. There was a rapid rate of recruitment in this trial, with 119 participants successfully enrolled in only 7 weeks. This is notable given that HD smokers have represented a small minority of participants in other clinical trials of web-based cessation interventions [39] and have been largely ignored in cessation studies. The results demonstrate that it would be feasible to conduct a large-scale trial of EX-HD recruiting solely from BecomeAnEX.

Follow-up rates ranged from just over 70% at 1 month to just over 50% at 6 months, similar to other trials of web-based smoking cessation programs [21], where the risk of bias is relatively low given roughly equal follow-up rates by condition. Additional methods to enhance completion of follow-up assessment, such as offering an additional bonus incentive for survey completion within 24 hours [39] or providing more frequent brief communications from the study between follow-up visits, would be important to utilize in a future trial. Consistent with prior research [41,42], there were large discrepancies between self-reported and biochemically verified smoking abstinence due to missing data rather than misreported smoking status. We agree with previous recommendations on the verification of smoking status, which suggest that using biochemical verification in large-scale web-based cessation trials is of limited value and is not feasible [43].

Overall, satisfaction with website content was high regardless of the treatment condition. However, although about two-thirds of participants set a quit date over the web, engagement with the website was low, with a median of 2 visits to the website, consistent with 2 large-scale trials on internet cessation, one of which was recruited from BecomeAnEX [39,44]. The lack of differences by condition in website engagement suggests that the additional alcohol content in EX-HD did not hinder engagement. However, only one-quarter of the participants in EX-HD completed a web-based activity specific to EX-HD. Given that website engagement was related to better smoking outcomes across conditions—echoing a recent study among smokers from EX with an alcohol use disorder [45]—testing ways to increase exposure to and engagement with website content remains a high priority. Identifying strategies to boost engagement with digital health behavior change programs has been noted as a priority for the field [46]. Tailoring the website experience to users' preferences may be one means of enhancing engagement. Indeed, among those assigned to EX-HD, exploratory analyses indicated that those who were strongly considering—but not definitely—changing their drinking while quitting smoking were the most likely to engage with the alcohol-related content. Future iterations of EX-HD could assess interest in changing drinking in an initial assessment and then provide more or less alcohol content based on user intentions. Such tailored and dynamic interventions may enhance engagement. Regardless of further efforts to enhance program engagement, future clinical trials of EX-HD should determine sample size requirements, accounting for the fact that only one-quarter of participants will have meaningful interaction with the intervention content, substantially reducing the potential effect size for EX-HD.

About half of the participants were enrolled in text messaging. Those in EX-HD were more likely to unenroll from text messaging. EX-HD participants tended to want fewer alcohol-focused messages, and those in EX-S tended to want more, which could account for the greater rate of unenrollment by EX-HD participants. Participants receiving text messaging in EX-HD, compared with those in EX-S, rated the messaging program as more helpful for managing drinking, although the effect did not reach the .05 significance level. Future text messaging interventions will need to balance the potential for overmessaging on alcohol use while ensuring that texts address alcohol in sufficient depth, perhaps by allowing some user choice over content. Again, allowing for user tailoring of text messaging content and frequency may obviate concerns about oversaturating users with content that they do not consider relevant. In this trial, participants signed up for text messaging before they knew that they would participate in a study of alcohol-focused intervention, and therefore, they did not sign up knowing that they would receive a substantial number of alcohol-related texts. Given that almost 1 in 5 messages were alcohol-focused, a more optimum proportion would likely be lower than that and could be tailored based on user preference.

EX-HD did not result in greater reductions in drinking compared with EX-S. However, reductions in drinking seen in both conditions in this study were larger than those reported in most cessation trials [12,14,17,47], on the order of about 50%

reductions in both HD days and drinks per week. Such changes could represent regression to the mean given that participants had to meet certain thresholds for drinking immediately before participation. However, they also could reflect the fact that over two-thirds of participants reported that they probably or definitely intended to cut down on drinking when trying to quit smoking. Before receiving either intervention, participants believed that changing their drinking habits when quitting smoking was quite important, with a mean of over 6 on a 0-10 measure of importance.

Although reductions in drinking were similar across conditions, EX-HD resulted in significantly reduced odds of alcohol-related smoking lapses compared with EX-S, mirroring results of an initial trial testing an in-person smoking cessation program targeting HD smokers [14]. These outcomes may reflect a heightened awareness of the risk of alcohol-related smoking lapses caused by either web or text messaging content. That this effect emerged was somewhat surprising, given the low exposure to alcohol-related content overall. Given that smoking lapses for HD smokers often occur in the context of drinking, such an effect may be important in improving smoking outcomes.

For all smoking outcomes, effect sizes were small, and none approached significance. Smoking outcomes were similar to those reported in a recent trial of smokers from EX [48], with abstinence rates ranging from 15% to 19% at 1- and 6-month follow-ups. The relatively small effect sizes align with those found in a study of quitline counseling with HD smokers, which showed a statistically significant 3.2% difference that favored the HD intervention [15]. A small effect size in the context of a website that reaches thousands of smokers each month could have a meaningful impact on public health with no added costs beyond the website development.

Limitations

The primary limitations of this pilot study include its modest sample size and follow-up rates, which make it inappropriate to draw conclusions about its efficacy and potential impact. The sample size also limits the degree of depth to which we can understand factors that might have impacted engagement and satisfaction with EX-HD content. When examining factors that might relate, for example, to the utilization of text messaging, the number of participants who opted into text messaging (n=63) makes in-depth analyses impractical. Likewise, because there was modest engagement in the website content, it was not possible to examine experimental outcomes among those who had substantial contact with intervention content. The lack of biochemical validation of the primary smoking outcomes is also a limitation that is common in many trials of web-based cessation programs.

Recruitment was done from the available pool of newly registered users of EX. Therefore, the sample was limited to those with computer literacy and access and reflected the demographics of EX users, who tend to be white women with relatively high education, demographics that are roughly consistent with other web-based cessation trials [20]. A more diverse sample could be achieved in a full-scale randomized controlled trial of EX-HD by deactivating the recruitment of overrepresented groups once targets for those groups are reached [37]. In addition, publicity for digital smoking cessation programs may need to be targeted to websites and social media outlets where minority smokers are more represented. Engaging smokers from underrepresented groups in the development of campaigns to promote digital cessation programs is likely to improve the reach of these programs and is an important area for future research and practice.

Conclusions

HD has been highlighted as an important predictor of smoking relapse and cessation failure, but it is rarely addressed in web-based smoking cessation platforms. As such, enhancing proven web-based cessation interventions with alcohol-related content is critical to leveraging the potential public health impact of a broad reach treatment modality for heavy drinkers who smoke. In our study, many new HD registrants to a digital smoking cessation program are willing to participate in a trial focused on smoking cessation targeting HD. EX-HD is acceptable and may be useful in reducing alcohol-involved smoking lapses. The feasibility and acceptability of EX-HD, coupled with a modest initial indication of its clinical promise, warrants testing its effects on both drinking and smoking outcomes in a fully powered large-scale clinical trial. A critical next step in intervention development efforts is to focus on methods for increasing exposure to and engagement with intervention content, including tailoring user experiences to their interest and intention in changing drinking.

Acknowledgments

The authors gratefully acknowledge Majesta Kitts, Timothy Souza, Suzanne Sales, Sarah Cha, and Jody Brookover for their assistance in conducting this research. This project was supported by a grant from the NIAAA, R34AA024593.

Conflicts of Interest

BT testifies as an expert witness on behalf of plaintiffs who have filed litigation against the tobacco companies. AG is employed by Truth Initiative, which licenses an enterprise version of BecomeAnEX to employers, health plans, and other tobacco control organizations.

Multimedia Appendix 1
CONSORT EHEALTH checklist (v.1.6.1).
[PDF File (Adobe PDF File), 105 KB - formative_v4i6e7570_app1.pdf]

References

1. Falk DE, Yi H, Hiller-Sturmhöfel S. An epidemiologic analysis of co-occurring alcohol and tobacco use and disorders: findings from the national epidemiologic survey on alcohol and related conditions. Alcohol Res Health 2006;29(3):162-171 [FREE Full text] [Medline: 17373404]
2. Kahler CW, Strong DR, Papandonatos GD, Colby SM, Clark MA, Boergers J, et al. Cigarette smoking and the lifetime alcohol involvement continuum. Drug Alcohol Depend 2008 Jan 11;93(1-2):111-120 [FREE Full text] [doi: 10.1016/j.drugalcdep.2007.09.004] [Medline: 17964082]
3. Augustson EM, Wanke KL, Rogers S, Bergen AW, Chatterjee N, Synder K, et al. Predictors of sustained smoking cessation: a prospective analysis of chronic smokers from the alpha-tocopherol beta-carotene cancer prevention study. Am J Public Health 2008 Mar;98(3):549-555. [doi: 10.2105/AJPH.2005.084137] [Medline: 17600267]
4. Kahler CW, Borland R, Hyland A, McKee SA, Thompson ME, Cummings KM. Alcohol consumption and quitting smoking in the international tobacco control (ITC) four country survey. Drug Alcohol Depend 2009 Mar 1;100(3):214-220 [FREE Full text] [doi: 10.1016/j.drugalcdep.2008.10.006] [Medline: 19056188]
5. Toll BA, Cummings KM, O'Malley SS, Carlin-Menter S, McKee SA, Hyland A, et al. Tobacco quitlines need to assess and intervene with callers' hazardous drinking. Alcohol Clin Exp Res 2012 Sep;36(9):1653-1658 [FREE Full text] [doi: 10.1111/j.1530-0277.2012.01767.x] [Medline: 22703028]
6. Cook JW, Fucito LM, Piasecki TM, Piper ME, Schlam TR, Berg KM, et al. Relations of alcohol consumption with smoking cessation milestones and tobacco dependence. J Consult Clin Psychol 2012 Dec;80(6):1075-1085 [FREE Full text] [doi: 10.1037/a0029931] [Medline: 22963593]
7. Hendricks PS, Delucchi KL, Humfleet GL, Hall SM. Alcohol and marijuana use in the context of tobacco dependence treatment: impact on outcome and mediation of effect. Nicotine Tob Res 2012 Aug;14(8):942-951 [FREE Full text] [doi: 10.1093/ntr/ntr312] [Medline: 22259148]
8. Lynch KL, Twesten JE, Stern A, Augustson EM. Level of alcohol consumption and successful smoking cessation. Nicotine Tob Res 2019 Jul 17;21(8):1058-1064 [FREE Full text] [doi: 10.1093/ntr/nty142] [Medline: 29986105]
9. Borland R. Slip-ups and relapse in attempts to quit smoking. Addict Behav 1990;15(3):235-245. [doi: 10.1016/0306-4603(90)90066-7] [Medline: 2378283]
10. Shiffman S, Paty JA, Gnys M, Kassel JA, Hickcox M. First lapses to smoking: within-subjects analysis of real-time reports. J Consult Clin Psychol 1996 Apr;64(2):366-379. [doi: 10.1037//0022-006x.64.2.366] [Medline: 8871421]
11. Kahler CW, Spillane NS, Metrik J. Alcohol use and initial smoking lapses among heavy drinkers in smoking cessation treatment. Nicotine Tob Res 2010 Jul;12(7):781-785 [FREE Full text] [doi: 10.1093/ntr/ntq083] [Medline: 20507898]
12. Kahler CW, Cioe PA, Tzilos GK, Spillane NS, Leggio L, Ramsey SE, et al. A double-blind randomized placebo-controlled trial of oral naltrexone for heavy-drinking smokers seeking smoking cessation treatment. Alcohol Clin Exp Res 2017 Jun;41(6):1201-1211 [FREE Full text] [doi: 10.1111/acer.13396] [Medline: 28401564]
13. Kahler CW, Metrik J, Spillane NS, Day A, Leventhal AM, McKee SA, et al. Acute effects of low and high dose alcohol on smoking lapse behavior in a laboratory analogue task. Psychopharmacology (Berl) 2014 Dec;231(24):4649-4657 [FREE Full text] [doi: 10.1007/s00213-014-3613-3] [Medline: 24858377]
14. Kahler CW, Metrik J, la Chance HR, Ramsey SE, Abrams DB, Monti PM, et al. Addressing heavy drinking in smoking cessation treatment: a randomized clinical trial. J Consult Clin Psychol 2008 Oct;76(5):852-862 [FREE Full text] [doi: 10.1037/a0012717] [Medline: 18837602]
15. Toll BA, Martino S, O'Malley SS, Fucito LM, McKee SA, Kahler CW, et al. A randomized trial for hazardous drinking and smoking cessation for callers to a quitline. J Consult Clin Psychol 2015 Jun;83(3):445-454 [FREE Full text] [doi: 10.1037/a0038183] [Medline: 25419583]
16. Ames SC, Pokorny SB, Schroeder DR, Tan W, Werch CE. Integrated smoking cessation and binge drinking intervention for young adults: a pilot efficacy trial. Addict Behav 2014 May;39(5):848-853. [doi: 10.1016/j.addbeh.2014.02.001] [Medline: 24583274]
17. Correa-Fernández V, Díaz-Toro EC, Reitzel LR, Guo L, Chen M, Li Y, et al. Combined treatment for at-risk drinking and smoking cessation among Puerto Ricans: a randomized clinical trial. Addict Behav 2017 Feb;65:185-192 [FREE Full text] [doi: 10.1016/j.addbeh.2016.10.009] [Medline: 27825036]
18. Graham A, Amato M. Twelve million smokers look online for smoking cessation help annually: health information national trends survey data, 2005-2017. Nicotine Tob Res 2019 Jan 4;21(2):249-252 [FREE Full text] [doi: 10.1093/ntr/nty043] [Medline: 29660037]
19. McCausland K, Curry L, Mushro A, Carothers S, Xiao H, Vallone D. Promoting a web-based smoking cessation intervention: implications for practice. Cases Public Health Commun Mark 2011;5:3-26 [FREE Full text]
20. Graham A, Carpenter K, Cha S, Cole S, Jacobs M, Raskob M, et al. Systematic review and meta-analysis of internet interventions for smoking cessation among adults. Subst Abuse Rehabil 2016;7:55-69 [FREE Full text] [doi: 10.2147/SAR.S101660] [Medline: 27274333]
21. Taylor G, Dalili M, Semwal M, Civljak M, Sheikh A, Car J. Internet-based interventions for smoking cessation. Cochrane Database Syst Rev 2017 Sep 4;9:CD007078 [FREE Full text] [doi: 10.1002/14651858.CD007078.pub5] [Medline: 28869775]
22. SmokeFree. URL: http://smokefree.gov/ [accessed 2020-04-13]

23. Cohn AM, Zhao K, Cha S, Wang X, Amato MS, Pearson JL, et al. A descriptive study of the prevalence and typology of alcohol-related posts in an online social network for smoking cessation. J Stud Alcohol Drugs 2017 Sep;78(5):665-673 [FREE Full text] [doi: 10.15288/jsad.2017.78.665] [Medline: 28930053]
24. Fridberg DJ, Cao D, Grant JE, King AC. Naltrexone improves quit rates, attenuates smoking urge, and reduces alcohol use in heavy drinking smokers attempting to quit smoking. Alcohol Clin Exp Res 2014 Oct;38(10):2622-2629 [FREE Full text] [doi: 10.1111/acer.12513] [Medline: 25335648]
25. Cunningham JA, Selby P, van Mierlo T. Integrated online services for smokers and drinkers? Use of the check your drinking assessment screener by participants of the stop smoking center. Nicotine Tob Res 2006 Dec;8(Suppl 1):S21-S25. [doi: 10.1080/14622200601039865] [Medline: 17491167]
26. Fiore M, Jaén C, Baker T, Tobacco Use and Dependence Guideline Panel. Treating Tobacco Use and Dependence: Quick Reference Guide for Clinicians: 2008 Update. Rockville, MD: US Department of Health and Human Services; 2008.
27. Rethinking Drinking - NIAAA - NIH. URL: https://www.rethinkingdrinking.niaaa.nih.gov/ [accessed 2020-04-23]
28. Free C, Knight R, Robertson S, Whittaker R, Edwards P, Zhou W, et al. Smoking cessation support delivered via mobile phone text messaging (txt2stop): a single-blind, randomised trial. Lancet 2011 Jul 2;378(9785):49-55 [FREE Full text] [doi: 10.1016/S0140-6736(11)60701-0] [Medline: 21722952]
29. Heatherton TF, Kozlowski LT, Frecker RC, Fagerström KO. The Fagerström test for nicotine dependence: a revision of the Fagerström tolerance questionnaire. Br J Addict 1991 Sep;86(9):1119-1127. [doi: 10.1111/j.1360-0443.1991.tb01879.x] [Medline: 1932883]
30. Tonigan J, Miller W. Assessment and validation of the drinker inventory of consequences (DrInC): a multi-site outpatient and aftercare clinical sample of problem drinkers. Alcohol Clin Exp Res 1993;17:513A [FREE Full text] [doi: 10.1037/e563232012-001]
31. Larsen DL, Attkisson C, Hargreaves WA, Nguyen TD. Assessment of client/patient satisfaction: development of a general scale. Eval Program Plann 1979;2(3):197-207. [doi: 10.1016/0149-7189(79)90094-6] [Medline: 10245370]
32. Toll BA, Leeman RF, McKee SA, O'Malley SS. A brief 7-day estimate of alcohol consumption for use in smoking cessation clinical trials. Addict Behav 2008 Dec;33(12):1606-1609 [FREE Full text] [doi: 10.1016/j.addbeh.2008.07.007] [Medline: 18722721]
33. Piper ME, Bullen C, Krishnan-Sarin S, Rigotti NA, Steinberg ML, Streck JM, et al. Defining and measuring abstinence in clinical trials of smoking cessation interventions: an updated review. Nicotine Tob Res 2019 Jul 4 epub ahead of print. [doi: 10.1093/ntr/ntz110] [Medline: 31271211]
34. Foulds J, Bryant A, Stapleton J, Jarvis MJ, Russell MA. The stability of cotinine in unfrozen saliva mailed to the laboratory. Am J Public Health 1994 Jul;84(7):1182-1183. [doi: 10.2105/ajph.84.7.1182] [Medline: 8017556]
35. Etter J, Neidhart E, Bertrand S, Malafosse A, Bertrand D. Collecting saliva by mail for genetic and cotinine analyses in participants recruited through the internet. Eur J Epidemiol 2005;20(10):833-838. [doi: 10.1007/s10654-005-2148-7] [Medline: 16283473]
36. SAS: Analytics, Artificial Intelligence and Data Management. 2012. SAS 9.4 URL: https://www.sas.com/en_ca/software/sas9.html [accessed 2020-04-22]
37. Richardson A, Graham AL, Cobb N, Xiao H, Mushro A, Abrams D, et al. Engagement promotes abstinence in a web-based cessation intervention: cohort study. J Med Internet Res 2013 Jan 28;15(1):e14 [FREE Full text] [doi: 10.2196/jmir.2277] [Medline: 23353649]
38. Graham AL, Bock BC, Cobb NK, Niaura R, Abrams DB. Characteristics of smokers reached and recruited to an internet smoking cessation trial: a case of denominators. Nicotine Tob Res 2006 Dec;8(Suppl 1):S43-S48 [FREE Full text] [doi: 10.1080/14622200601042521] [Medline: 17491170]
39. Bricker JB, Mull KE, McClure JB, Watson NL, Heffner JL. Improving quit rates of web-delivered interventions for smoking cessation: full-scale randomized trial of WebQuit.org versus Smokefree.gov. Addiction 2018 May;113(5):914-923 [FREE Full text] [doi: 10.1111/add.14127] [Medline: 29235186]
40. Fox S. Pew Research Center. 2011. Health Topics URL: https://www.pewresearch.org/internet/2011/02/01/health-topics-2/ [accessed 2020-04-22]
41. Cha S, Ganz O, Cohn AM, Ehlke SJ, Graham AL. Feasibility of biochemical verification in a web-based smoking cessation study. Addict Behav 2017 Oct;73:204-208 [FREE Full text] [doi: 10.1016/j.addbeh.2017.05.020] [Medline: 28551588]
42. Piñeiro B, Vidrine DJ, Wetter DW, Hoover DS, Frank-Pearce SG, Nguyen N, et al. Implementation of Ask-Advise-Connect in a safety net healthcare system: quitline treatment engagement and smoking cessation outcomes. Transl Behav Med 2020 Feb 3;10(1):163-167. [doi: 10.1093/tbm/iby108] [Medline: 30476236]
43. SRNT Subcommittee on Biochemical Verification. Biochemical verification of tobacco use and cessation. Nicotine Tob Res 2002 May;4(2):149-159. [doi: 10.1080/14622200210123581] [Medline: 12028847]
44. Graham AL, Papandonatos GD, Cha S, Erar B, Amato MS, Cobb NK, et al. Improving adherence to smoking cessation treatment: intervention effects in a web-based randomized trial. Nicotine Tob Res 2017 Mar 1;19(3):324-332 [FREE Full text] [doi: 10.1093/ntr/ntw282] [Medline: 27836982]

45. Cohn AM, Zhou Y, Cha S, Perreras L, Graham AL. Treatment engagement mediates the links between symptoms of anxiety, depression, and alcohol use disorder with abstinence among smokers registered on an internet cessation program. J Subst Abuse Treat 2019 Mar;98:59-65. [doi: 10.1016/j.jsat.2018.11.001] [Medline: 30665605]
46. Alkhaldi G, Hamilton FL, Lau R, Webster R, Michie S, Murray E. The effectiveness of prompts to promote engagement with digital interventions: a systematic review. J Med Internet Res 2016 Jan 8;18(1):e6 [FREE Full text] [doi: 10.2196/jmir.4790] [Medline: 26747176]
47. Przulj D, Hajek P, Snuggs S, McRobbie H. Changes in alcohol consumption during a stop-smoking attempt and differences between smokers using nicotine replacement and smokers using varenicline. Nicotine Tob Res 2018 Apr 2;20(5):583-588. [doi: 10.1093/ntr/ntx105] [Medline: 28521015]
48. Graham AL, Papandonatos GD, Cha S, Erar B, Amato MS. Improving adherence to smoking cessation treatment: smoking outcomes in a web-based randomized trial. Ann Behav Med 2018 Mar 15;52(4):331-341 [FREE Full text] [doi: 10.1093/abm/kax023] [Medline: 29878062]

Abbreviations

CSQ-8: client satisfaction questionnaire
EX: BecomeAnEX.org
EX-HD: EX with additional content specific to HD
EX-S: EX with standard content
GEE: generalized estimating equations
HD: heavy drinking
IP: Internet Protocol

Edited by G Eysenbach; submitted 21.08.19; peer-reviewed by M Mahoney, E Brett, V Vennu; comments to author 03.10.19; revised version received 29.01.20; accepted 29.03.20; published 08.06.20.

Please cite as:
Kahler CW, Cohn AM, Costantino C, Toll BA, Spillane NS, Graham AL
A Digital Smoking Cessation Program for Heavy Drinkers: Pilot Randomized Controlled Trial
JMIR Form Res 2020;4(6):e7570
URL: https://formative.jmir.org/2020/6/e7570
doi:10.2196/formative.7570
PMID:32348286

Original Paper

Low-Burden Mobile Monitoring, Intervention, and Real-Time Analysis Using the Wear-IT Framework: Example and Usability Study

Timothy R Brick[1], PhD; James Mundie[1], MA; Jonathan Weaver[1], BS; Robert Fraleigh[2], PhD; Zita Oravecz[3], PhD

[1]Department of Human Development and Family Studies, Real-Time Science Laboratory, The Pennsylvania State University, University Park, PA, United States

[2]Applied Research Laboratories, The Pennsylvania State University, University Park, PA, United States

[3]Department of Human Development and Family Studies, IMPEC Lab, The Pennsylvania State University, University Park, PA, United States

Corresponding Author:
Timothy R Brick, PhD
Department of Human Development and Family Studies
Real-Time Science Laboratory
The Pennsylvania State University
115 HHD Building
University Park, PA,
United States
Phone: 1 8148654868
Email: tbrick@psu.edu

Abstract

Background: Mobile health (mHealth) methods often rely on active input from participants, for example, in the form of self-report questionnaires delivered via web or smartphone, to measure health and behavioral indicators and deliver interventions in everyday life settings. For short-term studies or interventions, these techniques are deployed intensively, causing nontrivial participant burden. For cases where the goal is long-term maintenance, limited infrastructure exists to balance information needs with participant constraints. Yet, the increasing precision of passive sensors such as wearable physiology monitors, smartphone-based location history, and internet-of-things devices, in combination with statistical feature selection and adaptive interventions, have begun to make such things possible.

Objective: In this paper, we introduced *Wear-IT*, a smartphone app and cloud framework intended to begin addressing current limitations by allowing researchers to leverage commodity electronics and real-time decision making to optimize the amount of useful data collected while minimizing participant burden.

Methods: The *Wear-IT* framework uses real-time decision making to find more optimal tradeoffs between the utility of data collected and the burden placed on participants. *Wear-IT* integrates a variety of consumer-grade sensors and provides adaptive, personalized, and low-burden monitoring and intervention. Proof of concept examples are illustrated using artificial data. The results of qualitative interviews with users are provided.

Results: Participants provided positive feedback about the ease of use of studies conducted using the *Wear-IT* framework. Users expressed positivity about their overall experience with the framework and its utility for balancing burden and excitement about future studies that real-time processing will enable.

Conclusions: The *Wear-IT* framework uses a combination of passive monitoring, real-time processing, and adaptive assessment and intervention to provide a balance between high-quality data collection and low participant burden. The framework presents an opportunity to deploy adaptive assessment and intervention designs that use real-time processing and provides a platform to study and overcome the challenges of long-term mHealth intervention.

(JMIR Form Res 2020;4(6):e16072) doi:10.2196/16072

KEYWORDS

smartphone apps; ecological momentary assessment; real-time analysis; behavior change

Introduction

Background

One of the primary strengths of the mobile health (mHealth [1]) movement is its ability to deliver intervention and assessment where and when it is needed and to remain passive, imposing a minimal burden on the participant at other times. For example, ecological momentary interventions [2] and just-in-time adaptive interventions (JITAI [3]) are developed to deliver targeted, adaptive interventions to participants only when needed. This process minimizes the burden on the participants, removing an important barrier to engaging participants in interventions over extended periods and making it possible to efficiently target otherwise difficult long-term goals of maintenance, growth, and rare event prevention. Specifically, mHealth analyses and interventions have targeted maintenance goals such as sustained weight loss [4], treatment for chronic conditions such as generalized anxiety disorder [5], long-term processes such as drug addiction recovery [6], and protective and growth processes to promote thriving [7].

Another related strength of mHealth interventions is the ability to collect data about participants nearly continuously with minimal interruption to their everyday lives [8]. The success of these approaches depends on the ability to predict vulnerable states from these data by modeling the dynamic interactions of static and transient factors. As a wide array of behavioral and physiological data can only be collected via self-reports [9], repeated structured surveys remain the dominant data source [10].

Self-report surveys provide rich and valuable data, but they come at a nontrivial cost in terms of participant burden and dropout. A meta-analysis reported an average adherence level of 78% in children and adolescents [11], but variability is high depending on population, measurement, and engagement approaches. Although some recent examples show response rates over 90% [12], others report dropout of over half the sample [13]. Clearly related to this dropout is the balance between burden and incentive or engagement.

Although simple incentives such as participant payments may suffice for short-term studies, these do not scale to long-term interventions. Alternative approaches to improve engagement have used individualized feedback and visualization [14], badges [15], self-tracking [16], or self-experimentation [17]. Participant burden can be reduced, for example, via passive sensing tools [18]. Methods also exist to balance data cost with burden, for example, by reducing survey size through feature selection [19] or by modeling adherence propensity [20]. Research is still needed to optimize the interaction between participant and technology to maximize engagement and minimize burden while ensuring data quality.

Objectives

In this paper, we introduced the *Wear-IT* framework [21], a software system designed to study and overcome the challenges of long-term engagement in mHealth settings. Wear-IT uses a combination of passive and active sensors, novel computational architecture, visualization for individual feedback and self-monitoring, and real-time responsiveness to optimize participant and technology interaction. These tools allow the development of mHealth solutions for measurement and intervention to deliver treatment, derive scientific inferences, and promote individual thriving through lifelong engagement in mHealth apps.

Methods

Wear-IT

The *Wear-IT* framework [21] is a combination of web interfaces, cloud tools, and smartphone apps designed to carefully balance the data needs of researchers, medical doctors, and clinicians with the burden placed on participants. The overarching goal of the framework is to provide a testbed and deployment tool for the rapid iteration of novel approaches to measurement and intervention to understand human psychology, behavior, and health in real time. Wear-IT is currently deployed in targeted studies; a more general release is planned soon.

Approach

Wear-IT's general philosophy has been the distinctive focus on balancing the information received by a data source against the demands on the participant. The initial focus of the Wear-IT project is on increasing engagement and reducing the impact of the tools that are most likely to disrupt and interfere with the everyday lives of participants: structured surveys.

Wear-IT runs on participants' own smartphones (iOS or Android), avoiding the inconvenience of carrying an additional device and providing benefits for, for example, passive collection of phone usage data. As a result, power, network, drive space, and central processing unit (CPU) must be managed as limited resources and balanced, alongside participant burden, attention, and disengagement, against information gain.

Wear-IT concentrates on the following five primary themes to optimize the interaction between participant and technology: (1) intelligent and responsive scheduling to minimize interruption to daily life, (2) adaptive questionnaire design to reduce response time and effort, (3) custom question types to improve response precision, (4) individualized modeling to adapt to each participant, and (5) individualized visualization and feedback to improve intrinsic motivation and promote self-care and self-experimentation.

The *Wear-IT* framework follows a typical mobile app architecture consisting of four primary components: a smartphone app and sensor monitoring service; a web service for survey management, application programming interface (API) integration, and data collection; a cloud-based storage and computation service for real-time processing; and a visualization generation and delivery engine.

Smartphone App

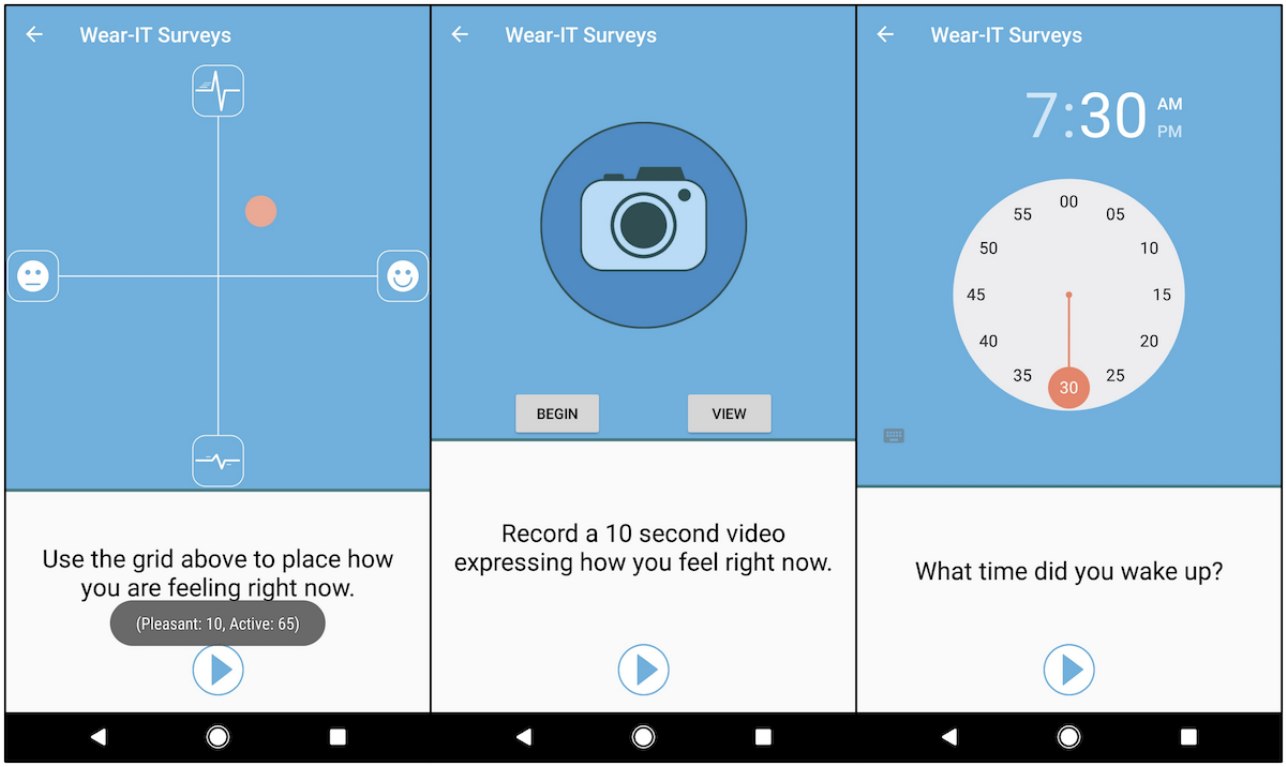
Wear-IT has been developed from the ground up to target the tradeoff between resource load and data collected. The smartphone app, therefore, accesses a combination of passive and active sensing tools and is designed to integrate and manage the results of those sensors in real time following a model inspired by concepts from the field of *edge computing* [22].

In our approach, lightweight models run continuously *at the edge of the cloud*—that is, directly on the smartphone or interface device. These models provide efficient real-time responsiveness when network is unavailable but are constrained to have minimal impact on device resources. Heavier computation (eg, individualized model parameter estimation) occurs on the cloud, and the results are used to update the lightweight models.

For example, the app might constantly read streaming data from a wrist-based heart rate sensor and deliver a stressor questionnaire every time some indicator (say, heart rate variability, HRV) crossed a threshold. Later, data would be delivered to a cloud server and combined with historical data to make updated predictions and update or adjust the threshold for that individual. Wear-IT thus delivers real-time, network-independent responsiveness using up-to-date data models without draining resources on the participant's smartphone.

As a smartphone-native app, Wear-IT allows researchers and clinicians to develop, test, and deliver questions in formats that are less burdensome or more efficient than traditional approaches. For example, Figure 1 shows examples of three nontraditional survey question types: a two-dimensional *core affect* rating space [23] (similar to the core affect grid in the study by Meers et al [24]), a multimedia recording prompt to collect *affective selfies* or video diaries [25], and an intuitive time picking question.

Figure 1. Three example questions from the Wear-IT framework. From left to right: core affect grid, affective video prompt, and time-picker.



Web Server and App Architecture

Wear-IT is designed to allow researchers to quickly create independent studies that customize how surveys and other interventions are delivered to the participants and to control how, when, where, and by what device(s) other data are collected. Researchers define the content of the study's surveys and provide a set of rules defining the conditions that trigger their delivery to the participant. Study rules also define what types of passive data are collected from the participant's smartphone and how those data are used. These rules are delivered from a central web server to participant devices in the form of JSON-encoded attributes from a RESTful web API and then interpreted by the Wear-IT smartphone *app* on the wearer's device. This configuration mimics typical mobile content delivery methods and effectively decouples survey content from survey presentation, allowing for a more robust delivery mechanism. Updates to the Wear-IT app for new versions of mobile operating systems (OSs) or new devices (eg, tablets, wearables that run apps, or internet-of-things [IoT] platforms) require only that an appropriate app be built for that device using that meets the modular requirements of the web service. Thus, Wear-IT is able to adapt to a wide variety of commercially available consumer electronic devices owned by participants. As of the time of writing, Wear-IT has successfully drawn data from devices manufactured by Google, Apple, Garmin, Oura, Fossil, and Empatica; development and testing are still in progress for devices from Fitbit and other third-party devices running Android's WearOS.

The web server is responsible for distributing survey definitions and other study-dependent information, including managing the rules for how and when the smartphone app delivers surveys and other interventions and the limits of customization for those rules. The app then interprets those rules within the mobile context of individual participants. For example, a study might allocate a time frame for an end-of-the-day survey to be delivered but might allow the participant to customize (within

the app) the latest possible survey time to correspond with their usual bedtime. In another study, interventions might be triggered when the participant enters a geographic region of interest but also depend on the participant's physiological markers, for example, if heart rate is higher than a person-specific threshold. The app is responsible for monitoring the specific conditions, given the hardware configuration available, and delivering the survey. That is, the app might rely on OS-provided geofencing capability to determine when the participant entered the appropriate region or might be required to periodically measure GPS location and compute entry itself. Survey rules also provide specifications about cases where the signal is not available. For example, if a participant opted out of location tracking, they might be removed from the study or might face altered rules for survey delivery that focus on different predictors. The modular balance between server and app and the ability to adapt to different situations makes Wear-IT a versatile platform for data collection and intervention delivery.

The overall architecture of the Wear-IT program is centered around data collection items of different types, called *item types*, which are collected into data collection events. These events include any case where participant data are recorded, including timed surveys, passive collection events, or participant-triggered *one-button* labeling responses. Conceptually, a common scheduling process can be used to organize regular passive collection, random or semirandom, according to triggered events or through some combination of those characteristics. For example, a study might request GPS or Bluetooth-based proximity every 5 min, deliver four surveys per day such that one appears in each of the morning, midday, afternoon, and evening, and trigger additional surveys if a participant reports a craving for alcohol or enters an area indicated as a likely place for a drink. The server architecture converts these conceptual arguments into a common JSON format; the app on each smartphone platform is responsible for implementing the specific requests.

Note that not all data sources are available on every platform. For example, at the time of writing, limitations on beacon libraries reduce the ability of devices running recent versions of iOS to respond to the proximity of other iPhones or to respond to the usage of different app types; by contrast, these services are available on most Android devices. Researchers and clinicians must, therefore, balance the importance of these data streams against the burden of providing participants with an additional smartphone and the data privacy cost of collecting such data.

Wearable, Passive, and Emplaceable Data Collection Devices

The Wear-IT framework can integrate with a variety of sensors and devices, including wearables, passive smartphone sensors, and emplaced IoT devices such as Bluetooth low energy (BLE) beacons or digital home assistants. At the time of writing, most ongoing studies rely on smartphone-based self-report or self-recording alongside passive tracking of one or more of activity, sleep, heart rate or HRV, stress, smartphone app usage, text messages, location, and proximity to other participants or emplaced objects. It is worth noting, however, that the Wear-IT framework is designed to rapidly incorporate additional measures. For example, several of the Garmin devices we have tested include pulse oximetry. Although this data stream is in theory already available for users of the Wear-IT framework, we have not yet identified a use case for that data stream and so its use within the system remains essentially unapplied and untested.

Wearable Physiological Monitors

Wear-IT is capable of integrating with a number of consumer-grade off-the-shelf wearable devices that participants may already own and use daily, such as Fitbit, Garmin, Apple, or Android Wear watches. Although these commercially available wearables may not provide scientific-quality data, they can reduce the cost of data collection and lower participant burden by using socially acceptable devices that participants may already own and use. The individualized modeling approach [12] taken by Wear-IT's responsive assessment engine is designed to make it possible to *tune* to the differences in predictive power of the device's data results. Note that consumer devices come with their own drawbacks, including differences in real-time access, privacy, and precision. When validated data collection is necessary, Wear-IT is also able to integrate data from scientific wearables (eg, the Empatica E4 device [26]).

Wear-IT's approach to integration attempts to meet each device according to its own capabilities and affordances as developed by its manufacturer. As a framework for research, we have designed Wear-IT so that it accesses each tool in a modular fashion, such that common data streams can be harmonized across different devices. In practice, each device has different limitations and a different means of access. For example, the Empatica E4 wristband has a released software development kit for real-time monitoring that permits Wear-IT to directly interpret real-time data about movement, electrodermal activity, skin temperature, and heart rate. By contrast, the Oura ring [27] currently has no published direct access. Although the ring's smaller form factor and longer battery life are benefits, data must first be uploaded to the Oura servers and then downloaded to Wear-IT servers via OAuth2-authenticated API calls authorized by the participant in the app. As a result, the Oura ring's data cannot be used to trigger just-in-time interventions. It might, however, be used to adapt an exercise intervention to account for the previous night's sleep, where the Empatica's shorter battery life might make such adaptation difficult. Another tradeoff is at the level of control. Some devices running Android's WearOS, such as many Fossil smartwatches [28], may be able to actively trade off data density with battery life by explicitly controlling sampling frequency using an on-device companion app, which may also provide wrist-based controls. For other devices, such as Garmin watches, Wear-IT relies on the manufacturer's proprietary smartphone app to collect data from the device and transmit it to a web server; again, the Wear-IT server accesses the data regularly via OAuth2-authenticated API calls. As Wear-IT accesses these devices indirectly, it is not possible to guarantee real-time access to data, although, in general, data are accessible within a short period (eg, a few hours). This later access may be sufficient for cases where, for example, the evening survey is altered based on a person's daytime location or their peak stress level

throughout the day, but may not be sufficient to trigger a mindfulness intervention in response to a stress event. These types of tradeoffs must be managed individually on a per-device basis.

In theory, common data types should yield identical results regardless of the device that makes up the source of the data. In practice, however, there are nontrivial differences in the precision, frequency, and preprocessing pipelines applied to data from different devices. Although the goal of Wear-IT is to seamlessly integrate with an array of devices, at present, data integration is performed on a case-by-case basis. In cases where scientific goals require specific data quality or where integration is difficult for other reasons, devices may need to be provided to participants to ensure consistency or equivalence or limits may need to be applied to the types of data collected based on available sensors.

Passive Smartphone Sensors

Owing to their always-on and often-connected nature, smartphones provide a convenient way to gather data that can be used to make automated inferences about the participant's context with little or no burden and a high degree of accuracy. Modern phone-based data sources such as GPS location, proximity to other participants, and phone usage can be collected through efficient on-phone APIs that have minimal impact on battery life or CPU. Note that some collection, such as raw accelerometry, may come with costs (eg, drive space) that may also need to be balanced.

Emplaced Devices

Not all devices that have therapeutic importance are wearable in nature. With the increasing availability of IoT devices, researchers have begun to use motion recording [29] and passive sensors such as BLE beacons [30] to provide vital sensor streams for research and clinical applications. Although most of these sensors must be handled with custom analysis, a few provide clear applications. For example, BLE beacons can provide quick identification of distance to other phones and locations, for example, to identify engagement with a 12-step program or to trigger assessment when a participant enters an area that may constitute a potential relapse trigger. IoT devices such as the Amazon Echo also may provide easy forms of interaction with the participant, whereas tools such as Bluetooth scales or bed-based weight monitors [31] can measure other important health characteristics with minimal participant burden. Again, the specific implementation of these various tools may depend on the specific application and use case. At present, the Wear-IT app framework has limited but increasing ability to interface with emplaced measurement devices.

Real-Time Adaptation for Responsive Assessment and Intervention

In addition to assembling information for later analysis, the large quantity of passive data collected by the Wear-IT framework is intended to serve as a basis for real-time decision making, for both the framework internally and for clinicians working with the participant. Specifically, Wear-IT aims to leverage low-burden data to determine when and how to collect data that exacts a higher participant burden. Wear-IT's computing architecture provides passive monitoring with high responsivity and can integrate the results of this monitoring with other active data collection to decide when and how to request new information.

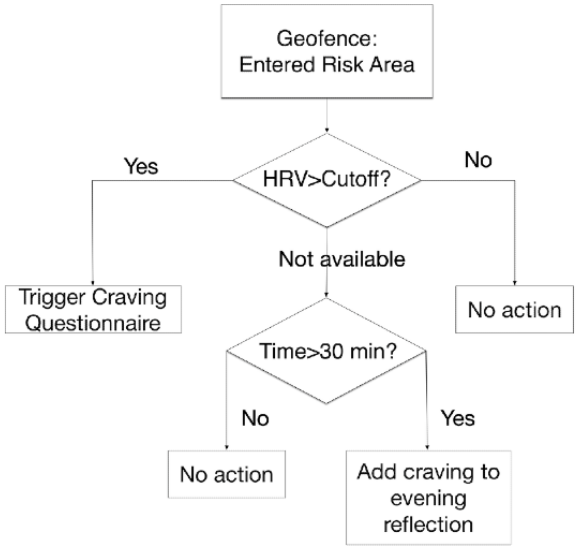
For example, in the beginning of addiction recovery, regular tracking of craving, affect, and sleep are helpful; an intensive assessment strategy may be needed to keep the patient on task. After several years of recovery, however, the patient may only rarely encounter states of high craving and risk, most of which are tied to encounters with, for example, contexts that trigger memories of substance use, such as places of previous high usage [32]. Constant four times a day Ecological Momentary Assessment (EMA) monitoring of craving state at this point would be needlessly burdensome. Instead, Wear-IT is designed to collect passive data to predict risk states and to use it to adapt assessment timing and intensity to match that risk. For example, an additional craving assessment might be delivered when the participant returns to a high-risk area (eg, a bar) and shows signs of physiological stress. The combined active and passive data can then be used to deliver interventions (eg, a mindfulness intervention and an evening recommendation to attend a 12-step meeting). Of particular note, these events and risk states may frequently occur where network access is limited, especially for rural participants.

To promote this type of adaptive responding, Wear-IT provides *contextually driven prompts*: assessments or interventions triggered based on a tree of decision rules. Typically, assessments delivered via contextual prompts are costly in terms of some finite resources such as participant burden, burnout, or test-retest reliability; risks such as disclosure, regulatory concerns, or privacy violations; or technical limitations such as battery, storage, or processing. Rules may combine transformations of contextual factors extracted from the continuous data stream, such as time of day, location, or proximity to other participants, with prior active or passive data and parameters of heavier computations from the cloud server. One simplified but common example might be an EMA data collection (costly in terms of participant burden), which is triggered by a randomized timer. More intricate collections are restricted by decision rules using arbitrary combinations of continuous or periodic data. For example, the app might request information about a location once the participant has visited the same place for a length of time that is determined by a predictive model from the cloud and adjusted regularly as updated estimates become available (eg, following an approach such as MIDDLE [33]). Simple linear models such as logistic regression or even support vector machine–based classification can be implemented in this same decision tree–like framework by simply replacing the single value (eg, heart rate is higher than 100) with a formula-based combination of values. For example, a survey might trigger when the weighted sum B_1×HeartRate+B_2×HRV is higher than a threshold T, where B_1, B_2, and T have been determined by fitting a logistic regression or support vector machine. Ongoing research has begun to incorporate more advanced models, such as Bayesian sequential updating models [34], vector autoregressive models [35], and automated feature selection results [19] into this process,

although integrating these more intricate tools into Wear-IT is still a work in process.

In brief, these models are kept relatively lightweight by using sequential updating methods whenever possible, by relying on existing OS-level functions when available (eg, for GPS), and by tuning the temporal precision of modeling whenever possible. For example, the rule set provided in Figure 2 uses a preselected set of identified *geofences*—that is, areas of the map that have been precomputed as spaces of interest. Rather than call upon constant GPS tracking (high in battery cost) and performing comparisons in the app itself, Wear-IT can rely on existing OS processes to add geofencing. As the OS manufacturers customize this process to the phone hardware already, we benefit immediately from the result. In this particular case, HRV is also a condition of triggered feedback. As the study more generally records HRV at regular intervals (eg, every time heart rate is sampled), the most recent version can be cached in the app for quick lookup when it is required. If the most recent sample is not sufficiently current, a new sample can be requested via Bluetooth (given availability). This type of active triggering would allow both the wearable and app to remain in a mostly passive state and reduce cost in terms of battery while still remaining responsive and contextually aware.

Figure 2. A hypothetical example of a decision tree for triggering different interventions. HRV: heart rate variability.



To provide an example, a responsive intervention (costly in participant burden) in recovery from substance use disorder might monitor proximity to areas where a participant previously reported high cravings (eg, bars or areas they previously used drugs [32]). When in proximity to those areas, the app might begin to monitor HRV from a wearable as a symptom of stress [36], and trigger an intervention if a threshold is crossed. An example of such a decision tree is shown in Figure 2. Notice that the threshold cutoff is not a fixed quantity: this might be changed for each individual and updated weekly from the server to reflect changes in baseline HRV. As another example, automated audio recording [37] (costly in terms of disclosure risk) for a client in addiction recovery might be suppressed when the participant was in a location used for private group therapy.

Evaluation Approach

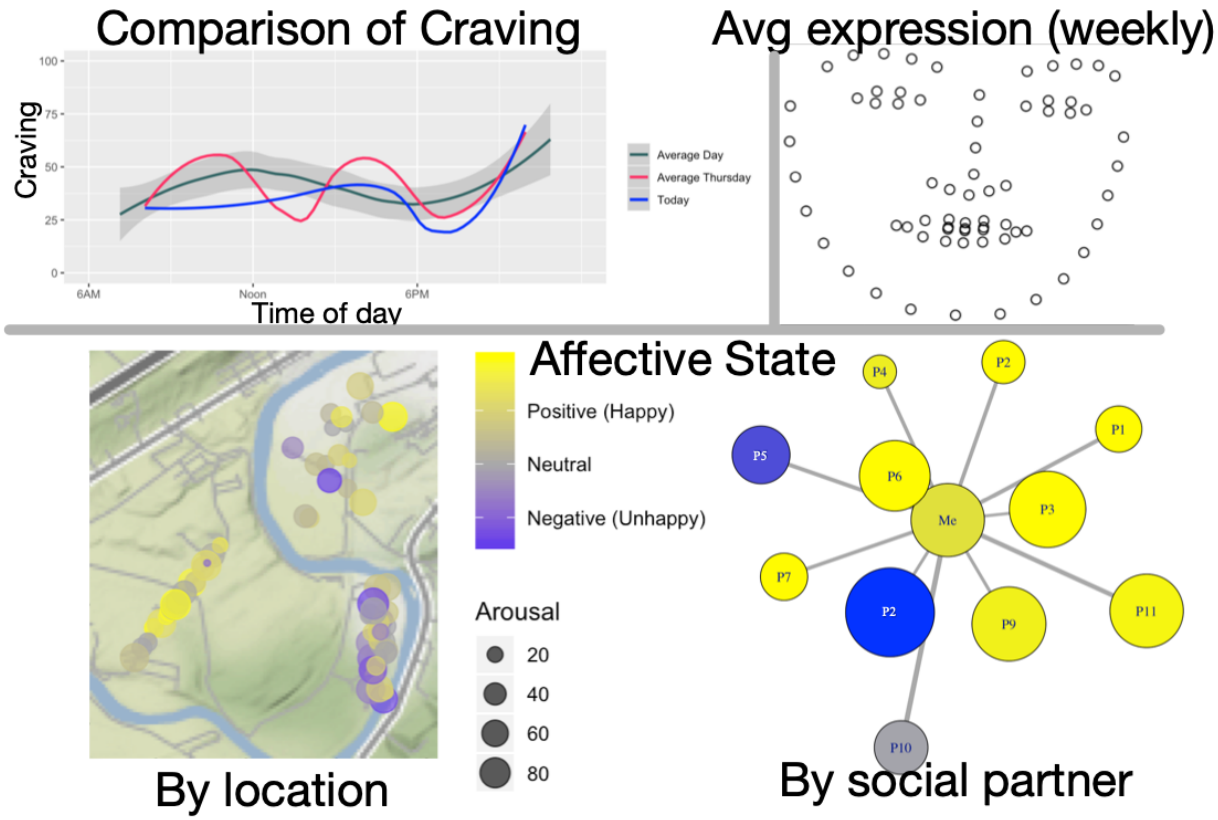
We provide two approaches to evaluation of the Wear-IT framework. First, we present a case study of visualization and feedback as a proof of concept and to demonstrate the types of data that Wear-IT can collect and process. Second, we held qualitative interviews with 4 users who are currently deploying studies using the Wear-IT framework. Interviewees included 3 faculty members and 1 graduate student in different departments across the health and behavioral sciences.

Results

Visualization and Feedback

Individualized visualization and feedback have been shown to be effective both as a means of engagement [15] and treatment [14]. Although these visualizations must be tailored to the specific study or treatment, a variety of tools are available to display different types of data, such as those displayed in Figure 3. Note that these data are provided for demonstration purposes only and are modified from a true dataset to protect privacy. These do not represent research data and should not be used to drawn generalizable inferences; rather they are intended only to display the types of visualizations in use.

Figure 3. Example visualization of momentary affect and craving from Wear-IT.



In the top left, a multiline comparison shows a psychological state (here craving, but potentially affect, anxiety, or some other psychological state of interest) across time. The current day (*today*) is compared with other similar days (*average Thursday*) and all days overall (*average day*). Top right shows an averaged facial expression of all emotional expressions (eg, based on selfies) recorded over the course of the week. The bottom two panels show affect (left: both valence and arousal) distributed across location and in response to different interactors over the week (right: closer nodes to the center indicate longer times spent). Together, these tools allow participants to examine the influences on their lives (eg, persons, places, and times). For example, the lower right area of the map indicates an area in which the participant experienced common negative affect; approaching this location might be the right time to deliver an intervention. In the lower left area, persons P2 and P5 on the network plot are persons who induce negative affect; the participant might seek out coping strategies to improve their relationship with these two people or learn to avoid them whenever possible. Similarly, rules might be added to Wear-IT to deliver appropriate JITAIs to assist the participant in applying appropriate coping strategies when they found themselves in such a location or near a given individual.

Qualitative Interview

In total, 3 faculty members and 2 graduate students agreed to provide feedback in a qualitative interview to relate their experiences using the Wear-IT framework. Note that the resulting data are for product evaluation, and not a part of a formal study. It is not intended to be used for generalizable inference but rather to provide evaluation of the framework in its status as of early 2020. Some researchers additionally reported anonymous quotations from participants engaged in their studies—these data were collected under institutional review board (IRB) approval for those studies using Wear-IT. For clarity, we use the term *users*, meaning researchers using the Wear-IT framework, and *participants* to indicate participants in their studies.

None of the users or participants whose comments are described here were part of the development team or authors on this paper.

Overall, users were highly positive about the ease with which the platform allowed them to pilot, test, update, and deploy new data collection instruments (eg, to adjust survey questions). Participants, too, reported the platform was easy to use. One participant volunteered that they always felt comfortable responding to the prompts provided.

A primary theme across the interviews was the benefit of real-time processing and adaptive responding, including real-time visualizations. A particular point of interest was the ability to create adaptive assessments, although the lack of established best practices for this type of deployment limited ease of use for these more intricate cases. For example, one user on the one hand applauded the ability to deliver adaptive assessments based on proximity triggers but on the other expressed concerns about the complexity required to specify these rule sets (eg, how long must a participant be near a target before they are considered to be *close*).

Although most participants were not informed about the adaptive nature of scheduling, they also highlighted the benefits of adaptive assessment. Users highlighted that their participants were satisfied with the timing of survey assessments, with one participant expressing that they never had to worry about remembering to answer the surveys in question. Although we did not formally evaluate it, we hope that this results in reductions in data missingness in the future.

A second theme that emerged from interviews was excitement about the opportunities that the Wear-IT framework provided them in the design of future studies. Most discussion around this topic focused on the benefits of integrating social and behavioral context into upcoming assessments, the burden reduction inherent in context-responsive assessment or intervention delivery, and the benefits of triggered assessments, especially for cases where substantive questions were related to social interaction or to activities related to real-world locations (eg, recovery communities).

Discussion

Principal Findings

The Wear-IT framework provides a new approach to combining passive and active sensing using real-time processing. Although other approaches provide mobile frameworks for survey scheduling and delivery (eg, Ohmage [38,39] and Effortless Assessment of Risk States [40]) and passive data collection, and some tools include the ability to trigger deployment of assessments based on simple decision rules (eg, Sensus [41]), Wear-IT expands on this same primary goal by providing a deeper integration of real-time processing from the outset. The use of arbitrarily complex decision rules to trigger assessment or intervention delivery enables contextual markers more intricate than simple boundary conditions. As real-time processing of contextual information is used by more scientists, we expect that the challenges of specifying these more complex deployments will be simplified by a reduction to common practice. Wear-IT also integrates with a newer wave of wearable devices, such as the Oura ring, to provide new options for data collection that may more easily fit into participants' lifestyles.

Participant Privacy

The large amount of data collected by the Wear-IT project raises serious concerns about participant privacy and data security. Wear-IT is designed to be as private as possible while maintaining scientific precision and verifiability. At present, we follow specific guidelines to ensure this type of privacy. First, we protect participant confidentiality by leaving the Wear-IT app in a *white labeled* state whenever possible. The Wear-IT app provides a relatively consistent look and feel across studies in which it is deployed. This allows us to provide surveys and perform data collection without alerting incidental users of the phone to the potential goals of the study. That is, a participant using Wear-IT as a part of a study on addiction recovery will not give away the reason for their participation simply because they use the app itself. Second, we limit data collection to only that data that are required and limit the data available on the phone itself to those elements that are immediately necessary. All scientific data collection must be approved by an IRB to ensure scientific oversight, and any data storage follows the latest security standards and meets all appropriate regulations (eg, applicable requirements of the US Health Insurance Portability and Accountability Act or the European General Data Protection Regulation in Europe). Third, we provide users with as much capability to customize data collection as reasonably possible. This is generally done by requesting permissions for each type of data to be used (eg, GPS and proximity), and by providing users with preferences inside the app that allow them to disable each type of data collection as needed. Of course, limiting GPS data collection may reduce the ability of the app to respond to location-based cues. Hybrid approaches, where they do not conflict with scientific or clinical goals, may also be possible. For example, it is possible to provide users with a map of their GPS-recorded travels for the day and allow them to remove specific locations or routes before their data are uploaded to the server, although this may conflict with real-time responsiveness goals. Again, our goal is to balance participant risks and burden with the scientific and clinical goals of each project. Note that in some cases, it may even be possible to generate responses to data without uploading it.

Finally, no matter how many security precautions are taken, there is some risk that data are disclosed. Although we take as many precautions as possible to reduce this risk, it must still be recognized and balanced against gains. As a result, we consider the risk of disclosure to be another specific type of burden that must be balanced with the others. That is, just as there is burden on the user incurred by taking the time to respond to a survey, there is burden in the risk of disclosure incurred storing the result on the server. We follow the same principals in balancing this cost as we do the other types of cost (eg, technical costs such as battery life and psychological costs such as questionnaire burden). As with the others, we use a combination of context-sensitive adaptation, specific instructions, and real-time processing to limit the identifiability of data whenever possible.

As an example of this last approach, consider a study in which participants record daily *video diaries*, which are then processed to understand a participant's affective state. This type of data collection involves a variety of privacy risks. First, facial expression video of the participant themselves contains sensitive information about their identity, and there is some risk of the participant disclosing information that is more sensitive than they wish to share or information that might be overheard by others nearby. Second, the background behind the person's face includes potentially identifying location information. Third, it is possible that other people in the area might also be recorded, possibly without their express knowledge or permission. As with other cases, it is important to balance the value of the data against the risk and cost to the participant. A variety of options exist to manage this type of risk. For example, if we know the specific information we want to collect from the video, for example, affective state variables, all three concerns could be mitigated by applying facial expression tracking to the video on the phone and sending only a processed stream of coded affective signals (eg, happy and sad) or a version cropped only to the participant's face, in lieu of raw video. This might carry additional burden in the form of processor time and battery life

in exchange for greater privacy protection. In cases where this is not possible, video recording might be triggered via geofencing so that it is requested only when the user was in their home or via beacon proximity when they were in, for example, a private area of a recovery community. In this case, the generalizability of the experiences recorded might be limited in exchange for greater privacy protection. If the clinical or scientific needs still do not allow this type of adaptation, instructions might simply be provided, asking the participant to check their surroundings and seek out a space that limits those risks in which they can answer the questions. In any of these cases, participants should be provided with the opportunity to review the video after they record it and to delete it before upload if desired.

The rise of mHealth methodology has provided a large number of new tools and processes for scientific data collection and clinical intervention. The combination of passive and active sensing approaches available to these tools makes them perfect choices to balance the informational needs of scientific inquiry and adaptive intervention with the challenges related to participant engagement and burden to target the long-term requirements of treating chronic and long timescale processes. However, mHealth tools are rarely designed to collect data or deliver interventions for these long-term projects. In this paper, we presented the Wear-IT framework, an app framework designed to leverage real-time processing of active and passive measurement to optimally balance resources, engagement, and data quality for clinical intervention and scientific inquiry.

Acknowledgments

Some research reported in this publication was supported by the Intensive Longitudinal Health Behavior Cooperative Agreement Program funded jointly by the Office of Behavioral and Social Sciences Research (OBSSR), National Institute on Alcohol Abuse and Alcoholism (NIAAA), and National Institute of Biomedical Imaging and Bioengineering (NIBIB) under Award Number 1U24AA027684-02. This study was supported by the Penn State Social Science Research Institute and the Penn State Fund for Innovation.

Authors' Contributions

TB and ZO conceptualized the project. Authors JM and JW developed the software and managed the data. RF assisted with visualization design. TB oversaw data collection/simulation/anonymization, visualization, and analysis. All authors participated in manuscript writing and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med 2015 Apr 15;7(283):283rv3 [FREE Full text] [doi: 10.1126/scitranslmed.aaa3487] [Medline: 25877894]
2. 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]
3. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
4. Martin CK, Gilmore LA, Apolzan JW, Myers CA, Thomas DM, Redman LM. Smartloss: a personalized mobile health intervention for weight management and health promotion. JMIR Mhealth Uhealth 2016 Mar 16;4(1):e18 [FREE Full text] [doi: 10.2196/mhealth.5027] [Medline: 26983937]
5. Lipschitz J, Miller CJ, Hogan TP, Burdick KE, Lippin-Foster R, Simon SR, et al. Adoption of mobile apps for depression and anxiety: cross-sectional survey study on patient interest and barriers to engagement. JMIR Ment Health 2019 Jan 25;6(1):e11334 [FREE Full text] [doi: 10.2196/11334] [Medline: 30681968]
6. Quanbeck A, Gustafson DH, Marsch LA, Chih M, Kornfield R, McTavish F, et al. Implementing a mobile health system to integrate the treatment of addiction into primary care: a hybrid implementation-effectiveness study. J Med Internet Res 2018 Jan 30;20(1):e37 [FREE Full text] [doi: 10.2196/jmir.8928] [Medline: 29382624]
7. Oravecz Z, Brick TR. Associations between slow- and fast-timescale indicators of emotional functioning. Soc Psychol Personal Sci 2018;10(7):864-873. [doi: 10.1177/1948550618797128]
8. Trull TJ, Ebner-Priemer U. Ambulatory assessment. Annu Rev Clin Psychol 2013;9:151-176 [FREE Full text] [doi: 10.1146/annurev-clinpsy-050212-185510] [Medline: 23157450]
9. Choe EK, Abdullah S, Rabbi M, Thomaz E, Epstein DA, Cordeiro F, et al. Semi-automated tracking: a balanced approach for self-monitoring applications. IEEE Pervasive Comput 2017;16(1):74-84. [doi: 10.1109/mprv.2017.18]
10. Smyth JM, Juth V, Ma J, Sliwinski M. A slice of life: Ecologically valid methods for research on social relationships and health across the life span. Soc Personal Psychol Compass 2017;11(10):e12356. [doi: 10.1111/spc3.12356]

11. Wen CK, Schneider S, Stone AA, Spruijt-Metz D. Compliance with mobile ecological momentary assessment protocols in children and adolescents: a systematic review and meta-analysis. J Med Internet Res 2017 Apr 26;19(4):e132 [FREE Full text] [doi: 10.2196/jmir.6641] [Medline: 28446418]
12. Osotsi A, Oravecz Z, Li Q, Smyth J, Brick TR. Individualized modeling to distinguish between high and low arousal states using physiological data. J Healthc Inform Res 2020;4(1):91-109. [doi: 10.1007/s41666-019-00064-1]
13. Shapiro JR, Bauer S, Andrews E, Pisetsky E, Bulik-Sullivan B, Hamer RM, et al. Mobile therapy: Use of text-messaging in the treatment of bulimia nervosa. Int J Eat Disord 2010 Sep;43(6):513-519. [doi: 10.1002/eat.20744] [Medline: 19718672]
14. Miller MB, Leavens EL, Meier E, Lombardi N, Leffingwell TR. Enhancing the efficacy of computerized feedback interventions for college alcohol misuse: An exploratory randomized trial. J Consult Clin Psychol 2016 Feb;84(2):122-133. [doi: 10.1037/ccp0000066] [Medline: 26689302]
15. Rabbi M, Philyaw-Kotov M, Lee J, Mansour A, Dent L, Wang X, et al. SARA: a mobile app to engage users in health data collection. Proc ACM Int Conf Ubiquitous Comput 2017 Sep;2017:781-789 [FREE Full text] [doi: 10.1145/3123024.3125611] [Medline: 29503985]
16. Larsen JE, Eskelund K, Christiansen TB. arXiv preprint. 2017 Mar 9. Active Self-Tracking of Subjective Experience with a One-Button Wearable: A Case Study in Military PTSD URL: https://arxiv.org/abs/1703.03437 [accessed 2020-03-04]
17. Karkar R, Zia J, Vilardaga R, Mishra SR, Fogarty J, Munson SA, et al. A framework for self-experimentation in personalized health. J Am Med Inform Assoc 2016 May;23(3):440-448 [FREE Full text] [doi: 10.1093/jamia/ocv150] [Medline: 26644399]
18. Chatterjee S, Hovsepian K, Sarker H, Saleheen N, al'Absi M, Atluri G, et al. mCrave: continuous estimation of craving during smoking cessation. Proc ACM Int Conf Ubiquitous Comput 2016 Sep;2016:863-874 [FREE Full text] [doi: 10.1145/2971648.2971672] [Medline: 27990501]
19. Brick TR, Koffer RE, Gerstorf D, Ram N. Feature selection methods for optimal design of studies for developmental inquiry. J Gerontol B Psychol Sci Soc Sci 2017 Dec 15;73(1):113-123 [FREE Full text] [doi: 10.1093/geronb/gbx008] [Medline: 28164232]
20. Liao P, Dempsey W, Sarker H, Hossain SM, Al'absi M, Klasnja P, et al. Just-in-time but not too much: determining treatment timing in mobile health. Proc ACM Interact Mob Wearable Ubiquitous Technol 2018 Dec;2(4):pii: 179 [FREE Full text] [doi: 10.1145/3287057] [Medline: 30801052]
21. Brick, Mundie, Weaver, & Oravecz. Sites at Penn State.: Pennsylvania State University Real-time Science Laboratory; 2020. Wear-IT: Real Time Science in Real Life URL: https://sites.psu.edu/realtimescience/wear-it/ [accessed 2019-08-30]
22. Lopez PG, Montresor A, Epema D, Datta A, Higashino T, Iamnitchi A, et al. Edge-centric computing. SIGCOMM Comput Commun Rev 2015;45(5):37-42. [doi: 10.1145/2831347.2831354]
23. Russell JA. A circumplex model of affect. J Pers Soc Psychol 1980;39(6):1161-1178. [doi: 10.1037/h0077714]
24. Meers K, Dejonckheere E, Kalokerinos EK, Rummens K, Kuppens P. mobileQ: A free user-friendly application for collecting experience sampling data. Behav Res Methods 2020 Jan 2 [Epub ahead of print]. [doi: 10.3758/s13428-019-01330-1] [Medline: 31898294]
25. Yue Z, Litt E, Stern J, Baxter KK, Guan Z, Sharma N, et al. Photographing Information Needs: The Role of Photos in Experience Sampling Method-Style Research. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014 Presented at: CHI'14; April 26 - May 1, 2014; Ontario, Toronto, Canada p. 1545-1554. [doi: 10.1145/2556288.2557192]
26. Garbarino M, Lai M, Tognetti S, Picard R, Bender D. Empatica E3 - A Wearable Wireless Multi-Sensor Device for Real-Time Computerized Biofeedback and Data Acquisition. In: Proceedings of the 4th International Conference on Wireless Mobile Communication and Healthcare.: IEEE; 2014 Presented at: MOBIHEALTH'14; November 3-5, 2014; Athens, Greece. [doi: 10.4108/icst.mobihealth.2014.257418]
27. Oura Ring: The Most Accurate Sleep Tracker Available. 2019. URL: https://ouraring.com/ [accessed 2019-11-22]
28. Fossil Group. Fossil - The Official Site for Fossil Watches, Handbags, Jewelry & Accessories. 2019. URL: https://www.fossil.com/us/en.html [accessed 2019-11-22]
29. Planinc R, Kampel M. Introducing the use of depth data for fall detection. Pers Ubiquit Comput 2012;17(6):1063-1072. [doi: 10.1007/s00779-012-0552-z]
30. Wu L, Waber BN, Aral S, Brynjolfsson E, Pentland A. SSRN Papers. 2008 May 7. Mining Face-to-Face Interaction Networks using Sociometric Badges: Predicting Productivity in an IT Configuration Task URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1130251 [accessed 2020-03-06]
31. Su BY, Ho KC, Skubic M, Rosales L. Pulse rate estimation using hydraulic bed sensor. Conf Proc IEEE Eng Med Biol Soc 2012;2012:2587-2590. [doi: 10.1109/EMBC.2012.6346493] [Medline: 23366454]
32. Meyer RE. The Heroin Stimulus: Implications for a Theory of Addiction. New York: Springer-Verlag; 2012.
33. Boker SM, Brick TR, Pritikin JN, Wang Y, von Oertzen T, Brown D, et al. Maintained Individual Data Distributed Likelihood Estimation (MIDDLE). Multivariate Behav Res 2015;50(6):706-720 [FREE Full text] [doi: 10.1080/00273171.2015.1094387] [Medline: 26717128]
34. Oravecz Z, Huentelman M, Vandekerckhove J. Sequential Bayesian updating for big data. In: Jones MN, editor. Big Data In Cognitive Science. Basingstoke: Psychology Press; 2016:21-40.

35. Ji L, Chen M, Oravecz Z, Cummings EM, Lu Z, Chow S. A Bayesian vector autoregressive model with nonignorable missingness in dependent variables and covariates: development, evaluation, and application to family processes. Struct Equ Modeling 2019:1-26. [doi: 10.1080/10705511.2019.1623681]
36. Hovsepian K, al'Absi M, Ertin E, Kamarck T, Nakajima M, Kumar S. cStress: towards a gold standard for continuous stress assessment in the mobile environment. Proc ACM Int Conf Ubiquitous Comput 2015 Sep;2015:493-504 [FREE Full text] [doi: 10.1145/2750858.2807526] [Medline: 26543926]
37. Mehl MR. The Electronically Activated Recorder (EAR): a method for the naturalistic observation of daily social behavior. Curr Dir Psychol Sci 2017 Apr;26(2):184-190 [FREE Full text] [doi: 10.1177/0963721416680611] [Medline: 28529411]
38. Tangmunarunkit H, Kang J, Khalapyan Z, Ooms J, Ramanathan N, Estrin D, et al. Ohmage: a general and extensible end-to-end participatory sensing platform. ACM Trans Intell Syst Technol 2015;6(3):1-21. [doi: 10.1145/2717318]
39. Ramanathan N, Alquaddoomi F, Falaki H, George D, Hsieh CK, Jenkins J, et al. ohmage: An Open Mobile System for Activity and Experience Sampling. In: Proceedings of the 6th International Conference on Pervasive Computing Technologies for Healthcare. 2012 Presented at: PervasiveHealth'12; May 21-24, 2012; San Diego, United States. [doi: 10.4108/icst.pervasivehealth.2012.248705]
40. Lind MN, Byrne ML, Wicks G, Smidt AM, Allen NB. The Effortless Assessment of Risk States (EARS) Tool: An Interpersonal Approach to Mobile Sensing. JMIR Ment Health 2018 Aug 28;5(3):e10334 [FREE Full text] [doi: 10.2196/10334] [Medline: 30154072]
41. Xiong H, Huang Y, Barnes LE, Gerber MS. Sensus: A Cross-platform, General-purpose System for Mobile Crowdsensing in Human-Subject Studies. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing. 2016 Presented at: UbiComp'16; September 12 - 16, 2016; Heidelberg Germany p. 415-426. [doi: 10.1145/2971648.2971711]

Abbreviations

API: application programming interface
BLE: Bluetooth low energy
CPU: central processing unit
EMA: Ecological Momentary Assessment
HRV: heart rate variability
IoT: internet-of-things
IRB: institutional review board
JITAI: just-in-time adaptive interventions
mHealth: mobile health
OS: operating system

Edited by G Eysenbach; submitted 30.08.19; peer-reviewed by A Baglione, S Haghayegh; comments to author 01.10.19; revised version received 25.11.19; accepted 24.01.20; published 17.06.20.

Please cite as:
Brick TR, Mundie J, Weaver J, Fraleigh R, Oravecz Z
Low-Burden Mobile Monitoring, Intervention, and Real-Time Analysis Using the Wear-IT Framework: Example and Usability Study
JMIR Form Res 2020;4(6):e16072
URL: https://formative.jmir.org/2020/6/e16072
doi:10.2196/16072
PMID:32554373

Original Paper

Adherence of the #Here4U App – Military Version to Criteria for the Development of Rigorous Mental Health Apps

Brooke Linden[1], PhD; Linna Tam-Seto[1], PhD; Heather Stuart[1], PhD

Health Services and Policy Research Institute, Queen's University, Kingston, ON, Canada

Corresponding Author:
Brooke Linden, PhD
Health Services and Policy Research Institute
Queen's University
21 Arch Street
Kingston, ON, K7L 3L3
Canada
Phone: 1 613 533 6387
Email: brooke.linden@queensu.ca

Abstract

Background: Over the past several years, the emergence of mobile mental health apps has increased as a potential solution for populations who may face logistical and social barriers to traditional service delivery, including individuals connected to the military.

Objective: The goal of the #Here4U App – Military Version is to provide evidence-informed mental health support to members of Canada's military community, leveraging artificial intelligence in the form of IBM Canada's Watson Assistant to carry on unique text-based conversations with users, identify presenting mental health concerns, and refer users to self-help resources or recommend professional health care where appropriate.

Methods: As the availability and use of mental health apps has increased, so too has the list of recommendations and guidelines for efficacious development. We describe the development and testing conducted between 2018 and 2020 and assess the quality of the #Here4U App against 16 criteria for rigorous mental health app development, as identified by Bakker and colleagues in 2016.

Results: The #Here4U App – Military Version met the majority of Bakker and colleagues' criteria, with those unmet considered not applicable to this particular product or out of scope for research conducted to date. Notably, a formal evaluation of the efficacy of the app is a major priority moving forward.

Conclusions: The #Here4U App – Military Version is a promising new mental health e-solution for members of the Canadian Armed Forces community, filling many of the gaps left by traditional service delivery.

(JMIR Form Res 2020;4(6):e18890) doi:10.2196/18890

KEYWORDS

mental health services; telemedicine; mHealth; chatbot; e-solutions; Canadian Armed Forces; military health; mobile phone

Introduction

Background

Recent advances in and increased access to technology have created opportunities to shift mental health support services away from traditional person-delivered models toward those that are more technologically based, including those that may be provided through smartphone apps. The ease of use and availability of these apps are ideal for populations who may face logistical or social barriers to traditional service delivery, including members of the Canadian Armed Forces (CAF).

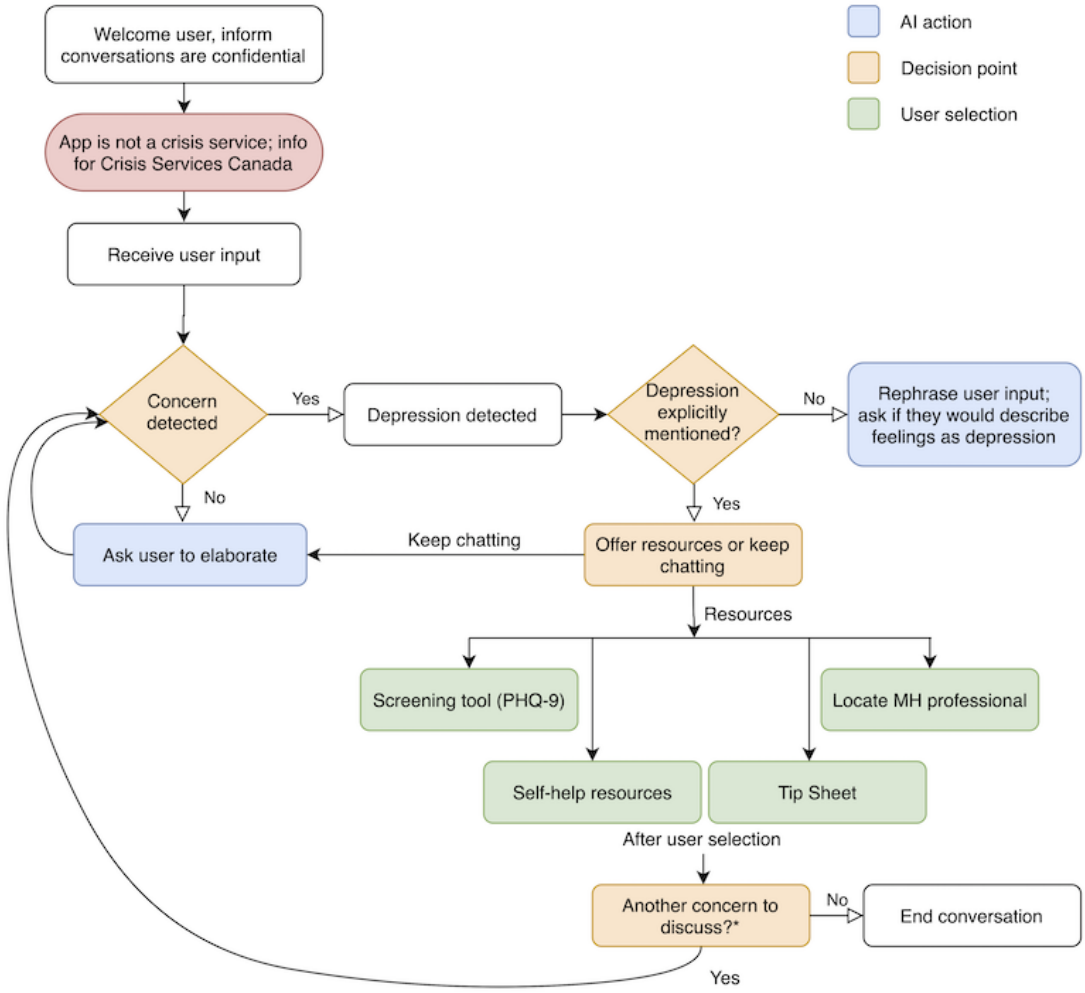
Military life is characterized by frequent relocations and family absences, and increased risk for illness, injury, and death [1]. Research has also identified how military life can affect health and well-being, including increasing risk for mental health difficulties and challenges with accessing health care services [2-4]. These events have been shown to impact all members of the community, including personnel [3,5,6], veterans [7], and their family members [8-10]. Despite the prevalence of mental health challenges faced by members of the CAF community, traditional mental health care continues to be underused. Commonly cited barriers include a preference for self-management [11], fear of impact on one's military career

[11], perceived stigma [12], and structural barriers such as living in remote locations and being frequently relocated [5]. As a result, a technology-based solution may better serve members of the military community.

The #Here4U App – Military Version (hereafter, #Here4U) is an app designed to provide evidence-informed mental health support to members of the CAF community, including serving personnel, veterans, and their adult family members. #Here4U uses artificial intelligence (AI) in the form of IBM's Watson Assistant to carry on unique text-based conversations with users, identify presenting mental health concerns, and refer users to self-help resources or recommend professional health care where appropriate. Users are invited to converse with the chatbot and select their end point (ie, self-help solution) from several options, facilitating a user-directed experience. Options include (1) tip sheets, which provide the user with brief mental health education on the selected topic; (2) use of validated screening tools, which provide the user with a self-assessment of the severity of psychological distress they may be experiencing; (3) information on how to contact local supports, including counsellors, psychologists, and psychiatrists; (4) frequently asked questions (FAQs) about common mental disorders, offered as downloadable PDF; and (5) self-help resources, ranging from cognitive behavioral therapy (CBT) education and activities to simple at-home exercises such as tactical breathing, meditation, and progressive muscle relaxation. The beta version of #Here4U has been designed to serve members of the CAF, veterans, and adult military family members, with plans to expand the app's applicability to additional target groups in the future. Figure 1 demonstrates the functionality of the app, using a flowchart design.

The #Here4U app was developed as an interdisciplinary project, in partnership with IBM Canada, the Queen's University Centre for Advanced Computing, Kingston, ON, Canada, and the Queen's University Health Services and Policy Research Institute. In addition to the research team, subject matter experts were engaged throughout the development process to share their experiences regarding common presenting mental health concerns among members of the CAF community, review resources provided by the app, and provide professional input and guidance on conversational flows. Subject matter experts included social workers, physicians (specializing in family medicine and psychiatry), psychologists, first responders, military family health researchers, child and youth workers, and members of the CAF community (active members, military spouses, family of adult military members, and veterans).

Figure 1. Demonstration of the basic functionality of the Here4U app, using depression as an example of a pressing concern. If the user has indicated an additional concern, the subsequent conversational flow (and end point options or resources provided) would change to reflect the new concern. *An additional concern may be expressed by the user at this point (AI prompt: "Is there anything else you wanted to chat about today?") or from appended additional concerns (eg, if more than one concern is detected at the beginning of the conversation). AI: artificial intelligence; MH: mental health; PHQ-9: Patient Health Questionnaire-9.



Objective

As the availability and use of mental health apps has increased, so too has the list of recommendations for efficacious development. In 2016, Bakker and colleagues published a set of 16 evidence-based recommendations to support the development of rigorous apps [13]. Our objective is to demonstrate the rigorous development and testing processes undertaken between 2018 and 2020 to produce the beta version of the #Here4U app and, in doing so, evaluate the degree to which the app complies with each of the criteria laid out by Bakker and colleagues.

Methods

Development of the App

The first year of development included background research to inform the content of the app, in addition to the initial development of conversational flows. First, we undertook a preliminary study to determine the most common presenting mental health challenges within the CAF community, as well as to derive particular language and slang used in the community to discuss mental health–related challenges. This was necessary to train IBM's Watson Assistant (IBM Corporation) to be able to recognize and interpret specific vernacular used within the community. We conducted a combination of online surveys (n=12), focus groups (n=17), and individual interviews (n=14) to achieve the goals of this phase of development, the results of which are described in detail elsewhere (available from the authors upon request). Ultimately, we identified 3 presenting concerns as the primary areas of focus: generalized anxiety, depression, and posttraumatic stress disorder (PTSD). We also identified several secondary concerns as priorities for later development (eg, substance use and addictive behaviors, general stress, difficulties related to sleep).

Next, we drew up initial logic trees to map likely conversational flows. These were then coded into the back end of the AI by members of the computing team. We initially designed

conversational flows for secondary concerns to be more streamlined (ie, to program the AI to recognize the concern expressed and immediately offer potential resources), whereas flows for primary concerns were more intricate. From here, we selected useful, efficacious end points for these conversations with input from subject matter experts (eg, validated screening tools, evidence-based self-help resources and activity recommendations, tip sheets to facilitate improved mental health education).

Beta Testing

During the second year of development, we used a snowball sampling method to gather participants to test the beta version of the app. Participants were a mix of subject matter experts involved in app development, attendees at the Canadian Institute for Military and Veteran Health Research Forum in 2019, members of the CAF community (eg, active members, veterans, military family members, staff at Military Family Resource Centres, mental health professionals working within the military community), and personal contacts of members of the research team who were connected to military research in some way (n=93). Users were asked to engage in conversations with the app, submit their comments regarding usability and suggestions for improvement, and identify any problems with the conversation dialogue. For example, testers identified instances in which the AI misunderstood a term or comment, or drew incorrect conclusions, such as placing the user into a conversational flow for anxiety when they were expressing concerns related to depression. This allowed us to refine and expand the AI's vocabulary, as well as to observe areas in need of improvement or expansion in the app. We made changes to the app in response to tester feedback using a waterfall approach, whereby we corrected conversational flows after receiving feedback from each tester.

Assessment of Rigorous Development

Finally, prior to conducting a formal evaluation of the app, we sought to assess whether we had met expectations regarding the recommendations for rigorous app development. We referred to Bakker and colleagues' 16 criteria, outlining the degree to which the #Here4U app met (or did not meet) each of these criteria, emphasizing relevant features and functions of the app in doing so.

Ethics

All stages of research conducted to support the development of the #Here4U app were approved by the Queen's University Health Sciences and Affiliated Teaching Hospitals Research Ethics Board. The research project was also approved by the Canadian Department of National Defence/CAF Social Science Research Review Board in accordance with Defense Administrative Orders & Directives 5062-0 and 5062-1 (#1172/18F).

Results

In this section, we address each of the 16 criteria put forward by Bakker and colleagues, highlighting where the #Here4U app met expectations and where selected criteria were not applicable.

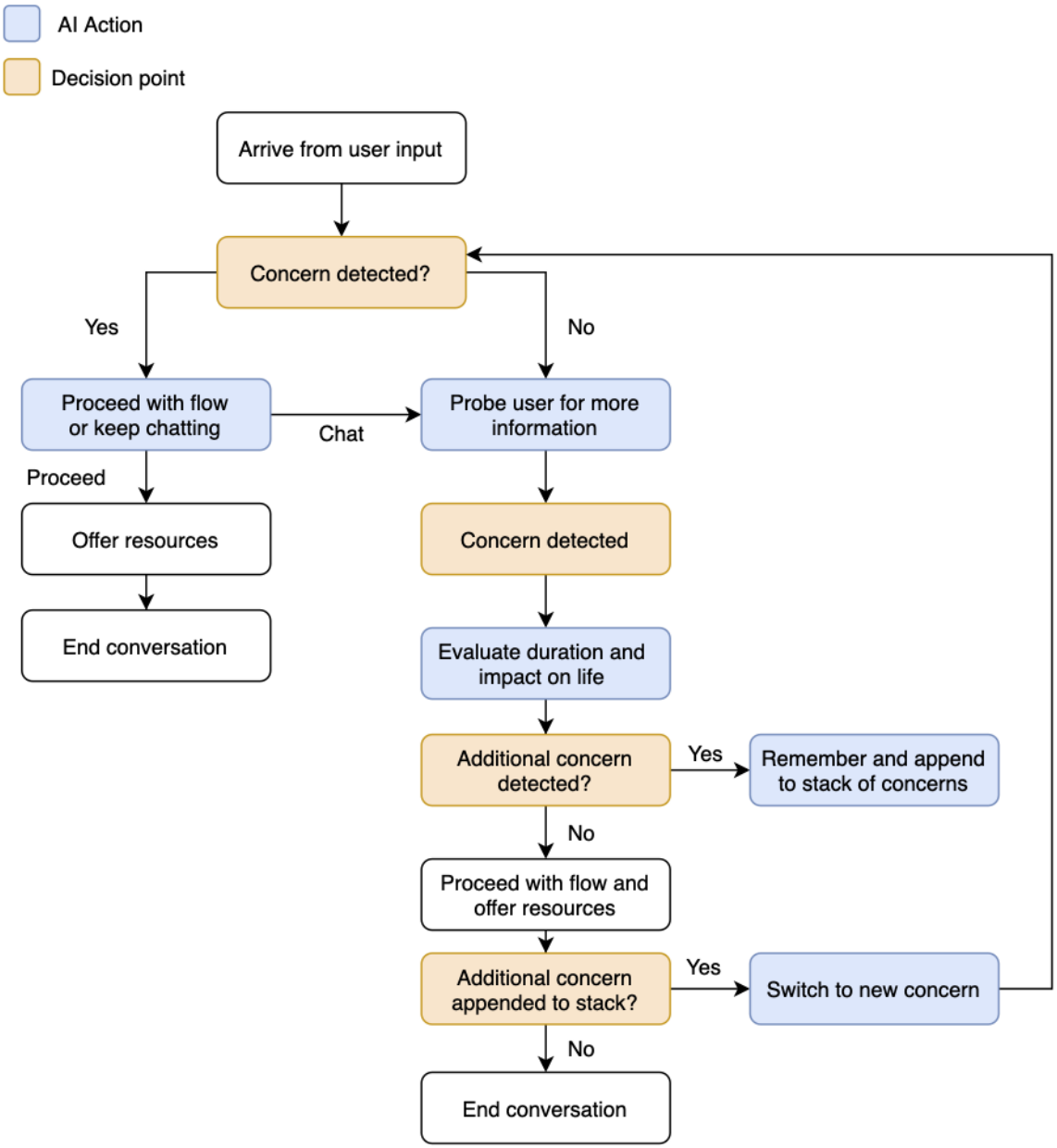
Recommendation 1: Cognitive Behavioral Therapy Based

CBT is a common therapeutic approach used to treat several mental illnesses, such as anxiety and depression, designed to help individuals develop more adaptive methods of thinking and processing emotions and behavior [14]. The #Here4U app met this recommended criterion by providing users with CBT-based self-help resources, in the form of exercises and self-help workbooks. Users are also invited to access psychoeducational resources regarding the purpose and principles of CBT, including what the cognitive behavioral model looks like, and how to challenge problematic thinking by using tools such as recognizing cognitive distortions, decatastrophizing, and reframing.

Recommendation 2: Address Both Anxiety and Low Mood

The #Here4U app met this criterion by addressing 3 major mental health concerns common among the CAF community: anxiety, depression, and PTSD. We selected these concerns after primary research conducted among members of the CAF community revealed these to be the most common mental health concerns. User-led conversational flows and self-help resources are provided for each of these areas of concern, in addition to validated screening measures for assessing degree of distress [15-17]. The app is also equipped to manage multiple presenting concerns at a given time, using a "loop-back" function. For example, anxiety, depression, and PTSD are often comorbid conditions, among both the general population [18] and the military population [19]. If a user were to express concerns associated with more than one mental health concern in a single conversation, the Here4U app would initially direct the conversation toward the first concern, later returning—or looping back—to discuss additional challenges. In the event that a user expresses both primary and secondary concerns, the app was designed to prioritize primary concerns (eg, depression, anxiety, and PTSD), prior to returning to address secondary concerns. Figure 2 depicts the loop-back function. The ability to address multiple concerns during a single conversation ensures that users receive the most complete mental health information, in addition to eliminating the need to juggle multiple apps or multiple sessions to address different concerns.

Figure 2. Demonstration of the loop-back function in the Here4U app. AI: artificial intelligence.



Recommendation 3: Designed for Use by Nonclinical Populations

The #Here4U app was specifically designed for use among nonclinical populations (ie, individuals who do not have a clinical diagnosis for a mental health disorder), with the app intended to be broadly applicable to a variety of individuals within the CAF community. While some of the screening tools contained in the app can be used to determine a clinical diagnosis when administered by a mental health professional (eg, the Patient Health Questionnaire-9), they are explicitly used here as measures of severity [16,17]. Apps that provide users with a likelihood of a clinical diagnosis have the potential to be harmful where users may be tempted to self-diagnose a mental illness when they may simply be experiencing low levels of distress [20]. Rather than providing a diagnosis or treatment, the #Here4U app was intended as an upstream measure (eg, with focus laid on mental health promotion and mental illness prevention). While the app can assist with directing users who may be experiencing more serious levels of psychological distress to a mental health professional for additional enquiry, the resources offered within the app were designed to provide users with the education to self-manage nonclinical, low levels of distress.

Recommendation 4: Automated Tailoring

The #Here4U app achieved automated tailoring through the use of IBM's Watson Assistant AI agent, which is capable of interpreting user concerns (eg, referred to technically as entities, or key words, and intents, or key phrases) and connecting users to the most appropriate conversational flow. With this function, users do not have to sort through menu items to locate the information that is most relevant or useful to them.

As previously noted, the #Here4U app was designed to respond to 3 primary mental health concerns (anxiety, depression, and PTSD), each with its own end point solutions offered, rather than using a one-size-fits-all approach. Secondary conversational flows within the app allow for even more detailed tailoring. For example, changes in appetite and sleep are commonly linked to depression and anxiety among members of the CAF community [21,22], while experiencing chronic stress or pain can similarly lead to mental health difficulties [23,24]. Additionally, substance use and addictive behaviors are often adopted as negative coping mechanisms [25-27]. Many members of the CAF community have reported experiencing family-related challenges, such as frequent absences, child rearing, supporting an injured family member, and frequent relocations. In addition to primary concerns, the #Here4U app also offers brief conversational flows that provide resources for users experiencing any of these secondary concerns.

Recommendation 5: Reporting of Thoughts, Feelings, or Behaviors

While the #Here4U app does not track or store user information (see recommendation 12), this criterion was met by providing users with self-help resources that encourage the reporting of thoughts, emotions, and behaviors to improve awareness of changes in one's mental health. For example, one option for users expressing concerns related to low mood is to begin using a daily gratitude journal, where users are invited to record three "good things" that happen every day. Users expressing challenges associated with stress management are provided with step-by-step guidelines on how to use progressive muscle relaxation as a stress reduction technique. All self-help resources provided through the app have been extracted or adapted from established programs (eg, Anxiety Canada) and have been reviewed by mental health professionals.

Recommendation 6: Recommend Activities

Many activities offered through the #Here4U app focus on mood improvement, such as encouraging users to engage in hobbies, leisure time, and physical activity and to spend time in nature. Behavioral activation is encouraged through CBT-based activities, including goal setting, cognitive distortion reframing, and decatastrophizing. Finally, users are invited to develop coping skills through improving their understanding of the importance of having a strong social support system and the adoption of coping activities such as progressive muscle relaxation, meditation, and tactical breathing.

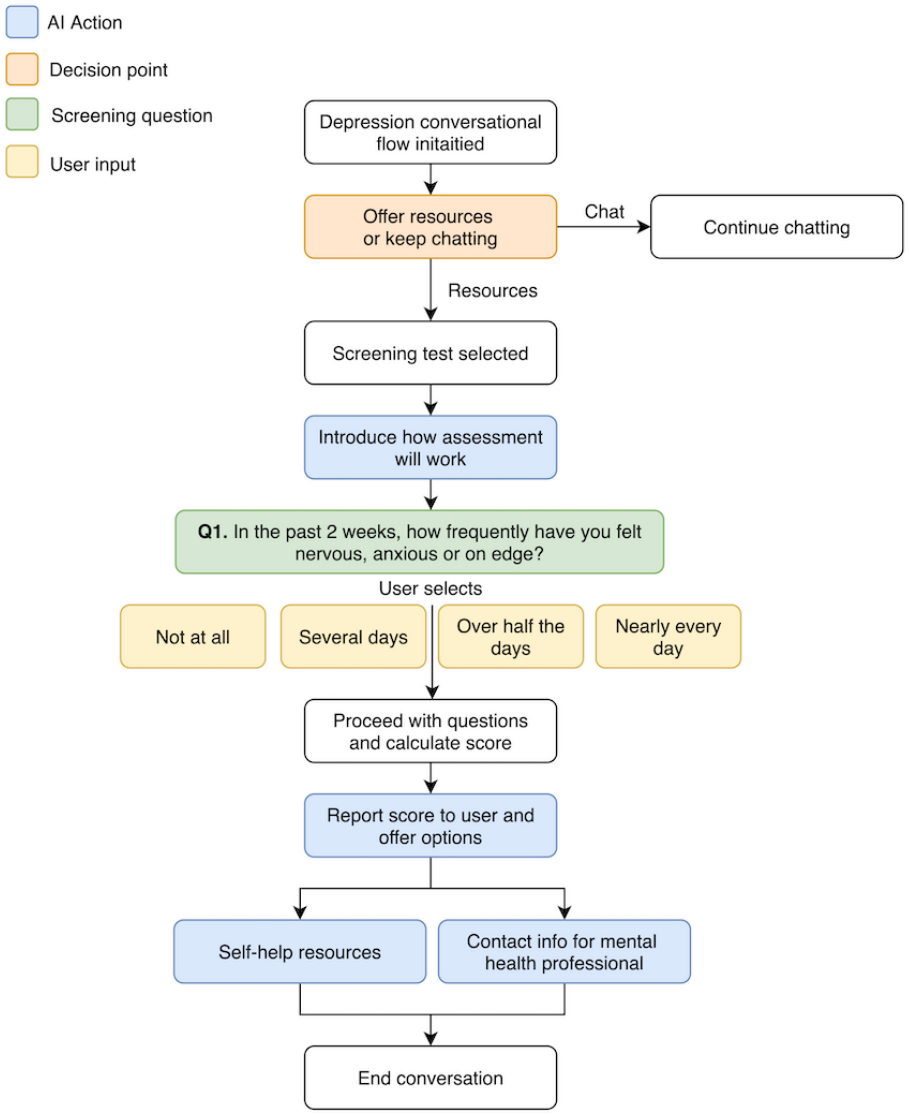
Recommendation 7: Mental Health Information

The #Here4U app met this criterion by providing users with psychoeducation by offering tip sheets for anxiety, depression, and PTSD, as well as more detailed FAQ documents that cover areas such as common symptoms, treatments, and how to help someone else who may be struggling. General education on CBT techniques for improving and managing mental health is also provided as an option for users. All materials were developed using evidence-based, reputable resources and have been reviewed by mental health professionals to ensure accuracy of information.

Recommendation 8: Real-Time Engagement

The #Here4U app capitalizes on the real-time capabilities of mobile devices by encouraging users to engage with the recommended self-help activities and resources immediately upon accessing them. Within the app, users can complete screening tests, which provides them with an idea of the degree to which they may be experiencing psychological distress, using an interactive, question-and-answer format (Figure 3). While these screening tests do not provide users with a diagnosis due to the risks outlined in recommendation 3, the app can also link users to the *Psychology Today* search engine to help them find a psychologist or counsellor in their local area [28] or to the Calian Military Family Health Portal [29], where users can get help finding a family physician. Finally, the 24/7 availability of the app encourages users to reach out and access resources in real time whenever they experience a change in their mental health.

Figure 3. Demonstration of the functionality of the question-and-answer screening test in the Here4U app. AI: artificial intelligence.



Recommendation 9: Activities Explicitly Linked to Specific Reported Mood Problems

With its tailoring capabilities, combined with the ability to address multiple presenting issues in a single session, the #Here4U app links reported mood challenges to relevant activities, rather than using a one-size-fits-all approach. While there is some overlap between concerns (eg, CBT educational materials are provided for all users), a user experiencing symptoms of depression would be directed toward activities and self-help resources specifically designed to address low mood. Bakker and colleagues argued that linking activities in this way can improve engagement and encourage habit-forming use of the app [13].

Recommendation 10: Encourage Nontechnology-Based Activities

The #Here4U app met this criterion by encouraging users to engage in a selection of nontechnology-based activities to improve their mental health (eg, engaging in regular physical activity, healthy nutrition, or self-care hobbies or spending time in nature). All of these recommended activities have been linked to reducing stress and improving low mood and anxiety [30-33].

Recommendation 11: Gamification and Intrinsic Motivation to Engage

Gamification, or the use of game-based mechanics to improve users' engagement, learning, and problem solving, can help to counteract problems related to lack of motivation and improve goal setting [13]. Currently, the #Here4U app does not contain any elements of gamification, but this may be an area for future development. Based on data derived from focus groups and interviews, gamification is a feature that potential users may find beneficial with respect to encouraging ongoing engagement with the app.

Recommendation 12: Log of Past App Use

During the first phase of research for the project, we consulted members of the CAF community and subject matter experts (eg, psychologists, counsellors, social workers) to determine, among other things, common barriers to using a mental health app. The most commonly cited barriers were the need for

privacy and confidentiality, most often driven by concerns regarding the potential for mental illness stigma resulting in damage to one's military reputation or career. For military families using the app, there was the concern that any kind of mental health challenge or family difficulty would reflect poorly on the military member. As a result of these findings, we made an informed decision to not implement any feature that would track app history. Bakker and colleagues mainly referred to this recommendation as a component of gamification, where users can track their progress week over week [13]. As the #Here4U app does not contain elements of gamification, logging past app use in this manner was not applicable.

Recommendation 13: Reminders to Engage

Though evidence suggests that providing users with reminders to engage (also known as push notifications) can help to improve and maintain engagement, too many reminders can also have the opposite effect, resulting in disengagement [13]. Given that users of the #Here4U app will be discussing sensitive topics related to mental health and emotional well-being, combined with the associated stigma within the military community [12] and a preexisting reluctance to seek help [3,5], we made an informed decision to not implement any push notifications.

Recommendation 14: Simple and Intuitive Interface and Interactions

The #Here4U app was designed to provide users with a sleek and simple interface. The app itself is driven by IBM's Watson Assistant AI, which carries on text-based conversations with users to determine areas of concern and suggest potential self-care resources and other solutions. All components of the app, including tip sheets and FAQs to drive mental health education, self-help resources, and other activities, are offered within a single interface. Conversations and access to self-help resources are user led, meaning that users choose how in-depth to go with content.

Recommendation 15: Links to Crisis Support Services

Because the #Here4U app was not intended for use as a crisis service, it was imperative to ensure that the AI had the ability to accurately recognize expressions related to self-harm and suicidality. In this case, users are reminded that the app is not equipped to assist with acute mental health crises and are directed to contact Crisis Services Canada, their primary health care provider, or local hospital.

Recommendation 16: Experimental Trials to Establish Efficacy

To date, no evaluation work has been conducted on the efficacy of the #Here4U app, and this is therefore a major priority moving forward.

Discussion

Principal Findings

To our knowledge, the #Here4U app met more of Bakker and colleagues' recommended criteria than any other available app targeting the mental health of military communities. Only 1 relevant criterion was not met: recommendation 16, to conduct experimental trials to establish efficacy of the app. While an evaluation study was out of scope for the development phase of the project, it is a recognizable gap for the #Here4U app and is therefore a major priority moving forward. Two other criteria were unmet, as they were not applicable to this project: recommendations 12 and 13, to maintain a log of past app use and to remind users to engage, respectively. Due to the sensitivity of the subject matter to be discussed within the app, combined with existing research regarding mental health–related stigma within the military community [12] and a preexisting reluctance to seek help [3,5], we made an informed decision early on to not implement any push notifications or information tracking within the #Here4U app.

Conclusion

Although not all criteria were met by the #Here4U app, Bakker and colleagues rightly noted that it may not be possible to design mental health apps that meet all suggested criteria. It is important to note, however, that those criteria that fall within the strongest ranks of evidence as determined by Bakker and colleagues (eg, recommendations 1-8) were all met by the #Here4U app [13]. Moving forward, an experimental evaluation study should be conducted to determine the efficacy of the app and user experiences more formally. The evaluation should incorporate assessments for effectiveness of the app in connecting users with mental health education and useful self-help resources, user interface performance and usability, and general user experience. Overall, the #Here4U app may be a promising new mental health e-solution for members of the CAF community, filling many of the gaps left by traditional service delivery.

Acknowledgments

The authors would like to acknowledge the herculean efforts of the entire #Here4U Development Team in developing and testing the app, as well as all subject matter experts who shared their time and expertise with us over the course of the project. Special thanks to IBM Canada Ltd, the Centre for Innovation in Military and Veteran Health Research, Mitacs, SOSCIP, and the Centre for Advanced Computing (Queen's University) for their support and contributions to the project.

The project was generously funded through the Canadian Institute for Military and Veteran Health Research Advanced Analytics Initiative. Funders included IBM Canada, Ltd and Mitacs. The future delivery method of the app and pricing decisions are at the discretion of IBM Canada.

Conflicts of Interest

BL and LTS were postdoctoral research fellows (funded by Mitacs) on the #Here4U project, and were actively involved in the development and testing processes. HS was the principal investigator on the project.

References

1. Daigle P. On the Homefront: Assessing the Well-Being of Canada's Military Families in the New Millennium. Ottawa, ON: Ombudsman, Department of National Defense, Canadian Armed Forces; 2013 Nov. URL: http://www.ombudsman.forces.gc.ca/assets/OMBUDSMAN_Internet/docs/en/mf-fm-eng.pdf [accessed 2020-05-25]
2. Cramm H, Norris D, Tam-Seto L, Eichler M, Smith-Evans K. Making military families in Canada a research priority. J Mil Veteran Fam Health 2015 Nov;1(2):8-12. [doi: 10.3138/jmvfh.3287]
3. Zamorski MA, Rusu C, Garber BG. Prevalence and correlates of mental health problems in Canadian Forces personnel who deployed in support of the mission in Afghanistan: findings from postdeployment screenings, 2009-2012. Can J Psychiatry 2014 Jun;59(6):319-326 [FREE Full text] [doi: 10.1177/070674371405900605] [Medline: 25007406]
4. Mahar AL, Aiken AB, Cramm H, Whitehead M, Groome P, Kurdyak P. Mental health services use trends in Canadian Veterans: a population-based retrospective cohort study in Ontario. Can J Psychiatry 2018 Jun;63(6):378-386 [FREE Full text] [doi: 10.1177/0706743717730826] [Medline: 28903578]
5. Pearson C, Zamorski M, Janz T. Mental Health of the Canadian Armed Forces. Ottawa, ON: Statistics Canada; 2014 Nov 25. URL: https://www150.statcan.gc.ca/n1/pub/82-624-x/2014001/article/14121-eng.pdf [accessed 2020-05-25]
6. Zamorski MA, Boulos D. The impact of the military mission in Afghanistan on mental health in the Canadian Armed Forces: a summary of research findings. Eur J Psychotraumatol 2014 Aug 14;5 [FREE Full text] [doi: 10.3402/ejpt.v5.23822] [Medline: 25206951]
7. Van Til L, Sweet J, Poirier A, McKinnon K, Sudom K, Dursun S. Well-Being of Canadian Regular Force Veterans, Findings from LASS 2016 survey. Charlottetown, PE: Veterans Affairs Canada. Research Directorate; 2017 Jun 23. URL: https://cimvhr.ca/vac-reports/data/reports/VanTil%202017_Well-Being%20of%20CAF%20Veterans%20LASS%202016.pdf [accessed 2020-05-25]
8. Cramm H, Mahar A, Tam-Seto L, Rowan-Legg A. Providing care to children and youth from military families. Ottawa, ON: Canadian Paediatric Society; 2018. Canadian Paediatric Surveillance Program 2017 Results URL: https://www.cpsp.cps.ca/uploads/publications/CPSP-2017-Results_1.pdf [accessed 2020-05-25]
9. Cramm H, Norris D, Schwartz KD, Tam-Seto L, Williams A, Mahar A. Impact of Canadian Armed Forces veterans' mental health problems on the family during the military to civilian transition. Milit Behav Health 2019 Aug 12;8(2):148-158. [doi: 10.1080/21635781.2019.1644260]
10. Manser L. State of Military Families in Canada: Issues Facing Regular Force Members and Their Families. Ottawa, ON: Canadian Forces Morale and Welfare Services; 2018 Aug. URL: https://www.cfmws.com/en/AboutUs/MFS/FamilyResearch/Documents/2018%20Research%20on%20Families/State%20of%20Military%20Families%20in%20Canada%20August%202018.pdf [accessed 2020-05-25]
11. Thériault F, Gabler K, Naicker K. Health and Lifestyle Information Survey of Canadian Forces Personnel 2013/2014. Regular Force Report. Ottawa, ON: Epidemiology Section, Directorate Force Health Protection Canadian Forces Health Services Group Headquarters; 2016 Sep. URL: https://www.canada.ca/content/dam/dnd-mdn/documents/health/health-and-lifestyle-survey-2013-2014.pdf [accessed 2020-04-22]
12. Weeks M, Zamorski MA, Rusu C, Colman I. Mental illness-related stigma in Canadian military and civilian populations: a comparison using population health survey data. Psychiatr Serv 2017 Jul 01;68(7):710-716. [doi: 10.1176/appi.ps.201600398] [Medline: 28245701]
13. Bakker D, Kazantzis N, Rickwood D, Rickard N. Mental health smartphone apps: review and evidence-based recommendations for future developments. JMIR Ment Health 2016;3(1):e7 [FREE Full text] [doi: 10.2196/mental.4984] [Medline: 26932350]
14. Fenn K, Byrne M. The key principles of cognitive behavioural therapy. InnovAiT 2013 Sep 06;6(9):579-585. [doi: 10.1177/1755738012471029]
15. Prins A, Bovin MJ, Smolenski DJ, Marx BP, Kimerling R, Jenkins-Guarnieri MA, et al. The Primary Care PTSD Screen for DSM-5 (PC-PTSD-5): development and evaluation within a veteran primary care sample. J Gen Intern Med 2016 Oct;31(10):1206-1211 [FREE Full text] [doi: 10.1007/s11606-016-3703-5] [Medline: 27170304]
16. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [Medline: 11556941]
17. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
18. Sartorius N, Ustün TB, Lecrubier Y, Wittchen HU. Depression comorbid with anxiety: results from the WHO study on psychological disorders in primary health care. Br J Psychiatry Suppl 1996 Jun(30):38-43. [Medline: 8864147]

19. Ginzburg K, Ein-Dor T, Solomon Z. Comorbidity of posttraumatic stress disorder, anxiety and depression: a 20-year longitudinal study of war veterans. J Affect Disord 2010 Jun;123(1-3):249-257. [doi: 10.1016/j.jad.2009.08.006] [Medline: 19765828]
20. Giles DC, Newbold J. Self- and other-diagnosis in user-led mental health online communities. Qual Health Res 2011 Mar;21(3):419-428. [doi: 10.1177/1049732310381388] [Medline: 20739589]
21. Smith TJ, White A, Hadden L, Young AJ, Marriott BP. Associations between mental health disorders and body mass index among military personnel. Am J Health Behav 2014 Jul;38(4):529-540. [doi: 10.5993/AJHB.38.4.6] [Medline: 24636115]
22. Taylor MK, Hilton SM, Campbell JS, Beckerley SE, Shobe KK, Drummond SPA, Behavioral Health Needs Assessment Team. Prevalence and mental health correlates of sleep disruption among military members serving in a combat zone. Mil Med 2014 Jul;179(7):744-751. [doi: 10.7205/MILMED-D-13-00551] [Medline: 25003859]
23. Vun E, Turner S, Sareen J, Mota N, Afifi TO, El-Gabalawy R. Prevalence of comorbid chronic pain and mental health conditions in Canadian Armed Forces active personnel: analysis of a cross-sectional survey. CMAJ Open 2018;6(4):E528-E536 [FREE Full text] [doi: 10.9778/cmajo.20180093] [Medline: 30389752]
24. Hourani LL, Williams TV, Kress AM. Stress, mental health, and job performance among active duty military personnel: findings from the 2002 Department of Defense Health-Related Behaviors Survey. Mil Med 2006 Sep;171(9):849-856. [doi: 10.7205/milmed.171.9.849] [Medline: 17036605]
25. Meshberg-Cohen S, Kachadourian L, Black AC, Rosen MI. Relationship between substance use and attitudes towards seeking professional psychological help among veterans filing PTSD claims. Addict Behav 2017 Nov;74:9-12. [doi: 10.1016/j.addbeh.2017.05.024] [Medline: 28558337]
26. Bray RM, Hourani LL. Substance use trends among active duty military personnel: findings from the United States Department of Defense Health Related Behavior Surveys, 1980-2005. Addiction 2007 Jul;102(7):1092-1101. [doi: 10.1111/j.1360-0443.2007.01841.x] [Medline: 17567397]
27. Tyler Boden M, Kimerling R, Kulkarni M, Bonn-Miller MO, Weaver C, Trafton J. Coping among military veterans with PTSD in substance use disorder treatment. J Subst Abuse Treat 2014 Aug;47(2):160-167. [doi: 10.1016/j.jsat.2014.03.006] [Medline: 24854218]
28. Psychology Today: find a therapist. New York, NY: Sussex Publishers, LLC; 2020. URL: https://www.psychologytoday.com/ca [accessed 2020-05-25]
29. Calian. Health solutions military family health portal. Ottawa, ON: Calian Group Ltd; 2020. URL: https://www.calian.com/en/services/health-services/mfhp [accessed 2020-05-25]
30. Adams K, Minogue V, Lucock M. Nutrition and mental health recovery. Ment Health Learn Disabil Res Pract 2010 Apr;7(1):43-57. [doi: 10.5920/mhldrp.2010.7143]
31. Maller C, Townsend M, Pryor A, Brown P, St Leger L. Healthy nature healthy people: 'contact with nature' as an upstream health promotion intervention for populations. Health Promot Int 2006 Mar;21(1):45-54. [doi: 10.1093/heapro/dai032] [Medline: 16373379]
32. Bratman GN, Hamilton JP, Daily GC. The impacts of nature experience on human cognitive function and mental health. Ann N Y Acad Sci 2012 Feb;1249:118-136. [doi: 10.1111/j.1749-6632.2011.06400.x] [Medline: 22320203]
33. Jacka FN, Berk M. Depression, diet and exercise. Med J Aust 2013 Sep 16;199(S6):S21-S23. [Medline: 25370279]

Abbreviations

AI: artificial intelligence
CAF: Canadian Armed Forces
CBT: cognitive behavioral therapy
FAQs: frequently asked questions
PTSD: posttraumatic stress disorder

Edited by J Torous, G Eysenbach; submitted 25.03.20; peer-reviewed by C Gissel, C El Morr; comments to author 06.04.20; revised version received 23.04.20; accepted 30.04.20; published 17.06.20.

Please cite as:
Linden B, Tam-Seto L, Stuart H
Adherence of the #Here4U App – Military Version to Criteria for the Development of Rigorous Mental Health Apps
JMIR Form Res 2020;4(6):e18890
URL: https://formative.jmir.org/2020/6/e18890
doi:10.2196/18890
PMID:32554374

Original Paper

Smartphone Self-Monitoring by Young Adolescents and Parents to Assess and Improve Family Functioning: Qualitative Feasibility Study

Dallas Swendeman[1*], PhD, MPH; Stephanie Sumstine[1*], MPH; Amber Brink[2*], MPH; Deborah Mindry[3*], PhD; Melissa Medich[4*], PhD, MPH; Michael Russell[5*], PhD

[1]Global Center for Children and Families, Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, University of California, Los Angeles, CA, United States

[2]Department of Community Health Sciences, Fielding School of Public Health, University of California, Los Angeles, CA, United States

[3]Center for Culture and Health, Department of Psychiatry and Biobehavioral Sciences, David Geffen School of Medicine, University of California, Los Angeles, CA, United States

[4]Department of Family Medicine, David Geffen School of Medicine, University of California, Los Angeles, CA, United States

[5]The Methodology Center, Department of Behavioral Health, The Pennsylvania State University, University Park, PA, United States

[*]all authors contributed equally

Corresponding Author:
Dallas Swendeman, PhD, MPH
Global Center for Children and Families, Department of Psychiatry and Biobehavioral Sciences
David Geffen School of Medicine
University of California
10920 Wilshire Blvd
Suite 350
Los Angeles, CA, 90024
United States
Phone: 1 3107948158
Email: dswendeman@mednet.ucla.edu

Abstract

Background: The natural integration of mobile phones into the daily routines of families provides novel opportunities to study and support family functioning and the quality of interactions between family members in real time.

Objective: This study aimed to examine user experiences of feasibility, acceptability, and reactivity (ie, changes in awareness and behaviors) of using a smartphone app for self-monitoring of family functioning with 36 participants across 15 family dyads and triads of young adolescents aged 10 to 14 years and their parents.

Methods: Participants were recruited from 2 family wellness centers in a middle-to-upper income shopping area and a low-income school site. Participants were instructed and prompted by alarms to complete ecological momentary assessments (EMAs) by using a smartphone app over 2 weeks 4 times daily (upon waking in the morning, afternoon, early evening, and end of day at bedtime). The domains assessed included parental monitoring and positive parenting, parent involvement and discipline, parent-child conflict and resolution, positive interactions and support, positive and negative affect, sleep, stress, family meals, and general child and family functioning. Qualitative interviews assessed user experiences generally and with prompts for positive and negative feedback.

Results: The participants were primarily white and Latino of mixed-income- and education levels. Children were aged 10 to 14 years, and parents had a mean age of 45 years (range 37-50). EMA response rates were high (95% to over 100%), likely because of cash incentives for EMA completion, engaging content per user feedback, and motivated sample from recruitment sites focused on social-emotional programs for family wellness. Some participants responded for up to 19 days, consistent with some user experience interview feedback of desires to continue participation for up to 3 or 4 weeks. Over 80% (25/31) of participants reported increased awareness of their families' daily routines and functioning of their families. Most also reported positive behavior changes in the following domains: decision making, parental monitoring, quantity and quality of time together, communication, self-regulation of stress and conflict, discipline, and sleep.

Conclusions: The results of this study support the feasibility and acceptability of using smartphone EMA by young adolescents and parents for assessing and self-monitoring family daily routines and interactions. The findings also suggest that smartphone self-monitoring may be a useful tool to support improvement in family functioning through functions of reflection on antecedents and consequences of situations, prompting positive and negative alternatives, seeding goals, and reinforcement by self-tracking for self-correction and self-rewards. Future studies should include larger samples with more diverse and higher-risk populations, longer study durations, the inclusion of passive phone sensors and peripheral biometric devices, and integration with counseling and parenting interventions and programs.

(JMIR Form Res 2020;4(6):e15777) doi:10.2196/15777

KEYWORDS

adolescents; parenting; conflict; self-monitoring; smartphones; mHealth; ecological momentary assessment; mobile phone

Introduction

Background

Research demonstrates that family processes in daily routines and settings have significant impacts on children's development and well-being [1-4]. The feelings, actions, and interpersonal interactions of individuals are structured by daily routines that influence the household and family. Thus, family routines provide a bridge between individual and systemic levels of the multilevel family system [3,4]. Key factors in daily family routines include parent-child communication and family interactions. Lack of parent-child communication has been associated with low life satisfaction for adolescents [5,6]. In contrast, parent-child conflict and perceived lack of support have been associated with negative psychological, social, and health risks for children (ie, depression) [7,8]. Conversely, positive family interactions have been linked to decreases in internalizing emotional distress [9,10]. Emotional states such as affect, conflict, and stress can also be transmitted between parents and their children [9,11,12]. Family stress can also negatively impact peer relationships of adolescents and school domains [13]. Fostering positive interactions, communication, support, and conflict resolution within families may better protect families from maladaptive outcomes such as depression, behavioral and school problems, lower self-esteem, and poor social skills [8].

Engaging families in therapeutic activities addressing family processes in real time during daily routines is a persistent challenge in interventions and research [14]. The broad proliferation of mobile phones creates novel opportunities for interventions and research modalities that are integrated into daily routines and are widely scalable. Self-monitoring is one strategy that can be easily implemented via smartphones. Early research on self-monitoring recognized reactivity to self-assessments as a means to support self-regulation and behavior change through feedback and goal-setting processes [15-18]. One form of self-monitoring is daily diaries and ecological momentary assessment (EMA). EMAs are repeated self-reports conducted multiple times throughout a day to assess behaviors, attitudes, states, and experiences in real time, in the natural environments of subjects [19]. EMA has greater ecological validity, fewer recall biases compared with observational or global questionnaire methods, and the capacity to elucidate within- and between-person processes and temporal dynamics [19]. For example, utilizing EMA in family interventions can allow researchers to examine the relationships between intrapersonal processes (ie, mood), interpersonal processes (ie, supportive or hostile exchanges), and broad family-level contexts (ie, family conflict, cohesion) that may address more complex and nuanced questions about sequential processes that influence behavior and affect in the daily lives of individuals [20].

EMA and diary methods have been used to study family experience in daily routines across multiple domains such as parent-child interactions [8], family relationships [21], family conflict [13], and stress [10,22]. Notably, the intensive nature of daily diaries and EMA may result in reactivity (ie, changes in awareness and behavior, particularly in populations motivated to change [19,23]). This is a methodological nuisance of basic behavioral research but presents a potential opportunity for self-monitoring as an ecological momentary intervention [23]. The little research done previously on reactivity has favored minimizing reactivity and its related effects [19], including in family research [24]. In general, EMA and diary research does not address reactivity routinely or robustly [23]. Most important to family research, EMA allows for real-time collection of data from multiple informants (ie, multiple family members) who often share the same natural environments, while also experiencing similar *events* (ie, family meals, arguments) [20]. Using EMA as an assessment tool in families allows different perceptions of the same experiences and the ability to identify discrepancies in perception.

Objectives

This paper examines the user experiences of families on feasibility, acceptability, and perceived benefits of self-monitoring and reactivity to smartphone EMA and daily diaries for assessment, self-monitoring, and as a potential tool for intervention to seed and support behavior change.

Methods

Sample and Recruitment

This study enrolled 36 participants across 15 families consisting of 15 children in 9 family dyads (all mother and child) and 6 family triads (mother, father, and child). The child participants included 6 boys and 9 girls aged between 10 and 14 years. Participants were recruited through a family wellness center's e-mail newsletter and website. The family wellness centers, funded by the Robert Wood Johnson Foundation, provided social and emotional learning and physical activities in a

metropolitan US community at a shopping area marketplace-based site (average income of US $67,000), and at a middle school located in a low-income neighborhood (average income of US $27,000) serving primarily Central American and Korean immigrant populations. Over one-third (n=6) of participant families came from the low-income site, 5 of which were Latino. Only 1 family from the middle-income site was Latino, whereas the rest were white, Asian, or African American. Prospective participants were informed that this was a pilot study to develop and test a smartphone app designed to enrich our understanding of ways to improve daily family routines and well-being.

Families who called the study contact were screened for the following eligibility criteria: parent and child coresided for some portion of the 2-week study period, at least one parent agreed to participate, the child was aged 10 to 14 years and gave assent to participate, and participants were fluent in English. Families with multiple children had the option to enroll again to participate with another child in the family (2 families exercised this option). Participants signed informed consent forms according to the university's institutional review board–approved protocols.

Procedures

After consent at an in-person meeting, participants were issued a smartphone (Samsung Galaxy S) for the study on which they completed EMA and daily diary surveys 4 times per day for 2 weeks. The study coordinator gave participants a brief training on how to use the smartphone and a step-by-step instructional manual on how to use the smartphone app platform to complete EMA surveys. All participants were given the study coordinator's phone number in case they had questions or experienced any difficulties while using the smartphone. At the end of the EMA period, qualitative interviews lasting approximately 40 min assessed the user experiences of parents and children, reactions to using the smartphone app, the obtrusiveness of the monitoring, any technical problems they encountered, relevance and usefulness of the EMA and diary questions, and perceived effects of study participation on them and their family. The semistructured interview guide first queried for general feedback and experiences, followed by prompts for "what was useful or helpful?" then "what was not helpful, or annoying?" and finally, suggestions for changes or improvements to the protocol and app. Participants also completed web-based questionnaires on demographic characteristics and family functioning at the start and end of their study participation. Participants received gift cards valued up to US $150 for completion of the different components of the study.

EMA and diary data were collected using Ohmage, an open-source mobile survey app supported by a web platform that supports the collection, storage, analysis, and visualization of EMA or self-monitoring data streams. Ohmage is a feature-rich and extensible platform that facilitates the collection of multidimensional, heterogeneous, and complex personal data streams. The software was programmed using time-based reminders to display question sequences and response choices on the smartphone screen. EMA survey responses were automatically timestamped, geotagged, and linked to the participant's assigned study identifier used as their login ID. Web interfaces were available for researchers to access and view participant data. The Ohmage user interface was designed based on feedback from behavioral and technology researchers focusing on group participants and end users of the system [25].

Participants were prompted to respond 4 times daily to EMA/diary surveys on the following domains: parental monitoring and positive parenting, parent involvement and discipline, parent-child conflict and resolution, positive interactions and support, positive and negative affect, sleep, stress, family meals, and general child and family functioning. Although many family assessment tools are widely available to researchers, clinicians, and families, none directly measure daily routines in real time. EMA/diary domains were chosen based on systematic reviews of standardized family functioning measures [26-29], which consistently assess communication, conflict, problem solving, cohesion or bonding, affect or emotion, organization, or regulation (eg, roles, rules, leadership, monitoring, and stress; see Table 1). EMA/diary questions were adapted from retrospective or global self-reported family measures. Domain and measure selection decisions were also informed by their use in intervention research with high-risk adolescents and the desire to balance with domains linked to resilience and wellness. Table 1 shows the EMA/diary domains and global/retrospective self-report measures that were adapted for EMA format. EMA/diary question contents are available as Multimedia Appendices 1-3.

The timing of the EMA vibration/ring prompts was scheduled by the participants and the study coordinator at times convenient for their individual schedules as follows: (1) morning upon awakening, (2) before school/work, (3) between the end of the school or work day and dinner, and (4) before bedtime. Upon hearing the reminder, participants were instructed to stop their current activity and complete a short (less than 5 min) EMA. Families received 1 phone call on the third day of the EMA period from the study coordinator to inquire about technical problems with the smartphone and app and answer any other study questions.

Table 1. Smartphone ecological momentary assessment and daily diary measures, sources, and schedule.

Domains	Measures and sources (parent and child, unless noted)	Schedule
Sleep	Number of hours and subjective quality (1=very poor, 5=very well)	Wake-up only
Stress	Single rating of current stress level (1=not, 5=very)	All
Affect and mood	Positive and negative affect schedule [30] and personal affect measure [31]	All
Monitoring/positive parenting	Stattin and Kerr parental monitoring questionnaire [32,33]	3 × (not wake-up)
Parent involvement and inconsistent discipline	Alabama parenting questionnaire [34,35]	3 × (not wake-up)
Parent-child conflict	Issues checklist [36,37] and network of relationships inventory (child) [38]	3 × (not wake-up) end of day only
Conflict resolution	Conflict tactics scale, resolution subscale [39]	3 × (not wake-up)
Positive interactions	Network of relationship inventory, companionship subscale [40]	End of day only
Family meals	Who do you eat with and doing other activities?	End of day only
Overall functioning	Outcome rating scale and child outcome rating scale [41]	End of day only

Data Analysis

Descriptive analyses for demographic characteristics were conducted using simple frequency distribution statistics in Stata 15.1 (StataCorp). The qualitative user-experience interviews were audio recorded and transcribed verbatim. Of the 36 participants, 31 had audio recorded interviews available for transcription (5 audio files were inadvertently erased before transcription), and transcripts were redacted to remove personal identifying information and uploaded to the Dedoose web-based mixed methods analysis platform (version 4.5.91, Sociocultural Research Consultants 2013). A grounded theory inductive approach was used to code the data to identify key themes that emerged from the data [42,43]. The coding scheme was developed by the lead anthropologist with 2 research assistants. The research assistants engaged in initial discussion around substantive codes emerging from the data and analytic categories that evolved into tangible themes; the generated codes were organized into broader, more conceptual themes. The lead anthropologist reviewed all themes identified by the research assistants for the discrepant cases. The codes were shared with the research team and revised over several iterations. Codes and excerpts were retained for analysis when there was agreement between the coders and authors.

Results

Children were on average aged 12 years (SD 1.44), mothers were 46.25 years (SD 3.81), and fathers were 40.33 years (SD 3.51). Approximately half were white (n=20), one-third were Latino (n=11), 9% (3/35) were Asian, and 3% (1/35) were black.

Tables 2 and 3 present more demographic results for children and parents, respectively. Response rates were high overall, including some participants who completed more EMAs than scheduled (prompted), either by reporting for more than 14 days or reporting more on some days to compensate for missed EMAs (typically for the previous day). Overall, the response rate excluding more than 4 EMAs in a day and more than fourteen days of reporting (ie, the on-time and per protocol response rate) was 96.2% (1941/2016), with children slightly lower at 95.1% (799/840) and parents slightly higher at 97.1% (1142/1176). Overall, 69% (25/36) of the participants had 100% or greater response rates; 60% of the children and 76% of the parents. The lowest response rate among children was 70% (39/56) and 79% (44/56) among parents. A total of 6 parents and 3 children responded for 16 to 19 days. In terms of missed EMAs, children tended to miss the morning and noontime EMAs, whereas parents tended to miss the late afternoon/early evening EMAs followed by the morning EMAs.

Qualitative results from the analysis of user-experience interviews are presented below based on 2 broad code themes and several subthemes that emerged from the data. The first broad code theme was feasibility, acceptability, and suggestions for the future, with the subcodes desire for feedback, seasonality, technical problems/challenges, survey burden, timing and frequency (weekends and duration), and global/recall web surveys. The second broad code theme was self-reflection, awareness, and seeds of change, with the subcodes decision making, parental monitoring, quality and quantity of time spent together, communication, self-regulation of stress and conflict, discipline, rewards and punishments, and sleep.

Table 2. Demographic characteristics of children at baseline (N=15).

Characteristic	Participants, n (%)
Age (years)	
10	3 (20)
11	2 (13)
12	4 (27)
13	3 (20)
14	3 (20)
Grade	
4	1 (7)
5	2 (13)
6	2 (13)
7	4 (27)
8	2 (13)
9	4 (27)
Ethnicity	
White	10 (67)
Latino or Hispanic	4 (27)
Black or African American	0 (0)
American Indian or Native American	0 (0)
Asian or Pacific Islander	2 (13)
Other	0 (0)

Table 3. Demographic characteristics of parents at baseline (N=20).

Characteristic	Participants
Age (years), mean (range)	
Informed	45 (37-50)
Missing	10 (50)
Gender, n (%)	
Female	14 (70)
Male	6 (30)
Marital status, n (%)	
Married	18 (90)
Separated	1 (5)
Living with a partner	1 (5)
Ethnicity, n (%)	
White	10 (50)
Latino or Hispanic	7 (35)
Black or African-American	1 (5)
Asian or Pacific Islander	2 (10)
Highest level of education, n (%)	
11th grade	1 (5)
12th grade	2 (10)
1-year trade, college, or university	1 (5)
3-year trade, college, or university	2 (10)
4-year trade, college, or university	4 (20)
Graduate education	9 (45)
Total family income (before taxes or other deductions; US $), n (%)	
10,000 to 29,999	2 (10)
30,000 to 49,999	2 (10)
50,000 to 69,999	4 (20)
70,000 to 89,999	1 (5)
90,000 to 99,999	1 (5)
100,000 and over	8 (40)
Missing	2 (10)
How difficult is it for you to live on your total household income right now?, n (%)	
Not at all difficult	9 (45)
A little difficult	6 (30)
Somewhat difficult	3 (15)
Very difficult	2 (10)
Do you have a spouse or partner with whom you share parenting responsibilities (for this child)?, n (%)	
Yes	19 (95)
No	0 (0)
Missing	1 (5)

Feasibility, Acceptability, and Suggestions for the Future

Participants found the smartphone self-monitoring feasible and acceptable and provided feedback for changes and improvements. Most participants reported enjoying their participation in the study:

> *Oh, it was, it was a really nice experience… it's like not every day you can talk to somebody about all the stuff they ask you in the app, in the surveys. Yeah… it was kind of a little fun to do 'cause I was like okay, just waiting to do it, you know? I've [sic] taken another kind of like testy thing. I always liked taking them and I hardly took 'em, so this was like, "Finally. Yay! I can do another one."* [male child (14 years), family 5]

The same participant also noted the convenience of smartphone surveys in comparison to web-based surveys on a computer:

> *I think it's better, kind of, to take the phone surveys because it's kind of way more convenient for people that are busy, 'cause they're shorter and you can take 'em wherever, anywhere you go. …Web surveys, you have to be on a computer, and you have to have time. And, so it may not be useful or good for people with a busy schedule.* [male child, 14 years, family 5]

Technical Problems and Challenges

Participants reported a few minor technical problems associated with the mobile phone, such as slow uploading of data or having to power the phone off and on to upload data. Some participants reported problems with the app freezing or force closing when they were trying to complete a survey. A few participants reported that they were not receiving reminders (alarms) to complete the surveys after a period.

Seasonality

Some participants noted that the survey questions needed to be geared toward the time of year. For example, for a number of participants who participated during the summer school break, questions about school and homework were inappropriate:

> *There were parts that didn't seem to apply because it seemed like the survey was sort of designed to assess children when they're in school. So, since it's summer, sometimes, you know, the questions didn't seem to apply, and then particularly with the, the final assessment, there was a lot of stuff about school. So, some things we didn't know how to respond to…* [mother, family 6]

Survey Burden: Timing and Frequency

With regard to the smartphone EMA, 10 participants reported finding the end of day bedtime survey burdensome because it was long, and several noted being tired:

> *The only thing that I would say is, I thought the evening one was really long. And, I think that the evening one should have been short. I think the length should have switched…[the] afternoon could have been the longer one, and the evening one could have been shorter, because I… I took it at like ten or whatever, and I was already really tired and I didn't really want to take a really long survey.* [male child, 13 years, family 8]

Some noted the value of the longer end of the day survey. For example, one mother noted the end of the day survey allowed her to reflect on her day:

> *Yeah, I mean, obviously, the last one of the day was longer, so that was the only thing. Like if I wasn't taking it at a decent hour, and I was super tired… that was the only, the only thing about that. But, I mean, I think that's when you have to have the long one because that's when you're really reflecting on everything that happened that day.* [mother, family 8]

Some participants also felt that there were just too many surveys to do each day:

> *Well, I think having to answer four times a day was too much. I would have preferred like breakfast, lunch and dinner. The wake up and the, you know, so that was a little annoying, just a little bit. And then, after so many days, it was like, "Oh my God. Did I do this?"* [mother, family 5]

Some participants noted the value of multiple surveys over the course of the day for transient states as a means of checking in and reflecting:

> *I think it was good timing 'cause, you know, in the morning, when you wake up, you have a different mood, and then in the afternoon, you have a different, and then you have different like every two hours or three.* [female child, 12 years, family 13]

Weekends

A few participants noted that it was difficult to complete surveys on weekends when daily activities and schedules are less routine:

> *I know that some days, it was more of a challenge to complete them kind of in the actual set time, especially on the weekends is really hard because by not being sort of on that same schedule.* [mother, family 1]

Duration

Most participants found the 2-week study period a good time period:

> *No, I think that's perfect [two weeks]. I think that was really, that was really good. It, it was just long enough to where it became a habit and you're makin' sure you aren't forgetting it. And then, but then towards the, near the end, you can just tell for, lookin' forward to just bein' finished 'cause it does become a burden…* [father, family 1]

However, a few participants suggested a longer duration, particularly to support behavior change:

> *I, to be honest with you, I found myself wanting to continue.…Two weeks would be a minimum. A month would be a maximum. You know, for me in my*

> *imaginary world - three weeks would be ideal, just because it, it gives you that, that first week - it was more about, "Hey," at least for me, it was like, "Okay. Let's just follow the behavior." The second week, it was like, "Okay. How am I gonna improve the behavior." But now, I don't have a chance to see if, to implement it, and to see how, how it's going. So, I was like, "Hey, okay. We have to give the phones back, alright." (laughter) So, that, that left, that left me wanting more.* [mother, family 2]

> *Probably about a week longer, but any more than that, I think it would have gotten like really repetitive and the answers would have all been the same.* [female child, 13 years, family 7]

Desire for Feedback

Participants noted that they would have liked the feedback from their participation in the study to know the findings from the survey and how these could help them and their families improve their relationships. Parents were very interested in getting more feedback from the study to improve their parenting skills and strategies. For example:

> *I think, eventually, if this was gonna be something more, intervention oriented, it would be interesting, that kind of assessment, even if it's something like, sort of the average, an average for the family and, and how that changes depending on, sleep or, I don't know. Changes in communication. I'm kind of more interested in just our overall picture as a family. And, where we need to improve…and I could sort of pick up on that based on the things that we were being asked.* [mother, family 6]

Self-Reflection, Awareness, and Seeds of Change

Over 80% (25/31) of the participants reported increased awareness of their relationship dynamics with their child/parent, their own behavior, or their communication styles. Of the 5 participants who did not report changes in awareness of their family routines, 4 were children. For many participants, study participation provided novel opportunities to reflect on their family routines in general:

> *I thought it was a very good exercise in terms of just self-reflection on things you do every day… you get on automatic pilot sometimes. You don't stop and really consider, "How, how did that go?" (laughter) So, I like that aspect because it sort of forced me to check in with myself and see how things were going.* [mother, family 1]

> *It's basically somethin' out of the ordinary, (laughter) to kind of express your feelings throughout the day. …But, to me, it was a productive kind of thing because, when you're thinkin' about what's goin' on for the day, you were able to reply too. I thought I was just holding it in, letting it out, you know, at the end of the week or somethin'.* [father, family 5]

One child noted increased awareness of his lack of contribution to household chores from 1 of the question prompts:

> *This question in particular, "I'm not as responsible as I should be." I answered that, I'm like, "You know what? I, I think I agree with that." And, that made me feel kind of bad. So, from, from that point on I started doing, now I unload the dishwasher.* [male child, 12 years, family 11]

These themes of reflection-seeding behavior change are represented throughout participants' feedback on their experiences in more specific domains (described below).

Decision Making

Several parents reflected on decision making in the family as a result of self-monitoring. For example, 1 noted:

> *Well, it made me think about how we decide is this something we're gonna give [daughter] input into? Or, is this something we just decide because, you know, we're the parents? It kind of made me think, sometimes do we give her too much input? (laughter)* [mother, family 10]

> *I see. You were becoming aware of how family decisions were…?* [interviewer]

> *Yeah, how family decisions were made I really wish my husband could have taken it. It would be good for him too. And, it was just kind of interesting to think about…* [mother, family 10]

Parental Monitoring

Children, in particular, described increased awareness of parental monitoring. For example, 1 child noted:

> *I realized my mom stays out of my life and I stay out of hers. (laughter)… 'Cause, I don't, I tell my mom some things….And, she never really asks what I'm doing. Like she, she pretty much knows like "he was outside," or "he was on the computer." She doesn't know specifically what I'm doing, but she's always got a vague idea, so….I think that's what I realized the most…some of the different questions like, "Did your parent know where you were at night?" …It made me realize, yeah, she pretty much always knows where I am, even when I'm not using my phone. So, I'm like, "I guess we are pretty connected."* [male child, 12 years, family 11]

Quality and Quantity of Time Spent Together

Participants reflected that the quantity and quality of time spent together were both important. One parent reported becoming aware of how seldom her family ate together:

> *It makes you think about stuff. Like, are we talking? Are we eating together? And, it's kind of embarrassing like, "Ugh," we eat alone, or so and so ate with so and so. You know? That was cool for us as a family to think, "Oh my gosh. We don't eat together."* [mother, family 5]

In addition to self-reflective functions, families noted how EMA/diary self-monitoring helped them to be more accountable or consistent with their values or goals for their families. For

example, 2 parents indicated an element of accountability from self-monitoring in regard to time with family:

> *The process and the idea behind it, I thought was actually really good, 'cause it helps you be more conscious about spending time with your child and in, especially 'cause you know you're gonna be reporting it later, so you want to make sure you have some good things to put in there, as opposed to bad. So, I think it helps people be conscious of that relationship.* [father, family 1]

This father noted how the anticipation of reporting influenced his behavior. One mother noted more direct reminder functions:

> *But… it does make you think that maybe you should… just be more in touch, communicate more. You know, one thing that struck me was… the leisure time, spending fun time with your child…I'm working full-time, I'm not gonna have that, but on the weekend, you know? … I think it does sort of, at least for me, remind me that I do need to spend fun time with my child….* [mother, family 6]

Her daughter also noted becoming aware of needing to spend more time with her parents and considering a change in her routine:

> *I'd think back on like how I reacted to some things or the interaction I had with my parents. Like maybe I should be out in the house more, not in my room.* [female child, 14 years, family 6]

Communication

Participants noted reflection, awareness, and some changes in patterns of communication in the family, including themes of limited time, positive or negative tone, praise or critique, and openness. For example, a mother stated:

> *I liked it [self-monitoring app]. It actually made me more aware of how my daughter and I communicate, and when we communicate.* [mother, family 10]

Another mother noted how she became aware of how little she communicated with her child:

> *But, in terms of the way I interact with my son, it was good because I didn't realize how little I actually communicated with him. So, in some ways, I felt good about myself because I felt like, "Oh, well, I think I really trust my son because I don't feel a need to constantly see what he's doing or talk to him." But then on the other hand, I thought, "Wow, two weeks has gone by and, really, we didn't make time to talk." So, it just made me a lot more aware.* [mother, family 11]

The son in this family also noted reactivity to self-monitoring and becoming aware of limited communication with his mother and that most often it involved giving him instructions:

> *I realized, I don't spend that much time communicating with her. Like it'll be like 10 minutes here, 20 minutes there, a minute there. Usually, it's just we talk for a second and say like, "You gotta do this." I'm like, "Okay." And then I go do that, and that's pretty much it.* [male child, 12 years, family 11]

Other families noted that self-monitoring reactivity led to more frequent communication:

> *I ended up speaking more to them - to my wife and my, my son, my other son too. Trying to eat dinner with them, trying to make more family time, you know?* [father, family 5]

A few participants noted increased awareness of needing to focus on positive communication and not just negative issues, and making changes based on this awareness:

> *I don't think that either S. [father] or I are big complimenters and,…it tends to run more negatively, which I think when you really think about it, you go, "Oh, that's not, that's not really how I want to be, and that's not how I feel, really." It's just trying to keep him afloat and making progress all the time, that you sort of get in these grooves. And so, I do think it was helpful, like I tried to… compliment him when he did something well.* [mother, family 8]

One father noted how the EMA questions also functioned as reminders for him:

> *I guess you don't always tell your kids that they're doin' a good job. Like I, after the surveys, I started makin' a point to tell him he was doin' a good job. (laughter) So, I guess that that's probably a good thing that the surveys tell you, is to remind you to remind, to let your kids know when they're doing well or not.* [father, family 1]

Another parent noted how reactivity to self-monitoring functioned by modeling questions that deepen levels of communication with her son around the domains assessed:

> *So, there is that element of both of us are doing these surveys and then, for me, part of it was, "Okay, I want to start asking you these questions since I don't really ask you these questions." So, I would actually ask him more about the ins and outs of his day more, and then he would talk to me more about it 'cause I was asking him for this information. …So, in a way, I was getting more information from him, and we were discussing more between the two of us than we would, normally, if I wasn't doing the survey.* [mother, family 3]

Similarly, her son reflected that their improved communication made him feel closer to his parent:

> *I enjoyed getting closer to my mom throughout the study because she'd always ask me questions that she wouldn't really normally ask me, and we got a little closer through that.* [male child, 13 years, family 3]

Self-Regulation of Stress and Conflict

EMA self-monitoring was also reported to support the self-regulation of stress and staying calm, including during conflicts and their resolution. For example, 1 parent noted that

self-monitoring helped her focus on trying to stay calm when interacting with her child:

> *Well, I tried to become more of being calm (laughter) and not yelling. And, that was really it. …I mean, obviously, if you're tired or you're, somethin' else is going on, it's hard to do that. But, when I was in a relaxed state, it made me mindful of, "Okay, when she does something that annoys you, just be calm about it, and try and work through it." It doesn't always happen. But, it did make me more cognizant to try and just be more patient and talk with her - depending on what it was.* [mother, family 10]

Similarly, a child noted that self-monitoring helped her and her mother identify what was making them angry and enabled them to resolve conflict:

> *When my mom and I would argue, and we would put it down [in the app], …we were like, "oh, we're actually mad about something," and then we would start thinking about like…Actually think about what happened.…and then we would talk and apologize.* [female child, 12 years, family 2]
>
> *And then you would talk and apologize. Is that something you did before?* [interviewer]
>
> *Not really because we wouldn't think about it, we'd just get mad (laughter).* [female child, 12 years, family 2]

One child became aware that she could not remember why she was angry with her parent:

> *I think before I took the survey, I just wouldn't think about why I was mad at her. I'd just be so mad. But then when I sat down and took the survey, and it was sayin' what was it about - I was like, "Wait. I don't even remember, anymore, what it was about."* [female child, 10 years, family 9]

One mother noted that the EMA self-monitoring helped her communicate more calmly with her child when they were in conflict by reflecting on and using conflict resolution strategies represented in the EMA response options:

> *Well, there's a certain question, for example, about, "Did you try to speak to your child calmly?"…The first time that I'm reading through them, I was like, "I don't know. Did I even try? (laughter) Did I just start yelling? Did I?" It…literally had me stop and, and take a step back and remember the whole scenario. You know, literally picture by picture, and break it down. And then, I caught myself, it's like, "I didn't even try." …and so after that, it was like, "Okay, let me try to speak to her calmly. Let me try and explain, you know, why the chair is yellow, and why the sky is blue." (laughter) ….And, towards the end of these last two weeks, I wasn't trying, I was speaking to her calmly.* [mother, family 2]

This mother also noticed her child using more positive reinforcement in her communication with a sibling, based on her initial reactivity to EMA and then her daughter modeling the behavior:

> *I noticed that she started it with her sister with the, "Hey, Sophie, I, thank you for helping me do the dishes for mom." You know, something like that out of the blue. Or, she had to throw the trash out and, "Oh, Sophie - thanks for taking the trash out. I, I really appreciate that."…It was a change. It was an actual, positive change, and it was after she caught on to what I was doing.* [mother, family 2]

Discipline, Rewards, and Punishments

Although parents were generally not very comfortable talking about disciplining their children, many discussed becoming more aware of patterns of discipline or rules in their families. Some parents were more comfortable talking about rewarding or praising their children and were consciously working on improving their positive reinforcement of their children's behavior. For example, a mother spoke about becoming more aware of limits and ground rules:

> *You know, and, and maybe because of the older one, she's testing limits. You know, and it's like, "Okay, but these are the ground rules. We have to establish ground rules… And yes, I can switch things around as you grow, and I can bend and lean, but, these are the ground rules and these are the consequences for these actions. And, vice versa - if you do good, you, you get good. If you do bad - you get bad." So, that's, I did it, and I hadn't even thought about that until right now that you asked me again. (laughter) But, I set them down and it's like, "Okay, you know what? Let's just be clear about that." And, I don't think I would have even thought about it if it wasn't asking me daily (laughter) about consequences and actions, and, and do I act out in anger, or if I'm tired, or you know?* [mother, family 2]

This mother also noted how reactivity to the EMA question and response content influenced her parenting behaviors. Another parent reported:

> *About the [questions], what happened when she did a good job or behaved well this morning? …About half-way through the week, I realized that, I guess I'm really hugging and kissing her when I'm feeling like she behaved well, but I'm not hugging and kissing her and saying, "Oh, you did such a good job."* [mother, family 9]

In this example, reactivity to self-monitoring resulted in moving away from awareness to changing her behavior to more actively reinforce her child's good behavior.

Sleep

Finally, both children and parents noted that the study made them realize the importance of sleep in their daily routines and self-monitoring:

> *It helped me 'cause it, with the questions…on the wake up, like how many hours of sleep, and I'd be tired, and I'd look at the time,...and then I'd be like, "Oh, I only got like seven hours." And then the day where I'd wake up better, I'd be like, "Oh, I had nine*

> *hours," so that's better for me.* [male child, 14 years, family 5]

Another parent gave an example of reflecting on the potential relationship between family conflict and sleep:

> *Like was there more conflict when I was sleep deprived? (laughter)* [mother, family 1]

In this example, self-monitoring reactivity first resulted in a strong awareness of sleep and functioning. One mother noted how responding to sleep questions seeded motivation:

> *I thought the sleep one was a very good question, 'cause then when I would answer it, I thought, "Ooh, yeah. I better get some more sleep."* [mother, family 11]

Overall, these results demonstrate how reactivity to EMA self-monitoring typically begins with increased awareness of behavior, followed by associations with antecedents or consequences. Then, some participants experienced motivation to change, with behavior changes supported by reminder functions of EMA prompts (alarms), accountability to subsequent reporting, and tracking of goal progress and outcomes.

Discussion

Principal Findings

The results of this study support the high feasibility and acceptability of using a smartphone EMA by young adolescents and parents for assessing and self-monitoring family daily routines and interactions over 2 weeks, as evidenced by high response rates of 95% and greater and in user-experience interviews. Some participants suggested that a third or fourth week of self-monitoring would further enhance the behavioral changes that they initiated. Some participants also reported preferring fewer surveys each day and fewer questions, particularly when considering a longer duration of self-monitoring beyond 2 to 3 weeks.

Our findings also suggest that smartphone self-monitoring may be a useful tool to support improvement in family functioning through functions of reflection on antecedents and consequences of situations, prompting positive and negative alternatives, seeding goals, and reinforcement by self-tracking for self-correction and self-rewards. These functions are core elements of self-regulation [14-17], which may now be enhanced by smartphone integration into daily routines. Reactivity in self-monitoring has been documented for a wide range of clinically relevant behaviors and may make an adjunctive contribution to intervention efforts [44]. The portability and convenience of smartphone integration into daily routines is creating novel opportunities to reinvigorate research on self-monitoring. Participants in this study reported increased awareness of their family routines, and many also reported behavioral changes in terms of decision making, parental monitoring, quantity and quality of time together, communication, self-regulation of stress and conflict, discipline, and sleep.

Our primary findings also suggest there was a potential indication of ethnic and/or income differences among parents in discussing discipline, rewards, and punishments that may warrant further exploration in future research. However, the small sample size does not warrant inferences as a primary result. White parents, from the higher-income site, seemed to be more comfortable discussing parenting practices with an emphasis on rewards and expectations and de-emphasizing punishments or negative reinforcements. Latino parents, from the low-income site, were more inclined to focus on parenting as a *job* to keep their children *safe* and seemed to be more comfortable discussing consequences or negative reinforcements. Some research has conceptualized that parenting practices and discipline may be moderated by the interaction of parental beliefs and ethnicity [45]. However, the small sample and recruitment sites confound not only ethnicity and income but also neighborhood safety, as the low-income neighborhood is noted for gang violence.

Limitations

This pilot study had several notable limitations. First, the sample size was small in terms of the number of families. Nonetheless, the overall number of participants is typical and acceptable for a qualitatively focused pilot study focused on user experience and feedback, and saturation of themes was achieved. The sample lacked representation from African American and Asian-American families, lower-income white families, and higher-income Latino families. Second, the family wellness center recruitment sites attracted families primed for motivation to improve their family functioning, as did the recruitment material framing the study as seeking support in developing and testing mobile apps to assess and improve family functioning. Third, our measure/domain selection reflects assumptions for successful parenting and interactions that are prevalent in family assessment tools and evidence-based interventions for risky adolescents, while also including domains reflecting resiliency. Fourth, it is important to note that the steep rise in mobile phone use among children and adolescents has also raised concerns about possible adverse effects such as addictive tendencies, depression, anxiety, sleep disruption, and cyberbullying [46,47]. Fifth, the smartphone app did not employ passive monitoring of smartphone usage, which was a separate module in the Ohmage smartphone app platform and was not used for both privacy and battery power preservation concerns. Notably, the study protocol and app did have an audio-sensing module, which continuously and passively monitored and classified the audio environment for speech versus nonspeech (eg, including discrimination of background audio from televisions) in small snippets of privacy-preserving nonaudio data, but the module failed to function outside of the laboratory because of data not being transferred off the phones quickly enough through a mobile connection (as opposed to Wi-Fi in the lab), causing the phones' operating system to crash. Only the first study family had Audiosens data, and for only several days before the phones crashed. Since the time of the study in 2013, improvements in smartphone technology, memory, and wireless data speeds would now make this component more feasible, and in fact, potentially advancing classification from speech/nonspeech to emotion detection. Finally, the study was

not powered for statistical analyses; the primary aims were feasibility, acceptability, user experience, and preliminary perceived efficacy of smartphone app self-monitoring for assessing and potentially improving family routines.

Future research should evaluate reactivity to EMA and diary self-monitoring as a tool to improve family routines in larger and more diverse samples of families, with statistical power to robustly examine behavioral, symptom, state, and functioning changes. Further research could also consider recruiting families coping with challenges such as chronic illnesses, substance abuse, or conduct problems. Future studies should also include longer follow-up periods to examine the sustainability of self-monitoring and decreasing burden over time and the significance of behavior changes indicated by participants in this study. Furthermore, future studies could also provide empirical and theoretical insight into how sibling relationships can serve as important contexts for individual development and family functioning. Due to participant feedback about frequency and duration of assessments, future studies using frequent assessments for highly transient states or frequent behaviors (eg, every 30 min) should be short in duration (eg, seconds to minutes) to minimize response burden over shorter assessment periods (ie, several days) [20]. Conversely, daily assessments may allow for a longer duration over longer periods of time (eg, weeks to months). Finally, future research assessing family functioning should ensure that assessments are adequately collecting data during times or situations of interest to families [20]. For example, if a study examines parent-child interactions, assessments should only occur when parents and children are together, such as mornings, evenings, and weekends (eg, not when at school/work). Future studies should include larger samples with more diverse and higher-risk populations, longer study durations, the inclusion of passive phone sensors and peripheral biometric devices, and integration with counseling and parenting interventions and programs.

Conclusions

Due to the increasing ease of implementing EMA and diary self-monitoring via smartphones, practitioners may reconsider using smartphones to enhance psychotherapy, parenting programs, and other counseling modalities. Conversely, researchers using EMA and diaries should examine reactivity more consistently and robustly. Real-time data visualization tools (eg, time trends, correlations, and maps) hold the potential to make self-monitoring more salient and actionable through use in counseling sessions for problem solving, feedback, praise, and goal refinement. Machine learning algorithms also hold promise for detecting patterns and anticipating change points to trigger automated in-the-moment or *just-in-time* interventions. Further research is needed on self-monitoring as a purely self-directed intervention activity and the potential for enhancing therapeutic relationships.

Acknowledgments

This research was supported by a grant from the William T Grant Foundation (number 180039) to DS. Other means of support were provided by the Center for HIV Identification, Prevention, and Treatment, National Institute of Mental Health (grant MH58107), the University of California, Los Angeles (UCLA), Center for AIDS Research (grant 5P30AI028697), the National Center for Advancing Translational Sciences through UCLA Clinical and Translational Science Institute (grant UL1TR000124), and the Beneventures Foundation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
EMA Parent Survey.
[DOCX File , 25 KB - formative_v4i6e15777_app1.docx]

Multimedia Appendix 2
EMA Child Survey.
[DOCX File , 21 KB - formative_v4i6e15777_app2.docx]

Multimedia Appendix 3
Post Interview Questions.
[DOCX File , 12 KB - formative_v4i6e15777_app3.docx]

References

1. Day RD, Padilla-Walker LM. Mother and father connectedness and involvement during early adolescence. J Fam Psychol 2009 Dec;23(6):900-904. [doi: 10.1037/a0016438] [Medline: 20001149]
2. Dickstein S. Family routines and rituals-the importance of family functioning: comment on the special section. J Fam Psychol 2002 Dec;16(4):441-444. [Medline: 12561289]

3. Friesen M, Woodward L, Horwood L, Fergusson D. Quality of parent-child relations in adolescence and later adult parenting outcomes. Soc Dev 2013 Aug;22(3):539-554. [doi: 10.1111/j.1467-9507.2012.00657.x]
4. Segrin C, Woszidlo A, Givertz M, Bauer A, Murphy M. The association between overparenting, parent-child communication, and entitlement and adaptive traits in adult children. J Fam Relat 2012 Apr;61(2):237-252. [doi: 10.1111/j.1741-3729.2011.00689.x]
5. Levin KA, Dallago L, Currie C. The association between adolescent life satisfaction, family structure, family affluence and gender differences in parent–child communication. Soc Indic Res 2011 Feb 12;106(2):287-305. [doi: 10.1007/s11205-011-9804-y]
6. Brown AM, Fitzgerald MM, Shipman K, Schneider R. Children's expectations of parent–child communication following interparental conflict: do parents talk to children about conflict? J Fam Viol 2007 Jun 17;22(6):407-412. [doi: 10.1007/s10896-007-9095-x]
7. Repetti RL, Robles TF, Reynolds B. Allostatic processes in the family. Dev Psychopathol 2011 Aug;23(3):921-938. [doi: 10.1017/S095457941100040X] [Medline: 21756442]
8. Sarigiani PA, Heath PA, Camarena PM. The significance of parental depressed mood for young adolescents' emotional and family experiences. J Early Adolescence 2016 Jul 26;23(3):241-267. [doi: 10.1177/0272431603254292]
9. Gerard JM, Krishnakumar A, Buehler C. Marital conflict, parent-child relations, and youth maladjustment. J Fam Issues 2016 Jun 30;27(7):951-975. [doi: 10.1177/0192513x05286020]
10. Story LB, Repetti R. Daily occupational stressors and marital behavior. J Fam Psychol 2006 Dec;20(4):690-700. [doi: 10.1037/0893-3200.20.4.690] [Medline: 17176205]
11. Larson RW, Richards MH. Family emotions: do young adolescents and their parents experience the same states? J Res Adolescence 1994 Oct;4(4):567-583. [doi: 10.1207/s15327795jra0404_8]
12. Sossin K, Birklein S. Nonverbal transmission of stress between parent and young child: considerations and psychotherapeutic implications of a study of affective movement patterns. J Infant Child Adolescent Psychother 2006;5(1):46-69.
13. Chung GH, Flook L, Fuligni AJ. Reciprocal associations between family and peer conflict in adolescents' daily lives. Child Dev 2011;82(5):1390-1396 [FREE Full text] [doi: 10.1111/j.1467-8624.2011.01625.x] [Medline: 21793820]
14. Bamberger KT. The application of intensive longitudinal methods to investigate change: stimulating the field of applied family research. Clin Child Fam Psychol Rev 2016 Mar;19(1):21-38 [FREE Full text] [doi: 10.1007/s10567-015-0194-6] [Medline: 26541560]
15. Bandura A. Social cognitive theory of self-regulation. Organ Behav Hum Decis Process 1991 Dec 21;50(2):248-287. [doi: 10.1016/0749-5978(91)90022-L]
16. Carver CS. A cybernetic model of self-attention processes. J Pers Soc Psychol 1979 Aug;37(8):1251-1281. [doi: 10.1037/0022-3514.37.8.1251]
17. Kanfer FH. Self-monitoring: methodological limitations and clinical applications. J Consult Clin Psychol 1970 Mar 21;35(2):148-152. [doi: 10.1037/H0029874]
18. Kazdin AE. Reactive self-monitoring: the effects of response desirability, goal setting, and feedback. J Consult Clin Psychol 1974 Oct;42(5):704-716. [doi: 10.1037/h0037050] [Medline: 4427011]
19. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008;4(1):1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
20. Smyth J, Heron K. Ecological momentary assessment (EMA) in family research. In: McHale SM, Amato P, Booth A, editors. National Symposium on Family Issues: Volume 4. Emerging Methods in Family Research. New York, USA: Springer International Publishing; 2014:145-161.
21. Ducharme J, Doyle AB, Markiewicz D. Attachment security with mother and father: associations with adolescents' reports of interpersonal behavior with parents and peers. J Soc Pers Relatsh 2016 Jun 29;19(2):203-231. [doi: 10.1177/0265407502192003]
22. Rusby JC, Westling E, Crowley R, Light JM. Concurrent and predictive associations between early adolescent perceptions of peer affiliates and mood states collected in real time via ecological momentary assessment methodology. Psychol Assess 2013 Mar;25(1):47-60 [FREE Full text] [doi: 10.1037/a0030393] [Medline: 23088201]
23. 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]
24. Johnson S, Bolstad O. Education Resources Information Center. 1974. Methodological Issues in Naturalistic Observation: Some Problems and Solutions for Field Research. Final Report URL: https://eric.ed.gov/?id=ED071749 [accessed 2020-04-29]
25. Ramanathan N, Alquaddoomi F, Falaki H, George D, Hsieh C, Jenkins J. ohmage: An Open Mobile System for Activity and Experience Sampling. In: Proceedings of the 6th International Conference on Pervasive Computing Technologies for Healthcare and Workshops. 2012 Presented at: PervasiveHealth'12; May 21-24, 2012; San Diego, CA, USA p. 203-204. [doi: 10.4108/icst.pervasivehealth.2012.248705]
26. Alderfer MA, Fiese BH, Gold JI, Cutuli JJ, Holmbeck GN, Goldbeck L, et al. Evidence-based assessment in pediatric psychology: family measures. J Pediatr Psychol 2008 Oct;33(9):1046-61; discussion 1062 [FREE Full text] [doi: 10.1093/jpepsy/jsm083] [Medline: 17905801]

27. Bloom BL, Naar S. Self-report measures of family functioning: extensions of a factorial analysis. Fam Process 1994 Jun;33(2):203-216. [doi: 10.1111/j.1545-5300.1994.00203.x] [Medline: 7925929]
28. Bray JH. Family assessment: current issues in evaluating families. J Fam Relat 1995 Oct;44(4):469. [doi: 10.2307/585001]
29. Grebelsky-Lichtman T. Parental patterns of cooperation in parent-child interactions: the relationship between nonverbal and verbal communication. Hum Commun Res 2013 Oct 3;40(1):1-29. [doi: 10.1111/hcre.12014]
30. Crawford JR, Henry JD. The positive and negative affect schedule (PANAS): construct validity, measurement properties and normative data in a large non-clinical sample. Br J Clin Psychol 2004 Sep;43(Pt 3):245-265. [doi: 10.1348/0144665031752934] [Medline: 15333231]
31. Pollak J, Adams P, Gay G. PAM: A Photographic Affect Meter for Frequent, In Situ Measurement of Affect. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011 Presented at: CHI'11; May 7-12, 2011; Vancouver, BC. [doi: 10.1145/1978942.1979047]
32. Stattin H, Kerr M. Parental monitoring: a reinterpretation. Child Dev 2000;71(4):1072-1085. [doi: 10.1111/1467-8624.00210] [Medline: 11016567]
33. Kerr M, Stattin H, Burk W. A reinterpretation of parental monitoring in longitudinal perspective. J Res Adolescence 2010;20(1):39-64. [doi: 10.1111/j.1532-7795.2009.00623.x]
34. Elgar FJ, Waschbusch DA, Dadds MR, Sigvaldason N. Development and validation of a short form of the Alabama parenting questionnaire. J Child Fam Stud 2006 Aug 22;16(2):243-259. [doi: 10.1007/s10826-006-9082-5]
35. Essau CA, Sasagawa S, Frick PJ. Psychometric properties of the Alabama parenting questionnaire. J Child Fam Stud 2006 May 18;15(5):595-614. [doi: 10.1007/s10826-006-9036-y]
36. Robin A, Foster S. Negotiating Parent-Adolescent Conflict: A Behavioral-Family Systems Approach. New York, USA: Guilford Press; 1989.
37. Bright Futures: A National Health Promotion Initiative. 2019. Issues Checklist (Abridged) URL: https://www.brightfutures.org/mentalhealth/pdf/professionals/ad/issues.pdf [accessed 2020-04-29]
38. Furman W, Buhrmester D. The network of relationships inventory: behavioral systems version. Int J Behav Dev 2009 Sep 1;33(5):470-478 [FREE Full text] [doi: 10.1177/0165025409342634] [Medline: 20186262]
39. Straus MA, Hamby SL, Boney-McCOY S, Sugarman DB. The revised conflict tactics scales (CTS2). J Fam Issues 2016 Jun 30;17(3):283-316. [doi: 10.1177/019251396017003001]
40. Furman W, Buhrmester D. The network of relationships inventory: behavioral systems version. Int J Behav Dev 2009 Sep 1;33(5):470-478 [FREE Full text] [doi: 10.1177/0165025409342634] [Medline: 20186262]
41. Bringhurst D, Watson C, Miller S, Duncan B. The reliability and validity of the Outcome rating scale: a replication Study of a brief clinical measure. J Brief Ther 2006;5(1):23-30 [FREE Full text]
42. Miles M, Huberman A. Qualitative Data Analysis: A Methods Sourcebook. Thousand Oaks, CA: Sage Publications; 1994.
43. Ryan GW, Bernard HR. Techniques to identify themes. Field Methods 2016 Jul 24;15(1):85-109. [doi: 10.1177/1525822x02239569]
44. Korotitsch WJ, Nelson-Gray RO. An overview of self-monitoring research in assessment and treatment. Psychol Assess 1999 Mar 21;11(4):415-425. [doi: 10.1037//1040-3590.11.4.415]
45. Pinderhughes EE, Dodge KA, Bates JE, Pettit GS, Zelli A. Discipline responses: influences of parents' socioeconomic status, ethnicity, beliefs about parenting, stress, and cognitive-emotional processes. J Fam Psychol 2000 Sep;14(3):380-400 [FREE Full text] [doi: 10.1037//0893-3200.14.3.380] [Medline: 11025931]
46. Sánchez-Martínez M, Otero A. Factors associated with cell phone use in adolescents in the community of Madrid (Spain). Cyberpsychol Behav 2009 Apr;12(2):131-137. [doi: 10.1089/cpb.2008.0164] [Medline: 19072078]
47. Thomée A, Härenstam A, Hagberg M. Mobile phone use and stress, sleep disturbances, and symptoms of depression among young adults-a prospective cohort study. BMC Public Health 2011 Jan 31;11:66 [FREE Full text] [doi: 10.1186/1471-2458-11-66] [Medline: 21281471]

Abbreviations

EMA: ecological momentary assessment
UCLA: University of California, Los Angeles

Edited by G Eysenbach; submitted 06.08.19; peer-reviewed by A Denisova, S Sittig, N Wang; comments to author 21.10.19; revised version received 14.01.20; accepted 29.03.20; published 23.06.20.

Please cite as:
Swendeman D, Sumstine S, Brink A, Mindry D, Medich M, Russell M
Smartphone Self-Monitoring by Young Adolescents and Parents to Assess and Improve Family Functioning: Qualitative Feasibility Study
JMIR Form Res 2020;4(6):e15777
URL: http://formative.jmir.org/2020/6/e15777/
doi:10.2196/15777
PMID:32574148

Original Paper

Barriers to Gestational Diabetes Management and Preferred Interventions for Women With Gestational Diabetes in Singapore: Mixed Methods Study

Sumali Hewage[1], PhD; Jananie Audimulam[2], MPH; Emily Sullivan[3], MPH; Claudia Chi[4], MBBS, MD; Tong Wei Yew[5], MBBS; Joanne Yoong[6], PhD

[1]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore
[2]Integrated Health Promotion Unit, Ministry of Health Transformation Office, Singapore, Singapore
[3]Family Planning 2020, United Nations Foundation, Washington, WA, United States
[4]Department of Obstetrics and Gynecology, National University Hospital, Singapore, Singapore
[5]Division of Endocrinology, Department of Medicine, National University Hospital, Singapore, Singapore
[6]Center for Economic and Social Research, University of Southern California, Los Angeles, CA, United States

Corresponding Author:
Sumali Hewage, PhD
Saw Swee Hock School of Public Health
National University of Singapore
Tahir Foundation Building
12 Science Drive 2, #10-01
Singapore, 117549
Singapore
Phone: 65 65164988
Email: sumali_hewage@u.nus.edu

Abstract

Background: Gestational diabetes mellitus (GDM) is associated with risks for both the mother and child. The escalated prevalence of GDM because of obesity and changes in screening criteria demands for greater health care needs than before.

Objective: This study aimed to understand the perception of patients and health care providers of the barriers to GDM management and preferred interventions to manage GDM in an Asian setting.

Methods: This mixed methods study used a convergent parallel design. Survey data were collected from 216 women with GDM, and semistructured interviews were conducted with 15 women and with 8 health care providers treating patients with GDM. Participants were recruited from 2 specialized GDM clinics at the National University Hospital, Singapore.

Results: The patients were predominantly Chinese (102/214, 47.6%), employed (201/272, 73.9%), with higher education (150/216, 69.4%) and prenatal attendance at a private clinic (138/214, 64.2%), already on diet control (210/214, 98.1%), and receiving support and information from the GDM clinic (194/215, 90.2%) and web-based sources (131/215, 60.9%). In particular, working women reported barriers to GDM management, including the lack of reminders for blood glucose monitoring, diet control, and insufficient time for exercise. Most women preferred getting such support directly from health care providers, whether at the GDM clinic (174/215, 80.9%) or elsewhere (116/215, 53.9%). Smartphone apps were the preferred means of additional intervention. Desirable intervention features identified by patients included more information on GDM, diet and exercise options, reminders for blood glucose testing, a platform to record blood glucose readings and illustrate or understand trends, and a means to communicate with care providers.

Conclusions: A GDM-focused smartphone app that is able to integrate testing, education, and communication may be a feasible and acceptable intervention to provide support to women with GDM, particularly for working women.

(JMIR Form Res 2020;4(6):e14486) doi:10.2196/14486

KEYWORDS

gestational diabetes; pregnancy; telemedicine; self-management; patient-centered care; mobile phone

Introduction

Background

Gestational diabetes mellitus (GDM) is a well-established risk factor for type 2 diabetes mellitus (T2DM). In parallel with the growing T2DM epidemic [1], an increasing number of pregnancies are being complicated by GDM [2]. The International Diabetes Federation reported that GDM currently affects 1 in 6 births globally [3]. GDM is defined as the onset or first diagnosis of high blood glucose concentrations during pregnancy, which usually resolves after childbirth [4]. GDM demands drastic lifestyle changes in pregnant women and additional medical attention to minimize detrimental fetal and maternal outcomes [5,6].

As with other types of diabetes, women with GDM receive medical advice on appropriate nutrition alongside directions to perform self-monitoring of blood glucose and to administer insulin therapy if required [5-7]. For optimal blood glucose control, women must be engaged in an intensive process from diagnosis until the baby's delivery, which usually spans 10 to 12 weeks. In conventional care, patients are expected to attend clinic appointments frequently so that health care providers can monitor glucose concentrations and patient behaviors closely. Patients often experience difficulties in adopting the required behavioral changes in a brief period, which can significantly compromise blood glucose control [8]. This is challenging for both patients and health systems and leads to exhausted capacities and burnout [9,10]. These challenges include patient barriers such as lack of reliable information, family and employment responsibilities, and social support and health system barriers such as lack of access to health care and inconsistent care delivery [11].

Asian populations are at a disproportionately higher risk of T2DM [12-15]. The prevalence of GDM in Singapore is above the global prevalence of 13.8% [3]. According to an analysis conducted with pregnant women who participated in a birth cohort study, *Growing Up in Singapore Towards Healthy Outcomes*, compared with high-risk GDM screening, universal screening diagnosed a significantly greater number of GDM cases (18.9% versus 9.8%, respectively) [16]. In addition, an analysis of the cost-effectiveness of GDM screening in the same study population showed that compared with no screening and high-risk screening, universal screening is cost-effective to reduce maternal and fetal complications due to GDM [17]. These results suggest that when it comes to policy implementation considerations for GDM care, universal screening has been shown to be significantly effective compared with high-risk screening, which is the current practice for GDM screening in Singapore [18].

It is widely accepted that patients' active collaboration with health care providers and their appropriate behavior changes are key factors in optimal GDM management [19,20]. As a result, acknowledging the patient's contribution is important in building a healthy patient-provider relationship. In addition, the US Institute of Medicine committee on quality of health care in America has identified patient-centered care as 1 of the 6 attributes of health care quality [21]. Therefore, focusing on patient-centric care has been the focus of providing quality health care.

Objectives

In response to the anticipated increased prevalence from universal screening, to make improvements in the health care system to serve the needs of women diagnosed with GDM, it is important to understand the patients' requirements. Therefore, this feasibility study explored 2 aims to understand (1) women's perceptions of knowledge and management of GDM and (2) women and care providers' perceptions and attitudes toward GDM control, related barriers, and potential interventions to overcome the barriers.

Methods

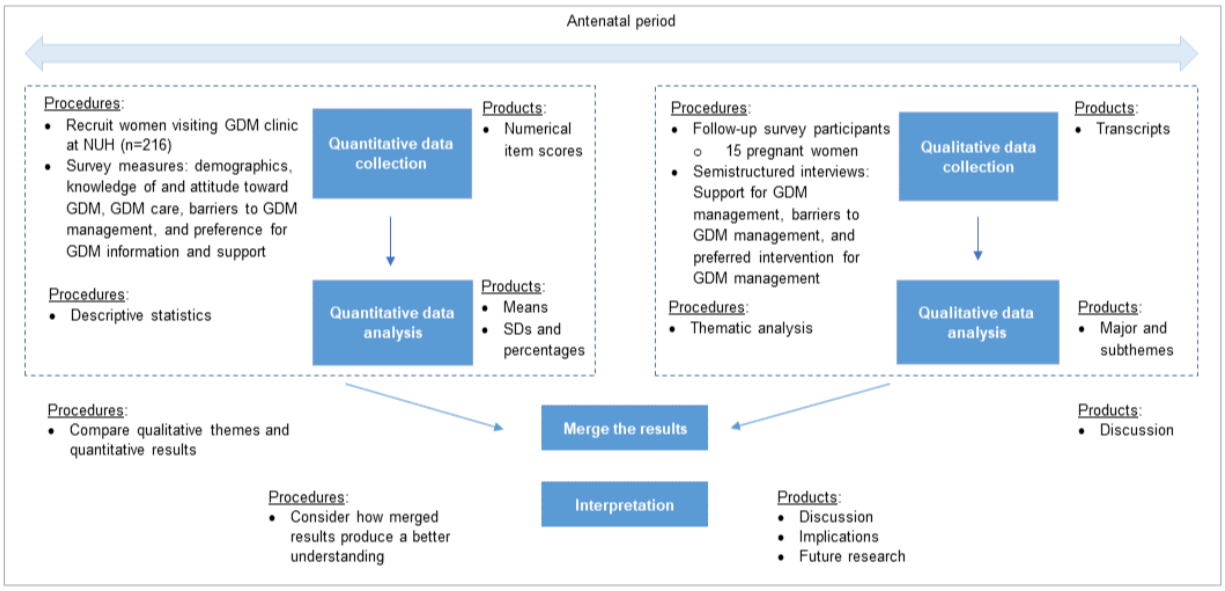
Design and Study Population

This study was undertaken with pregnant women who were diagnosed with GDM at the National University Hospital (NUH) in Singapore. Currently, all public hospitals in Singapore that provide prenatal medical care follow the latest World Health Organization (WHO) guidelines [22] for the diagnostic criteria of GDM and provide universal screening for GDM to pregnant women. Screening is typically offered between 24 weeks' and 28 weeks' gestation. Women diagnosed with GDM attend a comprehensive multidisciplinary educational session to help equip them with the knowledge required to manage GDM during pregnancy.

This study was conducted using mixed methods with quantitative and qualitative components and a convergent parallel design (Figure 1). A cross-sectional survey was undertaken in pregnant women attending 1 of 2 specialized GDM care clinics at NUH.

Inclusion criteria were being pregnant, aged 21 to 40 years, diagnosed with GDM during the index pregnancy, and attendance at the GDM clinic's workshop on GDM management (delivered by a nurse educator and a dietitian). Exclusion criteria were inability to speak English and known type 1 or type 2 diabetes before the current pregnancy. Recruitment took place over 2 years between May 2015 and May 2017, with 2 phases of data collection.

Figure 1. Convergent parallel mixed methods design. GDM: gestational diabetes mellitus; NUH: National University Hospital.



Study Tools

For the quantitative component, a 27-item cross-sectional survey (Multimedia Appendix 1) was used to collect demographic information, participant knowledge, control and attitude toward GDM, perceived barriers to GDM management, and current and preferred GDM support. Women were asked to rate their GDM knowledge before and after attending the GDM clinic, and the answer options included *excellent*, *good*, *fair*, *poor*. Questions related to GDM risk, perception, and attitudes were adapted from validated tools [23]. Attitude-related questions measured perceived personal control (2 questions), worry (2 questions), and optimism regarding recurrent GDM in future pregnancies (1 question). The 4 responses ranged from *strongly agree* to *strongly disagree*, and the scores were provided using a 4-point Likert scale (0-4). Average scores were calculated for each participant.

For the qualitative component, semistructured interviews were conducted using a topic guide based on published literature. This included questions on knowledge and attitudes regarding GDM, existing sources of support and coping mechanisms, current lifestyle practices, perceived barriers, ways of managing and monitoring GDM, and preferred intervention for GDM management.

Potential participants for the survey were approached in the waiting room of the GDM clinic. The study team refrained from approaching women who seemed tired or anxious. After getting informed consent, women were given a self-administered survey. No personal information was collected. In the first phase of the study, only a hard copy (paper) of the survey was offered. A web-based survey was added to the second phase to increase the recruitment rate.

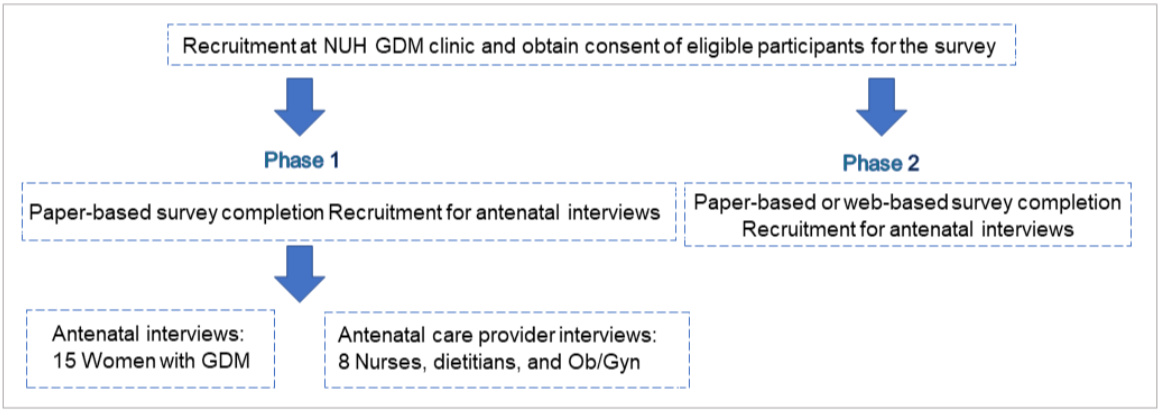
The web-based form was created using the SurveyMonkey survey tool (SurveyMonkey Inc) and tested for any technical difficulties. The data collection was carried out using a closed survey. A total of 30 questions were distributed over 3 screens. Those who elected to complete the survey on the web were sent the link to the survey via email, and 3 additional *reminder* emails were sent to women in the following 6 weeks, inviting them to complete the survey before they were designated as true *nonresponders*.

For pregnant women who were willing to participate in prenatal interviews, sessions were scheduled based on their convenience. In the second phase, to increase survey completion, the women were given an incentive of $25 (US $17.5). In the first phase, 50 patients participated in the survey.

Of these 50 patients, 35 gave verbal consent to participate in the interviews and 15 participated in the interviews, and 2 study members (JA and SH) conducted all the interviews in English. Most interviews were conducted at the health care facility or at the participants' homes, in a private space conducive for the participant to share thoughts effectively. Before the interview, participants provided informed consent, including consent for audio recording. Health care providers treating GDM patients from nursing, dietetics, and obstetrics and gynecology specialties were approached for interviews. The final number of interviews conducted was decided based on reaching thematic saturation as understood by the analysis, which was conducted in parallel with the data collection. Data saturation was reached at the 15th patient interview and the eighth care provider interview. Patient interview participants were from the first phase and were given Singapore $50 (US $35). Ethics approval was obtained from the National Healthcare Group (NHG) domain-specific review board (DSRB), Singapore (NHG DSRB Ref: 2015/00196).

The study flowchart outlines the steps involved in the recruitment, consent, and follow-up (Figure 2).

Figure 2. Study flow chart. GDM: gestational diabetes mellitus; NUH: National University Hospital; Ob/gyn: obstetrician/gynecologist.



Data Analysis

Univariate analysis was conducted to describe the study participants and GDM management. Results are presented using medians and IQRs for continuous variables with skewed distributions, whereas categorical variables are presented using frequencies and percentages. Perceived improvement in GDM knowledge after attending the GDM workshop was analyzed using the Fisher exact test. Those who received a score between 0 and 2 for GDM attitude-related factors, perceived GDM control, and perceived worry for GDM control were labeled as *low* and others were labeled as *high*. BMI was calculated using self-reported height and weight. The analysis was performed using the Stata statistical software (Stata Corp, 2013, Stata Statistical Software: Release 13).

The interviews were audio recorded, transcribed *verbatim*, and analyzed thematically [24]. The transcripts were not returned to participants for any further comments and/or corrections. First, transcription occurred verbatim, and the accuracy of the transcripts was verified by 1 of the researchers (SH). Following multiple readings of the verified transcripts, the main coder (SH) coded the transcripts using the Atlas Ti Software (QSR International Pty Ltd, 2012) and Microsoft Excel. Some themes were identified in advance, and some new themes were derived from the data. Reliability checks were performed by reading and checking transcripts, and a field expert (JY) verified the themes. The identified codes were used to derive the code structure that was needed to develop meaningful themes.

Results

Survey Completion

The average number of women attending the GDM clinic was 60 per month. In phase 1, the survey completion rate was 18.5%. In phase 2, 67.2% of the women visiting the clinic were interested in participating in the study, with 75.4% actually completing the survey, giving a completion rate of 44.0%. Most of the nonparticipants declined to participate without a reason, whereas some declined due to insufficient time. There were no significant differences between the 2 groups in terms of demographic characteristics (Multimedia Appendix 2).

Demographic and Pregnancy Information

Table 1 provides the demographic and health information of the survey participants (n=216). In brief, the median age of the women was 32 years; 47.6% (102/216) women were Chinese, 73.9% (201/216) were in full- or part-time employment, and 69.4% (150/216) were university graduates. At the time of survey completion, the median duration of pregnancy was 30 weeks, and 50.4% (109/216) of women were multiparous, and 64.2% (138/216) of women attended a private clinic for their prenatal care. Approximately half (116/216, 53.9%) of the women were either overweight or obese before pregnancy and a large majority, 83.3% (180/216) women, perceived their health as *good* or *excellent*.

The women who participated in interviews had a demographic profile similar to that of the survey questionnaire participants. Half of the 15 participants in the age group of 31 to 35 years were primiparous or had a family history of T2DM, and most of them were Chinese or Indian, employed full time, with a degree, and attending a private prenatal clinic. Eight prenatal care providers at the GDM clinic were interviewed. Table 2 provides the demographic and health information of the patients interview participants and Table 3 shows the information of the health care provider interview participants.

Table 1. Survey participants' demographic characteristics.

Variable	Values
Age (years) (n=209), median (IQR)	32 (22-40)
Ethnicity (n=214), n (%)	
Chinese	102 (47.6)
Malay	46 (21.5)
Indian	43 (20.1)
Other	23 (10.7)
Employment (n=214), n (%)	
Full time	146 (68.2)
Unemployed	55 (25.7)
Part time	13 (6.1)
Education (n=216), n (%)	
Degree or professional qualification	150 (69.4)
Secondary education	62 (28.7)
Lower than secondary education	4 (1.8)
Duration of pregnancy in weeks, median (IQR)	30 (8-39)
Parity—multiparous (n=215), n (%)	109 (50.4)
Prenatal care (n=215), n (%)	
Private clinic	138 (64.2)
Subsidized clinic	77 (35.8)
Prepregnancy BMI (n=208), n (%)	
<23 kgm^{-2}	96 (46.1)
23-27.5 kgm^{-2}	59 (28.4)
>27.5 kgm^{-2}	53 (25.5)
Health (n=216), n (%)	
Poor	1 (0.5)
Fair	35 (16.2)
Good	158 (73.1)
Excellent	22 (10.2)

Table 2. Interview participant profile—patients attending a gestational diabetes mellitus clinic (N=15).

Variable	Participants, n (%)
Age (years)	
26-30	3 (20)
31-35	8 (53)
36-40	4 (27)
Ethnicity	
Chinese	5 (33)
Malay	3 (20)
Indian	5 (33)
Filipino and Sri Lankan	2 (13)
Employment	
Full time	11 (73)
Unemployed	3 (20)
Part time	1 (7)
Education	
Degree or professional qualification	14 (93)
Secondary education	1 (7)
Parity—primiparous	8 (53)
Prenatal care	
Private clinic	12 (80)
Subsidized clinic	3 (20)
Family history of gestational diabetes mellitus	8 (53)

Table 3. Interview participant profile—health care providers serving at a gestational diabetes mellitus clinic (N=8).

Variable	Participants, n (%)
Professional qualification	
Obstetrics and gynecology	2 (25)
Diabetes care nurse educator	4 (50)
Dietitian	2 (25)
Years of total experience	
1 to 5	2 (25)
6 to 10	3 (37)
More than 10	3 (37)
Gender	
Female	6 (75)

Knowledge on Gestational Diabetes Mellitus

As reported by the participants, there was a significant improvement in perceived GDM knowledge after attendance at the GDM workshop (χ^2_9=54.0; P<.05). The vast majority, 93.5% of women, identified large for gestational age of the baby as a potential outcome of GDM, and 92% correctly identified the optimal range for pre- and postprandial capillary blood glucose levels.

As confirmed by the interviews, most of the women with a previous history of GDM were unaware or not fully aware of GDM before their clinical diagnosis. Those who had any knowledge of GDM mentioned that they had heard about it from peers with GDM. Most participants, even those with a family history of T2DM, felt that they did not have sufficient information on GDM. However, the participants felt that the clinic helped to *increase the understanding of GDM*, which is consistent with the survey results:

> *Actually, I didn't know anything about it... So, I... thought that if you were diabetic then you kind of get it. But then I didn't know something you can just develop during pregnancy as well. So, it was quite new to me.* [ID_02]

> *Now, the understanding level has gone high. I know like how to control and then how to manage my diet and then when to measure, what are the steps I need to take.* [ID_13]

Management of Gestational Diabetes Mellitus

Almost all of the participants, 98.1% (210/214), controlled their diet; half, 48.1% (103/214), were physically active; and 21 participants, one-tenth, used insulin to regulate blood glucose concentrations. Most participants, 85.6% (185/216), were able to follow typical recommendations from the GDM clinic to perform finger prick blood testing 7 times a day for 2 days each week (Table 4).

Similarly, women who were interviewed also managed their blood glucose levels using diet modifications and monitored blood glucose levels using finger prick blood testing. Employed women, mostly Chinese and Malay, are more likely to consume outside food rather than home-cooked food. Few of the interviewed women were physically active, and only 2 used insulin therapy:

> *Mainly diet. I really watch my diet. Yes, actually now I, there's only about a few choices that I can have every day.* [ID_08]

In addition, most women mentioned that they performed the test more frequently than required and mainly used test results to *interpret the effectiveness of GDM control* primarily diet. A few women believed that physical activity helped with blood glucose control:

> *So, from there I monitor. Let's say one weekend and one weekday I monitor. But, let's say if I want to feel, say in the morning I want to watch I just go on. So, it doesn't matter, two times a week. Let's say, watch three times or more, depends I feel want to check my sugar.* [ID_09]

Table 4. Management of gestational diabetes mellitus.

Variable	Value, n (%)
Gestational diabetes mellitus management (n=214)	
Diet management	210 (98.1)
Physical activity	103 (48.1)
Insulin use	21 (9.8)
Finger prick blood test (n=216)	
Record 7 readings on 2 days or more	185 (85.6)
Most of the time	20 (9.3)
Sometimes or never	11 (5.1)

Attitudes Toward Gestational Diabetes Mellitus Control

In the survey, 97.7% of participants received a high score for perceived GDM control, whereas 74.5% had a high score for the perceived worry of GDM control. Most women (72.3%) perceived that they would not be able to control getting GDM in future pregnancies.

Consistent with the survey findings, at the time of the interview (subsequent to the clinic), many reported feelings of control related to diet and that they were *on the right track* in monitoring using the finger prick test:

> *That, I was just grappling around, reading on the internet, and then I was like "Oh my god, what do I do?" So, I was panicking.* [ID_01]

> *Because I am actually monitoring the sugar level at the moment. And I think it's actually ok. So, far I am actually on the right track.* [ID_06]

Most women reported feeling worried or anxious when GDM was diagnosed, although the GDM clinic's workshop helped to lessen their anxiety.

Health care providers also felt that most women were motivated to control their blood glucose levels and also stressed the importance of *discipline*, especially with respect to their food intake during this period:

> *Generally, this group of patients, they are very motivated. So, if they don't have any language barriers, they should be able to understand that important for them... But in general, at least a good 80% of them seems to be quite receptive, I would say yeah...* [ID_04_ Dietitian]

Barriers to Gestational Diabetes Mellitus Control

Most survey participants did not feel that they faced any significant barriers to GDM control. However, a significant minority (30.4%) felt they had difficulties in remembering blood testing schedules, whereas 22% felt discouraged because of the lack of immediate effects from their lifestyle changes (ie, continuing to have abnormal test results even after implementing GDM management). One-fifth of participants (21.5%) agreed or strongly agreed to experiencing difficulties following the recommended diet and physical activity plan. On the other hand, most women reported that they had sufficient help from family

and friends and that family, cultural beliefs, or traditions did not interfere with GDM management (Table 5).

Conversely, based on the qualitative data, most employed women mentioned that they had barriers to GDM management mainly related to diet, increasing physical activity, and monitoring blood glucose levels. Women felt that they had a *limited (food) variety* because they had few food options that helped control blood glucose levels. These differences between quantitative and qualitative results may have been partly due to the lack of quantitative measures for assessing barriers to GDM control:

> *I have still maybe about three, four months or even more to go. So, I think I must expand the variety to make life easier.* [ID_08]

Only a few mentioned that they did exercise to control GDM. Although they had received advice to increase their physical activity, women felt unable to follow the recommendations mainly due to the *lack of time for exercise*:

> *I think one problem is that it's not always very easy to exercise. Because one of the advice(s) is that you should have like a 10-15 minutes' walk or some form of exercise after every meal. But it's not very possible to do it after your lunch, for example, if you are working.* [ID_02]

According to the clinical practice guidelines of the Ministry of Health, Singapore, women are required to self-monitor their blood glucose levels [25]. At the NUH GDM clinic, women were advised to submit 7 blood glucose readings for 1 weekday and 1 weekend day each week. However, women had difficulty performing the required 7 readings for each testing day, largely because they forgot to take the test:

> *So, it is very hard to get these seven readings [finger prick test reading]. So, one thing is that it is tiring, and the other thing is that you, kind of unconsciously forget[s] with work and [a] lot [of] things as well. So, maybe you had your lunch and you finish[ed] it at one [pm] and then you take one [finger prick test] at three [pm]. And then you forget because you are like completely into your work.* [ID_02]

Similarly, health care providers reported that working women experienced more barriers to managing GDM than nonworking women:

> *Of course, if let's say if they do work, then shift workers are quite difficult to tackle their meal timings and all. So, that is one of their barriers like limitation, their work commitments and all.[ID_04_ Dietitian]*

Table 5. Barriers to management of gestational diabetes mellitus.

Barrier	Strongly disagree, n (%)	Disagree, n (%)	Agree, n (%)	Strongly agree, n (%)
Difficult to remember to take medication or blood tests at the scheduled times	28 (13.4)	117 (56.0)	59 (28.2)	5 (2.4)
Feel discouraged due to lack of immediate results (eg, high blood sugar)	23 (10.7)	144 (67.3)	45 (21.0)	2 (1.0)
Difficult to follow a specific diet and physical activity plan	28 (13.1)	140 (65.4)	41 (19.2)	5 (2.3)
Busy with family or work to manage my GDM properly	34 (16.0)	138 (64.8)	38 (17.8)	3 (1.4)
Nonspecific educational resources/opportunities about GDM	31 (14.6)	147 (69.0)	29 (13.6)	6 (2.8)
Family's cultural beliefs/ traditional practices conflict with GDM management	69 (32.4)	125 (58.7)	17 (8.0)	2 (1.0)
Family and friends are not supportive of my efforts to eat right	74 (34.6)	127 (59.3)	12 (5.6)	1 (0.5)

Support for Gestational Diabetes Mellitus Management

Of the 216 survey participants, 194 (90.2%) reported receiving GDM-related information from a physician, nurse, or trained counselor at the GDM clinic and 131 (60.9%) from websites. Moreover, 6.1% (16/216) of participants reported in-person support groups for expectant mothers as the method used least frequently (Table 6).

The interview themes appeared concordant with the survey results. Women were most likely to rely on health care providers to guide them with medical advice. Furthermore, almost all mentioned that they used the internet to obtain additional information related to GDM management. However, most of the resources used were intended for Western populations. Most of the women received help from their families and peers. In addition, they either approached peers with previous experience in controlling GDM or online forums. Most of them felt they had *sufficient support* in GDM management, whereas a few needed extra help:

> *Internet and friends and doctors.* [ID_07]
>
> *I would say mainly like my husband and I would say my colleagues.* [ID_02]
>
> *Well, one is you know because my colleague is also a pregnant woman and we are good friends, so we discuss a lot about it.* [ID_14]

The following interview results further support the above findings. According to health care providers, support from physicians, nurse educators, and dietitians was available in the usual clinic setting, and women were followed up fortnightly. On occasions requiring additional medical assistance, women were advised to approach the GDM clinic at the hospital. Some providers mentioned that they offered additional reading resources. In contrast, others pointed out the need to regulate the quality of supplementary material, especially of internet sources. Health care providers also acknowledged that, during this time, women may need additional emotional assistance from family members, including husbands and friends:

I mean we give them resources to read. The Royal College of Obstetricians and Gynecologists in the UK have lots [of] patient education leaflets that we can refer them to read at home and at leisure. [ID_02_Ob/gyn]

And some of the database [are] based on different other countries like for an example [the] US. So, if you look at the US database, you realize that most of the calorie content, carbohydrate content slightly, maybe the portion size larger than us. So, that's why I always tell them, if you use a US database always to cut [portion size]. They need to reduce the amount of carbohydrate and amount [number] of calories. [ID_03_ Dietitian]

Table 6. Source of gestational diabetes mellitus information.

Source	Value, n (%)
Doctor, nurse or trained counselor at the GDM[a] clinic	194 (90.2)
Websites	131 (60.9)
Family members	66 (30.7)
Friends or colleagues	59 (27.4)
Web-based forums or support groups for new or expectant mothers	44 (20.5)
Doctor or nurse other than the GDM clinic	34 (15.8)
In-person support groups for expectant mothers	13 (6.0)

[a]GDM: gestational diabetes mellitus.

Preferred Intervention for Gestational Diabetes Mellitus Management

In the survey, among the preferred methods of information and support for GDM management, the preferred option for most participants 80.9% (174/216) was *from a doctor, nurse or trained counselor at the GDM clinic*. Among their ranking, support from a physician or nurse outside of the GDM clinic ranked second (116/216, 53.9%), and this replaced the current support option of *websites*, which ranked as the third preferred option (46/216, 21.3%). However, only a minority of participants (34/216, 15.8%) were receiving this kind of care at the time of the survey (Table 6). The other listed options of family members, friends/colleagues, in-person support groups for expectant mothers, and support from other women who have previously managed GDM related to social support were less preferred.

Additional information regarding preferred interventions for GDM management was provided during the interviews. Most of the interview participants thought a smartphone app would be convenient, primarily because smartphones are now commonplace. They pointed out that such an intervention would be more helpful for GDM management, mainly for recording blood glucose readings daily, than a paper and pen, or a computer. In addition, women discussed the importance of having the means to *understand trends* and the steps that should be taken, if any, to rectify abnormal readings immediately:

I mean the most useful, convenient is the app [smartphone application]. Because everyone has a smartphone, and everyone can access. [ID_10]

Maybe there's an indication, your reading is good, there's a comment to supplement your, because sometimes when you write down your reading, you do not know if it is on target, not on target, high risk, or low risk. [ID_11]

Furthermore, the women stated the need of reminders for finger prick test schedules, general information about GDM, calorie calculations, and information about physical activity. Concerns about a possible app were relatively infrequently mentioned, but they included the level of complexity, technical glitches, and the need to charge the phone to access the app. Health care providers stressed the importance of strict control of blood glucose concentrations, and they also agreed that an app would be a convenient platform to assist GDM control. As additional features, health care providers suggested that the app could automate the transfer of blood glucose test results from the device to the provider overseeing the patient's care. They felt that this step would increase the reliability of test results, including an automated reported option and the convenience of timely monitoring. According to them, the intervention would be effective only if it was user-friendly and affordable, especially for those of lower socioeconomic status:

They may forget to bring [blood glucose readings], that is what I was saying, they may forget to bring it [or] may not be accurate or they may not want to bring out to write it down. So, it will not be 100% accurate. (ID_05_nurse practitioner)

Discussion

Principal Findings

Our study was able to understand women's and care providers' perceptions on improving GDM care. Most women preferred the assistance of health care providers to control their blood glucose levels. Furthermore, the women, most of whom were employed, experienced barriers mainly due to limited reminders for monitoring blood glucose, difficulties in diet control, and inadequate time to be physically active. Smartphone apps appear to be preferred by women to assist the standard of care to better support blood glucose control. In addition, they anticipated that a *mobile app* can assist them to overcome their common

difficulties as well to acquire reliable information on GDM and understand trends in blood glucose control.

One systematic review indicated that women with GDM feel overwhelmed in the initial period post diagnosis and that they are more likely to overcome these difficulties with appropriate medical assistance [26]. Our findings are consistent with these reported observations. GDM interventions have been shown to be important in helping women to curb adverse clinical outcomes as well as to elevate their quality of life [27]. These interventions appear to be successful due to receptiveness among highly motivated women who essentially want to safeguard their pregnancy.

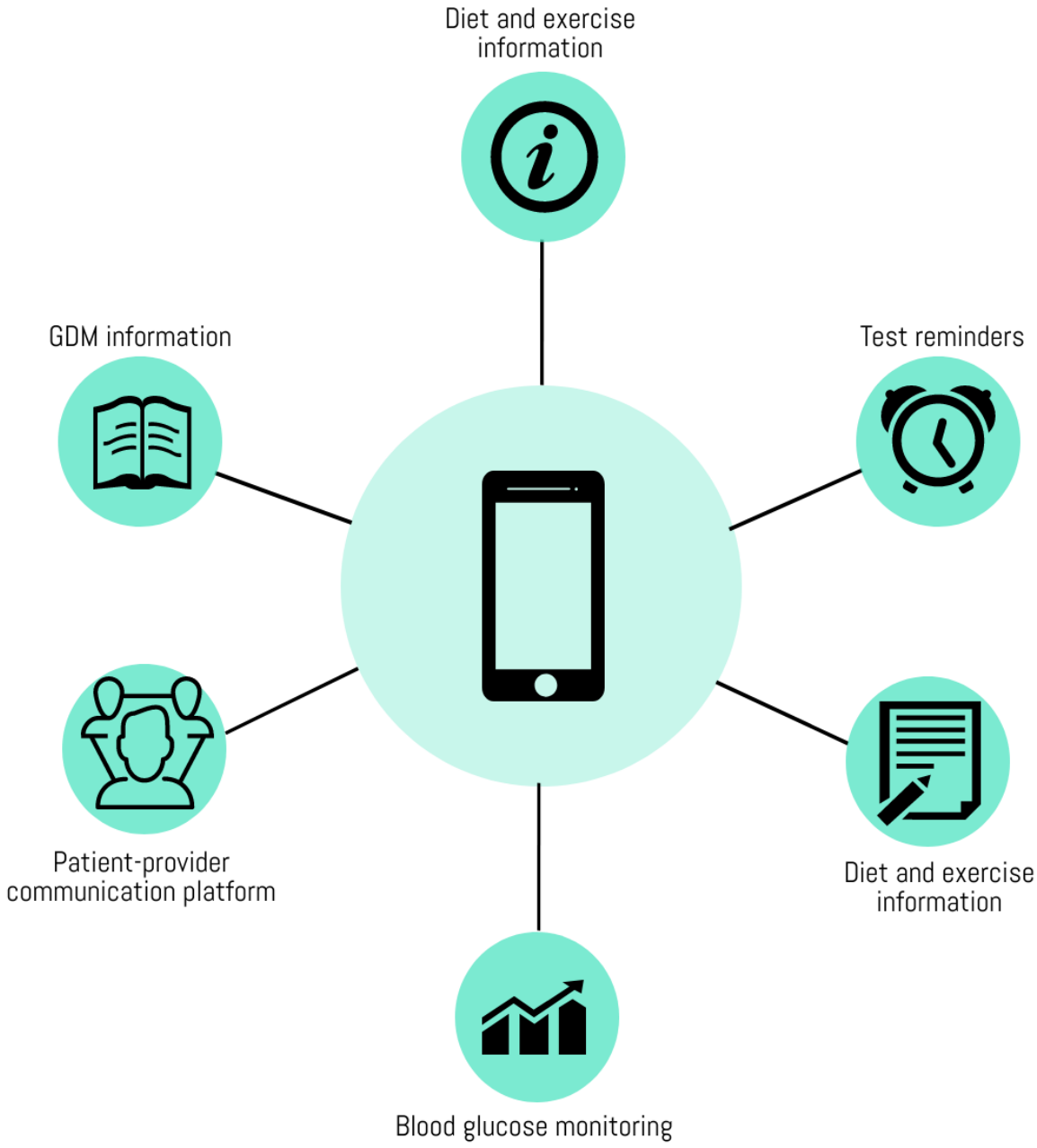
According to our qualitative findings, the working women in the study reported experiencing barriers, including a lack of reminders for blood glucose monitoring, issues related to diet control, and lack of time for recommended exercise. Although physical activity has been shown to be effective in regulating blood glucose concentrations among women with GDM [28], pregnant women in Singapore are less likely to be active, especially in the later stages of pregnancy [28]. This lack of a behavioral change may be further augmented by other commitments, such as work-related responsibilities. Identified gaps in GDM management highlight the need for appropriate interventions to integrate into busy lifestyles.

The participants in this study reported receiving advice primarily from specialized health care providers or web-based resources currently. However, among their preferences, the women conveyed the need for further assistance from health care providers other than the specialized GDM clinic. This highlights that women prefer to rely on medical personnel for advice, although web and smartphone usage among pregnant women is a common phenomenon with a significant ability to influence their health behaviors [29]. In the present environment, readily available health-related information from a wide variety of nonmedical resources may increase the possibility of erroneous information. Therefore, it is important to get the assistance of health care providers to critically review web-based content for medical accuracy and suitability [29,30].

As demands for health care are increasing, many parts of health systems, including diabetes care, are seeking help from telemedicine [31]. As defined by the WHO, telemedicine involves providing health care, including information on diagnosis, treatment, and disease prevention, where distance is a critical factor [32]. Similarly, several intervention studies have been published on telemedicine solutions for women with GDM [33-36]. Two meta-analyses concluded that telemedicine interventions, primarily mobile apps, may conveniently replace face-to-face clinic visits between women and health care providers without compromising the quality of care [37-41]. However, both reports pointed out the limited number of such interventions and the need for further investigation of possible cost evaluations.

In general, smartphone apps are considered to be patient-centric interventions [42]. As pointed out by study participants, a smartphone app would be a viable solution for most of the identified issues and could help improve lifestyle behaviors. The interventions undertaken included recording information on food, physical activities, blood glucose concentrations, and insulin regimens. To date, few studies have been conducted to understand the usability and acceptance of smartphone apps among women with GDM [37,38]. Neither of these reported studies assessed the perception and contribution of potential users before developing their programs. Having user input—both women with GDM and health care providers who treat women with GDM—is vital for the successful implementation of such interventions as it increases user acceptability and intervention sustainability (Figure 3) [43,44].

Figure 3. Preferred features of a smart phone app for gestational diabetes mellitus management. GDM: gestational diabetes mellitus.



Strengths and Limitations

The mixed methods approach used here has deepened our understanding of the needs that women with GDM have regarding GDM management. This study has a few limitations. Data collection was conducted in 1 public hospital, which may not be entirely representative of other centers in Singapore. However, in 2016, 85% to 90% of women in Singapore who were 25 to 34 years old were employed, and 56% of the women either had a degree, diploma, or professional qualification [45]. Our study sample has a similar demographic distribution; hence, the findings appear largely generalizable to women with GDM in Singapore. Second, we were not able to measure the change in participant knowledge before and after the GDM workshop or to evaluate the effectiveness of the information delivery via the workshop [46]. A further limitation was that only 1 researcher coded and analyzed the qualitative data. Although measures were taken to increase the response rate, the overall survey completion rate was low. In addition, only women who could speak and write English and were willing to participate were recruited. Therefore, it is possible that the results are biased toward those women who chose to participate. As suggested by a previously published meta-analysis, it is worth leveraging technology to facilitate behavior management and to make the intervention available to intended groups [37]. Recently, the National Health Service in the United Kingdom approved the use of a smartphone app in the management of GDM. It is anticipated that the intervention will result in fewer clinic visits

for working women and reduce inconvenience and unnecessary workplace absenteeism [47]. In Singapore, most women of reproductive age are employed and may benefit from such technological advances. Due to the increasing demand in the health care sector, telemedicine options have been gaining attention as a feasible option. The contribution of users is critical when designing a technological intervention, especially for pregnant women [48]. Our study identified the barriers experienced by women with GDM. These gaps may be addressed with a smartphone app. This was the commonly agreed intervention by the women and care providers to assist in optimal GDM management while easing the pressure on the local health system.

Conclusions

In conclusion, as informed by this study, a carefully planned randomized control trial is likely to be useful in assessing the effectiveness and cost-effectiveness of a smartphone app to minimize adverse maternal and fetal outcomes of GDM and the optimal use of health care resources in Singapore.

Acknowledgments

The authors thank the survey and interview participants and health care providers at the GDM clinic, NUH, Singapore. This research study was supported by the National University of Singapore (NUS) Global-Asia Institute NUS Initiative to Improve Health in Asia Research Grant (Grant No. NIHA-2013-1-002).

Authors' Contributions

JY, JA, and SH conceptualized and designed the study, established the research question, designed the questionnaire, and obtained necessary ethics approval. SH, JA, and ES collected the data. SH performed the data cleaning and analyses. SH performed the statistical analyses, prepared graphics, assisted in drafting the manuscript, and finalized the draft based on comments from other authors' feedback. CC and TY assisted in participant recruitment and critical review of the manuscript. JY assisted by critically reviewing the manuscript. All authors approved the final manuscript.

Conflicts of Interest

CC and TWY have received research funding from Jana Care project for the work related to the subject matters discussed in this manuscript. All other authors declared no conflicts of interest.

Multimedia Appendix 1
Cross-sectional survey questionnaire.
[DOCX File , 37 KB - formative_v4i6e14486_app1.docx]

Multimedia Appendix 2
Participants' characteristics by data collection phase.
[DOCX File , 16 KB - formative_v4i6e14486_app2.docx]

References

1. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. Lancet 2016 Apr 9;387(10027):1513-1530 [FREE Full text] [doi: 10.1016/S0140-6736(16)00618-8] [Medline: 27061677]
2. Ferrara A. Increasing prevalence of gestational diabetes mellitus: a public health perspective. Diabetes Care 2007 Jul;30(Suppl 2):S141-S146. [doi: 10.2337/dc07-s206] [Medline: 17596462]
3. International Diabetes Federation. 2019. IDF Diabetes Atlas: Ninth Edition URL: https://www.diabetesatlas.org/en/ [accessed 2020-04-28]
4. American Diabetes Association. 2. Classification and diagnosis of diabetes: standards of medical care in diabetes-2018. Diabetes Care 2018 Jan;41(Suppl 1):S13-S27. [doi: 10.2337/dc18-S002] [Medline: 29222373]
5. Brown J, Alwan NA, West J, Brown S, McKinlay CJ, Farrar D, et al. Lifestyle interventions for the treatment of women with gestational diabetes. Cochrane Database Syst Rev 2017 May 4;5:CD011970 [FREE Full text] [doi: 10.1002/14651858.CD011970.pub2] [Medline: 28472859]
6. Metzger BE, Buchanan TA, Coustan DR, de Leiva A, Dunger DB, Hadden DR, et al. Summary and recommendations of the fifth international workshop-conference on gestational diabetes mellitus. Diabetes Care 2007 Jul;30(Suppl 2):S251-S260. [doi: 10.2337/dc07-s225] [Medline: 17596481]
7. Martis R, Crowther CA, Shepherd E, Alsweiler J, Downie MR, Brown J. Treatments for women with gestational diabetes mellitus: an overview of Cochrane systematic reviews. Cochrane Database Syst Rev 2018 Aug 14;8:CD012327 [FREE Full text] [doi: 10.1002/14651858.CD012327.pub2] [Medline: 30103263]

8. Devsam B, Bogossian FE, Peacock AS. An interpretive review of women's experiences of gestational diabetes mellitus: proposing a framework to enhance midwifery assessment. Women Birth 2013 Jun;26(2):e69-e76. [doi: 10.1016/j.wombi.2012.12.003] [Medline: 23333029]
9. Carolan M. Diabetes nurse educators' experiences of providing care for women, with gestational diabetes mellitus, from disadvantaged backgrounds. J Clin Nurs 2014 May;23(9-10):1374-1384. [doi: 10.1111/jocn.12421] [Medline: 24313229]
10. Hui AL, Sevenhuysen G, Harvey D, Salamon E. Stress and anxiety in women with gestational diabetes during dietary management. Diabetes Educ 2014;40(5):668-677. [doi: 10.1177/0145721714535991] [Medline: 24874692]
11. Martis R, Brown J, McAra-Couper J, Crowther CA. Enablers and barriers for women with gestational diabetes mellitus to achieve optimal glycaemic control-a qualitative study using the theoretical domains framework. BMC Pregnancy Childbirth 2018 Apr 11;18(1):91 [FREE Full text] [doi: 10.1186/s12884-018-1710-8] [Medline: 29642898]
12. Chan JC, Malik V, Jia W, Kadowaki T, Yajnik CS, Yoon K, et al. Diabetes in Asia: epidemiology, risk factors, and pathophysiology. J Am Med Assoc 2009 May 27;301(20):2129-2140. [doi: 10.1001/jama.2009.726] [Medline: 19470990]
13. Hulman A, Simmons RK, Brunner EJ, Witte DR, Færch K, Vistisen D, et al. Trajectories of glycaemia, insulin sensitivity and insulin secretion in South Asian and white individuals before diagnosis of type 2 diabetes: a longitudinal analysis from the Whitehall II cohort study. Diabetologia 2017 Jul;60(7):1252-1260 [FREE Full text] [doi: 10.1007/s00125-017-4275-6] [Medline: 28409212]
14. Ma RC, Chan JC. Type 2 diabetes in east Asians: similarities and differences with populations in Europe and the United States. Ann N Y Acad Sci 2013 Apr;1281:64-91 [FREE Full text] [doi: 10.1111/nyas.12098] [Medline: 23551121]
15. Jenum AK, Richardsen KR, Berntsen S, Mørkrid K. Gestational diabetes, insulin resistance and physical activity in pregnancy in a multi-ethnic population–a public health perspective. Nor J Epidemiol 2013 Jun 3;23(1). [doi: 10.5324/nje.v23i1.1602]
16. Chong Y, Cai S, Lin H, Soh SE, Lee Y, Leow MK, GUSTO Study Group. Ethnic differences translate to inadequacy of high-risk screening for gestational diabetes mellitus in an Asian population: a cohort study. BMC Pregnancy Childbirth 2014 Oct 2;14:345 [FREE Full text] [doi: 10.1186/1471-2393-14-345] [Medline: 25273851]
17. Chen PY, Finkelstein EA, Ng MJ, Yap F, Yeo GS, Rajadurai VS, et al. Incremental cost-effectiveness analysis of gestational diabetes mellitus screening strategies in Singapore. Asia Pac J Public Health 2016 Jan;28(1):15-25. [doi: 10.1177/1010539515612908] [Medline: 26512030]
18. Goh S, Ang S, Bee Y, Chen Y, Gardner D, Ho E, et al. Ministry of health clinical practice guidelines: diabetes mellitus. Singapore Med J 2014 Jun;55(6):334-347 [FREE Full text] [doi: 10.11622/smedj.2014079] [Medline: 25017409]
19. Chipidza FE, Wallwork RS, Stern TA. Impact of the doctor-patient relationship. Prim Care Companion CNS Disord 2015;17(5) [FREE Full text] [doi: 10.4088/PCC.15f01840] [Medline: 26835164]
20. Viinamäki H, Niskanen L, Korhonen T, Tähkä V. The patient-doctor relationship and metabolic control in patients with type 1 (insulin-dependent) diabetes mellitus. Int J Psychiatry Med 1993;23(3):265-274. [doi: 10.2190/GLKJ-1G71-NCD6-DP6K] [Medline: 8270356]
21. Institute of Medicine, Committee on Quality of Health Care in America. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academies Press; 2001.
22. World Health Organization. 2013. Diagnostic Criteria and Classification of Hyperglycaemia First Detected in Pregnancy URL: https://www.who.int/diabetes/publications/Hyperglycaemia_In_Pregnancy/en/ [accessed 2019-03-03]
23. Walker EA, Caban A, Schechter CB, Basch CE, Blanco E, de Witt T, et al. Measuring comparative risk perceptions in an urban minority population: the risk perception survey for diabetes. Diabetes Educ 2007;33(1):103-110. [doi: 10.1177/0145721706298198] [Medline: 17272796]
24. Ritchie J. Qualitative Research Practice: A Guide for Social Science Students and Researchers. Thousand Oaks, CA: Sage Publications; 2013.
25. Goh SY, Ang SB, Bee YM, Chen YT, Gardner DS, Ho ET, et al. Ministry of health clinical practice guidelines: diabetes mellitus. Singapore Med J 2014 Jun;55(6):334-347 [FREE Full text] [doi: 10.11622/smedj.2014079] [Medline: 25017409]
26. Nielsen KK, Kapur A, Damm P, de Courten M, Bygbjerg IC. From screening to postpartum follow-up-the determinants and barriers for gestational diabetes mellitus (GDM) services, a systematic review. BMC Pregnancy Childbirth 2014 Jan 22;14(1):41 [FREE Full text] [doi: 10.1186/1471-2393-14-41] [Medline: 24450389]
27. Crowther CA, Hiller JE, Moss JR, McPhee AJ, Jeffries WS, Robinson JS, Australian Carbohydrate Intolerance Study in Pregnant Women (ACHOIS) Trial Group. Effect of treatment of gestational diabetes mellitus on pregnancy outcomes. N Engl J Med 2005 Jun 16;352(24):2477-2486. [doi: 10.1056/NEJMoa042973] [Medline: 15951574]
28. Brown J, Ceysens G, Boulvain M. Exercise for pregnant women with gestational diabetes for improving maternal and fetal outcomes. Cochrane Database Syst Rev 2017 Jun 22;6:CD012202 [FREE Full text] [doi: 10.1002/14651858.CD012202.pub2] [Medline: 28639706]
29. Wallwiener S, Müller M, Doster A, Laserer W, Reck C, Pauluschke-Fröhlich J, et al. Pregnancy ehealth and mhealth: user proportions and characteristics of pregnant women using web-based information sources-a cross-sectional study. Arch Gynecol Obstet 2016 Nov;294(5):937-944. [doi: 10.1007/s00404-016-4093-y] [Medline: 27084763]
30. Sayakhot P, Carolan-Olah M. Internet use by pregnant women seeking pregnancy-related information: a systematic review. BMC Pregnancy Childbirth 2016 Mar 28;16:65 [FREE Full text] [doi: 10.1186/s12884-016-0856-5] [Medline: 27021727]

31. El-Gayar O, Timsina P, Nawar N, Eid W. Mobile applications for diabetes self-management: status and potential. J Diabetes Sci Technol 2013 Jan 1;7(1):247-262 [FREE Full text] [doi: 10.1177/193229681300700130] [Medline: 23439183]
32. Ryu S. Telemedicine: opportunities and developments in member states: report on the second global survey on ehealth 2009 (global observatory for ehealth series, volume 2). Healthc Inform Res 2012;18(2):153. [doi: 10.4258/hir.2012.18.2.153]
33. Hirst JE, Mackillop L, Loerup L, Kevat DA, Bartlett K, Gibson O, et al. Acceptability and user satisfaction of a smartphone-based, interactive blood glucose management system in women with gestational diabetes mellitus. J Diabetes Sci Technol 2015 Jan;9(1):111-115 [FREE Full text] [doi: 10.1177/1932296814556506] [Medline: 25361643]
34. Mackillop L, Loerup L, Bartlett K, Farmer A, Gibson OJ, Hirst JE, et al. Development of a real-time smartphone solution for the management of women with or at high risk of gestational diabetes. J Diabetes Sci Technol 2014 Nov;8(6):1105-1114 [FREE Full text] [doi: 10.1177/1932296814542271] [Medline: 25004915]
35. García-Sáez G, Rigla M, Martínez-Sarriegui I, Shalom E, Peleg M, Broens T, et al. Patient-oriented computerized clinical guidelines for mobile decision support in gestational diabetes. J Diabetes Sci Technol 2014 Mar;8(2):238-246 [FREE Full text] [doi: 10.1177/1932296814526492] [Medline: 24876573]
36. Caballero-Ruiz E, García-Sáez G, Rigla M, Villaplana M, Pons B, Hernando ME. A web-based clinical decision support system for gestational diabetes: automatic diet prescription and detection of insulin needs. Int J Med Inform 2017 Jun;102:35-49. [doi: 10.1016/j.ijmedinf.2017.02.014] [Medline: 28495347]
37. Rasekaba TM, Furler J, Blackberry I, Tacey M, Gray K, Lim K. Telemedicine interventions for gestational diabetes mellitus: a systematic review and meta-analysis. Diabetes Res Clin Pract 2015 Oct;110(1):1-9. [doi: 10.1016/j.diabres.2015.07.007] [Medline: 26264410]
38. Ming W, Mackillop LH, Farmer AJ, Loerup L, Bartlett K, Levy JC, et al. Telemedicine technologies for diabetes in pregnancy: a systematic review and meta-analysis. J Med Internet Res 2016 Nov 9;18(11):e290 [FREE Full text] [doi: 10.2196/jmir.6556] [Medline: 27829574]
39. Kruger DF, White K, Galpern A, Mann K, Massirio A, McLellan M, et al. Effect of modem transmission of blood glucose data on telephone consultation time, clinic work flow, and patient satisfaction for patients with gestational diabetes mellitus. J Am Acad Nurse Pract 2003 Aug;15(8):371-375. [doi: 10.1111/j.1745-7599.2003.tb00410.x] [Medline: 14509102]
40. Dalfrà MG, Nicolucci A, Lapolla A, TISG. The effect of telemedicine on outcome and quality of life in pregnant women with diabetes. J Telemed Telecare 2009;15(5):238-242. [doi: 10.1258/jtt.2009.081213] [Medline: 19590029]
41. Pérez-Ferre N, Galindo M, Fernández MD, Velasco V, Runkle I, de la Cruz MJ, et al. The outcomes of gestational diabetes mellitus after a telecare approach are not inferior to traditional outpatient clinic visits. Int J Endocrinol 2010;2010:386941 [FREE Full text] [doi: 10.1155/2010/386941] [Medline: 20628517]
42. Chen Q, Carbone ET. Functionality, implementation, impact, and the role of health literacy in mobile phone apps for gestational diabetes: scoping review. JMIR Diabetes 2017 Oct 4;2(2):e25 [FREE Full text] [doi: 10.2196/diabetes.8045] [Medline: 30291088]
43. Garnweidner-Holme L, Borgen I, Garitano I, Noll J, Lukasse M. Designing and developing a mobile smartphone application for women with gestational diabetes mellitus followed-up at diabetes outpatient clinics in Norway. Healthcare (Basel) 2015 May 21;3(2):310-323 [FREE Full text] [doi: 10.3390/healthcare3020310] [Medline: 27417764]
44. Nicholson WK, Beckham AJ, Hatley K, Diamond M, Johnson L, Green SL, et al. The gestational diabetes management system (GooDMomS): development, feasibility and lessons learned from a patient-informed, web-based pregnancy and postpartum lifestyle intervention. BMC Pregnancy Childbirth 2016 Sep 21;16(1):277 [FREE Full text] [doi: 10.1186/s12884-016-1064-z] [Medline: 27654119]
45. Ministry of Manpower. 2017. Labour Market Statistical Information URL: http://stats.mom.gov.sg/Pages/Home.aspx [accessed 2019-03-20]
46. Alayoub H, Curran S, Coffey M, Hatunic M, Higgins M. Assessment of the effectiveness of group education on knowledge for women with newly diagnosed gestational diabetes. Ir J Med Sci 2018 Feb;187(1):65-68. [doi: 10.1007/s11845-017-1609-9] [Medline: 28477326]
47. The National Institute for Health and Care Excellence. 2017. Health App: GDm-Health for People With Gestational Diabetes URL: https://www.nice.org.uk/advice/mib131/chapter/Summary [accessed 2019-03-08]
48. Harrison TN, Sacks DA, Parry C, Macias M, Grant DS, Lawrence JM. Acceptability of virtual prenatal visits for women with gestational diabetes. Womens Health Issues 2017;27(3):351-355. [doi: 10.1016/j.whi.2016.12.009] [Medline: 28153743]

Abbreviations

DSRB: domain-specific review board
GDM: gestational diabetes mellitus
NHG: National Healthcare Group
NUH: National University Hospital
NUS: National University of Singapore
T2DM: type 2 diabetes mellitus
WHO: World Health Organization

Edited by G Eysenbach; submitted 28.04.19; peer-reviewed by J Most, W Na; comments to author 14.12.19; revised version received 04.02.20; accepted 21.02.20; published 30.06.20.

Please cite as:
Hewage S, Audimulam J, Sullivan E, Chi C, Yew TW, Yoong J
Barriers to Gestational Diabetes Management and Preferred Interventions for Women With Gestational Diabetes in Singapore: Mixed Methods Study
JMIR Form Res 2020;4(6):e14486
URL: http://formative.jmir.org/2020/6/e14486/
doi:10.2196/14486
PMID:32602845

Original Paper

Detecting Screams From Home Audio Recordings to Identify Tantrums: Exploratory Study Using Transfer Machine Learning

Rebecca O'Donovan[1], BSc; Emre Sezgin[1], PhD; Sven Bambach[1], PhD; Eric Butter[2], PhD; Simon Lin[1], MBA, MD

[1]The Abigail Wexner Research Institute, Nationwide Children's Hospital, Columbus, OH, United States

[2]Department of Psychology, Nationwide Children's Hospital, Columbus, OH, United States

Corresponding Author:
Emre Sezgin, PhD
The Abigail Wexner Research Institute
Nationwide Children's Hospital
700 Children's Drive
Columbus, OH, 43205
United States
Phone: 1 6143556814
Email: emre.sezgin@nationwidechildrens.org

Related Article:

This is a corrected version. See correction statement: https://formative.jmir.org/2020/7/e21591/

Abstract

Background: Qualitative self- or parent-reports used in assessing children's behavioral disorders are often inconvenient to collect and can be misleading due to missing information, rater biases, and limited validity. A data-driven approach to quantify behavioral disorders could alleviate these concerns. This study proposes a machine learning approach to identify screams in voice recordings that avoids the need to gather large amounts of clinical data for model training.

Objective: The goal of this study is to evaluate if a machine learning model trained only on publicly available audio data sets could be used to detect screaming sounds in audio streams captured in an at-home setting.

Methods: Two sets of audio samples were prepared to evaluate the model: a subset of the publicly available AudioSet data set and a set of audio data extracted from the TV show Supernanny, which was chosen for its similarity to clinical data. Scream events were manually annotated for the Supernanny data, and existing annotations were refined for the AudioSet data. Audio feature extraction was performed with a convolutional neural network pretrained on AudioSet. A gradient-boosted tree model was trained and cross-validated for scream classification on the AudioSet data and then validated independently on the Supernanny audio.

Results: On the held-out AudioSet clips, the model achieved a receiver operating characteristic (ROC)–area under the curve (AUC) of 0.86. The same model applied to three full episodes of Supernanny audio achieved an ROC-AUC of 0.95 and an average precision (positive predictive value) of 42% despite screams only making up 1.3% (n=92/7166 seconds) of the total run time.

Conclusions: These results suggest that a scream-detection model trained with publicly available data could be valuable for monitoring clinical recordings and identifying tantrums as opposed to depending on collecting costly privacy-protected clinical data for model training.

(JMIR Form Res 2020;4(6):e18279) doi:10.2196/18279

KEYWORDS

machine learning; scream detection; audio event detection; tantrum identification; autism; behavioral disorder; data-driven approach

Introduction

One of the challenges in studying and diagnosing children with (potential) behavioral disorders is the relative unreliability of the information available. Children with autism and other social or disruptive behavior disorders often report their own behavior inaccurately in interviews and on self-report questionnaires [1,2]. Parental reports of behavior, although found to be more accurate than child self-reports [1], can be influenced by parents' own adjustment [3] and their desire to either exaggerate or diminish familial problems [4]. Parents' reports can also be affected by poor recall of events and confusion about survey or interview questions. A quantitative, data-driven approach for evaluating behavioral problems would alleviate the need to rely on these potentially inaccurate reports.

With this goal in mind, our study focuses on detecting human screaming within continuous audio recording from inside a family's home. Real-life audio data streams collected from homes have already proven useful in the diagnosis and treatment of behavioral problems associated with autism [5-7]. Segmenting home audio to capture screams, with some postprocessing, could eventually allow researchers to use the scream as a proxy for temper tantrums or negative interactions between family members [8]. Analyzing these interactions would allow clinicians to assess family relationships in a more objective manner than direct self-report [9], but identifying the segments manually would be tedious and time intensive.

Although much research has been done on audio event classification, this study investigates the ability of a scream-detection model to make useful predictions for audio outside of the training data, which to our knowledge has not yet been addressed in previous work. As such, this work can be considered a proof-of-concept study, demonstrating that a model trained exclusively with publicly available audio data can be used to detect screams in clinical data that would otherwise be costly to collect.

Methods

The ideal way to train a machine learning model for clinical use would be to obtain large amounts of clinical data and hold out certain examples to use for testing and validation. However, relevant clinical data (eg, home audio recordings of families with children) can often be hard to obtain. We trained the classifier on a publicly available audio database, then approximated clinical data for validation using episodes of the TV show Supernanny.

Our training data is based on portions of the AudioSet data set, "an expanding ontology of 632 audio event classes and a collection of 2,084,320 human-labeled 10-second sound clips drawn from YouTube videos" [10]. Each AudioSet clip has been watched by a human annotator and labeled with any number of the 632 sound classes as clips can contain consecutive or overlapping audio events. For our training set, we compiled all clips annotated with the label "Scream." AudioSet defines this sound as, "A sharp, high-pitched human vocalization; often an instinctive action indicating fear, pain, surprise, joy, anger, etc. Verbal content is absent or overwhelmed, unlike Shout and Yell" [10], and does not make distinctions between the emotions producing the scream sound.

Additionally, validation of the AudioSet annotations was performed in two ways. First, several of the "Scream"-labeled clips did not contain screaming, or it was nearly indistinguishable, so these annotations were corrected. Second, annotating the on- and offset of a scream (on 1-second intervals) within each clip was necessary because most audio clips did not contain a full 10 seconds of their labeled sound. The resulting data set was composed of 3764 seconds of screaming audio tracks (and the same amount of randomly sampled nonscream audio) from 1158 different AudioSet clips. Additional data augmentation was done by duplicating each sound sample and adding white noise at a 20 dB signal-to-noise ratio (SNR), resulting in a final balanced set of ~15,000 seconds of audio.

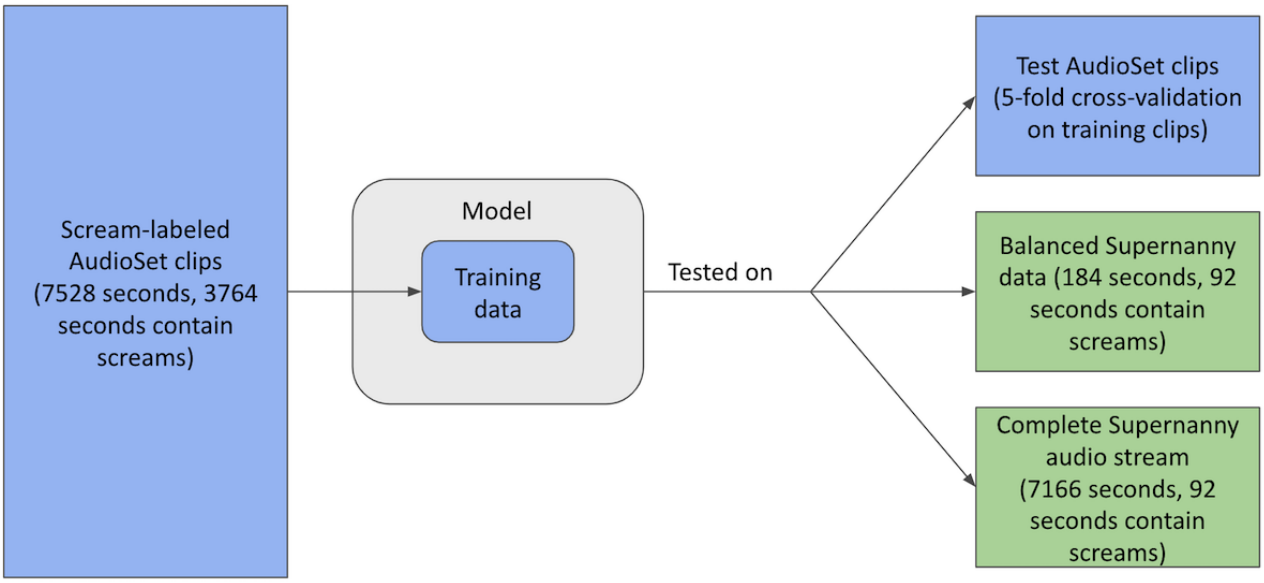
The target use case for the model is home audio recordings of families with children, which would likely contain normal conversation, arguing, screaming, and crying as well as silence, background noise, and household noises. In the absence of available clinical recordings of this type, episodes of the TV show Supernanny were used as stand-ins. Each episode focuses on a family in which the parents are struggling to control the behavior of their children. The "Supernanny" (Jo Frost) observes the family, then teaches the parents alternative strategies for discipline and parenting [11]. Supernanny contains several elements that would also be present in clinical data: genuine arguments, children screaming and crying, normal family conversations, and ambient noise. It also contains positive (scream) and negative (nonscream) data points from the same environment (the family's home) as clinical data would. Three Supernanny episodes were selected and annotated with the same process as the AudioSet clips; they were watched by a human annotator, and every second of audio data was classified as either scream or nonscream. Figure 1 depicts the role of Supernanny audio in relation to other data used in the study.

Two approaches were considered for audio feature extraction. The first was the OpenSMILE (audEERING GmbH) feature extraction software [12], which can extract a range of feature sets using different configuration files. For example, mel-frequency cepstral coefficients (MFCCs) for speech recognition can be extracted, as well as feature sets tuned specifically for emotion recognition [12]. The second approach considered was Google's open-source "VGGish" convolutional neural network (CNN) for feature extraction [13]. Google's VGGish network was chosen for two reasons: first, the research behind its feature extraction is more current, and second, it was used in similar recent classification tasks [14]. Functionally, the network analyzes audio and produces a 128-dimensional feature vector extracted by transforming the spectrogram (frequency information) of the audio sample. Since it processes audio data in 960 ms segments, we classified training and test data in 1-second intervals for compatibility.

The original CNN was trained on audio from YouTube-8M, a database of 8 million YouTube videos with over 30,000 class labels, so the features it produces are suitable for audio event

classification [13]. We did not fine-tune the VGGish network and used it only as a feature extractor. For classification, Scikit-learn's gradient boosting classifier was selected for practical reasons (speed of training and ease of use) [15]. This type of classifier iteratively creates an ensemble of simple decision trees such that each new tree tries to correct the previous trees' biases.

Figure 1. Study design. The model was built using the AudioSet clips by standard 5-fold cross-validation. The model was then tested on Supernanny data for transferability.



Results

Quantitative Evaluation

The classifier was trained and cross-validated on the class-balanced scream data set of AudioSet clips. The receiver operating characteristic (ROC) curve from a 5-fold cross-validation is shown in Figure 2. The area under the curve (AUC) is 0.86, indicating that our model successfully learned to classify screams using the out-of-the-box VGGish neural network as a feature extractor.

In the Supernanny data set, screams occur relatively rarely: only 92 seconds, or 1.3%, of the audio contained screaming out of the full 2 hours (7166 seconds) of annotated audio. Thus, two test sets of Supernanny audio were prepared. The first was artificially balanced by taking every positively marked second, then randomly sampling the same amount of negatively marked data points throughout the audio. The ROC curve for this test set is shown in Figure 3, achieving an AUC of 0.91. This result demonstrates that our model successfully transfers to correctly classifying screams in Supernanny episodes without the need of additional training.

Figure 2. Scream classification performance on the AudioSet data. AUC: area under the curve; ROC: receiver operating characteristic.

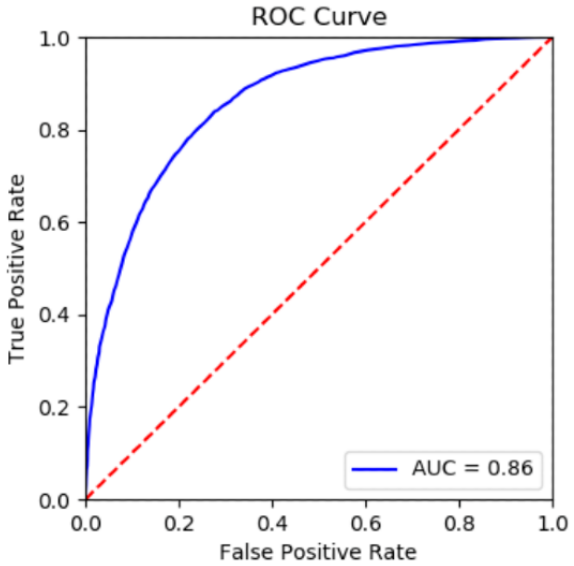
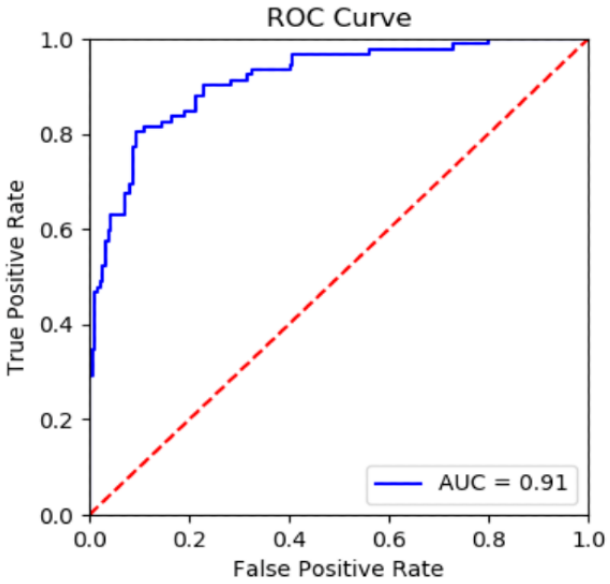


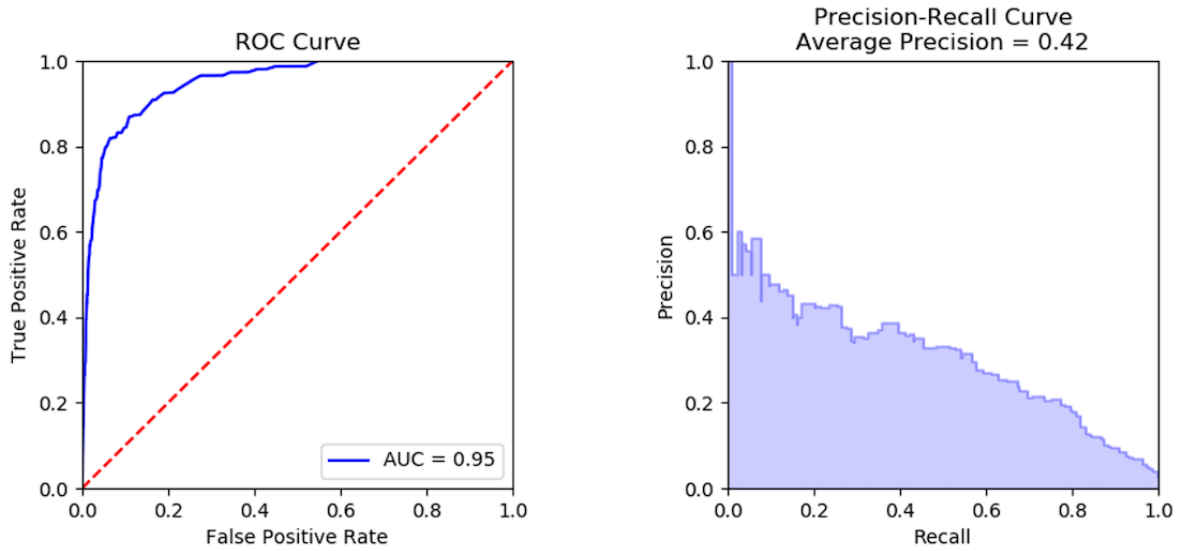
Figure 3. Results on balanced Supernanny data. AUC: area under the curve; ROC: receiver operating characteristic.



Additionally, we tested the model on the entire audio stream of all three Supernanny episodes. For this evaluation, we incorporated a 1-second margin of error for false positives. For example, if the ground truth annotations for an audio sample contained a scream at second 6 but the model predicted the scream at second 7, the model prediction for the scream would be counted as correct. There were two motivating factors for this approach. First, it accounts for small errors and imprecisions in the ground truth annotations (ie, if an annotator marked the onset of a scream slightly earlier or later than it actually occurred). Second, our model outputs predictions for 960 ms segments of audio, but our ground truth annotations are on a 1-second grid. To test the model's predictions, we match each (960 ms) prediction to the ground truth second it most overlaps with. Using a 1-second margin of error helps to not penalize the model for misalignment with our ground truth.

The ROC curve is shown in Figure 4 (left). The slight increase in AUC compared to Figure 3 is likely caused by our 1-second margin of error. Because ROC curves can be misleading in evaluating classifiers for rare events [16], we also provide the precision-recall curve in Figure 4 (right), indicating an average precision (positive predictive value) of 0.42.

Figure 4. Results on all Supernanny data. AUC: area under the curve; ROC: receiver operating characteristic.



Qualitative Evaluation

To evaluate the types of errors that our model made, we manually examined its predictions on a single episode of Supernanny. We found that many of the false negatives (actual screams that the model missed) were higher-frequency screams by children. We found that of the 9 false positives (seconds the model marked as screams but were not marked as screams by the human annotator) in the episode, 3 seemed to result from very short (<1 second) screams that we chose not to mark as positives during annotation, and 2 occurred on scream-like sounds like loud laughter and crying.

Discussion

Principal Results

The most significant finding of this study is that a scream detection model trained only on publicly available data sets had success making predictions on an independent data set containing child screams. Although we have not shown the ability to generalize directly to clinical data, we have shown generalizability from a public database to outside recordings containing home audio characteristics. One of the most significant quantitative results is the average precision value of 0.42 when evaluating the entire Supernanny audio stream, in which screams have a frequency of only 1.3%. A relatively high precision is important for a practical application of the model, especially as we expect scream events in real continuous home audio recordings to be rare as well.

Several characteristics of the proxy test data may bode well for the model's ability to classify clinical data. One of the most significant differences between the TV show and presumed clinical data is the presence of high-volume, overlapping sound events like voiceover narration and background music in the TV show. Intuitively, loud overlapping audio events would increase the difficulty of classifying the scream event. Sounds are harder to identify when paired with other louder sounds, and model performance on similar data has been shown to suffer as SNR decreases [17]. Thus, it is possible that the classifier would actually perform better on less noisy clinical recordings than it has on the Supernanny audio streams.

Generalizability from public to clinical data has value because it could allow classifiers for clinical recordings to train on public databases. Including patient data in model training sets causes significant risks and inconveniences; for example, if clinical data were directly obtained or reverse-engineered from the trained classifier, patients' home recordings would be vulnerable to third parties. In addition, consent procedures for inclusion of clinical recordings in training data would be extensive and time-consuming given the number of samples necessary for training in most cases.

Limitations

This study does not include any clinical data and instead relies on a proxy test data set. Thus, differences between the proxy data (the TV show Supernanny) and the "real-world" data we anticipate using (audio recordings from family homes) constitute limitations. Supernanny contains several audio elements that would not appear in clinical recordings (eg, pervasive background music, voiceover narration, scene cuts). In addition, home audio recordings would include long stretches of silence and everyday conversation that the Supernanny episodes do not include. It is possible that the AudioSet training data has more similarities to Supernanny audio samples than to clinical audio samples, and if so, we could observe a drop in model performance when evaluating the model on clinical data.

Additionally, for purposes of scream classification, this study does not discriminate between screams evoked by different emotions, a distinction that could be clinically useful. For example, an angry scream produced during a tantrum and a happy scream produced while playing would both be classified simply as "screams" by our model. Although both may be useful as emotional outbursts, clinical researchers may want to focus on a single type of scream, and such study would require review or postprocessing of the scream audio selected by the model.

The Supernanny scream annotations were not validated by a second researcher. Furthermore, Supernanny episodes are 35-40 minutes in length, but the audio recordings most likely to be clinically informative could be days or weeks in length. If the model's performance on clinical recordings is analogous to its performance on Supernanny, it may produce too many false positives for the raw output to be usable. However, postprocessing may help to reduce the number of false positives and isolate the true episodes of screaming.

Comparison With Prior Work

Technical Work

Earlier works in the area of scream detection have focused on evaluating different audio feature sets and determining which provide the most accurate classification results. Feature sets that have been tested include MFCCs [18], pitch-based features [19], a bag-of-audio-words feature representation [20], autocorrelation features [17], and various hand-constructed feature sets [21]. Studies before 2015 generally used Gaussian mixture models or support vector machines to classify audio data points [18,19,21]. More recent works have tested deep-learning classifiers like deep Boltzmann machines and deep belief networks [22,23]. Additionally, recent studies have tested robustness to noise, using both artificially generated noise and naturally occurring environmental noise [17,23].

Previous studies use a wide range of training data. Due to the scarcity of publicly available audio databases, training data sets generally fall into 1 of 2 categories: "self-compiled" [18-21] or "self-recorded" [17,22,23]. Self-compiled databases use audio samples from sound effect websites, movies, or other sources accessible to researchers. These sound samples may bear little acoustic resemblance to sounds detected "out in the wild." Databases recorded by the researchers themselves were more realistic but far smaller. These data sets also tend to be narrow in scope, limiting generalizability. For example, Nandwana et al [17] stitched together examples of speech and screaming from 6 male speakers to form 24 continuous recordings, and Zaheer et al [22] recorded 130 scream sounds and 110 "ah" sounds from 60 people in outdoor settings.

The two studies with the most similarities to our work are Nandwana et al [17] and Girin [23]. Nandwana et al [17] show the ability to segment continuous recordings into "scream" and "nonscream" portions, and train and test their model on noisy data. Girin [23] uses a realistic data set with 4 overlapping sound classes and naturally occurring (not artificially added) noise. This study builds on their work by using a larger, less homogenous, and more challenging training data set, and by testing the trained model on audio samples not originating from the training database.

Clinical Work

Although scream detection for clinical purposes does not seem to have been explored, we found several sources proposing methods for tantrums and aggression detection. Lefter et al [24] use a fusion of audio and visual streams for automated aggression detection on trains. Although this work has obvious applications in surveillance, a similar methodology could be used to detect aggression or anger in the home.

Current solutions for tantrum detection involve wearable devices combined with parent report [8,25]. The proposed approach for scream detection could be combined with either video-based aggression detection or wearable sensors to improve their accuracy in detecting tantrums or arguments. Another possible method for tantrum detection involves using scream detection to turn a video stream on and off, then allowing parents and caregivers to review the resulting videos and send to clinicians. This parental review process might also alleviate privacy concerns surrounding in-home video collection.

Our proposed method of information extraction may also be used to assist in behavioral management techniques like the "functional behavioral analysis." This type of clinical assessment has been proven effective in analyzing and designing treatment plans for problem behaviors but can require as many as 40-50 clinician observation sessions per child [26,27]. Video analysis and telecommunication have already been implemented to aid teachers in functional analyses at school with some success [28,29], and the same strategy could be applied using home video. Accurate home video segmentation could provide examples of the problem behavior if it involves or provokes screaming and reduce the need for clinician observation of the child.

The major barriers that remain for using scream detection analysis for routine clinical care relate to identifying conveniently ascertained but also relevant clinical samples that will have generalizability to real-world clinical applications. Screams and verbal outbursts are associated with behavioral disruption and are unpredictable, occurring infrequently but severely in some children while presenting more mildly with increased frequency in other children. The key to effective design of a clinical application will be the ability to ascertain video and audio samples triggered by child screams rather than collecting indeterminant hours of nondisruptive social interaction. The technology, privacy, and accuracy of this approach remains to be explored.

Conclusions

In this study, we present an approach for detecting screams in home audio recordings with the aim of segmenting those recordings to review clinically relevant portions. The proposed machine-learning approach draws training data from Google's AudioSet database, uses their associated CNN for feature extraction, and classifies data points with a gradient-boosted tree model. All of these tools are publicly available. The model has achieved an AUC of 0.86 on audio from the training data set, and an AUC of 0.91 on balanced data selected from episodes of the television show Supernanny. Furthermore, it has achieved an AUC of 0.95 and an average precision of 0.42 on the entire unmodified stream of Supernanny audio. These results suggest a possibility that a scream-detecting classifier for clinical audio recordings could be trained using a public database. Additionally, if scream information from audio is used to segment a home video stream, the resulting video clips could be relevant for behavioral disorder diagnosis or management in disorders like autism spectrum disorder or Prader-Willi syndrome. In future work, we plan to obtain and test on clinical data samples to further these hypotheses.

Acknowledgments

Authors acknowledge Brannon Oiler for his support in data retrieval and algorithm development, and Dr Yungui Huang, Dr Steve Rust, and the data science team at Nationwide Children's Hospital for helpful discussions and insight.

Conflicts of Interest

None declared.

References

1. Fisher MH, Mello MP, Dykens EM. Who reports it best? A comparison between parent-report, self-report, and the real life social behaviors of adults with Williams syndrome. Res Dev Disabil 2014 Dec;35(12):3276-3284 [FREE Full text] [doi: 10.1016/j.ridd.2014.08.011] [Medline: 25181359]
2. Finlay WML, Lyons E. Methodological issues in interviewing and using self-report questionnaires with people with mental retardation. Psychological Assess 2001;13(3):319-335. [doi: 10.1037/1040-3590.13.3.319]
3. Phares V, Compas BE, Howell DC. Perspectives on child behavior problems: comparisons of children's self-reports with parent and teacher reports. Psychological Assess: J Consult Clin Psychol 1989;1(1):68-71. [doi: 10.1037/1040-3590.1.1.68]
4. Bennett DS, Sullivan MW, Lewis M. Relations of parental report and observation of parenting to maltreatment history. Child Maltreat 2006 Feb;11(1):63-75 [FREE Full text] [doi: 10.1177/1077559505283589] [Medline: 16382092]
5. Oller DK, Niyogi P, Gray S, Richards JA, Gilkerson J, Xu D, et al. Automated vocal analysis of naturalistic recordings from children with autism, language delay, and typical development. Proc Natl Acad Sci U S A 2010 Jul 27;107(30):13354-13359 [FREE Full text] [doi: 10.1073/pnas.1003882107] [Medline: 20643944]

6. Dutheil F, Chambres P, Hufnagel C, Auxiette C, Chausse P, Ghozi R, et al. 'Do Well B.': design Of WELL Being monitoring systems. A study protocol for the application in autism. BMJ Open 2015 Feb 20;5(2):e007716 [FREE Full text] [doi: 10.1136/bmjopen-2015-007716] [Medline: 25710916]
7. Xu D, Gilkerson J, Richards J. Objective child behavior measurement with naturalistic daylong audio recording and its application to autism identification. 2012 Presented at: Conf Proc . Annu Int Conf IEEE Eng Med Biol Soc IEEE Eng Med Biol Soc Annu Conf; 2012; San Diego, CA p. 3708-3711. [doi: 10.1109/embc.2012.6346772]
8. Green JA, Whitney PG, Potegal M. Screaming, yelling, whining, and crying: categorical and intensity differences in vocal expressions of anger and sadness in children's tantrums. Emotion 2011 Oct;11(5):1124-1133 [FREE Full text] [doi: 10.1037/a0024173] [Medline: 21707157]
9. Schroeder JH, Cappadocia MC, Bebko JM, Pepler DJ, Weiss JA. Shedding light on a pervasive problem: a review of research on bullying experiences among children with autism spectrum disorders. J Autism Dev Disord 2014 Jul;44(7):1520-1534. [doi: 10.1007/s10803-013-2011-8] [Medline: 24464616]
10. Gemmeke J, Ellis D, Freedman D, Jansen A, Lawrence W, Moore R, et al. Audio set: an ontology and human-labeled dataset for audio events. : IEEE; 2017 Presented at: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2017; New Orleans, LA p. 776-780. [doi: 10.1109/icassp.2017.7952261]
11. YouTube. United Kingdom: Supernanny; 2012. Supernanny URL: https://www.youtube.com/channel/UCSkKAqmw8DZjU2A-Fupl1_Q [accessed 2019-09-09]
12. Eyben F, Weninger F, Gross F, Schuller B. Recent developments in openSMILE, the munich open-source multimedia feature extractor. In: MM '13: Proceedings of the 21st ACM international conference on Multimedia.: ACM; 2013 Presented at: 21st ACM international conference on Multimedia; 2013; Barcelona, Spain p. 835-838. [doi: 10.1145/2502081.2502224]
13. Hershey S, Chaudhuri S, Ellis D. CNN architectures for large-scale audio classification. : IEEE; 2017 Presented at: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2017; New Orleans, LA p. 131-135. [doi: 10.1109/icassp.2017.7952132]
14. Chan J, Rea T, Gollakota S, Sunshine JE. Contactless cardiac arrest detection using smart devices. NPJ Digit Med 2019;2:52 [FREE Full text] [doi: 10.1038/s41746-019-0128-7] [Medline: 31304398]
15. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Machine Learning Res 2011:2825-2830.
16. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLoS One 2015;10(3):e0118432 [FREE Full text] [doi: 10.1371/journal.pone.0118432] [Medline: 25738806]
17. Nandwana M, Ziaei A, Hansen J. Robust unsupervised detection of human screams in noisy acoustic environments. : IEEE; 2015 Presented at: 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2015; Brisbane, QLD, Australia p. 161-165. [doi: 10.1109/icassp.2015.7177952]
18. Huang W, Chiew T, Li H, Kok T, Biswas J. Scream detection for home applications. 2010 Presented at: Proc 5th IEEE Conf Ind Electron Appl ICIEA; 2010; Taichung, Taiwan p. 2115-2120. [doi: 10.1109/iciea.2010.5515397]
19. Uzkent B, Barkana B, Cevikalp H. Non-speech environmental sound classification using SVMs with a new set of features. Int J Innovative Computing Inf Control 2012;8(5):3511-3524.
20. Lim H, Kim M, Kim H. Robust sound event classification using LBP-HOG based bag-of-audio-words feature representation. In: Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH. Vol 2015-Janua; 2015 Presented at: 16th Annual Conference of the International Speech Communication Association; 2015; Dresden, Germany p. 3325-3329.
21. Valenzise G, Gerosa L, Tagliasacchi M, Antonacci F, Sarti A. Scream and gunshot detection and localization for audio-surveillance systems. : IEEE; 2007 Presented at: 2007 IEEE Conference on Advanced Video and Signal Based Surveillance; 2007; London, UK. [doi: 10.1109/avss.2007.4425280]
22. Zaheer Z, Young KJ, Kim H, You NS. A preliminary study on deep-learning based screaming sound detection. 2015 Presented at: 2015 5th International Conference on IT Convergence and Security (ICITCS); 2015; Kuala Lumpur, Malaysia. [doi: 10.1109/icitcs.2015.7292925]
23. Laffitte P, Sodoyer D, Tatkeu C, Girin L. Deep neural networks for automatic detection of screams and shouted speech in subway trains. : IEEE; 2016 Presented at: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2016; Shanghai, China p. 6460-6464. [doi: 10.1109/icassp.2016.7472921]
24. Lefter I, Rothkrantz L, Burghouts G. A comparative study on automatic audio–visual fusion for aggression detection using meta-information. Pattern Recognition Lett 2013 Nov;34(15):1953-1963. [doi: 10.1016/j.patrec.2013.01.002]
25. Yu X, Wu L, Liu Q, Zhou H. Children tantrum behaviour analysis based on Kinect sensor. : IEEE; 2011 Presented at: 2011 Third Chinese Conference on Intelligent Visual Surveillance; 2011; Beijing, China p. 49-52. [doi: 10.1109/ivsurv.2011.6157022]
26. Healy O, Brett D. Functional behavior assessments for challenging behavior in autism. In: Comprehensive Guide to Autism. New York, NY: Springer; 2014.

27. O'Reilly M, Rispoli M, Davis T, Machalicek W, Lang R, Sigafoos J, et al. Functional analysis of challenging behavior in children with autism spectrum disorders: a summary of 10 cases. Res Autism Spectrum Disord 2010 Jan;4(1):1-10. [doi: 10.1016/j.rasd.2009.07.001]
28. Machalicek W, O'Reilly M, Chan J, Lang R, Rispoli M, Davis T, et al. Using videoconferencing to conduct functional analysis of challenging behavior and develop classroom behavioral support plans for students with autism. Education Training Developmental Disabilities 2009;44(2):207-217.
29. Machalicek W, O'Reilly M, Rispoli M, Davis T, Lang R, Franco JH, et al. Training teachers to assess the challenging behaviors of students with autism using video tele-conferencing. Education Training Autism Developmental Disabilities 2010;45(2):203-215.

Abbreviations

AUC: area under the curve
CNN: convolutional neural network
MFCC: mel-frequency cepstral coefficient
ROC: receiver operating characteristic
SNR: signal-to-noise ratio

Edited by G Eysenbach; submitted 17.02.20; peer-reviewed by T Muto, C Jacob; comments to author 23.03.20; revised version received 17.04.20; accepted 19.04.20; published 16.06.20.

Please cite as:
O'Donovan R, Sezgin E, Bambach S, Butter E, Lin S
Detecting Screams From Home Audio Recordings to Identify Tantrums: Exploratory Study Using Transfer Machine Learning
JMIR Form Res 2020;4(6):e18279
URL: http://formative.jmir.org/2020/6/e18279/
doi:10.2196/18279
PMID:32459656

Original Paper

Patient Perception of Plain-Language Medical Notes Generated Using Artificial Intelligence Software: Pilot Mixed-Methods Study

Sandeep Bala[1], BS; Angela Keniston[2], MSPH; Marisha Burden[2], MD

[1]College of Medicine, University of Central Florida, Orlando, FL, United States

[2]Division of Hospital Medicine, University of Colorado School of Medicine, Aurora, CO, United States

Corresponding Author:
Sandeep Bala, BS
College of Medicine
University of Central Florida
6850 Lake Nona Blvd
Orlando, FL,
United States
Phone: 1 321 299 8429
Email: ucfbala@knights.ucf.edu

Abstract

Background: Clinicians' time with patients has become increasingly limited due to regulatory burden, documentation and billing, administrative responsibilities, and market forces. These factors limit clinicians' time to deliver thorough explanations to patients. OpenNotes began as a research initiative exploring the ability of sharing medical notes with patients to help patients understand their health care. Providing patients access to their medical notes has been shown to have many benefits, including improved patient satisfaction and clinical outcomes. OpenNotes has since evolved into a national movement that helps clinicians share notes with patients. However, a significant barrier to the widespread adoption of OpenNotes has been clinicians' concerns that OpenNotes may cost additional time to correct patient confusion over medical language. Recent advances in artificial intelligence (AI) technology may help resolve this concern by converting medical notes to plain language with minimal time required of clinicians.

Objective: This pilot study assesses patient comprehension and perceived benefits, concerns, and insights regarding an AI-simplified note through comprehension questions and guided interview.

Methods: Synthea, a synthetic patient generator, was used to generate a standardized medical-language patient note which was then simplified using AI software. A multiple-choice comprehension assessment questionnaire was drafted with physician input. Study participants were recruited from inpatients at the University of Colorado Hospital. Participants were randomly assigned to be tested for their comprehension of the standardized medical-language version or AI-generated plain-language version of the patient note. Following this, participants reviewed the opposite version of the note and participated in a guided interview. A Student *t* test was performed to assess for differences in comprehension assessment scores between plain-language and medical-language note groups. Multivariate modeling was performed to assess the impact of demographic variables on comprehension. Interview responses were thematically analyzed.

Results: Twenty patients agreed to participate. The mean number of comprehension assessment questions answered correctly was found to be higher in the plain-language group compared with the medical-language group; however, the Student *t* test was found to be underpowered to determine if this was significant. Age, ethnicity, and health literacy were found to have a significant impact on comprehension scores by multivariate modeling. Thematic analysis of guided interviews highlighted patients' perceived benefits, concerns, and suggestions regarding such notes. Major themes of benefits were that simplified plain-language notes may (1) be more useable than unsimplified medical-language notes, (2) improve the patient-clinician relationship, and (3) empower patients through an enhanced understanding of their health care.

Conclusions: AI software may translate medical notes into plain-language notes that are perceived as beneficial by patients. Limitations included sample size, inpatient-only setting, and possible confounding factors. Larger studies are needed to assess comprehension. Insight from patient responses to guided interviews can guide the future study and development of this technology.

(JMIR Form Res 2020;4(6):e16670) doi:10.2196/16670

KEYWORDS

artificial intelligence; patient education; natural language processing; OpenNotes; Open Notes; patient-physician relationship; simplified notes; plain-language notes

Introduction

Educating patients has been found to empower them to be involved participants in their health care and is an important principle of patient-centered care, yet clinicians are faced with increasing demands on their time, limiting opportunities to deliver thorough and understandable explanations to patients during encounters. Competing factors such as regulatory burden, documentation and billing, administrative responsibilities, and market forces that pressure clinicians to see more patients have increasingly limited time available to review medical information with patients. A study of ambulatory clinicians across four different specialties found that for each hour doctors spent at the bedside with patients, they spent 2 hours on desk work during the day, not including additional time spent documenting at home [1]. Another study of primary care physicians across 10 clinics found that more time is spent working in the electronic health record than on face-to-face interactions with patients [2].

Patients have rated communication and the quality of explanations about their health care among the most important factors for their satisfaction. In the inpatient setting, a recent focus group study of patients across four hospitals found that good communication and high-quality information provided at arrival and discharge were valued as important by patients [3].

Furthermore, improving patients' knowledge about their conditions has been shown to improve patient compliance and clinical outcomes in various chronic disease models by affecting patient behavior [4]. The quality of explanations and patient understanding have been shown to directly affect health outcomes [5]. As one of many examples, a 2-year prospective study with 4341 patients found that hurried communication and fewer explanations were highly correlated with poor insulin adherence and diabetes-related complications [6].

OpenNotes emerged as a potential strategy to help educate patients about their health beyond the office visit. OpenNotes was launched as a research initiative in 2010 exploring the sharing of medical notes with patients that included 105 volunteer primary care physicians and their patients in Boston, rural Pennsylvania, and inner-city Seattle. Of 13,564 patients with visit notes available, 11,797 opened at least one note, and of those, 5391 completed a postintervention survey. This initial study found that 99% of patients who completed the surveys wanted the practice of sharing medical notes with patients to continue, with 85% stating that they would factor this into their decision in choosing future health care. Patients cited an improved understanding of their health and more trust in their clinicians [7]. A follow-up to this study was published in 2019 reporting the results of an online survey of 28,782 patient respondents. Among the 22,947 who reported reading one or more notes, 73% rated the notes as being very important in taking care of their health [8].

Many additional studies with OpenNotes have since been published and further reported on the benefits patients and clinicians have experienced when medical notes have been shared with patients. In one such study, 6913 patients completed a survey, and shared decision making was quantitatively measured. Researchers found a correlation between note reading and patient perception that they were participating in shared decision making [9]. In another study, adult patients and their care partners were surveyed before and after 12 months of exposure to their doctors' notes electronically. At follow-up, patients stated they felt more confident in managing their health, felt better prepared for office visits, and had an improved understanding from baseline [10].

Despite the success of OpenNotes in research studies, a significant barrier to the widespread adoption of this approach has been clinician concerns that most patients may be confused by the extensive medical terms, abbreviations, and shorthand in medical-language notes [11]. Clinicians worry that the potential benefits of sharing medical notes may be offset by the need to spend more time clarifying these notes with patients or changing note-writing practices [12,13].

A strategy to bring OpenNotes to patients that does not significantly cost the clinician additional time is needed to encourage its adoption and may allow more patients to experience the benefits of having access to their medical notes.

In this study, we explored a method that may cost minimal time to clinicians through the application of recent advances in artificial intelligence (AI) technology. This approach harnesses machine learning and natural language processing (NLP) to simplify clinician written notes into plain-language notes. NLP is a branch of computer science that allows computer programs to interpret and manipulate human language and was shown to simplify medical texts as early as 2010 [14]. In contrast to methods such as the manual simplification of notes, which would require clinicians to spend time writing additional notes for patients, AI software can directly simplify medical language from existing systems. This may allow clinicians to preserve their note-writing practices, while creating a line-by-line plain-language version of patient notes. Such a note can be quickly reviewed by clinicians for correctness and may be more accessible for patients.

The application of NLP to simplify patient education materials has been an emerging field of research in recent years. Prototype models and applications of this technology have been created and conducted by various groups. A 2013 group in Romania created a model that increased the accessibility of medical language in a tele-care program by simplifying texts assigned by medical personnel into plain language [15]. A 2018 multisystem study assessed physician perception of the usability and quality of a web-based NLP system that linked medical notes in electronic health records to lay definitions. Physicians found that the system was easy to use with adequate lay definitions and recommended further development [16].

An important next step in establishing the utility of this approach has been to assess patient perceptions of plain-language notes. It has been important to identify whether patients see benefits of plain-language notes over medical-language (unsimplified) notes at all, and if so, how to best implement this practice. Therefore, our pilot study was designed to assess patient perceptions of AI simplified notes. We looked at whether patients find such notes to be useful and the benefits, concerns, or suggestions they may identify.

Methods

Materials Development

First, a synthetic patient note was generated with Synthea, an open-source, validated, software that generates synthetic patient records for research purposes [17]. The decision to use a synthetic note was made as synthetic notes offer the same utility of real patient notes without the associated privacy concerns. The note produced with Synthea simulated a real patient's hospital discharge summary. Sections generated in this note included the admission diagnosis, other diagnoses, past medical history, hospital course, discharge disposition, plan, and discharge medications. This note served as the medical-language note for all parts of the study.

Proprietary AI software was then developed by a private company, AIPiphany, was used to simplify this medical-language note into a plain-language version [18]. The software replaced complex medical language with plain language equivalents and corrected for grammar and syntax. As an example, the statement "A chest computed tomography (CT) was done on 6/23/04 which showed no aortic dissection" in the original note was simplified by the program to "A chest CT was done on 6/23/04. This showed no tear in the inner layer of the large blood vessel of the heart." This served as the plain-language note for all parts of the study. The parts of the note that were simplified included the history of present illness, past medical history, and hospital course. The decision to not simplify other parts of the note listed in the discharge summary was made as these portions were already generated with a patient as the intended reader, and further alteration of this text was deemed unnecessary and potentially confusing.

Notes were simplified by the AI software to a target Flesch-Kincaid measure of 5th grade level reading language. Although this was set as a target for the program, the actual Flesch-Kincaid measure of the note, not including the medication list, was 8th grade level reading language. The Flesch-Kincaid scale is a gold standard measure of readability that uses average sentence length (ASL) and average number of syllables per word (ASW) to approximate a grade level score using the following formula: (.39 × ASL) + (11.8 × ASW) – 15.59 [19]. Word version 16.0 (Microsoft Corp) was used to determine the Flesch-Kincaid measure of the notes. A 5th grade level reading language was selected as a target for the software as the average Medicare beneficiary reads at or below a 5th grade reading level [20]. This level may have made the simplified note accessible to as many participants as possible. However, while this was set as a target, this was not set as a hard constraint for simplifications. Under these conditions, the software attempted to simplify the note as close to a 5th grade level reading score as possible without eroding the meaning of the note. This produced a note of around an 8th grade level reading score as measured by the Flesch-Kincaid scale. This is likely both a limitation of the Flesch-Kincaid scale and the software used. While simplification of medical notes may decrease the average number of syllables per word, it requires using more words overall to convey the same meaning. This increases the average sentence length, reflected as an increase of the Flesch-Kincaid score. Additionally, simplifying notes further while still preserving the meaning and grammatical correctness of the medical note may require more robust changes to syntax and sentence structure than was possible with the software at the time of study. As the average US resident reads at or below an 8th grade reading level, this level of simplification was found to still be useful and relevant for our study purposes [20]. The simplified plain-language note was reviewed by an attending physician to check that the medical facts represented in the original note were not misrepresented in the simplified version. Other than editing font and formatting for print, no changes were made to the content of the simplified note.

It is important to note that participating patients were not tested with their own medical notes but rather those of the standardized synthetic patient generated with Synthea. This decision was made to prevent differences in complexities of patients' health care from confounding the readability of the notes. By using the same pair of notes for all participants, standardization of the test material was achieved.

Two attending hospitalists reviewed the medical-language note generated with Synthea and shared a list of points they felt would be the most important for patients to understand. This was used to draft 7 reading comprehension questions through discussion and review by the hospitalists. Questions were written in multiple choice format with 4 options per question. Each question had one correct answer choice, two incorrect choices, and one choice stating, "I don't know." Patients were instructed to select "I don't know" rather than guessing when unsure in order to reduce randomly correct answers. As an example, a question asks, "What is a concern if this patient suddenly stops taking coumadin?" Choices for this question were "The patient could have a blood clot that blocks a vessel," "The patient could have a severe bleed," "The patient would have problems with high blood pressure," and "I don't know." The assessment questionnaire was also targeted to be at or below a 5th grade level by Flesch-Kincaid measure. The final questionnaire was found to have a Flesh-Kincaid score of 4.2, which meets our criteria. For reference, the comprehension assessment questions that were used are listed in Multimedia Appendix 1.

A demographic survey was also prepared, with selections for participants to record their gender, age, ethnicity, highest level of education, and health literacy as measured by a 2-item literacy screener (TILS). The TILS score is a self-reported score of health literacy from 1 to 5, with 5 reflecting the greatest level of comfort with reading medical documents and 1 being the lowest [21]. This method of assessing health literacy is quick to complete, minimizes participant embarrassment, and predicts health literacy on par with other commonly used but lengthier

measurements [22]. These metrics were measured to assess for possible confounding effect on the comprehension score. All patient materials including the notes, comprehension questionnaire, and demographic survey were formatted similarly and printed on letter size paper in 12-point Times New Roman font.

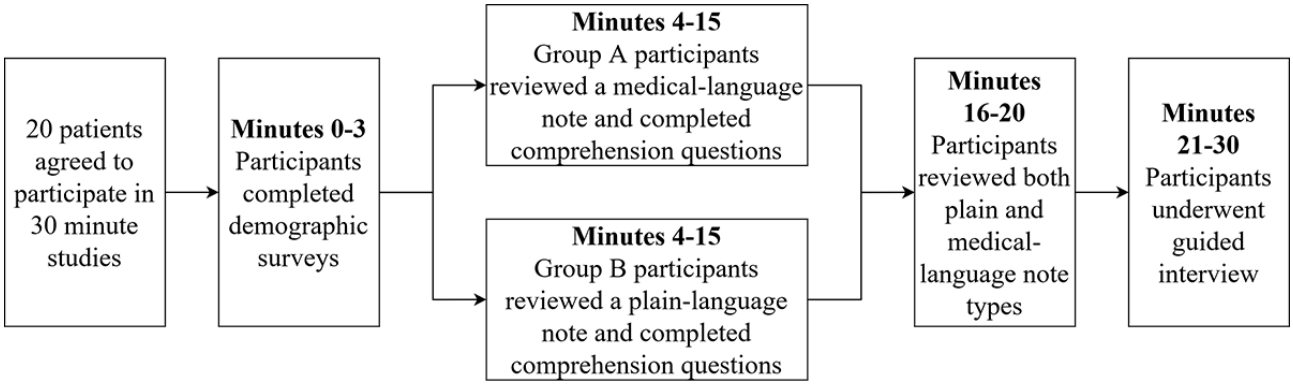
Last, a set of questions and prompts was prepared for use by the researcher during a guided interview. Questions were designed to understand patients' past exposure to OpenNotes, determine usability and usefulness of simplified notes, and identify suggestions patients may have to improve such notes. Three broad open-ended questions were written to be asked of all patients: "Have you ever read the notes that your doctor took about your visit?" "What did you think of these [simplified] notes?" "What else would you like to see in these [simplified] notes?" The complete list of guided interview questions with prompts can be found in Multimedia Appendix 2.

Recruitment

Potential study participants were identified from patients hospitalized at the University of Colorado Hospital in July 2018. This study was approved by the Colorado institutional review board. Inclusion criteria were age 18 to 89 years, admitted to an internal medicine service, able to provide informed consent, and had been hospitalized for at least 48 hours. Exclusion criteria included incarcerated patients, pregnant patients, and patients who were not fluent in English to avoid language as a confounding factor in the study.

Patients were consented, and studies were conducted in patients' private hospital rooms. The purpose, procedures, risks, and benefits of the study were explained to participants, who were given sufficient time to review consent forms and ask any questions. To be respectful of participants' time, the duration of the interview was limited to 30 minutes, enough time for the quantitative and qualitative sections of the study. A breakdown of times is seen in Figure 1. Three minutes were given at the start to allow participants to complete the demographic survey. Participants self-reported their gender, age group, ethnicity, highest level of education, and health literacy as measured by the TILS. These metrics were measured to assess if they had ancillary effect on the comprehension score and therefore could be confounding variables.

Figure 1. Participant assessment and interview flowchart.



Participants were randomly assigned to one of two equal sized groups via computer generated sequence of note types. Group A participants read the medical-language note first and were tested on that note. Group B read the AI-generated plain-language version first and were tested on that simplified version. Participants were given 12 minutes to read their respective note and answer 7 reading comprehension questions.

All participants were then given both versions of the note and allowed 5 minutes to view the differences between the medical-language note and the simplified plain-language note. This was followed by a 10-minute guided interview to elicit their perception of the simplified note. A list of guided interview questions with corresponding prompts was used. When necessary, additional questions were asked to clarify patient statements. Participant responses to interview questions were transcribed electronically in text format as verbatim quotes or summarized statements. Statements by participants were only summarized if they were difficult to understand if transcribed exactly as spoken.

Statistical Analysis

All quantitative analyses were performed using SPSS Statistics version 24 (IBM Corp). A Student *t* test was performed to assess if there was a significant difference in the mean comprehension assessment score between the group evaluated with the original medical-language note and the group evaluated with the simplified plain-language note. A multiple linear regression was performed to determine if the collected demographic variables may have been associated with the number of comprehension questions answered correctly.

Thematic Analysis

Responses to guided interviews were coded as perceived benefits of simplified notes, perceived concerns about simplified notes, patient suggestions regarding simplified notes, and other insight. These responses were analyzed for recurrent themes.

Results

Characteristics of Participants

Of 34 patients approached about participation in the study, 20 agreed to participate. Those who agreed to participate completed all parts of the study, including the comprehension assessment and guided interviews. Reasons shared by participants for declining to participate included tiredness, disinterest, and a

feeling of insufficient English language proficiency. Participant demographics are shown in Table 1.

Statistical Analysis

The mean comprehension assessment score was found to be higher in the plain-language note group at 4.7 questions answered correctly compared with 3.9 questions answered correctly in the medical-language note group with a maximum possible score of 7. However, statistical analysis of this difference was limited by insufficient power of the Student *t* test (16%, α=.05) to detect if this difference was significant or not. Only one participant from each group answered all questions correctly. A multiple linear regression was performed to determine if the collected demographic variables of age group, gender, ethnicity, highest level of education, and health literacy may have been associated with the number of comprehension questions answered correctly. In multivariate modeling, health literacy measured as TILS score (*P*=.003), age (*P*=.03), and ethnicity (*P*=.03) were significantly associated with the comprehension assessment score (Multimedia Appendix 3).

Table 1. Study demographics (n=10).

Characteristics	Tested with unsimplified notes	Tested with simplified notes
Gender, n (%)		
Male	7 (70)	2 (20)
Female	3 (30)	8 (80)
Age in years, n (%)		
18-34	7 (70)	7 (70)
35-55	1 (10)	2 (20)
55-64	2 (20)	0 (0)
65+	0 (0)	0 (0)
Ethnicity, n (%)		
White	7 (70)	7 (70)
Black	1 (40)	2 (20)
Hispanic	2 (20)	0 (0)
Asian	0 (0)	0 (0)
Other	0 (0)	1 (10)
Highest education, n (%)		
Some high school	0 (0)	0 (0)
High school diploma	2 (20)	1 (10)
Some college (some credit with no degree or trade, vocational, or technical training)	5 (50)	6 (60)
College graduate (associate degree, bachelor's degree, or equivalent)	2 (20)	3 (30)
Postgraduate (master's degree, doctoral degree, or equivalent)	1 (10)	0 (0)
Health literacy, mean (SD)		
TILS[a] score	3.85 (1.05)	4.2 (0.82)

[a]Two-item literacy screener.

Thematic Analysis

Three major themes of perceived potential benefits of simplified notes were found in guided interview responses: (1) enhanced usability of simplified notes compared with unsimplified notes, (2) improved patient-clinician relationship, and (3) empowerment of patients through an enhanced understanding of their conditions and management.

Regarding usability, participants shared that having access to a simplified note after a medical encounter would help them to better understand the information discussed during their visit.

> *...sometimes doctors use language that they don't realize is above our level, so having something simple to take home would help.*

Participants suggested that such notes may also help patients retain information discussed during the visit.

> *Things can get lost during a medical visit because you might be anxious or they [the clinician] may be talking too fast. Having a note like this to reference later is very useful.*

Participants reported that while they often only skim through papers with medical terminology, they would take the time to read through a note that is written in plain language.

> *I only skim through or glance my own medical notes, but I would actually read through a simplified note.*
>
> *The simplified version was much easier to understand. It was less confusing and helped me understand the procedure.*

Regarding improved patient-clinician relationship, participants expressed that being given plain language notes would help them to have more trust and confidence in their clinician.

> *This [simplified] note would help because it would show me that the physician cared that I understood what was going on.*

Patients stated that they would also be more likely to follow through with treatment recommendations due to this improved relationship.

> *...if I understood why my doctor wanted me to come to make an appointment, I would go.*

Regarding patient empowerment, participants suggested that such notes would help them meet needs related to daily living by allowing them to more effectively communicate about their limitations with family members or employers. A participant shared that while he currently depends on family members to understand his medical management, having a plain-language note may reduce the need to ask others for assistance and therefore make him feel like less of a burden. Participants also shared that such notes would reduce the number of online searches they may need to conduct to understand their conditions. Another participant reported that such notes can also help empower patients to talk with their families about their problems.

> *It's hard to talk to family when it's in the doctor's language but would be a lot easier with this.*

Participants shared that employers may be more likely to understand the circumstances of patients and be responsive to their needs if patients are better able to explain their medical conditions.

> *This may help in getting paid time off work or breaks during shifts for medical reasons.*

Another patient stated that these notes may allow him to overcome his hearing disability. He stated that he finds any form of text-based communication to be preferable to calling or talking in person due to his difficulty hearing spoken language. He stated that he would find it easier to clarify his questions by reading a text at his own pace rather than asking the clinician questions.

Other insight gained from interviews highlighted patients' previous experiences with OpenNotes or reflected opportunities to improve plain-language notes. A patient shared that she has had previous exposure to OpenNotes as her primary care physician posts a plain-language summary alongside her postencounter documents on an online portal to which she has access, and she states that she found this useful to remember information from and better understand her medical visits.

Asked if they would prefer the original medical language, the simplified plain language, or both notes provided to them during visits, many participants reported that they wanted to have both copies. A participant shared that this may allow him to compare the two notes to see if he had missed anything by only reading the plain-language note. Another patient stated that she would also like to have both the medical-language and plain-language notes so she could use the simplified plain-language note for her own understanding but have the medical-language note to share with specialists or new clinicians.

Concerns reported by participants included possible oversimplification of medical language. Participants reported concerns that if notes are oversimplified, the physician's intent may be lost. Patients also suggested that the level of simplification of notes would be best if adjusted to their level of experience with medical language. Participants additionally noted the opportunity to enhance the readability of simplified notes by reducing the length or repetition of some plain-language phrases.

Discussion

Principal Findings

The key findings of this pilot study are that (1) the main benefits noted by participants were an improved relationship with the clinician, increased usability of OpenNotes, and empowerment in their daily life; (2) patients desired access to their medical records and felt that simplified open medical notes would help them to better manage their health; and (3) while there was insufficient power to detect a significant difference in mean comprehension assessment scores, AI-simplified medical notes had been well received by patients.

Limitations

As this was a pilot study with a small sample size (n=20), statistical analysis was limited by insufficient power of the Student *t* test (16%, α=.05) to detect a significant difference in comprehension scores between the medical-language and plain-language note groups.

Additionally, multivariate modeling had found that the TILS score, race, and ethnicity had a significant impact on the mean comprehension scores. These factors may have been possible confounders and should be controlled in future studies if assessing comprehension.

Given the small sample size, it is also possible that complete thematic saturation for guided interviews may not have been achieved. Additionally, although most Americans read at an 8th grade reading level, most patients enrolled in Medicare read at a 5th grade level or below. This may explain why only 2 out of 20 patients answered all questions correctly; it is possible that even the simplified notes were beyond participant reading levels.

Participants were also not provided their own medical notes but rather those of a standardized patient. This may have impaired participants' ability to comprehend the notes due to a baseline lack of knowledge regarding the medical problems presented in the standardized note. As the plan and medications list from the patient discharge note were left unsimplified in this study, simplifying these portions of the note in future studies may be of further benefit for patients. Additionally, only one pair of

notes was tested in this study. As future studies test additional notes, it is possible they may find note-specific variances that affect patient comprehension and perception.

Only an English-language note was tested, so our findings may not be generalizable to non-English applications. Last, all notes were simplified to the same reading level in this study. Individualizing the level of simplification of the notes to participant level of health literacy may enhanced reader comprehension. Another possible source of confounding that has not been directly measured is patients' prior exposure to OpenNotes, as having had practice with reading notes may better equip them to read medical notes.

Comparison With Previous Work

Institutions around the country have begun large-scale implementation of OpenNotes with encouraging responses from patients and clinicians. Preliminary studies have shown that computer technology may be used to simplify medical language to assist in patient understanding of medical terminology. However, this is the first known study to assess patients' perception of medical notes that have been simplified with AI software. Patients' responses in this study may be useful for the development of this work and offer a glimpse into the future of how patients may be able to improve their understanding outside of the medical visit.

Future Studies

This is an interesting and emerging field of research with many opportunities for future study. As this initial pilot study found that patients perceived AI-generated simplified medical notes as desirable and useful, future studies should be conducted with larger sample sizes and take advantage of patients' insights and suggestions as mentioned here. In future studies, patients' own notes should be tested, and patients of different levels of care complexity should be included, as it is possible that patients with more complex management may derive a greater benefit from plain-language explanations. Studies should be repeated in the outpatient setting with longer exposure and follow-up times. The value of simplifying additional parts of the note such as the plan and medications list should be explored. Physicians should be surveyed on their perceptions of plain-language notes to assess if such notes would make them more likely to use OpenNotes in their own practices.

Suggestions from this study can be used to improve the NLP software used to simplify notes for future studies. As AI technology and use in this area evolves, methods should be developed to match the patient's health literacy, cultural and demographic background, and level of health care experience with the simplification level of the note. Methods should be developed to reduce lengthy plain-language phrases. At the same time, care should be taken to ensure that the simplification process does not lead to omission of information that physicians perceive as valuable for patient understanding and utility. Long-term studies should evaluate the impact of plain-language OpenNotes on clinical outcomes in various settings.

Conclusions

In conclusion, this study suggests that AI software may be used to generate plain-language medical notes that patients desire and find useful. Such notes may empower patients to better communicate and make decisions, increase adoption of OpenNotes, enhance the patient-clinician relationship, and improve clinical outcomes. The findings in this study can be used to optimize delivery and generation of simplified notes. Researchers in this field may particularly find the patient responses in the guided interviews in this study to be interesting and useful for the further development and application of NLP software in this field.

Acknowledgments

We would like to thank Dr Michelle Archuleta, CEO of AIPiphany, for her insight regarding the AI software used to simplify the notes in this study.

Conflicts of Interest

MB owns stock in and serves as an advisor to AIPiphany.

Multimedia Appendix 1
Comprehension assessment questionnaire.
[DOCX File , 14 KB - formative_v4i6e16670_app1.docx]

Multimedia Appendix 2
Guided interview questions.
[DOCX File , 14 KB - formative_v4i6e16670_app2.docx]

Multimedia Appendix 3
Statistical analyses.
[DOCX File , 13 KB - formative_v4i6e16670_app3.docx]

References

1. Sinsky C, Colligan L, Li L, Prgomet M, Reynolds S, Goeders L, et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. Ann Intern Med 2016 Sep 6;165(11):753-760. [doi: 10.7326/M16-0961] [Medline: 27595430]
2. Young RA, Burge SK, Kumar KA, Wilson JM, Ortiz DF. A time-motion study of primary care physicians' work in the electronic health record era. Fam Med 2018 Feb;50(2):91-99 [FREE Full text] [doi: 10.22454/FamMed.2018.184803] [Medline: 29432623]
3. Rapport F, Hibbert P, Baysari M, Long JC, Seah R, Zheng WY, et al. What do patients really want? An in-depth examination of patient experience in four Australian hospitals. BMC Health Serv Res 2019 Jan 15;19(1):38 [FREE Full text] [doi: 10.1186/s12913-019-3881-z] [Medline: 30646962]
4. Gold DT, McClung B. Approaches to patient education: emphasizing the long-term value of compliance and persistence. Am J Med 2006 Apr;119(4 Suppl 1):S32-S37. [doi: 10.1016/j.amjmed.2005.12.021] [Medline: 16563940]
5. Marcus C. Strategies for improving the quality of verbal patient and family education: a review of the literature and creation of the EDUCATE model. Health Psychol Behav Med 2014 Jan 01;2(1):482-495 [FREE Full text] [doi: 10.1080/21642850.2014.900450] [Medline: 25750796]
6. Linetzky B, Jiang D, Funnell MM, Curtis BH, Polonsky WH. Exploring the role of the patient-physician relationship on insulin adherence and clinical outcomes in type 2 diabetes: insights from the MOSAIc study. J Diabetes 2017 Jun;9(6):596-605 [FREE Full text] [doi: 10.1111/1753-0407.12443] [Medline: 27368146]
7. Delbanco T, Walker J, Bell SK, Darer JD, Elmore JG, Farag N, et al. Inviting patients to read their doctors' notes: a quasi-experimental study and a look ahead. Ann Intern Med 2012 Oct 2;157(7):461-470 [FREE Full text] [doi: 10.7326/0003-4819-157-7-201210020-00002] [Medline: 23027317]
8. Walker J, Leveille S, Bell S, Chimowitz H, Dong Z, Elmore JG, et al. OpenNotes after 7 years: patient experiences with ongoing access to their clinicians' outpatient visit notes. J Med Internet Res 2019 May 06;21(5):e13876 [FREE Full text] [doi: 10.2196/13876] [Medline: 31066717]
9. Fossa AJ, Bell SK, DesRoches C. OpenNotes and shared decision making: a growing practice in clinical transparency and how it can support patient-centered care. J Am Med Inform Assoc 2018 Sep 01;25(9):1153-1159. [doi: 10.1093/jamia/ocy083] [Medline: 29982659]
10. Wolff JL, Darer JD, Berger A, Clarke D, Green JA, Stametz RA, et al. Inviting patients and care partners to read doctors' notes: OpenNotes and shared access to electronic medical records. J Am Med Inform Assoc 2016 Aug 07;24(e1):e166-e172. [doi: 10.1093/jamia/ocw108] [Medline: 27497795]
11. Bell SK, Mejilla R, Anselmo M, Darer JD, Elmore JG, Leveille S, et al. When doctors share visit notes with patients: a study of patient and doctor perceptions of documentation errors, safety opportunities and the patient-doctor relationship. BMJ Qual Saf 2016 May 18;26(4):262-270. [doi: 10.1136/bmjqs-2015-004697] [Medline: 27193032]
12. Delbanco T, Walker J, Darer JD, Elmore JG, Feldman HJ, Leveille SG, et al. Open notes: doctors and patients signing on. Ann Intern Med 2010 Jul 20;153(2):121-125. [doi: 10.7326/0003-4819-153-2-201007200-00008] [Medline: 20643992]
13. Alpert JM, Morris BB, Thomson MD, Matin K, Geyer CE, Brown RF. OpenNotes in oncology: oncologists' perceptions and a baseline of the content and style of their clinician notes. Transl Behav Med 2019 Mar 01;9(2):347-356 [FREE Full text] [doi: 10.1093/tbm/iby029] [Medline: 29596633]
14. Kandula S, Curtis D, Zeng-Treitler Q. A semantic and syntactic text simplification tool for health content. AMIA Annu Symp Proc 2010:366-370 [FREE Full text] [Medline: 21347002]
15. Topac V, Stoicu-Tivadar V. Patient empowerment by increasing the understanding of medical language for lay users. Methods Inf Med 2013;52(5):454-462. [doi: 10.3414/ME12-02-0006] [Medline: 24072079]
16. Chen J, Druhl E, Polepalli Ramesh B, Houston TK, Brandt CA, Zulman DM, et al. A natural language processing system that links medical terms in electronic health record notes to lay definitions: system development using physician reviews. J Med Internet Res 2018 Jan 22;20(1):e26 [FREE Full text] [doi: 10.2196/jmir.8669] [Medline: 29358159]
17. Walonoski J, Kramer M, Nichols J, Quina A, Moesel C, Hall D, et al. Synthea: an approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. J Am Med Inform Assoc 2017 Aug 30. [doi: 10.1093/jamia/ocx079] [Medline: 29025144]
18. AIpiphany. AIpiphany Main Page URL: https://www.aipiphany.tech/ [accessed 2020-05-09]
19. Dubay W. The principles of readability. 2004 Aug 25. URL: http://www.impact-information.com/impactinfo/readability02.pdf [accessed 2020-05-09]
20. Stossel LM, Segar N, Gliatto P, Fallar R, Karani R. Readability of patient education materials available at the point of care. J Gen Intern Med 2012 Sep;27(9):1165-1170 [FREE Full text] [doi: 10.1007/s11606-012-2046-0] [Medline: 22528620]
21. Brice JH, Foster MB, Principe S, Moss C, Shofer FS, Falk RJ, et al. Single-item or two-item literacy screener to predict the S-TOFHLA among adult hemodialysis patients. Patient Educ Couns 2014 Jan;94(1):71-75. [doi: 10.1016/j.pec.2013.09.020] [Medline: 24169024]
22. Morris NS, MacLean CD, Chew LD, Littenberg B. The Single Item Literacy Screener: evaluation of a brief instrument to identify limited reading ability. BMC Fam Pract 2006 Mar 24;7:21 [FREE Full text] [doi: 10.1186/1471-2296-7-21] [Medline: 16563164]

Abbreviations

AI: artificial intelligence
ASL: average sentence length
ASW: average number of syllables per word
CT: computed tomography
NLP: natural language processing
TILS: two-item literacy screener

Edited by G Eysenbach; submitted 13.10.19; peer-reviewed by S Kim, S Wolver, R Zowalla; comments to author 11.11.19; revised version received 05.01.20; accepted 29.03.20; published 05.06.20.

Please cite as:
Bala S, Keniston A, Burden M
Patient Perception of Plain-Language Medical Notes Generated Using Artificial Intelligence Software: Pilot Mixed-Methods Study
JMIR Form Res 2020;4(6):e16670
URL: https://formative.jmir.org/2020/6/e16670
doi:10.2196/16670
PMID:32442148

Original Paper

Association Between Electroencephalogram-Derived Sleep Measures and the Change of Emotional Status Analyzed Using Voice Patterns: Observational Pilot Study

Hirotaka Miyashita[1,2], MD; Mitsuteru Nakamura[3], PhD; Akiko Kishi Svensson[1,4,5], MD, PhD; Masahiro Nakamura[1], MD, PhD; Shinichi Tokuno[3,6], MD, PhD; Ung-Il Chung[1,6,7], MD, PhD; Thomas Svensson[1,4,6], MD, PhD

[1]Precision Health, Department of Bioengineering, Graduate School of Engineering, University of Tokyo, Tokyo, Japan

[2]Internal Medicine, Mount Sinai Beth Israel, New York, NY, United States

[3]Voice Analysis of Pathophysiology, Graduate School of Medicine, University of Tokyo, Tokyo, Japan

[4]Department of Clinical Sciences, Lund University, Skåne University Hospital, Malmö, Sweden

[5]Department of Diabetes and Metabolic Diseases, University of Tokyo, Tokyo, Japan

[6]School of Health Innovation, Kanagawa University of Human Services, Kawasaki-ku, Kawasaki-shi, Kanagawa, Japan

[7]Clinical Biotechnology, Center for Disease Biology and Integrative Medicine, Graduate School of Medicine, University of Tokyo, Tokyo, Japan

Corresponding Author:
Thomas Svensson, MD, PhD
Precision Health, Department of Bioengineering
Graduate School of Engineering
University of Tokyo
7-3-1 Hongo, Bunkyo-ku,
Tokyo, 113-8656
Japan
Phone: 81 3 5841 4737
Email: t-svensson@umin.ac.jp

Abstract

Background: Measuring emotional status objectively is challenging, but voice pattern analysis has been reported to be useful in the study of emotion.

Objective: The purpose of this pilot study was to investigate the association between specific sleep measures and the change of emotional status based on voice patterns measured before and after nighttime sleep.

Methods: A total of 20 volunteers were recruited. Their objective sleep measures were obtained using a portable single-channel electroencephalogram system, and their emotional status was assessed using MIMOSYS, a smartphone app analyzing voice patterns. The study analyzed 73 sleep episodes from 18 participants for the association between the change of emotional status following nighttime sleep (Δvitality) and specific sleep measures.

Results: A significant association was identified between total sleep time and Δvitality (regression coefficient: 0.036, *P*=.008). A significant inverse association was also found between sleep onset latency and Δvitality (regression coefficient: –0.026, *P*=.001). There was no significant association between Δvitality and sleep efficiency or number of awakenings.

Conclusions: Total sleep time and sleep onset latency are significantly associated with Δvitality, which indicates a change of emotional status following nighttime sleep. This is the first study to report the association between the emotional status assessed using voice pattern and specific sleep measures.

(JMIR Form Res 2020;4(6):e16880) doi:10.2196/16880

KEYWORDS

voice analysis; emotional status; vitality; sleep; mobile phone

Introduction

The evaluation of emotion is a challenge in many studies. Questionnaires have been the gold standard to evaluate emotion [1], but they cannot avoid reporting bias [2]. To solve this problem, a novel approach to analyze emotional status has been developed: Voice analysis [3], which analyzes the acoustic features of voice to estimate the involuntary effect of autonomic nerves on vocal cord movement and calculates vitality as a measure of mental status [4]. Primarily, high vitality could be translated into positive moods, as Hagiwara et al [4] revealed an inverse association between vitality index and the Beck Depression Inventory. Vitality can be used to evaluate the therapeutic effect of continuous positive airway pressure as a treatment of sleep apnea syndrome [5].

Sleep is one of the essential activities for humans. Extensive studies have been conducted to elucidate the relationship between sleep and affect [6]. A systematic review by Konjarski et al [7] revealed the reciprocal relationship between positive mood and better quality, shorter latency, and regular duration of sleep. There is also some evidence on the association between sleep loss and emotion in a general population [8], albeit with some discrepancies between research articles.

Previous studies have revealed that sleep deprivation is associated with altered autonomic nervous activities [9,10]. The vocal cord is partially innervated with autonomic nerves—the superior laryngeal nerve and the recurrent laryngeal nerve. Therefore, it can be hypothesized that sleep insufficiency may negatively affect voice quality, and further that sleep may influence vitality. Good sleep may be associated with the change in vitality from evening to morning (Δvitality), given, based on previous studies, that good sleep improves mood positively, which in turn might bring about the increment of vitality overnight. Therefore, it was hypothesized that an appropriate duration of sleep, 7-9 hours for adults [11], is associated with a greater recovery of vitality, that is, a higher Δvitality.

Methods

Participants

The participants were 20 healthy Japanese adults (10 men and 10 women; age 25-67 years) recruited from Kanagawa and Tokyo. For the purpose of this study, only participants with repeated measures of the outcome variable (n=18, described in detail later) were included.

Participation in the study was entirely voluntary, and all participants were informed about the purpose of the study and that participation could be discontinued at any point without any disadvantage or imposed penalty for the participants. All research was performed in accordance with relevant guidelines/regulations. All participants provided written informed consent, and the study was approved by the Ethical Committee of the Department of Bioengineering, Graduate School of Engineering, the University of Tokyo (Approval number: KE18-7).

Study Design

This is an observational pilot study to investigate the association between electroencephalogram (EEG)-derived sleep measures and the change of emotional status analyzed from voice. The aim of the overall study from which the data were obtained was to validate the sleep-tracking functions of a consumer wearable device against a validated portable single-channel EEG system in naturalistic conditions. Details of the validation study can be found in the study by Svensson et al [12]. No restrictions for sleep or activity were imposed on the participants. Sleep measures in this study were obtained using the validated portable single-channel EEG system (SleepScope, SleepWell Co, Ltd.; the device is described in greater detail later). Participants were asked to use the device every night throughout the study period. Participants were also provided with an internet-connected Android smartphone, which had preinstalled an app (MIMOSYS; the app is described in greater detail later) that utilizes voice analysis for assessing overall mental wellness. During the study period, participants were asked to use MIMOSYS twice per day to obtain morning and evening vitality measures.

Sleep data were available for a total of 7 nights, equivalent to a total of 138 sleep episodes (person-nights). Details of sleep data retrieval and retention are described by Svensson et al [12].

Exposures

The main exposure of this study was total sleep time (minutes) converted into sleep hours (total sleep time/60). Total sleep time was obtained using EEG. Details of the single-channel EEG system can be found in the validation study by Yoshida et al [13]. In brief, the portable single-channel EEG system (SleepScope) is manufactured by SleepWell Co, Ltd. and is a registered medical device (Japanese Medical Device Certification number: 225ADBZX00020000) suitable for home use. The SleepScope has shown acceptable agreement with polysomnography and records waveforms appropriate for sleep staging (wake [W], R, N1, N2, and N3) as defined by version 2.3 of the American Academy of Sleep Medicine Scoring Manual [14].

Secondary endpoints (sleep efficiency, sleep onset latency, and number of awakenings during sleep) were chosen on the basis of a meta-analysis conducted by Baglioni et al [15]. This meta-analysis revealed that a problem in sleep continuity exists in most psychiatric disorders and that total sleep time, sleep efficiency, sleep onset latency, and number of awakenings are components of sleep continuity. Sleep efficiency was defined as the ratio of total sleep time to time in bed × 100 (%). Sleep onset latency was the time from going to bed to sleep onset. Number of awakenings was defined as the number of continuous records of W after sleep onset.

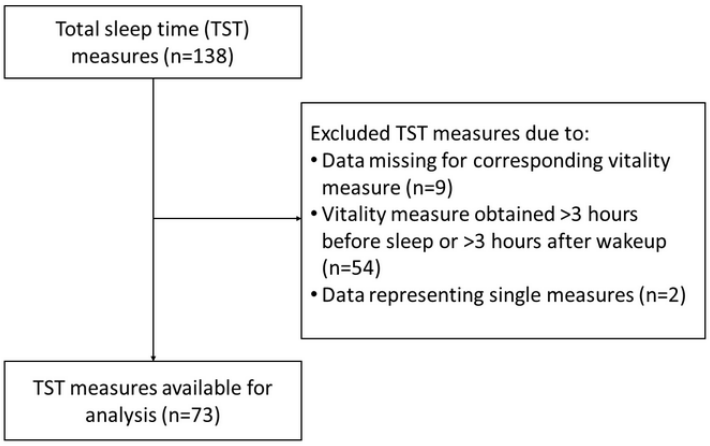
Vitality

The main outcome measure using the MIMOSYS voice analysis software was the change in vitality score from evening to morning (ie, Δvitality calculated as morning vitality – evening vitality). The algorithm of the MIMOSYS was originally developed to enable screening of major depression using only voice analysis. The basic idea of MIMOSYS is to check for

continuously depressed mood, one of the diagnostic criteria for major depression [16]. The software is developed and provided by PST Inc. In the algorithm, vitality was derived from proportions of elemental emotions evaluated from single utterance delimited by breathing using the Sensibility Technology Software Development Kit (PST Inc). The vitality for a single measurement was acquired by averaging vitality acquired from sequential utterances within the same measurement. This study used an app implemented for Android OS with additional functions such as voice recording, monitoring of execution status for each participant, and questionnaires. The app was installed on smartphones (ZenFone 2 Laser, ASUSTeK Computer Inc), which were provided to each participant of the study.

During the study period, there were 138 measures of evening vitality and 137 measures of morning vitality (Figure 1). One evening measure was excluded, as it would not allow for the calculation of Δvitality. In addition, 8 sleep measures were excluded, as they did not have a corresponding morning vitality measure (ie, the participant did not use the MIMOSYS app in the morning). A total of 54 sleep measures were excluded, as there were no corresponding vitality measures in the 3 hours preceding sleep or in the 3 hours following awakening as determined by EEG. Finally, the sole observations of two participants were excluded as the chosen statistical methods focus on repeated measures analysis. Consequently, there were 73 person-nights' data of sleep from 18 participants with corresponding Δvitality measures available for analysis.

Figure 1. Flowchart of participants included in the study.



Statistical Analyses

A total of 18 participants with at least two measures (range, 2-7 measures) of sleep hours with a corresponding value of Δvitality for the sleep period were included in the analysis. Generalized estimating equations (GEEs) were used to estimate the change in vitality score from evening to morning as a function of the exposure. Given the normally distributed outcome measure (Δvitality), the GEE model family was set as *gaussian*, with the link function specified as *identity*, the correlation structure specified as *unstructured*, and the variance-covariance matrix of the estimators specified as *robust*. The analyses on the associations between total sleep time, sleep efficiency, sleep onset latency, and number of awakenings and Δvitality were conducted in separate statistical models. Models investigating the association of Δvitality with total sleep time and sleep onset latency were adjusted for age, sex, and evening vitality score. Models investigating the association between Δvitality and sleep efficiency and number of awakenings were additionally adjusted for total sleep time. All statistical analyses were performed using Stata/MP version 15.1 (StataCorp LLC). *P* values were two-tailed and considered significant if <.05.

Data Availability Statement

We cannot publicly provide individual data due to participant privacy, according to ethical guidelines in Japan. Additionally, the written informed consent we obtained from study participants does not include a provision for publicly sharing data. Qualifying researchers may apply to access a minimal dataset by contacting the Principal Investigator for this study (TS).

TS had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Results

Participants Characteristics

Participants' baseline characteristics are summarized in Table 1. Of the 18 participants, 10 were male and the age of participants ranged from 25 to 67 years. The mean sleep duration was 5.4 (SD 1.23) hours. The means of sleep efficiency, sleep onset latency, and number of awakenings were 87.4%, 12.2 minutes, and 27.1 counts, respectively. The means of evening vitality, morning vitality, and Δvitality were 0.42, 0.37, and –0.05, respectively.

Table 1. Participants' baseline characteristics.[a,b]

Variable	Value
Age (years), mean (SD); range	42 (11.8), 25-67
Male sex, n (%); range	10 (56), N/A[c]
Total sleep time (h), mean (SD); range	5.42 (1.23), 2.35-8.58
Sleep efficiency (%), mean (SD); range	87.36 (8.22), 49.2-95.8
Sleep onset latency (min), mean (SD); range	12.18 (15.39), 0.5-89.5
Number of awakenings, n (%); range	27.1 (12.35), 3-75
Evening vitality, mean (SD); range	0.42 (0.15), 0.004-0.85
Morning vitality, mean (SD); range	0.37 (0.13), 0.08-0.69
ΔVitality, mean (SD); range	–0.05 (0.17), –0.48 to 0.41

[a]N=18, which was the total number for age and male sex.

[b]N=73, which was the total number for the remaining variables.

[c]N/A: not applicable.

Associations Between Exposures and ΔVitality

When adjusted for age, sex, and evening vitality score, total sleep time showed a significant positive association with Δvitality (regression coefficient: 0.036, *P*=.008). A significant inverse association was also found between sleep onset latency and Δvitality (regression coefficient: –0.026, *P*=.001). The exposures sleep efficiency and number of awakenings showed no significant association with Δvitality even after additional adjustment for total sleep time (Table 2). The scatter plots expressing the exposures and Δvitality are shown in Figure 2.

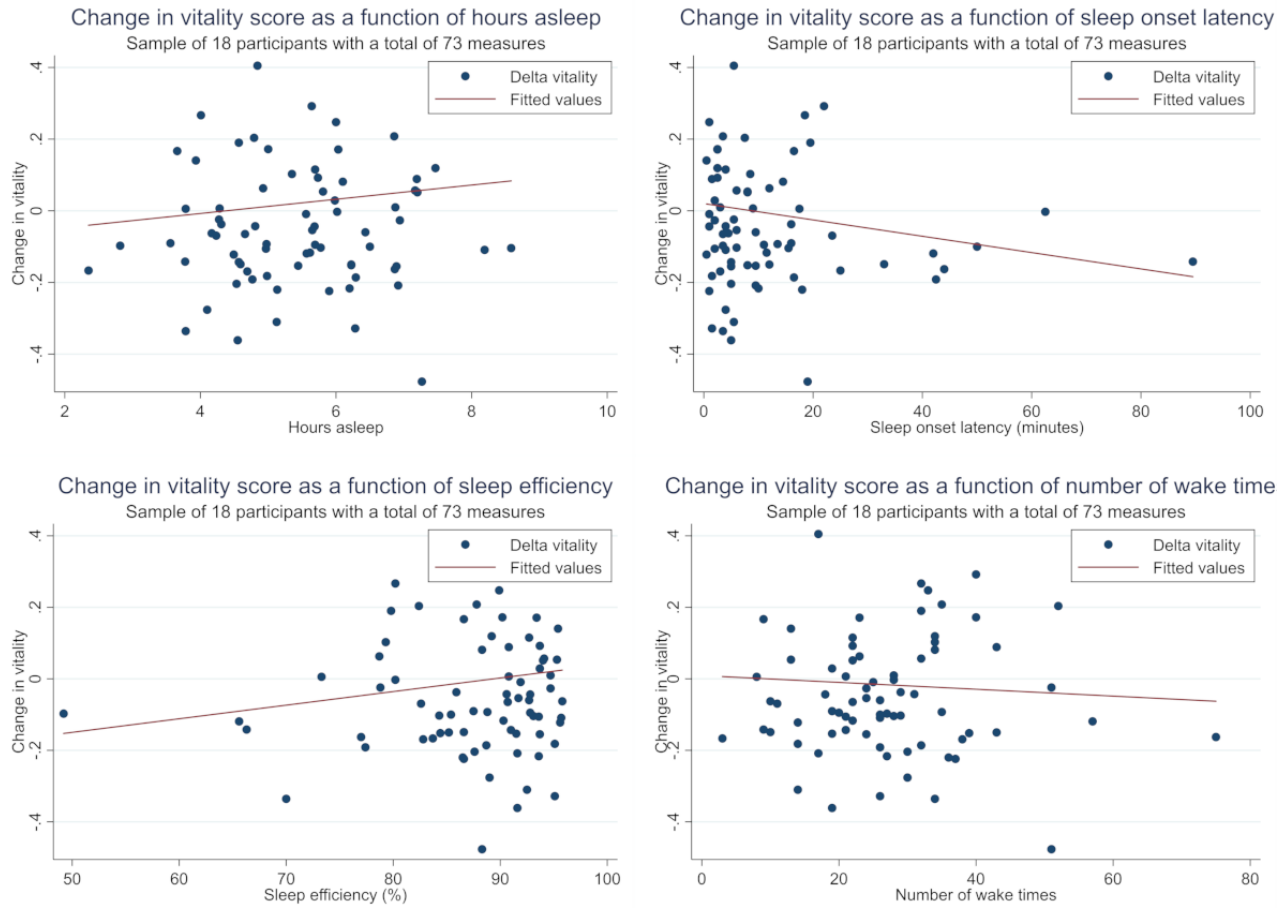
Table 2. Regression coefficients for each exposure obtained from separate statistical models.

Exposure	Regression coefficient (95% CI)	*P* value
Total sleep time (hours)[a]	0.036 (0.0094 to 0.062)	.008
Sleep onset latency (min)[a]	–0.0026 (–0.0042 to –0.0010)	.001
Sleep efficiency (%)[b]	0.0022 (–0.0025 to 0.0069)	.36
Number of awakenings (n)[b]	0.00092 (–0.0015 to 0.0033)	.45

[a]Statistical model adjusted for age, sex, and evening vitality score.

[b]Statistical model adjusted for age, sex, evening vitality score, and total sleep time.

Figure 2. Scatter plots of exposures and Δvitality. The y-axis represents change of vitality before and after sleep. The x-axis represents each exposure (total sleep time, sleep onset latency, sleep efficiency, and number of awakenings). The solid line refers to fitted values. The closed circles refer to Δvitality.



Discussion

Principal Findings

In this analysis of 73 measures of sleep from 18 adults, a significant positive association was found between total sleep time and Δvitality. Sleep onset latency was also found to be inversely associated with Δvitality. By contrast, there was no significant association between sleep efficiency and number of awakenings during sleep and Δvitality.

The magnitude of the regression coefficient (0.036) for the association between total sleep time and Δvitality is, despite its statistical significance, difficult to put into context given that there are no prior studies to compare with. As such, this is considered to be a pilot study with regard to the effect of sleep on the vitality measure obtained through voice analysis. The study hypothesis was that a sleep duration of 7-9 hours, which is considered appropriate for adults [11], is associated with a greater recovery of vitality, that is, a higher Δvitality. Indeed, the regression analyses indicated a statistically significant (P=.008) association between total sleep time and Δvitality, which is compatible with our hypothesis. There are some transient factors such as alcohol intake [17] that can affect voice patterns and, by extension, Δvitality. Although we did not adjust for these factors in study analyses, daily excessive alcohol intake was considered an exclusion criterion for participation in the study. The mechanisms for the found associations lie beyond the scope of this study; however, previous studies investigating the effect of sleep on the autonomic nervous system have found that short sleep duration on EEG is associated with low cardiac vagal regulation [18]. The result of this study also suggests that sleep duration is associated with autonomic nervous activity.

Very long sleep duration is known to be associated with adverse health outcomes [19], and sleeping for longer than 9 hours may be associated with psychiatric disorders [20]. In this study, the participants' sleep durations ranged between 2.4 and 8.6 hours, with no participant recording a sleep duration longer than 9 hours. Consequently, no association between very long sleep duration and reduced Δvitality could be analyzed. Future studies are encouraged to investigate whether the recovery of vitality diminishes with excessive sleep duration.

Poor sleep continuity (low sleep efficiency, long sleep onset latency, and large number of awakenings) has been revealed to be associated with various psychiatric disorders. For example, Baglioni et al [15] showed statistically significant inferiority of sleep continuity in patients with affective disorders, anxiety disorders, and schizophrenia. Thus, it was hypothesized that sleep efficiency is positively, and sleep onset latency and number of awakenings are inversely associated with Δvitality. This study found an inverse association between sleep onset latency and Δvitality. This result suggests that a shorter sleep

onset latency is associated with improved mental activity in the next morning. A previous article suggested that prolonged sleep onset latency is associated with autonomic nervous dysfunction as identified by reduced heart rate fluctuations during the sleep onset period [21]. This would be compatible with results of this study, given that the vocal cord is also partially controlled by the autonomic nervous system.

Sleep efficiency and number of awakenings are regarded as factors of sleep continuity and reported to be associated with some psychiatric disorders [15]. Furthermore, low sleep efficiency is reported to be associated with low levels of cardiac parasympathetic tone [18]. Thus, it was hypothesized that low sleep efficiency and large number of awakenings are associated with low Δvitality. However, an association between sleep efficiency or number of awakenings and Δvitality was not identified. This could be caused by actual null association; however, considering that the prior study was based on more than 500 measurements of sleep continuity [18], this study might have not had enough power to detect the association. In addition, the effect size of sleep efficiency and number of awakenings could be too small to override other unadjusted factors (eg, alcohol intake) contributory to Δvitality. More measurements and adjustments may be needed to address the association of Δvitality with sleep efficiency and number of awakenings.

MIMOSYS was used as a measure of mental status. The evaluation of mental status based on voice analysis is a new concept, whose validity should be carefully assessed. Currently, some evidence exists of voice analysis as a measure to assess mental status. Tokuno et al [22] showed that the voice analysis system had a sensitivity of 0.897 to detect patients with poor emotional hygiene who require medical intervention or counseling. Moreover, voice analyses had a comparable sensitivity when compared with the 30-item General Health Questionnaire (GHQ-30), a traditional questionnaire-based screening instrument to identify psychiatric conditions [23]. It was also reported that vitality determined from voice recordings is negatively correlated (–0.208) with the Beck Depression Inventory [4]. The study hypothesis was based on the assumption that voice patterns reflect mental status, which could be supported by these findings.

Limitations

This study has a few limitations. First, not all data from MIMOSYS were utilized, because analyses in this study a priori excluded any vitality measures taken more than 3 hours before falling asleep or after awakening. However, this was done to capture the vitality measures on which sleep was believed to have the most influence. Any measures occurring several hours before or after a sleep period could be confounded by other lifestyle or environmental factors. Irrespectively, the measurement of vitality requires regular and repeated measurements by the participants, which may render the study susceptible to reporting bias. Second, a single-channel EEG system was used in this study to obtain sleep measures. The current gold standard of sleep monitoring is polysomnography [24]. However, polysomnography requires the participants to sleep in a special examination room, which may affect the results [25]. The single-channel EEG system is a relatively new technology, and its validation may not be fully established for certain sleep parameters [26]. Portable EEGs can be used in research participants' natural sleep environments, which in turn may result in increased effectiveness as compared with polysomnography. Third, none of the participants had a sleep duration longer than 8.6 hours. This could result in an underestimation of the association between long sleep duration and Δvitality. Fourth, there is no available reference value to determine a healthy Δvitality score. Consequently, it is difficult to ascertain the clinical relevance of the regression coefficients obtained in this study. The main aim of this study, however, was not to establish a clinical cutoff, but to investigate whether sleep parameters were associated with a change in vitality. Finally, although Hagiwara et al [4] showed a correlation between the absolute value of vitality and the Beck Depression Inventory, no assessment has been conducted regarding Δvitality and emotion or well-being. Further studies are thus warranted to validate the findings of this pilot study.

Strengths

Despite some limitations, this study has a number of strengths. First, this is the first study to show an association between sleep parameters and vitality as recorded through voice, which has the potential of a brand-new marker of emotional conditions and recovery from sleep. Notably, this study focused on Δvitality, a new variable defined as the difference of vitality before and after sleep. Second, this is a naturalistic study that obtained sleep and vitality measures in free-living conditions. As such, the results may be more readily generalizable to a daily living environment. Third, this study has used GEEs to investigate the association between sleep measures and Δvitality. Owing to the repeated measures of both exposure and outcome available in the data set, this method allows for the maximum utilization of the available information of correlated measures. Lastly, MIMOSYS, the app used to obtain vitality measures through voice analysis, is inexpensive, thereby allowing the study design to be easily replicated in future studies.

Conclusion

A positive association was identified between total sleep time and Δvitality and an inverse association between sleep onset latency and Δvitality. These results suggest that specific sleep parameters are associated with the overnight change of vitality.

Acknowledgments

We thank all staff members at the Center of Innovation, the University of Tokyo, for their extensive efforts and help to conduct the study. We would also like to thank all members of Precision Health, the University of Tokyo, for their invaluable assistance, in particular, Fumiya Tatsuki, for his help with protocol writing. This research was supported by the Center of Innovation Program from Japan Science and Technology Agency, JST, under Grant Number JPMJCE1304. The sponsor had no role in study design;

in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

Authors' Contributions

AKS and TS were responsible for the conception and design of the study. TS analyzed and interpreted the data. HM, MiN, AKS, MaN, ST, and UC assisted with the interpretation of the data. HM drafted the manuscript. MiN, AKS, MaN, ST, UC, and TS critically revised the manuscript for important intellectual content. All authors approved the final version of the manuscript.

Conflicts of Interest

HM, AKS, MaN, UC, and TS have no competing interests to report. MiN reports grants from PST Inc; ST had received financial support from PST Inc until 2019 and currently reports non-financial support from PST Inc.

References

1. Patel NK, Nivethitha L, Mooventhan A. Effect of a Yoga Based Meditation Technique on Emotional Regulation, Self-compassion and Mindfulness in College Students. Explore 2018 Nov;14(6):443-447. [doi: 10.1016/j.explore.2018.06.008]
2. Buchan H, Sunderland M, Carragher N, Louie E, Batterham PJ, Slade T. Investigating factors that bias the reporting of depression symptomatology among older Australian adults. The American Journal of Geriatric Psychiatry 2015 Oct;23(10):1046-1055. [doi: 10.1016/j.jagp.2015.04.003]
3. Tokuno S. Stress evaluation by voice: from prevention to treatment in mental health care. Econophys Sociophys Multidiscip Sci J 2015;5(1):30-35.
4. Hagiwara N, Omiya Y, Shinohara S, Nakamura M, Higuchi M, Mitsuyoshi S, et al. Validity of mind monitoring system as a mental health indicator using voice. Adv Sci Technol Eng Syst J 2017 May;2(3):338-344. [doi: 10.25046/aj020343]
5. Nakamura M. Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE at Jeju. 2017. Relation between severity of sleep apnea syndrome and mental healthiness evaluated from voice URL: https://www.researchgate.net/publication/320891714_Relation_between_Severity_of_Sleep_Apnea_Syndrome_and_Mental_Healthiness_Evaluated_from_Voice [accessed 2020-05-13]
6. Palmer CA, Alfano CA. Sleep and emotion regulation: an organizing, integrative review. Sleep Med Rev 2017 Feb;31:6-16. [doi: 10.1016/j.smrv.2015.12.006]
7. Konjarski M, Murray G, Lee VV, Jackson ML. Reciprocal relationships between daily sleep and mood: a systematic review of naturalistic prospective studies. Sleep Med Rev 2018 Dec;42:47-58. [doi: 10.1016/j.smrv.2018.05.005]
8. Watling J, Pawlik B, Scott K, Booth S, Short MA. Sleep loss and affective functioning: more than just mood. Behav Sleep Med 2016 May 09;15(5):394-409. [doi: 10.1080/15402002.2016.1141770]
9. Dettoni JL, Consolim-Colombo FM, Drager LF, Rubira MC, Cavasin de Souza SBP, Irigoyen MC, et al. Cardiovascular effects of partial sleep deprivation in healthy volunteers. J Appl Physiol 2012 Jul 15;113(2):232-236. [doi: 10.1152/japplphysiol.01604.2011]
10. Virtanen I, Kalleinen N, Urrila AS, Leppänen C, Polo-Kantola P. Cardiac autonomic changes after 40 hours of total sleep deprivation in women. Sleep Med 2015 Feb;16(2):250-257. [doi: 10.1016/j.sleep.2014.10.012]
11. Hirshkowitz M, Whiton K, Albert SM, Alessi C, Bruni O, DonCarlos L, et al. National Sleep Foundation's sleep time duration recommendations: methodology and results summary. Sleep Health 2015 Mar;1(1):40-43. [doi: 10.1016/j.sleh.2014.12.010] [Medline: 29073412]
12. Svensson T, Chung U, Tokuno S, Nakamura M, Svensson AK. A validation study of a consumer wearable sleep tracker compared to a portable EEG system in naturalistic conditions. J Psychosom Res 2019 Nov;126:109822. [doi: 10.1016/j.jpsychores.2019.109822]
13. Yoshida M, Kashiwagi K, Kadotani H, Yamamoto K, Koike S, Matsuo M, et al. Validation of a portable single-channel EEG monitoring system. J Oral Sleep Med 2015;1:140-147 [FREE Full text]
14. Berry R. The AASM Manual for the Scoring of Sleep and Associated Events. 2015. AASM Scoring Manual Version 2.2 URL: https://aasm.org/resources/pdf/scoring-manual-preface.pdf [accessed 2020-05-12]
15. Baglioni C, Nanovska S, Regen W, Spiegelhalder K, Feige B, Nissen C, et al. Sleep and mental disorders: a meta-analysis of polysomnographic research. Psychol Bull 2016 Sep;142(9):969-990. [doi: 10.1037/bul0000053]
16. American Psychiatric Association. Major depression disorder. In: Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition. Washington, DC: American Psychiatric Publishing; 2013:160-168.
17. Watanabe H, Shin T, Matsuo H, Okuno F, Tsuji T, Matsuoka M, et al. Studies on vocal fold injection and changes in pitch associated with alcohol intake. J Voice 1994 Dec;8(4):340-346. [doi: 10.1016/s0892-1997(05)80282-6]
18. Castro-Diehl C, Diez Roux AV, Redline S, Seeman T, McKinley P, Sloan R, et al. Sleep duration and quality in relation to autonomic nervous system measures: the Multi-Ethnic Study of Atherosclerosis (MESA). Sleep 2016 Nov 01;39(11):1927-1940 [FREE Full text] [doi: 10.5665/sleep.6218] [Medline: 27568797]

19. Jike M, Itani O, Watanabe N, Buysse DJ, Kaneita Y. Long sleep duration and health outcomes: a systematic review, meta-analysis and meta-regression. Sleep Med Rev 2018 Jun;39:25-36. [doi: 10.1016/j.smrv.2017.06.011]
20. Park S, Cho MJ, Chang SM, Bae JN, Jeon HJ, Cho SJ, et al. Relationships of sleep duration with sociodemographic and health-related factors, psychiatric disorders and sleep disturbances in a community sample of Korean adults. J Sleep Res 2010 Dec;19(4):567-577 [FREE Full text] [doi: 10.1111/j.1365-2869.2010.00841.x] [Medline: 20477953]
21. Tsai H, Kuo TB, Lin Y, Yang CC. The association between prolonged sleep onset latency and heart rate dynamics among young sleep-onset insomniacs and good sleepers. Psychiatry Res 2015 Dec;230(3):892-898. [doi: 10.1016/j.psychres.2015.11.030]
22. Tokuno S, Mitsuyoshi S, Suzuki G, Tsumatori G. Stress evaluation by voice: a novel stress evaluation technology. 2014 Presented at: 9th International Conference on Early Psychosis (Tokyo); 2014; Tokyo.
23. Huppert FA, Walters DE, Day N, Elliott BJ. The factor structure of the General Health Questionnaire (GHQ-30). Br J Psychiatry 2018 Jan 02;155(2):178-185. [doi: 10.1192/bjp.155.2.178]
24. Hirshkowitz M. Polysomnography challenges. Sleep Med Clin 2016 Dec;11(4):403-411. [doi: 10.1016/j.jsmc.2016.07.002]
25. Matsuo M, Masuda F, Sumi Y, Takahashi M, Yamada N, Ohira MH, et al. Comparisons of portable sleep monitors of different modalities: potential as naturalistic sleep recorders. Front Neurol 2016 Jul 15;7. [doi: 10.3389/fneur.2016.00110]
26. Lucey BP, Mcleland JS, Toedebusch CD, Boyd J, Morris JC, Landsness EC, et al. Comparison of a single-channel EEG sleep study to polysomnography. J Sleep Res 2016 Jun 02;25(6):625-635. [doi: 10.1111/jsr.12417]

Abbreviations

EEG: electroencephalogram
GEE: generalized estimating equation
GHQ-30: 30-item General Health Questionnaire

Edited by M Focsa; submitted 07.11.19; peer-reviewed by K O'Keeffe, C Wu; comments to author 20.02.20; revised version received 01.04.20; accepted 02.04.20; published 09.06.20.

Please cite as:
Miyashita H, Nakamura M, Svensson AK, Nakamura M, Tokuno S, Chung UI, Svensson T
Association Between Electroencephalogram-Derived Sleep Measures and the Change of Emotional Status Analyzed Using Voice Patterns: Observational Pilot Study
JMIR Form Res 2020;4(6):e16880
URL: https://formative.jmir.org/2020/6/e16880
doi:10.2196/16880
PMID:32515745

Original Paper

Technical and Psychosocial Challenges of mHealth Usage for Antiretroviral Therapy Adherence Among People Living With HIV in a Resource-Limited Setting: Case Series

Kennedy Michael Ngowi[1,2], Msc; Furaha Lyamuya[3,4], MD; Blandina T Mmbaga[1,4], MD, DPhil; Eva Muro[3,4], MSc, PhD; Zawadiel Hillu[3], BSc; Mary Shirima[5], BSc; Rob E Aarnoutse[6], DPhil; Mirjam AG Sprangers[2], DPhil; Pythia T Nieuwkerk[2], DPhil; Peter Reiss[7,8], MD, DPhil; Marion Sumari-de Boer[1], DPhil

[1]Kilimanjaro Clinical Research Institute, Kilimanjaro Christian Medical Centre, Moshi, United Republic of Tanzania

[2]Department of Medical Psychology, Amsterdam University Medical Centers, Amsterdam, Netherlands

[3]Kilimanjaro Christian Medical Centre, Moshi, United Republic of Tanzania

[4]Kilimanjaro Christian Medical University College, Moshi, United Republic of Tanzania

[5]Majengo Dispensary, Moshi, United Republic of Tanzania

[6]Radboud Institute for Health Sciences & Department of Pharmacy, Radboud University Medical Center, Nijmegen, Netherlands

[7]Department of Global Health and Division of infectious Diseases, Amsterdam institute of Global Health and Development, Amsterdam University Medical Centers, Amsterdam, Netherlands

[8]Stichting HIV Monitoring, Hogeschool van Amsterdam, Amsterdam, Netherlands

Corresponding Author:
Kennedy Michael Ngowi, Msc
Kilimanjaro Clinical Research Institute
Kilimanjaro Christian Medical Centre
2236
Moshi
United Republic of Tanzania
Phone: 1 2754201
Email: k.ngowi@kcri.ac.tz

Abstract

Background: Mobile communication has been found to improve antiretroviral therapy (ART) adherence among people living with HIV. In an ongoing randomized clinical trial, 2 mobile communication strategies (ie, sending SMS text messages and real-time medication monitoring [RTMM]) were used to improve adherence to ART among people living with HIV in Tanzania. We noticed remarkable discrepancies between self-reported adherence and adherence recorded by SMS text messaging or RTMM among some of the first trial participants.

Objective: Our objective was to describe these cases and the observed discrepancies in more detail, to serve as a useful illustration of some of the challenges in using mobile health in resource-limited settings.

Methods: In an ongoing randomized trial, adults living with HIV from two HIV treatment centers in Tanzania who were suspected of low levels of adherence were randomly assigned in a 1:1:1 ratio to receive (1) SMS text message reminders, (2) an RTMM device, or (3) no additional intervention to standard HIV care. During bimonthly study visits, the participants self-reported their level of adherence, received feedback about their level of adherence based on SMS text messaging or RTMM, and discussed strategies to overcome adherence problems with nurses providing HIV care. For the purpose of this report, we selected people living with HIV who had completed 5 follow-up visits and consistently reported more than 95% adherence, while SMS text messaging or RTMM recorded lower than 75% adherence. The participants were invited for a short, face-to-face in-depth interview to explore reasons for this discrepancy.

Results: At the time of this analysis, 26 participants had completed follow-up. Six of these evidenced the above-mentioned discrepancies, with an average adherence of 46% based on SMS text messaging or RTMM, while self-reported adherence was 98%. Five of these 6 participants insisted that their adherence to ART was good, with 4 reporting that their adherence to properly using the monitoring device was low. Three participants mentioned concerns about involuntary disclosure of HIV status as a

main reason for low adherence to using the device. Two participants were still depending on other reminder cues despite receiving SMS text message or RTMM reminders. Poor network coverage caused low adherence in 1 participant.

Conclusions: Psychosocial barriers were reported as importantly contributing to low adherence, both with respect to use of ART and proper use of the adherence-monitoring device. This case series illustrates that when introducing new digital adherence monitoring technology, researchers should consider psychosocial barriers and distinguish between adherence to device use and adherence to treatment.

Trial Registration: Pan African Clinical Trials Registry PACTR201712002844286; https://tinyurl.com/y98q4p3l

(JMIR Form Res 2020;4(6):e14649) doi:10.2196/14649

KEYWORDS

mHealth; case series; adherence; HIV; real-time medication monitoring; SMS; antiretroviral therapy

Introduction

The Kilimanjaro region in Tanzania has an HIV prevalence rate of 2.6%, which is lower than the country's national prevalence of 4.5% [1,2]. However, in this region, of those who receive antiretroviral treatment (ART), only 47% of people living with HIV have viral load suppression [1]. This suggests that adherence to ART is often not optimal. Due to limited resources, viral load is not routinely monitored, and consequently, timely information about potential poor adherence to ART is not available to HIV health care providers. Therefore, alternative means of providing information about and interventions to improve adherence to ART among people living with HIV are needed [3]. Mobile phone technology can potentially help to fill this need by using digital tools.

Several digital tools exist to monitor and intervene on adherence to treatment. Real-time medication monitoring (RTMM) records the date and time of every opening of a pillbox, a so-called event. It sends this information in real time to a web platform via the mobile network and is one of the existing technologies for monitoring adherence. Reminder SMS text messages may be sent to people living with HIV if the pillbox is not opened on time [4]. Text messaging can also be used on its own as another mobile phone technology to monitor and improve adherence. The feasibility of text messages was found to have mixed results, especially in rural areas [5,6]. Previous studies have described several challenges of delivering RTMM and SMS text messaging in resource-limited settings, such as poor network coverage, power failure, and lack of interoperability among network providers [5,7]. However, previous studies did not describe in detail how such challenges affect the delivery and use of mobile health (mHealth) strategies for treatment adherence. Furthermore, despite these technical challenges, we found that it was feasible and acceptable to monitor ART adherence using RTMM among people living with HIV who reside in the Kilimanjaro region [8].

Currently, we are conducting a randomized trial in which we investigate the effect of mHealth strategies to improve adherence to treatment among people living with HIV in the Kilimanjaro region. People living with HIV are randomly assigned in a 1:1:1 ratio to receive (1) SMS text message reminders, (2) an RTMM device, or (3) no additional intervention to standard HIV care. During the trial, we encountered major discrepancies between self-reported adherence and adherence reports generated by the assigned mHealth intervention in several of our trial participants. The objective of this report is to describe these cases and the observed discrepancies in more detail to serve as a useful illustration of some of the challenges researchers and health care providers need to consider when intending to use mHealth interventions to enhance adherence to ART in resource-limited settings.

Methods

Patients Enrolled in the Randomized Controlled Trial

We first describe the trial and the interventions from where we selected our patients. Our ongoing, parallel-group, 3-arm randomized controlled trial (REMIND Study) has enrolled adults living with HIV from the Kilimanjaro Christian Medical Centre and the Majengo Health Centre, 2 specialized HIV care and treatment centers in Moshi, Tanzania. By 2018, the Kilimanjaro Christian Medical Centre served a total of 2800 people living with HIV, while Majengo served 1200 people living with HIV. In our larger randomized controlled trial study, 264 people living with HIV were screened from both health centers. The study nurses played a vital role in ensuring that potential participants would understand the study. Written informed consent was requested from participants, who were subsequently randomly assigned equally to the study arms SMS text messaging, RTMM, and control. The study was approved by the Kilimanjaro Christian Medical College Research Ethics and Review Committee and the National Health Research Ethics Sub-Committee of Tanzania.

The main eligibility criteria for enrollment in the trial were that the participant must currently be receiving ART and that there is a suspicion of low adherence levels. The suspicion of low adherence was based on the following criteria, as subjectively judged by nurse counselors: (1) self-reported limited adherence, (2) missed medication refill visit, (3) return of leftover medication, or (4) other signs of nonadherence**,** including continuous high viral loads. Exclusion criteria were hospital admission or participation in other trials.

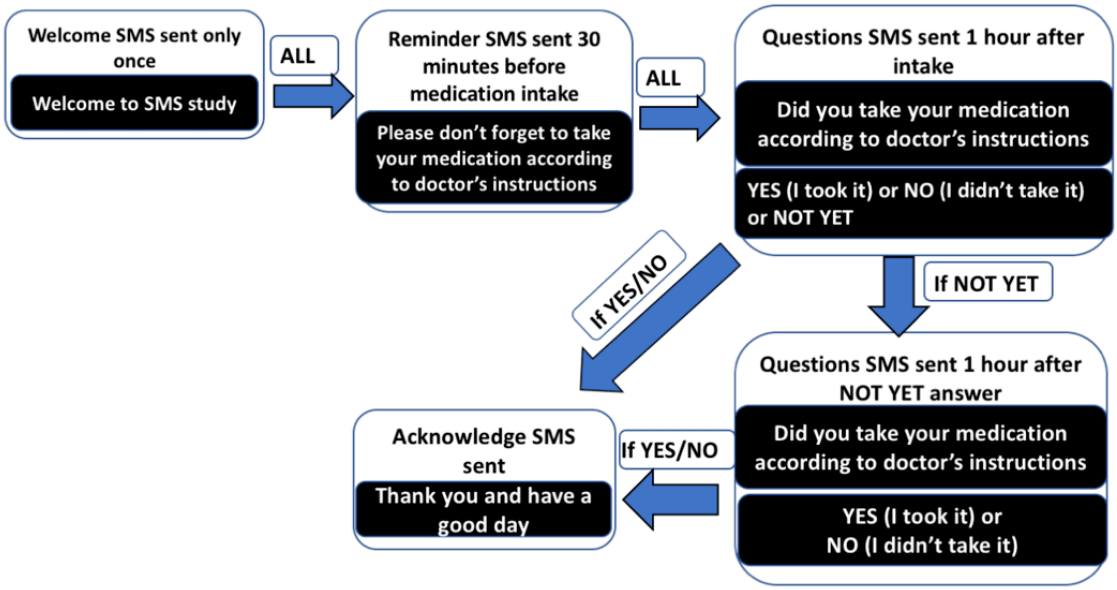
Intervention Arms

SMS Text Messaging

All participants who were enrolled in the SMS arm received 3 messages weekly on random days. The messages were in Swahili and were similar for all patients. Thirty minutes before the usual time of intake, an SMS text message was sent to

remind patients to take their medication. One hour after the usual time of intake, another SMS text message was sent with the question "Did you take your medication?" The participant was asked to reply, choosing from the following 3 options: (1) Yes, I took it, (2) No, I did not take it, or (3) Not yet. The flow of SMS text messages is depicted in Figure 1.

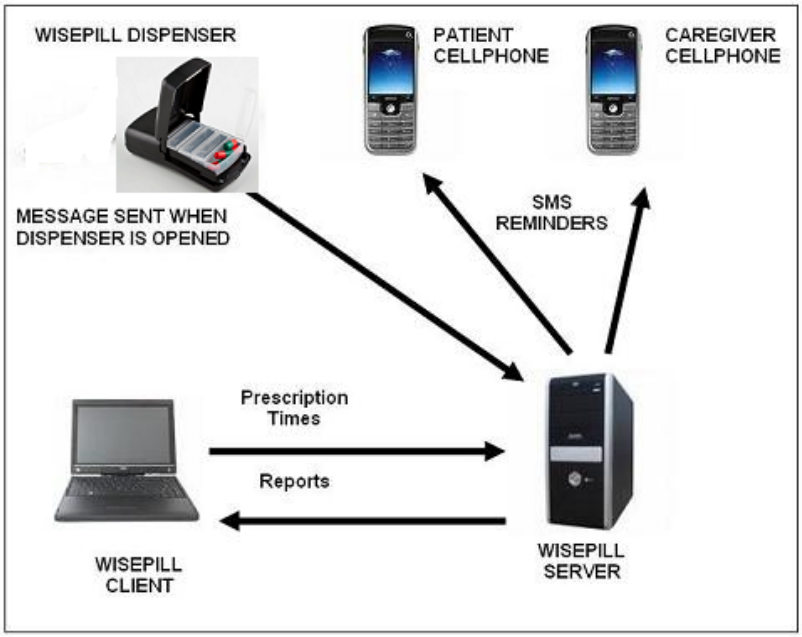
Figure 1. Flow of SMS text messages.



RTMM

All participants enrolled in the RTMM arm received a Wisepill device (Wisepill Technologies) containing a third-generation SIM card that sends signals to a central server each time the device is opened, a so-called medication event and proxy for medication intake. The signal contains information about the time the device is opened, identification number of the device, and technical information about the battery and strength of the signal. If the device was not opened within 1 hour and 15 minutes after the usual time of intake, the participant receives an SMS text message with a question about whether they took their medication. The respondent could not respond to the text message because the device does not have a function to support incoming SMS text messages from users. Authorized members of the study team could access the adherence report generated by the device by signing in through the protected web portal. Figure 2 presents the flow of communication between the pillbox, central server, and a patient's phone.

Figure 2. Wisepill device architecture. (Source: Wisepill Technologies).



Feedback Session Procedure

Participants attended the care and treatment clinics every 2 months to receive standard HIV care and ART medication refills, as per Tanzanian HIV treatment guidelines [6]. During the visit, participants in the 2 intervention arms (SMS and RTMM) received tailored feedback from the nurse counselor about their medication adherence according to SMS or RTMM. First, nurses asked participants whether they skipped any doses of ART and to what extent they had been able to take all doses of their ART in the past period. Next, the nurse provided participants with a graph (generated from SMS or RTMM) showing their level of adherence in percentages. Participants were asked to comment on their adherence behavior displayed on the graphs. Participants' responses to the questions, their answers, comments, and any explanations for potential discrepancies between reported adherence and adherence based on SMS text messaging or RTMM were registered. The feedback was obtained using semistructured questionnaires and recorded on case report forms.

Viral load was measured at enrollment in all study participants. The lower detection limit of the viral load test was 20 copies/ml. According to Tanzanian HIV treatment guidelines, viral loads above 1000 copies/ml should be repeated after 3 months of intensified counseling on adherence to treatment and those below 1000 copies/ml should be repeated after a year. For participants defined as "cases," we retrieved the viral load from their medical records at the beginning of the trial and at the end of the follow-up period of the trial.

Case Selection and Interview

In this paper, we define cases as trial participants for whom there was a major discrepancy between self-reported adherence at the bimonthly study visit with the nurse and the adherence as indicated by SMS or RTMM (ie, patients reporting good adherence, 95% or higher, to the nurse, while SMS or RTMM indicated poor adherence, 75% or lower). All participants defined as cases were invited to an in-depth interview to explore whether they had encountered any technical or practical problems using SMS or RTMM. The interviews were conducted by the study investigator provided with a topic guideline, including perception of being monitored, technical issues, and privacy concerns (ie, concerns that others would see the RTMM device or read the content of the SMS text message, which could disclose the participant's HIV status). The topic list was prepared by the study team.

Analysis

For each participant defined as a case, we calculated the average level of adherence reported to the nurse at bimonthly study visits and the participant's level of adherence according to SMS text messaging or RTMM over the 5 visits during the follow-up period. For self-reported adherence, we directly asked how many doses were missed, and from there we could calculate the participant's adherence level. For adherence according to SMS text messaging or RTMM, we calculated the number of correct intakes (measured by correct openings of the pillbox or by the number of YES replies to the text message) divided by the number of prescribed intakes. We report the participants' comments on discrepancies between self-reported and SMS/RTMM-based adherence, the explanations participants gave for the discrepancies, and possible technical problems, privacy concerns, or other issues that emerged during the interviews.

Results

Case Study Participants

By November 2018, 26 participants of 249 who started with the clinical trial had completed all 5 study clinic visits. Of the 26 participants, 12 were enrolled from Majengo sites and 14 from the Kilimanjaro Christian Medical Centre. As recorded by SMS text messaging or RTMM, adherence rates were higher than 75% in 20 participants. Adherence rates were lower than

75% in 6, but these participants self-reported adherence rates of 95% or higher and were thus considered to have major discrepancies between self-reported and SMS/RTMM-based adherence at each of the 5 study visits. For all cases, a median of 46% adherence based on SMS or RTMM was recorded. Three of these participants (all female) were enrolled in the SMS arm, and 3 (1 female and 2 male) were enrolled in the RTMM arm. All participants were on a twice-daily ART regimen. Two participants in the RTMM arm and 1 in the SMS arm had suppressed viral loads (Multimedia Appendix 1).

Case 1

The first case was a female living in urban Moshi who responded to 73 (43%) of 168 SMS text messages with the question of whether she took her medication. During study visits 1 to 5, the participant confirmed that she received SMS text message reminders on time, but she often did not respond to them; she said, "I think I lacked commitment in responding to SMS."

She felt depressed because she was separated from her family. She had not expected that the nurses would find out that she was not responding to the SMS text messages. After she was confronted by the nurses about her low response rate to SMS text messages, she tried to reply to them more often. She mentioned that she had no difficulties taking her pills, despite not replying to the SMS text messages. To improve adherence to ART, she said,

> *I will set my phone alarm as extra reminder and involve the caregiver as well.*

She did not have concerns about privacy and answered,

> *I have a smartphone and it's easy to disable SMS notification as well as put a password. So, no one can see or have access to my SMS.*

Therefore, this participant reported high levels of adherence to ART despite a low response rate to SMS text messages, and the low response rate appeared to be related to feelings of depression.

Case 2

Case 2 was an urban Moshi woman. Throughout visits 1 to 5, 186 SMS text messages were sent and delivered with the question about whether she took her medication. However, only 95 (51%) SMS text messages were responded to with, "Yes, I took it" and 91 (49%) SMS text messages were not responded to.

When confronted by the nurses about her low response rate to text messages during the feedback sessions, she mentioned that the main challenge that hindered her in replying to SMS text messages were her friends, who normally spent a lot of time with her, especially in the evening. She explained,

> *My friends are always in my room, especially in the evening; therefore, I am not feeling comfortable to respond to SMS.*

When asked whether she had disclosed her HIV status to them, she said,

> *I'm scared to tell them, and with my age I still need them around. Also, none of my friends will believe I acquired HIV when I was born.*

Of interest, she said, "Despite not responding to the SMS, I still take my medication."

This participant also reported high levels of adherence to ART despite having a low response rate to SMS text messages. The main reason for the low response rate to SMS text messages was related to the fear of unwanted disclosure of her HIV status.

Case 3

In 2012, a 29-year-old woman living in urban Moshi started ART. During visits 1 to 5, 186 SMS text messages were sent and received, asking whether she took her medication. She only replied to 53 (28%) text messages and did not respond to 133 (72%).

During feedback sessions with the nurses, she insisted, "I never missed the medication intakes and I replied to the SMS." She also said,

> *I am surprised to see that I did not reply to all SMS. I am wondering whether you are missing something as I believe that I managed to take about 80% of my pills.*

However, in the interview with the study coordinator, she confessed,

> *I was not honest during the feedback sessions with the nurses. The reason for that is that I was depressed because my sister found out I was receiving SMS to remind me to take medication for HIV and then it became even worse when she started telling my neighbors that I will die soon.*

When we asked if she thought the content of the SMS text messages should be changed, she said,

> *I don't have a problem with the content. Nowadays I delete the SMS once I have replied to it so that no one will read them in case they gain access to my phone.*

For this participant, the main reason for not replying to SMS was related to unwanted disclosure of HIV status and stigma surrounding HIV.

Case 4

Case 4 was a 58-year-old female living in rural Moshi who started ART in 2012 and was enrolled in the RTMM arm. According to RTMM, her level of adherence during the first 2 study months was 47%. After the nurses showed her this level of adherence during a feedback session, she said,

> *I never missed the medication intakes. I think the RTMM was not working properly. Therefore, I want to be provided with a new device.*

The study nurse exchanged the old device with a new one. In the next feedback session, her level of adherence according to RTMM was still low, at 20%, despite her insistence that she never missed a medication intake.

During the interview with the study investigator, the participant mentioned that the device was easy to use, but charging it was

a challenge because it did not have an alarm to indicate a low battery. At the beginning of the study, this participant lived in a rural area, where her adherence level was low, according to RTMM. After 10 months, she moved to an urban area and the device started to send daily signals, showing device openings indicative of a high level of adherence. The participant explained,

> *I have never missed my medication intakes; therefore, I was surprised why the device was not recording the openings. After I moved from my old house in the village, the device started to record the openings, so I thought the problem was network coverage.*

The participant was happy to be monitored in real time, as it made her feel cared for. Privacy concerns were not an issue for this participant as she had disclosed her HIV status to her family.

For this participant, the main reason for the discrepancy between self-reported adherence and adherence generated by the device seemed to be related to adequate power, charging, and availability of network coverage.

Case 5

Case 4 was a 33-year-old male who lives in urban Moshi. He was enrolled in the RTMM arm and started ART in 2013. According to RTMM, his level of adherence to ART was 57%, on average, during the follow-up period of the study. This participant explained that his main problem using the RTMM device was concern about disclosure of his HIV status:

> *I have no problem with using the device except that I did not disclose my HIV status to my children and my co-workers. So sometimes it is difficult to open the device when I am with them. However, I always take my pills, which I kept outside the device.*

This participant further explained,

> *I had difficulties carrying the device to my workplace and since I'm working late sometimes, I sometimes missed my evening dose of medication.*

Thus, the main problems for this participant were related to unwanted disclosure of HIV status and difficulties incorporating the use of the device in daily activities.

Case 6

The sixth case was a 21-year-old man. With an average of 60% ART adherence during the follow-up of the study, he acknowledged missing his medication due to the lack of an alarm on the RTMM device. He explained,

> *When I'm home with the device, I normally forget to open it…Sometimes I wish the device should have an alarm to notify me to open it for my medication intake.*

He also said,

> *Currently I depend on other reminders, such as news time hours and the alarm of the wall clock.*

Before this participant enrolled in the study, he used to set an alarm on his mobile phone as reminder to take his medication. Now he says,

> *Since the device does not give an alarm, sometimes I missed my medication intakes as there is nothing to ring as alarm. However, as soon I receive the SMS reminder, I take my medication.*

Therefore, for this participant, the main reason for poor adherence recorded by the device was that the normal strategy that he used to remind himself to take his medication was interrupted when entering the study, and he needed to adopt a new strategy. As the SMS text message reminder comes late, the participant might indeed be adherent but not on time.

Discussion

Principal Findings

This case series explored, in detail, 6 participants who had major discrepancies between their self-reported and digitally monitored adherence. This paper illustrates the challenges in using mHealth, which emerged during an intervention trial using SMS text message reminders and RTMM to improve adherence to ART among people living with HIV in the Kilimanjaro region in Tanzania. Interviews were conducted to determine whether the discrepancies were triggered by difficulties in using the interventions, poor network, power failures, potential stigma, or other reasons.

One remarkable observation was that it is important to distinguish between adherence to proper use of the digital adherence monitoring device and adherence to treatment. Patients enrolled in the SMS arm were expected to respond to each question they received by SMS text message. However, of all the SMS text messages that were delivered, on average, only 46% postintake text messages were responded to. The results showed that the postintake SMS text message was not triggering the trial participants to respond well to the messages. This finding seems to be in line with a previous study reporting that implementation of mHealth interventions in low-income countries can be complicated if end users have difficulties adapting to new technologies due to lack of experience or psychosocial barriers, which may lead to inadequate use of the technology (ie, limited responding to messages or incorrect use of the device for medication intake) [9].

Another remarkable finding was that several participants considered using additional ways to remind themselves to take their medication, such as using a phone alarm. We found this striking, as sending electronic reminders to participants was part of both intervention arms, yet participants needed additional tools to help them remember to take their medicine on time.

One participant insisted that she took her medication even when the RTMM device was replaced with a new one, but the new device continued to indicate poor adherence. This discrepancy turned out to be caused by the device not finding a network, and it was therefore unable to properly transmit the signal to indicate the medication intake event. As such, each time the device was opened, it was not recorded in the system. However, after the participant moved to an area with a network, the device was able to send the signal on a daily basis. This underlines that adequate network coverage is a prerequisite for the feasibility of mHealth.

In our previous study in the Kilimanjaro region [8], we also encountered the problem of network instability due to bad coverage that led to delays and failures in delivering reminder SMS text messages among participants residing in several remote villages. Despite the technical challenges that were addressed by interviewing participants, we believe that using mHealth and integrating it with the existing health system is possible, but it would require stability of several factors such as network coverage and power supply [9,10]. We therefore suggest that future studies on mHealth interventions in resource-limited settings may benefit from involving other stakeholders, especially mobile network providers, since they have a vital role in providing stable network coverage necessary for delivering such interventions.

Psychosocial barriers, including mental health issues and lack of social support, have been shown to result in poor use of interventions due to fear of HIV disclosure within the family or at workplaces. Findings from a previous study in Kenya showed that SMS text messaging had better effect on treatment adherence for those who received SMS text messages and who reported a high level of social support [11]. Therefore, it is important to take into account mental health issues when applying digital tools by, for example, providing support through messages or using the tool to increase communication about mental health issues during clinic visits.

Privacy was a main issue for participants who did not disclose their HIV status to others. This finding is consistent with that of studies in South Africa and a study in Kenya that showed that some participants had concerns that monitoring was intruding on their privacy [12,13]. The use of neutral messages that do not refer to topics related to sickness or medication could overcome the problem of unwanted disclosure. In addition, participants in the South African study had concerns that the size and design of the pillbox would reveal their HIV status to the community, especially when they needed to open it in the vicinity of others. However, our pilot study in Kilimanjaro [8] showed a different result, as most participants were happy with the color and size of the device.

Case 6 illustrates the importance of ensuring that an intervention intended to improve adherence does not interfere with a person's strategy to enhance their level of adherence. Despite that we told our participants to keep using their normal ways of reminding themselves to take their medication, this participant decided to quit using his phone alarm. Although he declared that the reminder SMS text message served as an alarm to him, it did not lead to improved medication intake according to the device. We cannot say that using the device hampered medication intake, as this participant was judged to be nonadherent at enrollment. Unfortunately, it was not clear why this patient was not adherent. It is possible that the participant was still not willing to disclose to the interviewer that he was not taking his medication or not taking it on time.

Two participants had suppressed viral loads after being exposed to the intervention for several months, despite having low levels of adherence as indicated by responses to text messages or by recorded RTMM device openings. It remains unclear whether these participants were adherent to taking their medication but nonadherent to the proper use of the intervention (ie, responding to SMS text messages or opening the RTMM device). This difficulty in distinguishing between adherence to treatment and adherence to proper use of an intervention or device is not unique to the 2 mHealth interventions that were used in this study. Similar findings have been reported in 2 studies conducted in Kenya [11,12]. Another explanation for the suppressed viral loads might be that levels of adherence were incorrectly recorded due to technical problems. To be able to distinguish between true low adherence to treatment and low adherence to the intervention combined with technical problems, drug levels in the blood could give more insight. Because it was not possible to measure plasma concentrations of antiretroviral drugs in this study, we could not verify whether these participants were nonadherent to their medication intake or nonadherent to using their device.

Conclusion

This case series explored the reasons underlying major discrepancies between self-reported adherence and digitally monitored adherence that occurred in 6 participants in our mHealth intervention trial. Based on our findings, we can make a number of recommendations. First, network coverage at the participant's home must be ensured before implementation. Second, neutral messages (eg, "Hello, this is your friend from the REMIND Study. I hope you are doing well.") should be used to avoid unwanted disclosure of HIV status. Third, participants should be advised to continue using their usual reminder cues. Fourth, a triage system could be developed to determine whether a patient is ready to use such mobile interventions to prevent the paradoxical situation that an intervention intended to improve treatment adherence actually worsens adherence. Fifth, researchers should attempt to distinguish between adherence to the use of a monitoring device and adherence to treatment. Last, our findings serve as a reminder of the paramount importance of psychosocial support in the context of providing HIV treatment and care.

Acknowledgments

We thank the participants for their involvement in our study and for allowing us to study potentially sensitive data. Furthermore, we thank the administration of the Kilimanjaro Christian Medical Centre and Majengo Dispensary for allowing us to conduct this study in the hospital and the nurse counselors for recruiting the participants. Last, we thank the European and Developing Countries Clinical Trials Partnership (EDCTP) for the financial support to conduct the study and The East African Consortium for Clinical Research (EACCR) for the financial support of article processing charges (publication fee).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of case characteristics.
[DOCX File , 13 KB - formative_v4i6e14649_app1.docx]

References

1. PHIA Project. Tanzania HIV Impact Survey. Summary Sheet: Preliminary Findings URL: https://phia.icap.columbia.edu/wp-content/uploads/2017/11/Tanzania_SummarySheet_A4.English.v19.pdf [accessed 2020-05-21]
2. UNAIDS. United Republic of Tanzania URL: https://www.unaids.org/en/regionscountries/countries/unitedrepublicoftanzania [accessed 2020-05-21]
3. Lyimo RA, van den Boogaard J, Msoka E, Hospers HJ, van der Ven A, Mushi D, et al. Measuring adherence to antiretroviral therapy in northern Tanzania: feasibility and acceptability of the Medication Event Monitoring System. BMC Public Health 2011 Feb 09;11(1):92 [FREE Full text] [doi: 10.1186/1471-2458-11-92] [Medline: 21306643]
4. Bionghi N, Daftary A, Maharaj B, Msibi Z, Amico KR, Friedland G, et al. Pilot evaluation of a second-generation electronic pill box for adherence to Bedaquiline and antiretroviral therapy in drug-resistant TB/HIV co-infected patients in KwaZulu-Natal, South Africa. BMC Infect Dis 2018 Apr 11;18(1):171 [FREE Full text] [doi: 10.1186/s12879-018-3080-2] [Medline: 29642874]
5. Friederici N, Hullin C, Yamamichi M. mHealth. In: Information and Communications for Development 2012: Maximizing Mobile. Washington, DC, United States: International Bank for Reconstruction and Development; 2012:45-58.
6. United Republic of Tanzania. Ministry of Health, Community Development, Gender, Elderly and Children. 2017. National Guidelines on HIV and Data Management URL: http://www.nacp.go.tz/guidelines/ [accessed 2020-05-21]
7. Haberer JE, Musiimenta A, Atukunda EC, Musinguzi N, Wyatt MA, Ware NC, et al. Short message service (SMS) reminders and real-time adherence monitoring improve antiretroviral therapy adherence in rural Uganda. AIDS 2016;30(8):1295-1299. [doi: 10.1097/qad.0000000000001021]
8. de Sumari-de Boer IM, van den Boogaard J, Ngowi KM, Semvua HH, Kiwango KW, Aarnoutse RE, et al. Feasibility of Real Time Medication Monitoring Among HIV Infected and TB Patients in a Resource-Limited Setting. AIDS Behav 2016 May 24;20(5):1097-1107. [doi: 10.1007/s10461-015-1254-0] [Medline: 26604004]
9. Wallis L, Blessing P, Dalwai M, Shin SD. Integrating mHealth at point of care in low- and middle-income settings: the system perspective. Glob Health Action 2017 Jun;10(sup3):1327686 [FREE Full text] [doi: 10.1080/16549716.2017.1327686] [Medline: 28838302]
10. Thondoo M, Strachan DL, Nakirunda M, Ndima S, Muiambo A, Källander K, InSCALE Study Group. Potential Roles of Mhealth for Community Health Workers: Formative Research With End Users in Uganda and Mozambique. JMIR Mhealth Uhealth 2015 Jul 23;3(3):e76 [FREE Full text] [doi: 10.2196/mhealth.4208] [Medline: 26206419]
11. van der Kop ML, Muhula S, Nagide PI, Thabane L, Gelmon L, Awiti PO, et al. Effect of an interactive text-messaging service on patient retention during the first year of HIV care in Kenya (WelTel Retain): an open-label, randomised parallel-group study. The Lancet Public Health 2018 Mar;3(3):e143-e152. [doi: 10.1016/s2468-2667(17)30239-6]
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. van der Straten A, Montgomery E, Pillay D, Cheng H, Naidoo A, Cele Z, et al. Feasibility, performance, and acceptability of the Wisebag™ for potential monitoring of daily gel applicator use in Durban, South Africa. AIDS Behav 2013 Feb;17(2):640-648 [FREE Full text] [doi: 10.1007/s10461-012-0330-y] [Medline: 23054042]

Abbreviations

ART: antiretroviral therapy
mHealth: mobile health
RTMM: real-time medication monitoring

Edited by G Eysenbach; submitted 01.10.19; peer-reviewed by H Apers, JJ Mira; comments to author 11.11.19; revised version received 20.02.20; accepted 21.02.20; published 10.06.20.

Please cite as:
Ngowi KM, Lyamuya F, Mmbaga BT, Muro E, Hillu Z, Shirima M, Aarnoutse RE, AG Sprangers M, Nieuwkerk PT, Reiss P, Sumari-de Boer M
Technical and Psychosocial Challenges of mHealth Usage for Antiretroviral Therapy Adherence Among People Living With HIV in a Resource-Limited Setting: Case Series
JMIR Form Res 2020;4(6):e14649
URL: http://formative.jmir.org/2020/6/e14649/
doi:10.2196/14649
PMID:32519966

Original Paper

A Mind-Body Physical Activity Program for Chronic Pain With or Without a Digital Monitoring Device: Proof-of-Concept Feasibility Randomized Controlled Trial

Jonathan Greenberg[1,2], PhD; Paula J Popok[1], BA; Ann Lin[1], BS; Ronald J Kulich[2,3], PhD; Peter James[2,4], MHS, SCD; Eric A Macklin[2,5], PhD; Rachel A Millstein[1,6], MHS, PhD; Robert R Edwards[2,7], PhD; Ana-Maria Vranceanu[1,2], PhD

[1]Integrated Brain Health Clinical and Research Program, Massachusetts General Hospital, Boston, MA, United States

[2]Harvard Medical School, Boston, MA, United States

[3]Center for Pain Medicine, Massachusetts General Hospital, Boston, MA, United States

[4]Department of Population Medicine, Harvard TH Chan School of Public Health, Boston, MA, United States

[5]Biostatistics Center, Massachusetts General Hospital, Boston, MA, United States

[6]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, United States

[7]Pain Management Center, Brigham and Women's Hospital, Boston, MA, United States

Corresponding Author:
Ana-Maria Vranceanu, PhD
Integrated Brain Health Clinical and Research Program
Massachusetts General Hospital
One Bowdoin Square, 1st Floor, Suite 100
Boston, MA,
United States
Phone: 1 617 724 4977
Email: avranceanu@mgh.harvard.edu

Abstract

Background: Chronic pain is associated with poor physical and emotional functioning. Nonpharmacological interventions can help, but improvements are small and not sustained. Previous clinical trials do not follow recommendations to comprehensively target objectively measured and performance-based physical function in addition to self-reported physical function.

Objective: This study aimed to establish feasibility benchmarks and explore improvements in physical (self-reported, performance based, and objectively measured) and emotional function, pain outcomes, and coping through a pilot randomized controlled trial of a mind-body physical activity program (*GetActive*) with and without a digital monitoring device (*GetActive-Fitbit*), which were iteratively refined through mixed methods.

Methods: Patients with chronic pain were randomized to the *GetActive* (n=41) or *GetActive-Fitbit* (n=41) programs, which combine relaxation, cognitive behavioral, and physical restoration skills and were delivered in person. They completed in-person assessments before and after the intervention. Performance-based function was assessed with the 6-min walk test, and step count was measured with an ActiGraph.

Results: Feasibility benchmarks (eg, recruitment, acceptability, credibility, therapist adherence, adherence to practice at home, ActiGraph wear, and client satisfaction) were good to excellent and similar in both programs. Within each program, we observed improvement in the 6-min walk test (mean increase=+41 m, SD 41.15; P<.001; effect size of 0.99 SD units for the *GetActive* group and mean increase=+50 m, SD 58.63; P<.001; effect size of 0.85 SD units for the *GetActive-Fitbit* group) and self-reported physical function (P=.001; effect size of 0.62 SD units for the *GetActive* group and P=.02; effect size of 0.38 SD units for the GetActive-Fitbit group). The mean step count increased only among sedentary patients (mean increase=+874 steps for the *GetActive* group and +867 steps for the *GetActive-Fitbit* group). Emotional function, pain intensity, pain coping, and mindfulness also improved in both groups. Participants rated themselves as much improved at the end of the program, and those in the *GetActive-Fitbit* group noted that Fitbit greatly helped with increasing their activity.

Conclusions: These preliminary findings support a fully powered efficacy trial of the two programs against an education control group. We present a model for successfully using the Initiative on the Methods, Measurement, and Pain Assessment in Clinical

Trials criteria for a comprehensive assessment of physical function and following evidence-based models to maximize feasibility before formal efficacy testing.

Trial Registration: ClinicalTrial.gov NCT03412916; https://clinicaltrials.gov/ct2/show/NCT03412916

(JMIR Form Res 2020;4(6):e18703) doi:10.2196/18703

KEYWORDS

chronic pain; meditation; walking; feasibility studies; actigraphy

Introduction

Background

Chronic musculoskeletal pain is costly [1] and associated with substantial emotional and functional limitations [2-5]. Current treatment recommendations support nonpharmacological approaches for pain management (Centers for Disease Control and Prevention guidelines) [6]. Evidence-based treatments such as cognitive behavioral therapy [7], acceptance and commitment therapy–based approaches [8], mindfulness-based approaches [9], and aerobic exercise [10] are efficacious in improving emotional and physical function. However, across all approaches, randomized controlled trials (RCTs) have tended to report mostly short-term improvements with relatively small effect sizes [9,11-16]. Novel interventions are needed to sustainably improve emotional and physical function in this population.

Using the International Classification of Functioning, Disability and Health (ICF) guidelines [17]; Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) recommendations [18,19]; and guidelines for intervention development and clinical trials [20-22], we developed and iteratively refined a multimodal mind-body physical activity program that combines evidence-based pain management skills with increased physical activity through both quota-based pacing (gradual increase in activity that is noncontingent on pain levels) and linking increased walking to activities of daily living [23]. Our guiding hypothesis was that the most effective and efficient way to sustainably improve physical function among patients with chronic pain is to combine evidence-based mind-body skills with physical activity. In line with IMMPACT guidelines [18], the goal of the program is to increase not only self-reported but also performance-based (eg, the 6-min walk test [24]) and objectively measured (eg, step count as measured by accelerometers) physical function, which together provide a critical and comprehensive snapshot of an individual's abilities and function. As previous research has shown that increasing activity in this population is challenging because of improper pacing [23,25], we hypothesized that Fitbit, an inexpensive, commercially available wrist-worn digital activity monitor with visible display, can be a useful aid by providing real-time self-monitoring and reinforcement of activity consistent with a predetermined weekly step goal. To date, several studies have used Fitbit alone, not paired with coping skills training, and did not find meaningful step count increases among patients with chronic pain [26,27]. We currently do not know whether daily monitoring of activity with a digital monitoring device such as Fitbit would aid participants in increasing step count.

Our long-term goal is to conduct a fully powered RCT comparing 2 identical mind-body physical activity programs, one with a Fitbit device (*GetActive-Fitbit*) and the other without (*GetActive*), with an attention educational placebo control. This future RCT will seek to (1) determine the efficacy of the 2 programs in sustainably improving self-reported, performance, and objectively measured physical function and to (2) understand whether program-dependent improvement in physical function (self-reported, performance based, and objectively measured) is enhanced by using a Fitbit. However, in accordance with the Obesity Related Behavioral Intervention Trials (ORBIT) [21] and the National Center for Complementary and Integrative Health (NCCIH) [22] models of intervention development and optimization, multiple program iterations are necessary to maximize methodology and feasibility before a large efficacy clinical trial. This study represents stage IIa and IIb of the ORBIT model (proof of concept and pilot testing) and stage 3 of the NCCIH treatment development model (feasibility and pilot studies). The goals of stages IIa and IIb of the ORBIT model are to test for clinically meaningful changes and determine the source of the treatment effect [21]. The goals of stage 3 of NCCIH's development model are to determine whether a subsequent larger study of the refined intervention can be successfully implemented and can provide clinically meaningful evidence for efficacy [22].

We used a sequential approach to the development and initial testing of our mind-body physical activity program focused on multiple iterations to optimize it, meet a-priori set markers of feasibility, and establish a signal of improvement in outcomes (Multimedia Appendix 1) [21,28]. First, we conducted qualitative focus groups with adults with heterogeneous chronic pain to gather their feedback, needs and preferences, and barriers and facilitators to program participation, increasing activity and using a Fitbit [23]. Next, we developed a mind-body physical activity program that incorporates mind-body skills adapted from the Relaxation Response Resiliency Program [29], for example, eliciting the relaxation response (mindfulness meditation), pain-specific cognitive behavioral skills (eg, goal setting, behavioral activation techniques, adaptive restructuring of pain-related misconceptions such as catastrophizing and avoidance of fear), and physical restoration skills (eg, quota-based pacing noncontingent on pain). From this initial version, we developed the first version of *GetActive* (8 sessions), where participants increase activity using time-based goals, and *GetActive with Fitbit* (8 sessions), where participants increase activity using step count–based goals with real-time reinforcement aided by a Fitbit. Both programs focused on walking, as this was the preferred activity within our focus groups [23]. Using a nonrandomized, controlled, small pilot

trial (ORBIT phase IIa; n=6 and n=7), we found that both programs had good to excellent feasibility and acceptability markers and relatively similar signals of improvement in physical, emotional function, and intervention targets [23]. However, exit interviews with group participants informed additional program modifications, including the use of self-compassion when goals are not met, language clarifications, and increasing the number of sessions to facilitate skill acquisition.

Objectives

Here, we report on feasibility markers and within-group improvements in outcomes after a pilot RCT of the refined *GetActive* (10 group sessions) and *GetActive-Fitbit* (10 group sessions; Fitbit to self-monitor activity) programs. Our primary hypothesis was that both programs will meet a-priori set feasibility benchmarks (feasibility of recruitment, program acceptability, credibility and expectancy, therapist adherence to the manual, feasibility of quantitative measures, adherence to practice at home, adherence to ActiGraph accelerometer and Fitbit, and safety) necessary before an efficacy trial. Our secondary hypothesis was that for both programs, we will observe within-group improvements in physical function (self-reported, 6-min walk test, and ActiGraph step count), emotional function (anxiety and depression), pain-specific outcomes (intensity and coping with pain), as well as adaptive coping and mindfulness.

Methods

Participants

We recruited patients with heterogeneous musculoskeletal chronic pain via direct referrals from the Massachusetts General Hospital Pain Clinic as well as flyers and hospital-wide email lists advertising our study. Recruitment occurred between July 2018 and September 2019. The study was funded by the NCCIH and approved by Massachusetts General Hospital's institutional review board (IRB). The inclusion criteria were as follows: (1) ≥18 years old, (2) self-reported nonmalignant chronic pain for >3 months, (3) ability to walk unassisted for at least 6 min, (4) access to a mobile device with Bluetooth version 4.0, (5) no change in psychotropic or pain medications for the past 3 months, and (6) cleared for participation by a physician. The exclusion criteria were as follows: (1) medical illness expected to worsen in the next 6 months, (2) serious and untreated psychiatric illness or active suicidality, (3) current untreated substance use disorder, (4) practicing meditation/yoga or relaxation response skills for >45 min a week in the past 6 months, (5) using a Fitbit device in the past 6 months, and (6) engaged in regular physical exercise for >30 min daily by self-report.

Procedure

Screening, Enrollment, and Randomization

We screened all referrals via a phone call using a standard, IRB-approved scripted checklist. All participants were informed that the goal of the program was to increase physical activity rather than decrease pain. We documented all reasons for ineligibility (Multimedia Appendix 2). All screenings were reviewed at team meetings with the study principal investigator. Eligible participants who were able to meet during the designated group times (a 3-hour block that would accommodate randomization to *GetActive* or *GetActive-Fitbit* 90-min groups) were immediately scheduled for the next available group sessions. The rest were placed on a waitlist and contacted for future groups. The date and time for each group cohort was flexibly determined based on the availability of most eligible potential participants. For each group cohort, participants were asked to come to the clinic on the same day and time to undergo informed consent, complete self-report assessments, complete a 6-min walk test [24], and start wearing a wGT3X-BT ActiGraph accelerometer (ActiGraph, LLC). Participants received detailed instructions (1) to wear the ActiGraph over their right hip using an elastic belt for 1 week during all waking hours, except while in water (bathing or swimming), (2) to maintain their regular levels of activity, and (3) to fill out a daily device wear and physical activity log. Participants were asked their preferred method to receive daily reminders to wear the ActiGraphs (eg, text messages, phone calls, or emails). At the end of the assessment session (approximately 90 min), participants were randomized to either the *GetActive* or *GetActive-Fitbit* program based on a 1:1 ratio via a sequence generated by *sealedenvelope.com* in blocks of 12. Participants were compensated with US $30 for completion of baseline assessments. They were instructed to return for the first session the following week at their assigned time and return the ActiGraph and wear log. In the first session, participants received the *GetActive* or *GetActive-Fitbit* treatment manuals. Those randomized to *GetActive-Fitbit* also received a Fitbit that was paired with each participant's phone. The average step count recorded by the ActiGraph was set as the initial step goal on the participants' Fitbit devices.

GetActive and GetActive-Fitbit Programs

Development of the original 8-week programs, details on program skills, and exit interviews to inform the current program versions have been previously reported [23]. The final *GetActive* and *GetActive-Fitbit* programs have 10 weekly 90-min sessions (Multimedia Appendix 3). The programs teach 4 core skills: (1) weekly SMART goal setting (defined as goals that are specific, measurable, achievable, relevant and time-based) [29] for a gradual increase in physical activity paired with activities of daily living that are meaningful and important to participants (ie, walk instead of drive to the store and walk to the park with kids) and the daily practice of mind-body skills (eg, engage in meditation before going to bed and when walking), (2) individualized quota-based pacing (eg, walk for 30 min or meet a step goal of 5000), (3) mind-body skills (diaphragmatic breathing to manage intense pain flares and pain anxiety, body scan to increase body awareness and reduce reactivity to pain sensations, mindfulness exercises to understand the transience of pain and change one's relationship with it, and self-compassion when falling short of set goals), and (4) understand the disability spiral (eg, how reducing activity perpetuates pain and disability) and correct myths about pain or automatic pain-related thoughts that interfere with meeting program goals. At each weekly session, the group leader reviewed home practice, including adherence to activity goals,

and helped participants solve barriers to adherence. Participants who missed group sessions were immediately contacted by the study staff and scheduled for a make-up session.

The *GetActive* and *GetActive-Fitbit* programs are identical in content and structure. However, in the *GetActive-Fitbit* program, the study staff instructed participants how to consistently wear and charge the Fitbit (session 1), uploaded an individualized step goal onto each participant's Fitbit during each weekly session through Fitbit.com, monitored Fitbit wear in real time through the Fitabase website (Fitabase) [30], immediately called nonadherent participants to solve problems with adherence, and encouraged participants to focus on meeting the daily activity SMART goals.

Procedure for Fitbit Step Count Assessments and Pacing

Details can be found in the study by Greenberg et al [23]. Briefly, before each session, a staff member downloaded the participants' Fitbit data from the past week, calculated the participants' step goals for the upcoming week, and provided the study clinician with individual adherence data for each participant, which were discussed during the session. If participants reached their weekly goal, they could choose whether to repeat it for the following week or increase it by 10% [31,32]. If they did not reach their goal, they would repeat it. If they did not reach their goal 2 weeks in a row, they could choose whether to repeat it or set a new goal that was 10% higher than their actual step count during that week (rather than their step count goal).

Postintervention Assessments

In the last group session, participants were handed the ActiGraph and were instructed to complete the activity log again for 1 week. They were asked to return the following week as a group to complete the postintervention assessment. Participants were again compensated with US $30 for the assessment session.

Feasibility Assessments

Feasibility markers were determined to be consistent with guidelines for intervention development [21,28]. Feasibility benchmarks were set a-priori. We assessed the following feasibility markers:

1. *Feasibility of recruitment:* This was assessed as the proportion of potential participants successfully contacted who agreed to participate. We considered a proportion of 80% excellent and a proportion of 70% good.
2. *Program acceptability:* This was assessed via the proportion of participants who attended at least 7 out of 10 sessions. Feasibility was considered excellent when >80% of the participants attended 7 out of 10 sessions and good if at least 70% of the participants did.
3. *Credibility and expectancy:* This was assessed using the credibility and expectancy questionnaire [33], a 6-item questionnaire assessing both the degree to which participants find the program logical and convincing (credibility) and the degree to which they believe they will benefit (expectancy). Credibility and expectancy were considered excellent when >75% of the participants scored above the scale midpoint and good if at least 70% of the participants did.
4. *Therapist adherence to the program manual:* This was assessed in 2 ways. First, the study clinician completed a content checklist at the end of each of the sessions, marked all session content that was covered within each session, and wrote a progress note. Therapist adherence to the program manual was determined as excellent if 100% of the therapist checklists covered 100% of the session skills and good if at least 75% of the checklists covered 100% of the session skills. Next, 20% of the audio recorded sessions were coded for content against the therapist checklists by 2 independent coders who were trained by the principal investigator. An agreement (Cohen κ) of at least 60% was considered good and higher than 80% was considered excellent [34].
5. *Feasibility of quantitative measures:* This was measured by assessing the questionnaire completion and internal consistency reliability of the measures. This was considered to be acceptable if the internal reliability of the questionnaires (Cronbach α) was higher than .70 and none of the questionnaires were fully missing in more than 25% of the participants.
6. *Adherence to home practice:* Adherence to home practice was assessed via the number of days per week participants self-reported that they practiced meditation and appreciation exercises. We considered adherence to be excellent if 3 home practice components (eg, relaxation response practice, SMART goal activity, and appreciations) were completed at least 3 out of 7 days per week or if they included at least one of the 3 components at least 5 out of 7 per week on average, in 80% of home practice logs handed in. We considered adherence to activity and home practice to be good if these criteria were met in 70% of logs.
7. *Adherence to ActiGraph and Fitbit:* This was calculated by the proportion of participants who wore the ActiGraph at least 5 out of 7 days per week [35,36] for at least seven hours a day (≥80% excellent and ≥70% good) [23]. Similar criteria were used for wearing the Fitbit in the *GetActive-Fitbit* group.
8. *Adherence to Fitbit:* For the *GetActive-Fitbit* group, adherence to Fitbit was calculated by the proportion of participants who wore a charged Fitbit for 5 out of 7 days per week on average (≥80% excellent and ≥70% good).
9. *Program satisfaction:* This was assessed using the client satisfaction scale questionnaire (CSQ-3) [37], which includes 3 items capturing the degree to which the program met the participants' needs and their satisfaction from it. Satisfaction was considered excellent if the proportion of participants who scored above the scale midpoint was ≥75% and good if ≥70%.
10. *Program safety:* Safety was assessed via self-reported adverse events. Safety was considered excellent if there were no adverse events linked to program participation and good if there were minimal and mild adverse events (eg, muscle soreness) linked to program participation, which occurred in no more than 10% of participants.

Physical Function Assessments

We assessed objectively measured, performance-based, and self-reported physical function per the IMMPACT criteria [18,19] and the ICF guidelines [17].

Objectively Measured Physical Function

We measured the actual number of steps taken daily by each participant using the *wGT3X-BT ActiGraph accelerometer* device [38]. We asked participants to wear the ActiGraph during all waking hours for 7 days before and after the programs and calculated the participants' average daily steps. We used ActiLife software (ActiLife LLC) to store, clean, and analyze data using the settings used in our nonrandomized controlled trial [23]. Briefly, the ActiGraphs were set to record counts in 30-second epochs [39]. Nonwear time was defined as ≥90 consecutive minutes of 0 activity counts [40]. Up to 2 min of activity counts between 0 and 100 were allowed [41], and we ignored wear periods of <10 min. A staff member checked for valid wear time (7 hours per day) and manually checked each participant's recorded activity with self-reported wear times logged in their activity diaries to ensure consistency. We used a minimum clinically important difference (MCID) [42] of 800 ActiGraph measured steps [43].

Performance-Based Physical Function

We measured performance-based physical function via the 6-min walk test [44], which has an MCID of 54 m [44]. We recorded the distance in meters each participant covered by walking on a flat surface for 6 min.

Self-Reported Physical Function

We used 3 self-report measures recommended by IMMPACT [18] that assess different aspects of physical function targeted within our program. The *World Health Organization Disability Assessment Schedule (WHODAS) 2.0* is a 36-item questionnaire assessing difficulties in 6 main areas of function: cognition (understanding and communication), mobility (moving and getting around), self-care (hygiene, dressing, eating, and staying alone), getting along with others (interacting with other people), life activities (domestic responsibilities, leisure, and work), and participation (joining community activities) [45]. WHODAS 2.0 does not yet have an established MCID. The *Patient-Reported Outcomes Measurement Information System (PROMIS) physical function*, version 1.2.8b, is an 8-item questionnaire that assesses the ability to perform various physical tasks (self-care to complex tasks) [46]. Scores are expressed as T scores (mean 50, SD 10). The MCID is 5.48 [47]. The self-reported *physical activity scale for individuals with physical disabilities (PASIPD)* is a 13-item measure assessing engagement in leisure, household, and work-related physical activities [48]. Internal reliability in the current sample was excellent for the PROMIS physical function and WHODAS (Cronbach α=.94 and .97, respectively). For the PASIPD, reliability was acceptable (α=.62).

Emotional Function

We measured anxiety using the PROMIS anxiety scale [49] (*version 1.08a*, MCID=4.28) and depression using the PROMIS depression scale [50] (*version 1.08b*, MCID=5.19) [47,49,50]. These 8-item measures assess the frequency of anxiety and depression symptoms, respectively, over the past week on a 1-5 Likert scale. Scores are expressed as T scores with mean 50 (SD 10). Internal reliability was excellent for PROMIS anxiety (Cronbach α=.95) and depression (Cronbach α=.96).

Pain Intensity

We assessed pain intensity at rest and with activity using the *numerical rating scale* (NRS) [51,52]. This is a 0-10 scale with high scores depicting higher pain. The NRS has an MCID of 1 [53].

Pain-Related Coping

We assessed pain-related coping using the *pain catastrophizing scale* [54], which assesses hopelessness, helplessness, and magnification of pain; the *Tampa kinesiophobia scale* (MCID=6) [55], which assesses fear of experiencing pain during activity; and the *pain resilience scale* (no MCID) [56], which measures the ability to regulate emotions and engage in activities despite pain. Internal reliability was excellent for pain catastrophizing (Cronbach α=.94) and good for both kinesiophobia (Cronbach α=.87) and pain resilience (Cronbach α=.89).

Adaptive Coping and Mindfulness

We assessed adaptive coping skills using the *measure of current status* (MOCS-A) [57], a 13-item questionnaire measuring the ability to utilize healthy coping skills such as relaxation, awareness of tension, assertiveness, and coping confidence. We assessed mindfulness using the *cognitive and affective mindfulness scale-revised* (CAMS-R), a 12-item questionnaire measuring participants' ability to pay attention to the present moment in a nonjudgmental manner [58]. Internal reliability was good for both measures (Cronbach α=.89 for MOCS-A and Cronbach α=.85 for CAMS-R).

Patients' Perception of Improvement

We used the *modified patient global impression of change* (MPGIC) to assess participants' overall perception of improvement in key program areas [59]. Participants used a 1-7 Likert scale to rate how much they perceived the program to have improved their physical function, activity levels, pain, emotional function, pain resilience, and the degree to which the Fitbit helped them or not in increasing physical activity (only the *GetActive-Fitbit* group). Lower scores indicate *higher* perceived improvement.

Analysis Plan

We first assessed the sample characteristics using descriptive statistics. We then analyzed feasibility markers based on the proportion of participants who achieved each benchmark, as detailed earlier in the *Feasibility Assessments* section. For the rest of the quantitative assessments, we used descriptive statistics to characterize the sample and paired sample two-tailed *t* tests to assess within-group changes between baseline and postprogram. Cohen d was used to determine effect sizes using conventional standards (small effect sizes of 0.2 SD units, medium effect sizes of 0.5 SD units, and large effect sizes of 0.8 SD units) [60]. When available, we report clinical significance based on MCID and refer to comparisons with population norms. Consistent with recommendations for

analyses for pilot studies [61,62], we did not perform between-group analyses of efficacy.

Sample Size Consideration

In line with recommendations for pilot RCTs [63,64], the target sample size for this study was 80 enrolled participants to achieve 60 completers. This sample size is sufficient to determine feasibility and acceptability markers and is typical in randomized controlled feasibility trials [65-67].

Results

Sample Characteristics

A total of 265 participants were referred and assessed for eligibility, and 82 participants were randomized (41 in each group; Multimedia Appendix 2). The sample characteristics are summarized in Table 1. Participants were predominantly female (54/82, 66%), white (66/82, 80%), and non-Hispanic (72/82, 88%). Approximately half of all participants completed either 4 years of college (17/82, 21%) or obtained a graduate or professional degree (28/82, 34%).

Table 1. Demographic characteristics of participants.

Demographic characteristics	GetActive (n=41)	GetActive-Fitbit (n=41)
Age (years), mean (SD)	54.46 (14.5)	49.07 (14.2)
Gender, n (%)		
Male	18 (43.9)	10 (24.4)
Female	23 (56.1)	31 (75.6)
Ethnicity, n (%)		
Hispanic or Latino/Latina	4 (9.8)	4 (9.8)
Not Hispanic or Latino/Latina	35 (85.4)	37 (90.2)
Missing	2 (4.9)	N/A[a]
Race, n (%)		
American Indian/Alaskan Native	2 (4.9)	0 (0.0)
Asian	1 (2.4)	2 (4.9)
Black/African American	5 (12.2)	2 (4.9)
Native Hawaiian/Pacific Islander	0 (0.0)	0 (0.0)
White	32 (78.0)	34 (82.9)
More than one race	1 (2.4)	3 (7.3)
Marital status, n (%)		
Single, never married	13 (31.7)	15 (36.6)
Living with significant other	5 (12.2)	6 (14.6)
Married	15 (36.6)	8 (19.5)
Separated/divorced	5 (12.2)	11 (26.8)
Widowed	3 (7.3)	1 (2.4)
Education, n (%)		
High school (12 years)	5 (12.2)	6 (14.6)
Some college/associate degree (<16 years)	15 (36.6)	11 (26.8)
Completed college (16 years)	9 (22.0)	8 (19.5)
Graduate/professional degree (>16 years)	12 (29.3)	16 (39.0)

[a]N/A: data not applicable.

Feasibility and Acceptability Markers

Program acceptability, feasibility of quantitative measures, therapist adherence, patient adherence to activity and home practice, adherence to wearing the ActiGraph (as well as Fitbit for the *GetActive-Fitbit* program), and program satisfaction were good to excellent and similar in both groups (Table 2). Feasibility of recruitment was excellent, with 265 participants out of 307 (86.3%) successfully contacted agreeing to participate). 9 out of 10 multi-item measures had internal reliability (Cronbach α) >.70, with the mean reliability=0.89, SD 0.11. Of the 82 participants randomized, baseline ActiGraph data were obtained from 72 participants (68 with valid data: 31 in the *GetActive* program and 37 in the *GetActive-Fitbit*

program) and postprogram data from 60 participants (56 with valid data: 25 in the *GetActive* program and 31 in the *GetActive-Fitbit* program). The 6-min walk test data were obtained from 82 participants at baseline (41 in each group) and 61 postprogram (28 in the *GetActive* program and 33 in the *GetActive-Fitbit* program). Complete data from all self-reported measures were obtained from 81 participants at baseline (41 from the *GetActive* program and 40 in the *GetActive-Fitbit* program) and from 70 (34 from the *GetActive* program and 36 in the *GetActive-Fitbit* program) participants postprogram. The benchmark criteria are detailed in the *Feasibility Assessments* section.

Table 2. Feasibility and acceptability of the programs.

Outcomes[a]	GetActive	GetActive-Fitbit
Program acceptability	31 out of 41 participants (76%) attended ≥7 out of 10 group or make-up sessions: good	34 out of 41 participants (83%) attended ≥7 out of 10 group or make-up sessions: excellent
Credibility and expectancy	27 out of 41 participants (66%) scored above the scale midpoint for expectancy: acceptable; 37 out of 41 participants (90%) scored above the scale midpoint for credibility: excellent	38 out of 41 participants (93%) scored above the scale midpoint for credibility: excellent; 22 out of 41 participants (54%) scored above the scale midpoint for expectancy: acceptable
Therapist adherence to the manual	Rater agreement (κ)=98%, therapist adherence to the manual was 98%: good	Rater agreement (κ)=97%, therapist adherence to the manual was 94%: good
Feasibility of quantitative measures	41 out of 41 (100%) were not fully missing questionnaires on quantitative measures at baseline: excellent; 34 out of 35 (97%) were not fully missing questionnaires on quantitative measures at posttest: excellent	40 out of 41 (97.56%) were not fully missing questionnaires on quantitative measures at baseline: excellent; 37 out of 37 (100%) were not fully missing questionnaires on quantitative measures at posttest: excellent
Adherence to homework	98% of logs handed in met adherence criteria (ie, 3 home practice components completed 3 out of 7 days per week or 1 component completed 5 out of 7 days per week): excellent	98% of logs handed in met adherence criteria (ie, 3 home practice components completed 3 out of 7 days per week or 1 component completed 5 out of 7 days per week): excellent
Adherence to ActiGraphs and Fitbit	31 out of 33 participants (94%) who received the ActiGraph at baseline wore it for ≥5 out of 7 days: excellent; 25 out of 27 participants (93%) who received the ActiGraph at posttest wore it for 5 out of 7 days: excellent; 30 out of 33 (91%) participants who received the ActiGraph at baseline had at least 5 out of 7 valid days (minimum of 7 wear hours): excellent; 19 out of 27 (70%) participants who received the ActiGraph at posttest had at least 5 out of 7 days (minimum of 7 wear hours): good	36 out of 39 participants (92%) who received the ActiGraph at baseline wore it for ≥5 out of 7 days: excellent; 29 out of 33 participants (88%) who received the ActiGraph at posttest wore it for 5 out of 7 days: excellent; 28 out of 39 participants (72%) who received the ActiGraph at baseline had at least 5 out of 7 valid (minimum of 7 wear hours): good; 25 out of 33 participants (76%) who received the ActiGraph posttest had at least 5 out of 7 valid (minimum of 7 wear hours): good; 34 out of 41 participants (83%) wore the Fitbit for at least 5 out of 7 days for 8 out of 10 weeks of the program: good
Client satisfaction	34 out of 34 participants (100%) scored above the scale midpoint: excellent	35 out of 36 participants (97%) scored above the scale midpoint: excellent
Program safety and adverse events	3 participants were hospitalized for reasons unrelated to the program (1 for a lung infection, 1 for a pain flare, and 1 for a stroke); 1 patient reported Sciatica: excellent	7 participants were hospitalized for reasons unrelated to the program (1 for sickle cell anemia flare-up; 1 for diverticulitis attack; 1 for chemotherapy; 1 for a pain flare; 1 for unknown reasons; 1 for falling; and 1 admitted twice for elevated heartrate, and then hurting leg); 1 patient reported a sprained quadricep muscle: good

[a]Feasibility of recruitment and internal-reliability portion of the feasibility of quantitative measures are reported in the text for the entire sample.

Physical Function

Objective Measured Physical Function

As a whole, neither the *GetActive* group (–191 steps; *P*=.68; effect size of 0.09 SD units) nor the *GetActive-Fitbit* group (+8 steps; *P*=.98; effect size of 0 SD units) exhibited increases in step count from pre- to postprogram (Table 3). A total of 48% (11/23) of participants in the *GetActive* group and 53% (16/30) of participants in the *GetActive-Fitbit* group exhibited a higher step count following the program compared with baseline, with mean improvements exceeding the MCID of 800 steps (+1341 steps for the *GetActive* group and +1441 for the *GetActive-Fitbit* group). There was wide variability in the baseline step count across participants (mean=5432 SD~2942 steps; range 1197-13,643 steps). Although our goal was to enroll individuals who were inactive, 47% of individuals in our sample recorded >5000 steps at baseline (48% in the *GetActive* group and 46% in the *GetActive-Fitbit* group). Within the subgroup of participants who were sedentary (ie, recorded <5000 steps at baseline), the mean change in number of steps was above the MCID in both the *GetActive* and *GetActive-Fitbit* groups (+874 steps; *P*=.25; effect size of 0.39 SD units for the *GetActive* group and +867 steps; *P*=.03; effect size of 0.56 SD units for the *GetActive-Fitbit* group). In all, 50% of the sedentary individuals in the *GetActive* group and 47% in the

GetActive-Fitbit group increased the number of steps over MCID.

Performance-Based Physical Function

Participants in both the *GetActive* group (+41 m; $P<.001$; effect size of 0.99 SD units) and *GetActive-Fitbit* group (+50 m; $P<.001$; effect size of 0.85 SD units) improved significantly with large effect sizes on the 6-min walk test, and these were just at the cusp of the MCID for the *GetActive-Fitbit* group.

Self-Reported Physical Function

Both groups exhibited significant improvements with medium effect size in both the WHODAS ($P=.001$; effect size of 0.62 SD units for the *GetActive* group and $P=.03$; effect size of 0.38 SD units for the *GetActive-Fitbit* group) and PROMIS physical function ($P=.01$; effect size of 0.49 SD units for the *GetActive* group and $P=.02$; effect size of 0.40 SD units for the *GetActive-Fitbit* group). Differences in the PROMIS measure (for which the MCID is available) did not reach the MCID. Groups exhibited no change in self-reported physical activity measured by the PASIPD ($P=.32$; effect size of 0.17 SD units for the *GetActive* group and $P=.56$; effect size of 0.09 SD units for the *GetActive-Fitbit* group).

Emotional Function

The *GetActive* group exhibited nonsignificant improvements in small effect size for the *PROMIS depression* ($P=.11$; effect size of 0.27 SD units) and *PROMIS anxiety* ($P=.08$; effect size of 0.30 SD units). The *GetActive-Fitbit* group exhibited significant improvements in medium effect sizes in the *PROMIS depression* ($P=.003$; effect size of 0.54 SD units) and *PROMIS anxiety* ($P=.02$; effect size of 0.40 SD units).

Pain Intensity

Both groups exhibited clinically meaningful and significant reductions in medium effect size in pain at rest ($P=.007$; effect size of 0.49 SD units for the *GetActive* group and $P=.001$; effect size of 0.58 SD units for the *GetActive-Fitbit* group) and with activity ($P<.001$; effect size of 0.70 SD units for the *GetActive* group and $P=.001$; effect size of 0.59 SD units for the *GetActive-Fitbit* group).

Pain-Related Coping

Both groups showed significant and medium-sized effects for improvements in pain catastrophizing ($P<.001$; effect size of 0.72 SD units for the *GetActive* group and $P=.01$; effect size of 0.43 SD units for the *GetActive-Fitbit* group) and kinesiophobia ($P<.001$; effect size of 0.74 SD units for the *GetActive* group and $P=.001$; effect size of 0.61 SD units for the *GetActive-Fitbit* group). The *GetActive* group exhibited significant medium-sized improvements in pain resilience ($P=.001$; effect size of 0.64 SD units), whereas improvements in the *GetActive-Fitbit* group had a small effect size and did not reach significance ($P=.09$; effect size of 0.28 SD units).

Adaptive Coping and Mindfulness

Both groups exhibited significant medium-sized effects for improvements in mindfulness ($P=.002$; effect size of 0.58 SD units for the *GetActive* group and $P=.01$; effect size of 0.44 SD units for the *GetActive-Fitbit* group) and significant medium-to-large effects for improvement in adaptive coping ($P<.001$; effect size of 0.76 SD units for the *GetActive* group and $P<.001$; effect size of 0.83 SD units for the *GetActive-Fitbit* group).

Patients' Perception of Improvement

Perceived improvement in pain, physical activity, and physical and emotional function were high for both programs. Improvement due to Fitbit was high in the *GetActive-Fitbit* group (Table 4).

Table 3. Outcome measures.

Measures	GetActive						GetActive-Fitbit					
	Baseline, mean (SD)	Posttest, mean (SD)	Pretest-posttest change, mean (95% CI)	*t* test value (df)	*P* value	Cohen *d*	Baseline, mean (SD)	Posttest, mean (SD)	Pretest-posttest change, mean (95% CI)	*t* test value (df)	*P* value	Cohen *d*
ActiGraph average steps	6330.57 (3556.10)	6139.25 (3424.39)	191.32 (−758.56 to 1141.19)	0.42 (22)	.68	0.088	5447.55 (2597.42)	5455.48 (2320.37)	−7.93 (−705.02 to 689.16)	−0.023 (29)	.98	−0.00
6-min walk test distance (m)	358.393 (85.11)	399.14 (77.19)	−40.75 (−56.71 to −24.79)	−5.24 (27)	<.001	−0.99	343.83 (77.56)	393.58 (71.88)	−49.74 (−70.53 to −28.95)	−4.87 (32)	<.001	−0.85
Physical function (PROMIS[a])	40.55 (6.63)	42.97 (7.46)	−2.41 (−4.10 to −0.72)	−2.90 (34)	.006	−0.49	38.10 (7.86)	40.76 (8.41)	−2.65 (−4.86 to −0.44)	−2.44 (36)	.02	−0.40
Physical function (WHO-DAS[b])	25.02 (15.52)	17.45 (13.72)	7.56 (3.39 to 11.73)	3.684 (34)	.001	0.62	34.65 (21.02)	27.13 (16.70)	7.52 (0.96 to 14.09)	2.32 (36)	.03	0.38
Self-reported physical activity	9.60 (7.04)	11.15 (8.51)	−1.55 (−4.64 to 1.55)	−1.02 (34)	.32	−0.17	15.37 (18.65)	17.13 (14.93)	−1.76 (−7.89 to 4.37)	−0.583 (36)	.56	−0.09
Depression	51.63 (10.34)	49.34 (9.29)	2.29 (−0.58 to 5.16)	1.62 (34)	.11	0.27	59.47 (9.53)	55.60 (7.72)	3.87 (1.45 to 6.28)	3.25 (35)	.003	0.54
Anxiety	52.48 (10.14)	50.81 (8.52)	2.87 (−0.40 to 6.14)	1.78 (34)	.08	0.30	58.87 (9.07)	56.12 (9.50)	2.95 (0.49 to 5.42)	2.43 (36)	.02	0.40
Pain at rest	4.77 (2.51)	3.57 (2.5)	1.2 (0.35 to 2.04)	2.89 (34)	.007	0.49	5.78 (2.18)	4.56 (2.16)	1.22 (0.51 to 1.94)	3.47 (35)	.001	0.58
Pain with activity	6.63 (1.99)	4.97 (2.54)	1.66 (0.85 to 2.46)	4.17 (34)	<.001	0.70	7.24 (2.23)	6.08 (2.55)	1.16 (0.51 to 1.81)	3.62 (36)	.001	0.59
Pain resilience	37.57 (9.65)	43.49 (9.99)	−5.91 (−9.10 to −2.73)	−3.77 (34)	.001	−0.64	33.16 (10.07)	36.46 (11.11)	−3.30 (−7.24 to 0.64)	−1.70 (36)	.09	−0.28
Pain catastrophizing	18.34 (11.06)	9.77 (7.66)	8.57 (4.47 to 12.67)	4.24 (34)	<.001	0.72	21.41 (11.94)	16.36 (10.84)	5.03 (1.13 to 8.93)	2.61 (36)	.01	0.43
Kinesiophobia	37.23 (7.72)	30.76 (7.36)	6.46 (3.41 to 9.52)	4.30 (33)	<.001	0.74	38.86 (8.87)	34.78 (8.25)	4.08 (1.84 to 6.32)	3.70 (36)	.001	0.61
Adaptive coping	28.49 (1.56)	36.37 (1.44)	−7.89 (−11.43 to −4.34)	−4.52 (34)	<.001	0.76	27.08 (1.65)	34 (1.38)	−6.92 (−9.69 to −4.15)	−5.07 (36)	<.001	0.83
Mindfulness	32.57 (6.58)	36.31 (6.48)	−3.74 (−5.96 to −1.53)	−3.43 (34)	.002	−0.58	30.43 (7.02)	33.14 (6.18)	−2.70 (−4.76 to −0.64)	−2.66 (36)	.012	−0.44

[a]PROMIS: Patient-Reported Outcomes Measurement Information System.

[b]WHODAS: World Health Organization Disability Assessment Schedule.

Table 4. Perceived improvement of change (1=very much improved and 7=very much worse).

Group	Impression of change in pain	Impression of change in physical activity	Impression of change in physical function	Impression of change in emotional function	Impression of change in resiliency	Impression of change from use of Fitbit
GetActive, mean (SD)	2.51 (1.17)	2.31 (1.18)	2.57 (1.19)	2.40 (1.01)	2.14 (0.91)	N/A[a]
GetActive-Fitbit, mean (SD)	2.70 (1.20)	2.30 (1.10)	2.78 (0.95)	2.54 (0.84)	2.32 (1.03)	1.86 (0.95)

[a]N/A: not applicable.

Discussion

Principal Findings

Previous intervention research in chronic pain has not comprehensively assessed improvement in physical function and has yielded modest improvements in self-reported emotional and physical function. To address this problem, we used evidence-based frameworks for intervention development and adaptation [21,22] and recent recommendations for assessment of physical function in chronic pain clinical trials [17-19] to iteratively develop *GetActiveand GetActive-Fitbit*, 2 identical mind-body physical activity programs (8 sessions each) aimed at improving physical (self-reported, performance based, and objectively measured) and emotional function among patients with heterogeneous chronic pain by teaching them pain-specific and mind-body skills, and to gradually increase their activity in a manner noncontingent on pain levels [23]. In this study, we report on a feasibility RCT of the final iterations of these programs (10 sessions each) necessary before a future RCT to determine the efficacy of the 2 programs compared with a control and usefulness of the Fitbit in improving activity.

Consistent with our first hypothesis and the goals of both the ORBIT and NCCIH intervention development stage models, both programs met or exceeded the a-priori set feasibility benchmarks. Retention was considerably higher than other mind-body trials for chronic pain [68]. The establishment of these feasibility benchmarks is critical before efficacy testing to ensure the scientific rigor of the future efficacy trial, per the ORBIT and NCCIH stage models of intervention development. Leaping to efficacy testing before establishing such feasibility markers may have numerous, substantial, and negative unintended consequences, including having insufficient power to detect change, inadequate fit to the target population, and lack of identification or inclusion of those who are most likely to be most responsive to the intervention [69,70]. These results indicate that *GetActive* and *GetActive-Fitbit* are ideally poised for efficacy testing.

Consistent with our second hypothesis, we found that participation in both programs is associated with improvement in physical function. As this is not an efficacy trial, we limited analyses to effect sizes for within-group changes between pretest and posttest and refrained from between-group comparisons [61]. Both groups exhibited significant and large effect sizes for improvements in performance-based physical function (6-min walk test) and significant medium effect sizes in improvements in self-reported physical function. For objectively measured physical function assessed by ActiGraph, as a group, participants did not exhibit step count increases following the programs. However, individuals who were sedentary at baseline increased their step count over the MCID. Although any statements about efficacy are spurious with this small sample, the results suggest that sedentary individuals may be more likely to meaningfully increase step count compared with their nonsedentary counterparts. Indeed, only 2 individuals in the *GetActive* group and 1 in the *GetActive-Fitbit* group, who had step counts higher than 5000 at baseline, exhibited clinically meaningful improvements at posttest. It is also possible that the results were affected by weather variations between the 2 assessment periods, as we ran groups in early fall or winter, and posttreatment assessment was conducted during worse weather than baseline. Although 1 week is the recommended assessment time, a longer monitoring assessment may better capture inherent weather variations. In addition, a clear plan for maintaining activity during bad weather with identification of specific places to walk (eg, gym and mall) should be emphasized in the future efficacy trial. Participants in both groups exhibited medium effect sizes for improvements in emotional function, and medium-to-large effect sizes for improvements in pain intensity (above the MCID for changes in both pain at rest and during activity), pain catastrophizing, fear of pain due to movement, mindfulness, and adaptive coping. Importantly, participants rated their perceived improvement on our main outcomes as *much improved*.

Limitations and Strengths

This study has several strengths. First, we used evidence-based frameworks and mixed methods to iteratively develop our mind-body program and refine our methodology. The study utilized strong scientific rigor and an RCT design that minimized the risks of selection bias and confounding. The emphasis on feasibility markers in preparation for efficacy is an additional strength of the study, which helps ensure the scientific rigor of the next step of the efficacy trial. Second, this study is the first RCT that follows recommendations from recent IMMPACT [18,19] and ICF [17] criteria to comprehensively assess physical function using objective, performance-based, and self-reported measures in chronic pain trials [71]. Finally, this study found improvements in physical function [71], pain coping [72,73], and pain intensity [11,73], which are similar to or larger than those found in other mind-body interventions, suggesting that our mind-body physical activity program shows strong potential for efficacy. Certain limitations should also be considered. First, our screening criteria for level of activity allowed many participants with relatively high baseline ActiGraph step counts to enroll (Table 3). These relatively active participants seem to have benefitted less in terms of ActiGraph measured step gains,

although they did benefit from improvement in other aspects, including the 6-min walk test, self-reported physical function, and other intervention targets. Furthermore, although patients were instructed not to make changes in their activity during the baseline assessment, several shared in the group sessions that they did push themselves to be more active because they were excited about participation. Providing patients with a clearer rationale on why it is important to maintain their regular activity during the baseline assessment will be important to accurately capture objective activity. The last and largest cohort in our study experienced continuous rain during the postprogram ActiGraph assessment period, and all participants exhibited no or limited increase in step count, which may have skewed results. Fitbit step count data from *GetActive-Fitbit* further support this; Fitbit step counts have significantly increased above the MCID between week 1 and the final week of the program (+~1500 steps; *P*=.01; effect size of 0.52 SD units), but less so between week 1 and the week after program completion, when ActiGraphs were worn (~843 steps; *P*=.26; effect size of 0.23 SD units). In addition, 35% of the participants did not wear their ActiGraph at both baseline and posttest.

Conclusions

The results of this study provide strong evidence that 2 novel mind-body and physical activity programs for patients with chronic pain are feasible, acceptable, credible, and yield high satisfaction. Furthermore, the programs show potential for improvement in physical function, pain and related coping, and other psychosocial variables. This study supports future testing of *GetActive* and *GetActive-Fitbit* as well as an educational control group in a fully powered 3-arm RCT to establish the efficacy of the programs in targeting physical and emotional function and determine whether reinforcing physical activity with a wearable digital monitoring device is beneficial in patients with chronic pain. This study provides a model for successfully following the IMMPACT criteria for a comprehensive assessment of physical function and following evidence-based models to maximize feasibility before formal efficacy testing.

Acknowledgments

This study was funded by an R34 grant from the National Center for Complementary and Integrative Health (1R34AT009356-01A1) to the senior author.

Conflicts of Interest

EM served as a Data Safety Monitoring Board member for Novartis Pharmaceuticals and Shire Human Genetic Therapies; served on an advisory committee for Biogen; consulted for Cerevance, Intrance, Inventram, Lavin Consulting, and Myolex; and his institution received grants on his behalf from Amylyx Pharmaceuticals, GlaxoSmithKline, and Mitsubishi Tanabe Pharmaceuticals. All other authors declare that they have no relevant conflicts of interest.

Multimedia Appendix 1
GetActive and GetActive-Fitbit program development.
[DOCX File , 69 KB - formative_v4i6e18703_app1.docx]

Multimedia Appendix 2
Participant flow.
[DOCX File , 42 KB - formative_v4i6e18703_app2.docx]

Multimedia Appendix 3
Program content.
[DOCX File , 13 KB - formative_v4i6e18703_app3.docx]

Multimedia Appendix 4
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 330 KB - formative_v4i6e18703_app4.pdf]

References

1. Schappert SM. Ambulatory care visits to physician offices, hospital outpatient departments, and emergency departments: United States, 1996. Vital Health Stat 13 1998 Feb(134):1-37 [FREE Full text] [Medline: 9564280]
2. Vranceanu A, Bachoura A, Weening A, Vrahas M, Smith RM, Ring D. Psychological factors predict disability and pain intensity after skeletal trauma. J Bone Joint Surg Am 2014 Feb 5;96(3):e20. [doi: 10.2106/JBJS.L.00479] [Medline: 24500592]
3. Kamper S, Apeldoorn A, Chiarotto A, Smeets R, Ostelo R, Guzman J, et al. Multidisciplinary biopsychosocial rehabilitation for chronic low back pain. Cochrane Database Syst Rev 2014 Sep 2(9):CD000963. [doi: 10.1002/14651858.CD000963.pub3] [Medline: 25180773]

4. Lungu E, Vendittoli P, Desmeules F. Preoperative determinants of patient-reported pain and physical function levels following total knee arthroplasty: a systematic review. Open Orthop J 2016;10:213-231 [FREE Full text] [doi: 10.2174/1874325001610010213] [Medline: 27398109]
5. Bair MJ, Robinson RL, Katon W, Kroenke K. Depression and pain comorbidity: a literature review. Arch Intern Med 2003 Nov 10;163(20):2433-2445. [doi: 10.1001/archinte.163.20.2433] [Medline: 14609780]
6. Dowell D, Haegerich TM, Chou R. CDC guideline for prescribing opioids for chronic pain--United States, 2016. J Am Med Assoc 2016 Apr 19;315(15):1624-1645 [FREE Full text] [doi: 10.1001/jama.2016.1464] [Medline: 26977696]
7. Ehde DM, Dillworth TM, Turner JA. Cognitive-behavioral therapy for individuals with chronic pain: efficacy, innovations, and directions for research. Am Psychol 2014;69(2):153-166. [doi: 10.1037/a0035747] [Medline: 24547801]
8. Wetherell JL, Afari N, Rutledge T, Sorrell JT, Stoddard JA, Petkus AJ, et al. A randomized, controlled trial of acceptance and commitment therapy and cognitive-behavioral therapy for chronic pain. Pain 2011 Sep;152(9):2098-2107. [doi: 10.1016/j.pain.2011.05.016] [Medline: 21683527]
9. Hilton L, Hempel S, Ewing BA, Apaydin E, Xenakis L, Newberry S, et al. Mindfulness meditation for chronic pain: systematic review and meta-analysis. Ann Behav Med 2017 Apr;51(2):199-213 [FREE Full text] [doi: 10.1007/s12160-016-9844-2] [Medline: 27658913]
10. O'Connor SR, Tully MA, Ryan B, Bleakley CM, Baxter GD, Bradley JM, et al. Walking exercise for chronic musculoskeletal pain: systematic review and meta-analysis. Arch Phys Med Rehabil 2015 Apr;96(4):724-34.e3. [doi: 10.1016/j.apmr.2014.12.003] [Medline: 25529265]
11. Anheyer D, Haller H, Barth J, Lauche R, Dobos G, Cramer H. Mindfulness-based stress reduction for treating low back pain: a systematic review and meta-analysis. Ann Intern Med 2017 Jun 6;166(11):799-807. [doi: 10.7326/M16-1997] [Medline: 28437793]
12. Ball EF, Sharizan E, Franklin G, Rogozińska E. Does mindfulness meditation improve chronic pain? A systematic review. Curr Opin Obstet Gynecol 2017 Dec;29(6):359-366. [doi: 10.1097/GCO.0000000000000417] [Medline: 28961631]
13. Garland EL, Brintz CE, Hanley AW, Roseen EJ, Atchley RM, Gaylord SA, et al. Mind-body therapies for opioid-treated pain: a systematic review and meta-analysis. JAMA Intern Med 2019 Nov 4 epub ahead of print. [doi: 10.1001/jamainternmed.2019.4917] [Medline: 31682676]
14. Hurwitz EL, Morgenstern H, Chiao C. Effects of recreational physical activity and back exercises on low back pain and psychological distress: findings from the UCLA low back pain study. Am J Public Health 2005 Oct;95(10):1817-1824. [doi: 10.2105/AJPH.2004.052993] [Medline: 16186460]
15. Kroll HR. Exercise therapy for chronic pain. Phys Med Rehabil Clin N Am 2015 May;26(2):263-281. [doi: 10.1016/j.pmr.2014.12.007] [Medline: 25952064]
16. Geneen LJ, Moore RA, Clarke C, Martin D, Colvin LA, Smith BH. Physical activity and exercise for chronic pain in adults: an overview of Cochrane reviews. Cochrane Database Syst Rev 2017 Jan 14;1:CD011279 [FREE Full text] [doi: 10.1002/14651858.CD011279.pub2] [Medline: 28087891]
17. World Health Organization. International Classification of Functioning, Disability and Health: ICF. Geneva, Switzerland: World Health Organization; 2001.
18. Taylor AM, Phillips K, Patel KV, Turk DC, Dworkin RH, Beaton D, et al. Assessment of physical function and participation in chronic pain clinical trials: IMMPACT/OMERACT recommendations. Pain 2016 Sep;157(9):1836-1850. [doi: 10.1097/j.pain.0000000000000577] [Medline: 27058676]
19. Gewandter J, Dworkin R, Turk D, Devine E, Hewitt D, Jensen M, et al. Improving study conduct and data quality in clinical trials of chronic pain treatments: IMMPACT recommendations. J Pain 2019 Dec epub ahead of print. [doi: 10.1016/j.jpain.2019.12.003] [Medline: 31843583]
20. Onken LS, Carroll KM, Shoham V, Cuthbert BN, Riddle M. Reenvisioning clinical science: unifying the discipline to improve the public health. Clin Psychol Sci 2014 Jan 1;2(1):22-34 [FREE Full text] [doi: 10.1177/2167702613497932] [Medline: 25821658]
21. Czajkowski SM, Powell LH, Adler N, Naar-King S, Reynolds KD, Hunter CM, et al. From ideas to efficacy: the ORBIT model for developing behavioral treatments for chronic diseases. Health Psychol 2015 Oct;34(10):971-982 [FREE Full text] [doi: 10.1037/hea0000161] [Medline: 25642841]
22. National Center for Complementary and Integrative Health. 2017. Framework for Developing and Testing Mind and Body Interventions URL: https://nccih.nih.gov/grants/mindbody/framework [accessed 2019-06-27]
23. Greenberg J, Lin A, Zale E, Kulich R, James P, Millstein R, et al. Development and early feasibility testing of a mind-body physical activity program for patients with heterogeneous chronic pain; the GetActive study. J Pain Res 2019;12:3279-3297 [FREE Full text] [doi: 10.2147/JPR.S222448] [Medline: 31849515]
24. ATS Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories. ATS statement: guidelines for the six-minute walk test. Am J Respir Crit Care Med 2002 Jul 1;166(1):111-117. [doi: 10.1164/ajrccm.166.1.at1102] [Medline: 12091180]
25. Cane D, McCarthy M, Mazmanian D. Obstacles to activity pacing: assessment, relationship to activity and functioning. Pain 2016 Jul;157(7):1508-1514. [doi: 10.1097/j.pain.0000000000000553] [Medline: 26963845]

26. Gordon R, Bloxham S. Influence of the Fitbit Charge HR on physical activity, aerobic fitness and disability in non-specific back pain participants. J Sports Med Phys Fitness 2017 Dec;57(12):1669-1675. [doi: 10.23736/S0022-4707.17.06688-9] [Medline: 28181773]
27. Amorim AB, Pappas E, Simic M, Ferreira ML, Jennings M, Tiedemann A, et al. Integrating mobile-health, health coaching, and physical activity to reduce the burden of chronic low back pain trial (IMPACT): a pilot randomised controlled trial. BMC Musculoskelet Disord 2019 Feb 11;20(1):71 [FREE Full text] [doi: 10.1186/s12891-019-2454-y] [Medline: 30744606]
28. Rounsaville B, Carroll K, Onken L. A stage model of behavioral therapies research: getting started moving on from stage I. Clin Psychol Sci Pract 2006;8(2):133-142. [doi: 10.1093/clipsy.8.2.133]
29. Park ER, Traeger L, Vranceanu A, Scult M, Lerner JA, Benson H, et al. The development of a patient-centered program based on the relaxation response: the relaxation response resiliency program (3RP). Psychosomatics 2013;54(2):165-174. [doi: 10.1016/j.psym.2012.09.001] [Medline: 23352048]
30. Fitabase - Research Device Data and Analytics. URL: https://www.fitabase.com/ [accessed 2019-07-02]
31. Mendoza JA, Baker KS, Moreno MA, Whitlock K, Abbey-Lambertz M, Waite A, et al. A Fitbit and Facebook mhealth intervention for promoting physical activity among adolescent and young adult childhood cancer survivors: a pilot study. Pediatr Blood Cancer 2017 Dec;64(12). [doi: 10.1002/pbc.26660] [Medline: 28618158]
32. Hayes LB, van Camp CM. Increasing physical activity of children during school recess. J Appl Behav Anal 2015 Sep;48(3):690-695. [doi: 10.1002/jaba.222] [Medline: 26119136]
33. Devilly GJ, Borkovec TD. Psychometric properties of the credibility/expectancy questionnaire. J Behav Ther Exp Psychiatry 2000 Jun;31(2):73-86. [doi: 10.1016/s0005-7916(00)00012-4] [Medline: 11132119]
34. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med (Zagreb) 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
35. Winckers AN, Mackenbach JD, Compernolle S, Nicolaou M, van der Ploeg HP, de Bourdeaudhuij I, et al. Educational differences in the validity of self-reported physical activity. BMC Public Health 2015 Dec 26;15:1299 [FREE Full text] [doi: 10.1186/s12889-015-2656-7] [Medline: 26856811]
36. Carl J, Grüne E, Popp J, Pfeifer K. Physical activity promotion for apprentices in nursing care and automotive mechatronics-competence counts more than volume. Int J Environ Res Public Health 2020 Jan 28;17(3):793 [FREE Full text] [doi: 10.3390/ijerph17030793] [Medline: 32012835]
37. Attkisson CC, Zwick R. The client satisfaction questionnaire. Psychometric properties and correlations with service utilization and psychotherapy outcome. Eval Program Plann 1982;5(3):233-237. [doi: 10.1016/0149-7189(82)90074-x] [Medline: 10259963]
38. Cain KL, Conway TL, Adams MA, Husak LE, Sallis JF. Comparison of older and newer generations of ActiGraph accelerometers with the normal filter and the low frequency extension. Int J Behav Nutr Phys Act 2013 Apr 25;10:51 [FREE Full text] [doi: 10.1186/1479-5868-10-51] [Medline: 23618461]
39. Knuth ND, Chen KM, Schrack JA. Objectively measured physical activity varies by task and accelerometer location in younger and older adults. Med Sci Sport Exerc 2016;48:1061. [doi: 10.1249/01.mss.0000488192.95209.a5]
40. Evenson KR, Wen F. Performance of the ActiGraph accelerometer using a national population-based sample of youth and adults. BMC Res Notes 2015 Jan 17;8:7 [FREE Full text] [doi: 10.1186/s13104-014-0970-2] [Medline: 25595702]
41. Aadland E, Ylvisåker E. Reliability of the actigraph GT3X+ accelerometer in adults under free-living conditions. PLoS One 2015;10(8):e0134606 [FREE Full text] [doi: 10.1371/journal.pone.0134606] [Medline: 26274586]
42. Copay AG, Subach BR, Glassman SD, Polly DW, Schuler TC. Understanding the minimum clinically important difference: a review of concepts and methods. Spine J 2007;7(5):541-546. [doi: 10.1016/j.spinee.2007.01.008] [Medline: 17448732]
43. Motl RW, Pilutti LA, Learmonth YC, Goldman MD, Brown T. Clinical importance of steps taken per day among persons with multiple sclerosis. PLoS One 2013;8(9):e73247 [FREE Full text] [doi: 10.1371/journal.pone.0073247] [Medline: 24023843]
44. Redelmeier DA, Bayoumi AM, Goldstein RS, Guyatt GH. Interpreting small differences in functional status: the six minute walk test in chronic lung disease patients. Am J Respir Crit Care Med 1997 Apr;155(4):1278-1282. [doi: 10.1164/ajrccm.155.4.9105067] [Medline: 9105067]
45. World Health Organization. Measuring Health and Disability: Manual for WHO Disability Assessment Schedule WHODAS 2.0. Geneva, Switzerland: World Health Organization; 2010.
46. Stone AA, Broderick JE, Junghaenel DU, Schneider S, Schwartz JE. PROMIS fatigue, pain intensity, pain interference, pain behavior, physical function, depression, anxiety, and anger scales demonstrate ecological validity. J Clin Epidemiol 2016 Jun;74:194-206. [doi: 10.1016/j.jclinepi.2015.08.029] [Medline: 26628334]
47. Selvarajah S, Neuman BJ, Skolasky RL. Using PROMIS health domains to identify clinically meaningful change in lumbar degenerative spine disease: concurrent validity and responsiveness. Spine J 2016 Oct;16(10):S369-S370. [doi: 10.1016/j.spinee.2016.07.304]
48. Washburn RA, Zhu W, McAuley E, Frogley M, Figoni SF. The physical activity scale for individuals with physical disabilities: development and evaluation. Arch Phys Med Rehabil 2002 Feb;83(2):193-200. [doi: 10.1053/apmr.2002.27467] [Medline: 11833022]

49. Outcomes Measurement Information System. 2015. A Brief Guide to the PROMIS Anxiety Instruments URL: http://www.healthmeasures.net/images/PROMIS/manuals/PROMIS_Anxiety_Scoring_Manual.pdf [accessed 2018-10-29]
50. The Patient-Reported Outcomes Measurement Information System. 2015. A Brief Guide to the PROMIS Depression Instruments URL: http://www.healthmeasures.net/images/PROMIS/manuals/PROMIS_Depression_Scoring_Manual.pdf [accessed 2018-10-29]
51. Rodriguez CS. Pain measurement in the elderly: a review. Pain Manag Nurs 2001 Jun;2(2):38-46. [doi: 10.1053/jpmn.2001.23746] [Medline: 11706769]
52. Farrar JT, Young JP, LaMoreaux L, Werth JL, Poole RM. Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. Pain 2001 Nov;94(2):149-158. [doi: 10.1016/s0304-3959(01)00349-9] [Medline: 11690728]
53. Salaffi F, Stancati A, Silvestri CA, Ciapetti A, Grassi W. Minimal clinically important changes in chronic musculoskeletal pain intensity measured on a numerical rating scale. Eur J Pain 2004 Aug;8(4):283-291. [doi: 10.1016/j.ejpain.2003.09.004] [Medline: 15207508]
54. Sullivan ML, Bishop SR, Pivik J. The pain catastrophizing scale: development and validation. Psychol Assess 1995;7(4):524-532. [doi: 10.1037//1040-3590.7.4.524]
55. Monticone M, Ambrosini E, Rocca B, Foti C, Ferrante S. Responsiveness and minimal clinically important changes for the Tampa Scale of Kinesiophobia after lumbar fusion during cognitive behavioral rehabilitation. Eur J Phys Rehabil Med 2017 Jun;53(3):351-358 [FREE Full text] [doi: 10.23736/S1973-9087.16.04362-8] [Medline: 27827516]
56. Slepian PM, Ankawi B, Himawan LK, France CR. Development and initial validation of the pain resilience scale. J Pain 2016 Apr;17(4):462-472. [doi: 10.1016/j.jpain.2015.12.010] [Medline: 26748044]
57. Carver CS. Measurement Instrument Database for the Social Sciences. 2006. Measure of Current Status (MOCS) URL: https://www.midss.org/content/measure-current-status-mocs [accessed 2019-06-19]
58. Feldman G, Hayes A, Kumar S, Greeson J, Laurenceau JP. Mindfulness and emotion regulation: the development and initial validation of the cognitive and affective mindfulness scale-revised (CAMS-R). J Psychopathol Behav Assess 2006 Nov 7;29(3):177-190. [doi: 10.1007/s10862-006-9035-8]
59. Geisser ME, Clauw DJ, Strand V, Gendreau RM, Palmer R, Williams DA. Contributions of change in clinical status parameters to patient global impression of change (PGIC) scores among persons with fibromyalgia treated with milnacipran. Pain 2010 May;149(2):373-378. [doi: 10.1016/j.pain.2010.02.043] [Medline: 20332060]
60. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Earlbaum Associates; 1988.
61. Leon AC, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. J Psychiatr Res 2011 May;45(5):626-629 [FREE Full text] [doi: 10.1016/j.jpsychires.2010.10.008] [Medline: 21035130]
62. Kistin C, Silverstein M. Pilot studies: a critical but potentially misused component of interventional research. J Am Med Assoc 2015 Oct 20;314(15):1561-1562 [FREE Full text] [doi: 10.1001/jama.2015.10962] [Medline: 26501530]
63. Browne RH. On the use of a pilot sample for sample size determination. Stat Med 1995 Sep 15;14(17):1933-1940. [doi: 10.1002/sim.4780141709] [Medline: 8532986]
64. Shih WJ, Ohman-Strickland PA, Lin Y. Analysis of pilot and early phase studies with small sample sizes. Stat Med 2004 Jun 30;23(12):1827-1842. [doi: 10.1002/sim.1807] [Medline: 15195318]
65. Morone NE, Rollman BL, Moore CG, Li Q, Weiner DK. A mind-body program for older adults with chronic low back pain: results of a pilot study. Pain Med 2009 Nov;10(8):1395-1407 [FREE Full text] [doi: 10.1111/j.1526-4637.2009.00746.x] [Medline: 20021599]
66. Koszycki D, Benger M, Shlik J, Bradwejn J. Randomized trial of a meditation-based stress reduction program and cognitive behavior therapy in generalized social anxiety disorder. Behav Res Ther 2007 Oct;45(10):2518-2526. [doi: 10.1016/j.brat.2007.04.011] [Medline: 17572382]
67. Ma SH, Teasdale JD. Mindfulness-based cognitive therapy for depression: replication and exploration of differential relapse prevention effects. J Consult Clin Psychol 2004 Feb;72(1):31-40. [doi: 10.1037/0022-006X.72.1.31] [Medline: 14756612]
68. Veehof MM, Trompetter HR, Bohlmeijer ET, Schreurs KM. Acceptance- and mindfulness-based interventions for the treatment of chronic pain: a meta-analytic review. Cogn Behav Ther 2016;45(1):5-31. [doi: 10.1080/16506073.2015.1098724] [Medline: 26818413]
69. Lancaster G, Dodd S, Williamson P. Design and analysis of pilot studies: recommenDesign and analysis of pilot studies: recommendations for good practicedations for good practice. J Eval Clin Pract 2004 May;10(2):307-312. [doi: 10.1111/j..2002.384.doc.x] [Medline: 15189396]
70. Weber W. The National Center for Complementary and Integrative Health. 2015. The R34: An Opportunity for Critical Groundwork in Clinical Trials Research URL: https://nccih.nih.gov/research/blog/r34 [accessed 2019-10-02]
71. Jackson W, Zale EL, Berman SJ, Malacarne A, Lapidow A, Schatman ME, et al. Physical functioning and mindfulness skills training in chronic pain: a systematic review. J Pain Res 2019;12:179-189 [FREE Full text] [doi: 10.2147/JPR.S172733] [Medline: 30655687]
72. de Giorgio A, Padulo J, Kuvačić G. Effectiveness of yoga combined with back school program on anxiety, kinesiophobia and pain in people with non-specific chronic low back pain: a prospective randomized trial. Muscle Ligaments and Tendons J 2019 Jan;8(1):104-112. [doi: 10.32098/mltj.01.2018.13]

73. la Cour P, Petersen M. Effects of mindfulness meditation on chronic pain: a randomized controlled trial. Pain Med 2015 Apr;16(4):641-652. [doi: 10.1111/pme.12605] [Medline: 25376753]

Abbreviations

CAMS-R: cognitive and affective mindfulness scale-revised
ICF: International Classification of Functioning, Disability and Health
IMMPACT: Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials
MCID: minimum clinically important difference
MOCS: measures of current status
NCCIH: National Center for Complementary and Integrative Health
NRS: numerical rating scale
ORBIT: Obesity Related Behavioral Intervention Trials
PASIPD: self-reported physical activity scale for individuals with physical disabilities
PROMIS: Patient-Reported Outcomes Measurement Information System
RCT: randomized controlled trial
SMART: Specific, Measurable, Achievable, Relevant and Time-based
WHODAS: World Health Organization Disability Assessment Schedule

Edited by G Eysenbach; submitted 16.03.20; peer-reviewed by HW Mak; comments to author 06.04.20; revised version received 13.04.20; accepted 27.04.20; published 08.06.20.

Please cite as:
Greenberg J, Popok PJ, Lin A, Kulich RJ, James P, Macklin EA, Millstein RA, Edwards RR, Vranceanu AM
A Mind-Body Physical Activity Program for Chronic Pain With or Without a Digital Monitoring Device: Proof-of-Concept Feasibility Randomized Controlled Trial
JMIR Form Res 2020;4(6):e18703
URL: https://formative.jmir.org/2020/6/e18703
doi:10.2196/18703
PMID:32348281

Corrigenda and Addenda

Correction: Using Natural Language Processing to Examine the Uptake, Content, and Readability of Media Coverage of a Pan-Canadian Drug Safety Research Project: Cross-Sectional Observational Study

Hossein Mohammadhassanzadeh[1], PhD; Ingrid Sketris[1], PharmD; Robyn Traynor[1], MSc; Susan Alexander[2], MHI; Brandace Winquist[3], PhD; Samuel Alan Stewart[1], PhD

[1]Dalhousie University, Halifax, NS, Canada

[2]Nova Scotia Health Authority, Halifax, NS, Canada

[3]University of Saskatchewan, Saskatoon, SK, Canada

Corresponding Author:
Samuel Alan Stewart, PhD
Dalhousie University
5790 University Ave, Room 408
Halifax, NS, B3H 4R2
Canada
Phone: 1 902 494 6287
Fax: 1 902 494 1597
Email: sam.stewart@dal.ca

Related Article:

Correction of: https://formative.jmir.org/2020/1/e13296/

(JMIR Form Res 2020;4(6):e20211) doi:10.2196/20211

In "Using Natural Language Processing to Examine the Uptake, Content, and Readability of Media Coverage of a Pan-Canadian Drug Safety Research Project: Cross-Sectional Observational Study" (JMIR Form Res 2020;4(1):e13296), several errors were noticed.

Susan Alexander's degree was incorrectly noted in the original article; it has been revised from *MSc* to *MHI*.

Figures 3 and 4 were incorrectly positioned in the original article. Figure 3 was displaying in the position of Figure 4 and Figure 4 was displaying in the position of Figure 3. This has now been corrected in the manuscript. Figure 3 is the circle graph, depicting the similarity of all articles to the three original articles; Figure 4 is the distribution chart, depicting the distribution of readability levels of articles (y-axis) based on text-standard measures.

The figure legends for Figures 3 and 4 were also incorrect in the original article.

Figure 3's legend has been revised from:

> *Steady trend of similarity (cosine similarity) between the media articles and the CNODES publications: CMAJ article, podcast, and press release*

to:

> *Trend of similarity (cosine similarity) between the media articles and the CNODES publications: CMAJ article, podcast, and press release*

Figure 4's legend has been revised from:

> *Similarity of all articles to three original articles (press release, podcast, and CMAJ article)*

to:

> *Distribution of readability levels of articles based on text-standard measure*

Additionally, due to a technical error, there was a discrepancy between the HTML and PDF versions of the original article. In the HTML version, the inline equation incorrectly displayed as a figure and was assigned a figure number, causing the figures to be numbered incorrectly, and the placement of the circle graph was also incorrect. This has now been corrected.

Finally, there were errors in Multimedia Appendix 1 and Multimedia Appendix 2 in the original article.

In Multimedia Appendix 1, the title has been revised from:

> *Appendix 2. List of articles (26 media articles, 3 CNODES reference publications)*

to:

Multimedia Appendix 1. List of articles (26 media articles, 3 CNODES reference publications)

In Multimedia Appendix 2, the title has been revised from:

Appendix 1. Readability scales

to:

Multimedia Appendix 2: Readability scales

The correction will appear in the online version of the paper on the JMIR website on June 23, 2020, 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.

Multimedia Appendix 1
List of articles (26 media articles, 3 CNODES reference publications).
[DOCX File , 26 KB - formative_v4i6e20211_app1.docx]

Multimedia Appendix 2
Readability scales.
[DOCX File , 16 KB - formative_v4i6e20211_app2.docx]

Submitted 19.05.20; this is a non–peer-reviewed article;accepted 20.05.20; published 23.06.20.

Please cite as:
Mohammadhassanzadeh H, Sketris I, Traynor R, Alexander S, Winquist B, Stewart SA
Correction: Using Natural Language Processing to Examine the Uptake, Content, and Readability of Media Coverage of a Pan-Canadian Drug Safety Research Project: Cross-Sectional Observational Study
JMIR Form Res 2020;4(6):e20211
URL: https://formative.jmir.org/2020/6/e20211
doi:10.2196/20211
PMID:32574152

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org