JMIR mHealth and uHealth

Impact Factor (2024): 5.4
Volume 7 (2019), Issue 12 ISSN 2291-5222 Editor in Chief: Lorraine Buis, PhD, MSI

Contents

Viewpoint

Lessons Learned: Recommendations For Implementing a Longitudinal Study Using Wearable and Environmental Sensors in a Health Care Organization (e13305)
Michelle L'Hommedieu, Justin L'Hommedieu, Cynthia Begay, Alison Schenone, Lida Dimitropoulou, Gayla Margolin, Tiago Falk, Emilio Ferrara, Kristina Lerman, Shrikanth Narayanan. . . . 3

Review

Digital Health Behavior Change Technology: Bibliometric and Scoping Review of Two Decades of Research (e13311)
Fawad Taj, Michel Klein, Aart van Halteren. . . . 16

Original Papers

Clustering Insomnia Patterns by Data From Wearable Devices: Algorithm Development and Validation Study (e14473)
Sungkyu Park, Sang Lee, Sungwon Han, Meeyoung Cha. . . . 37

Objectively Monitoring Amyotrophic Lateral Sclerosis Patient Symptoms During Clinical Trials With Sensors: Observational Study (e13433)
Luis Garcia-Gancedo, Madeline Kelly, Arseniy Lavrov, Jim Parr, Rob Hart, Rachael Marsden, Martin Turner, Kevin Talbot, Theresa Chiwera, Christopher Shaw, Ammar Al-Chalabi. . . . 48

Adults' Preferences for Behavior Change Techniques and Engagement Features in a Mobile App to Promote 24-Hour Movement Behaviors: Cross-Sectional Survey Study (e15707)
Ann DeSmet, Ilse De Bourdeaudhuij, Sebastien Chastin, Geert Crombez, Ralph Maddison, Greet Cardon. . . . 63

Medical Videography Using a Mobile App: Retrospective Analysis (e14919)
Julia Cambron, Kirk Wyatt, Christine Lohse, Page Underwood, Thomas Hellmich. . . . 74

A Hospital-Community-Family–Based Telehealth Program for Patients With Chronic Heart Failure: Single-Arm, Prospective Feasibility Study (e13229)
Xiaorong Guo, Xiang Gu, Jiang Jiang, Hongxiao Li, Ruoyu Duan, Yi Zhang, Lei Sun, Zhengyu Bao, Jianhua Shen, Fukun Chen. . . . 87

Usefulness of Modern Activity Trackers for Monitoring Exercise Behavior in Chronic Cardiac Patients: Validation Study (e15045)
Cyrille Herkert, Jos Kraal, Eline van Loon, Martijn van Hooff, Hareld Kemps. . . . 100

Improving Pacific Adolescents' Physical Activity Toward International Recommendations: Exploratory Study of a Digital Education App Coupled With Activity Trackers (e14854)
Olivier Galy, Kalina Yacef, Corinne Caillaud. . . . 111

A Stress Relief App Intervention for Newly Employed Nursing Staff: Quasi-Experimental Design (e15785)
I-Chiu Chang, Wei-Chen Cheng, Wen-Chuan Kung. . . . 122

Analysis of the Implementation, User Perspectives, and Feedback From a Mobile Health Intervention for Individuals Living With Hypertension (DREAM-GLOBAL): Mixed Methods Study (e12639)
Jordan Barsky, Rebekah Hunter, Colin McAllister, Karen Yeates, Norm Campbell, Peter Liu, Nancy Perkins, Diane Hua-Stewart, Marion Maar, Sheldon Tobe. . . . 132

Development and Local Contextualization of Mobile Health Messages for Enhancing Disease Management Among Community-Dwelling Stroke Patients in Rural China: Multimethod Study (e15758)
Enying Gong, Wanbing Gu, Erdan Luo, Liwei Tan, Julian Donovan, Cheng Sun, Ying Yang, Longkai Zang, Peng Bao, Lijing Yan. . . . 142

Use of Mobile Health Apps and Wearable Technology to Assess Changes and Predict Pain During Treatment of Acute Pain in Sickle Cell Disease: Feasibility Study (e13671)
Amanda Johnson, Fan Yang, Siddharth Gollarahalli, Tanvi Banerjee, Daniel Abrams, Jude Jonassaint, Charles Jonassaint, Nirmish Shah. . .
1 6 3

Tutorial

Strategies for the Successful Implementation of a Novel iPhone Loaner System (iShare) in mHealth Interventions: Prospective Study (e16391)
William Yang, Erin Spaulding, David Lumelsky, George Hung, Pauline Huynh, Kellen Knowles, Francoise Marvel, Valerie Vilarino, Jane Wang, Lochan Shah, Helen Xun, Rongzi Shan, Shannon Wongvibulsin, Seth Martin. . . . 155

Viewpoint

Lessons Learned: Recommendations For Implementing a Longitudinal Study Using Wearable and Environmental Sensors in a Health Care Organization

Michelle L'Hommedieu[1], PhD; Justin L'Hommedieu[1], MS; Cynthia Begay[2], MPH; Alison Schenone[1], BS; Lida Dimitropoulou[1], PhD; Gayla Margolin[3], PhD; Tiago Falk[4], PhD; Emilio Ferrara[1], PhD; Kristina Lerman[1], PhD; Shrikanth Narayanan[1], PhD

[1]Information Sciences Institute, University of Southern California, Los Angeles, CA, United States

[2]Department of Human Resources, Keck Medicine of University of Southern California, Los Angeles, CA, United States

[3]Department of Psychology, University of Southern California, Los Angeles, CA, United States

[4]Institut national de la recherche scientifique, University of Québec, Montreal, QC, Canada

Corresponding Author:
Michelle L'Hommedieu, PhD
Information Sciences Institute
University of Southern California
3740 McClintock Ave
EEB 413
Los Angeles, CA, 90089
United States
Phone: 1 2137402318
Email: mhasan@isi.edu

Abstract

Although traditional methods of data collection in naturalistic settings can shed light on constructs of interest to researchers, advances in sensor-based technology allow researchers to capture continuous physiological and behavioral data to provide a more comprehensive understanding of the constructs that are examined in a dynamic health care setting. This study gives examples for implementing technology-facilitated approaches and provides the following recommendations for conducting such longitudinal, sensor-based research, with both environmental and wearable sensors in a health care setting: pilot test sensors and software early and often; build trust with key stakeholders and with potential participants who may be wary of sensor-based data collection and concerned about privacy; generate excitement for novel, new technology during recruitment; monitor incoming sensor data to troubleshoot sensor issues; and consider the logistical constraints of sensor-based research. The study describes how these recommendations were successfully implemented by providing examples from a large-scale, longitudinal, sensor-based study of hospital employees at a large hospital in California. The knowledge gained from this study may be helpful to researchers interested in obtaining dynamic, longitudinal sensor data from both wearable and environmental sensors in a health care setting (eg, a hospital) to obtain a more comprehensive understanding of constructs of interest in an ecologically valid, secure, and efficient way.

(JMIR Mhealth Uhealth 2019;7(12):e13305) doi:10.2196/13305

KEYWORDS

research; research techniques; Ecological Momentary Assessment; wearable electronic devices

Introduction

Background

Facilitating a healthy and high-performing workforce has long been a topic of interest to both researchers and organizations [1] and is particularly relevant to employees working in health care settings. A significant fraction of an individual's time is spent at work, often managing dynamic and competing cognitive, emotional, social, and physical demands. The determinants of healthy and productive work behavior and performance are, however, not well understood. Moreover, the determinants of well-being and work performance are influenced by factors outside work, and vice versa. For example, past

research has found a significant relationship between employee well-being and job performance [2]. In addition, well-being is a predictor of job performance, even when controlling for job satisfaction, age, gender, and tenure [3]. Furthermore, meta analytic data show that that one's satisfaction has a positive relationship with one' job performance [4]. Taken together, these studies suggest that employee well-being is important to understanding job performance, and therefore more effort ought to be used to understand it.

Numerous studies examining worker wellness or job performance rely solely on self-report survey data [1,5], which are relatively easy to obtain. Such measures, however, are prone to response bias, as participants tend to underreport behaviors deemed inappropriate and overreport behaviors that are considered appropriate [6]. This bias is heightened in organizational settings, as employees often believe that their employers can see their responses or may be privy to such information in the future [6]. Consequently, obtaining accurate, dynamic data from health care employees in their workplace can pose a challenge to researchers who aim to better understand factors affecting employee wellness and workplace behaviors and outcomes.

A Brief Overview of Sensors

Converging advances in distributed biobehavioral sensing, wireless and wearable technologies, and machine learning promise novel ways for illuminating knowledge gaps about work behavior and facilitating productive and healthy job performance in the field of health care. Wearable sensors are devices that can be worn or momentarily attached to a person's body. They typically rely on wireless, miniature circuits embedded in patches or bandages, wristbands, rings, or shirts [7]. They can be used in tandem with hand-held devices (eg, a mobile phone) to temporarily store physiological or behavioral data and upload those data to a database server [7]. Examples of wearable sensors include smartwatches, heart rate monitors, and smart glasses [8].

Nonwearable trackers such as video cameras and gaze trackers [9] are devices that are placed in the environment and indirectly collect data from individuals. Environmental sensors are devices that are placed in the environment and collect environmental data such as temperature, humidity, and noise levels. They may pose less of a burden to participants than wearable sensors, as little to no participant interaction with nonwearable sensors is needed. However, nonwearable sensors do provide challenges for consent and privacy if data from identifiable participants are recorded without the participants' knowledge.

Advantages of Using Sensors in Research

Advances in sensor technology make continuous collection of behavioral and physiological data possible. Although traditional methods of data collection often yield cross-sectional data obtained at a specific point in time, both wearable and nonwearable sensors can provide a continuum of automated data [10]. The use of sensors as data collection instruments also allows researchers to overcome the limitations of traditional data collection tools, such as surveys [11], as sensor data are considered to be more objective and accurate [12,13]. Thus, sensor data can be used in tandem with self-report survey data and can help validate self-report data (eg, checking Fitbit data to identify the number of hours an individual slept the previous night and comparing this with the self-reported number of hours). Wearable sensors provide an advantage to research participants, as data can be collected unobtrusively, in real time and in natural settings. This advantage may be particularly relevant to health care employees, as they may be unable to take a break at work to complete a survey when they are busy providing critical patient care.

Sensor data can help answer diverse research questions related to working in a health care setting. Prior research has employed sensors to measure performance and personality in a postanesthesia care unit [14] and to measure physician workload to improve the safety and efficiency of emergency rooms [15]. Previous research has also used sensors to better understand employees' mood throughout the workday [16]. Researchers in one study used the Toshiba Silmee Bar Type chest sensor to measure heart rate and pulse rate. Participants also reported their mood (ie, excited, happy, calm, tired, bored, sad, stressed, and angry) every 2 hours over 11 workdays through a mobile app called HealthyOffice. This method improved predictions of mood. Another option is the mobile sensing platform EmotionSense, developed by Jason Rentfrow and Cecilia Mascolo. EmotionSense has been used to analyze and classify participants' voices as happy, sad, fearful, anger, and neutral [17]. Other researchers have used Moodscope, an app developed by Adrian Hosford and Caroline Ashcroft, to infer mood based on how participants used their mobile phones [18]. These researchers used text messages, emails, phone calls, app usage data, Web browsing, and location to predict mood intensity of 2 dimensions (pleasure and activeness). Sensors can also be used to examine leadership emergence and group structure [11], which may be of interest to health care leadership.

This study provides recommendations for researchers conducting sensor-based research in a health care setting. In addition, we demonstrate how we successfully implemented these recommendations and the outcomes of doing so by providing examples from a large-scale, sensor-based, longitudinal study of hospital employees. These recommendations may be useful to researchers who are interested in obtaining dynamic data in a health care setting to paint a more comprehensive picture of factors affecting job performance and worker well-being.

Tracking Individual Performance With Sensors as a Case Study

In early 2018, we conducted the *Tracking Individual Performance with Sensors(TILES)* project, a large-scale study examining physiological, environmental, and behavioral variables affecting employee wellness and job performance. Over 200 volunteer hospital employees of a large hospital in Los Angeles, California, enrolled in the study; participant characteristics are described in Table 1. Participants enrolled in 1 of 3 waves of participation, each with different start and end dates. Participants in each wave were asked to wear sensors and respond to brief daily surveys for 10 weeks, starting on the first day of participation for the wave.

Table 1. Participant characteristics (N=212).

Demographics	Value
Gender, n (%)	
Male	66 (31.1)
Female	146 (68.9)
Education, n (%)	
High school or some college	40 (18.9)
Bachelor's degree	126 (59.4)
Some graduate school or graduate degree	46 (21.7)
Work status, n (%)	
Full-time	210 (99.1)
Part-time	2 (0.9)
Job title, n (%)	
Registered nurse	113 (54.3)
Certified nursing assistants	25 (12.0)
Monitor technicians	11 (5.3)
Physical therapists	6 (2.9)
Occupational therapists	2 (1.0)
Respiratory therapists	3 (1.4)
Other	48 (23.1)
Type of work, n (%)	
Direct patient care	155 (73.1)
Lab	25 (11.8)
Administrative	2 (0.9)
Age (years), median (range)	36 (21-65)

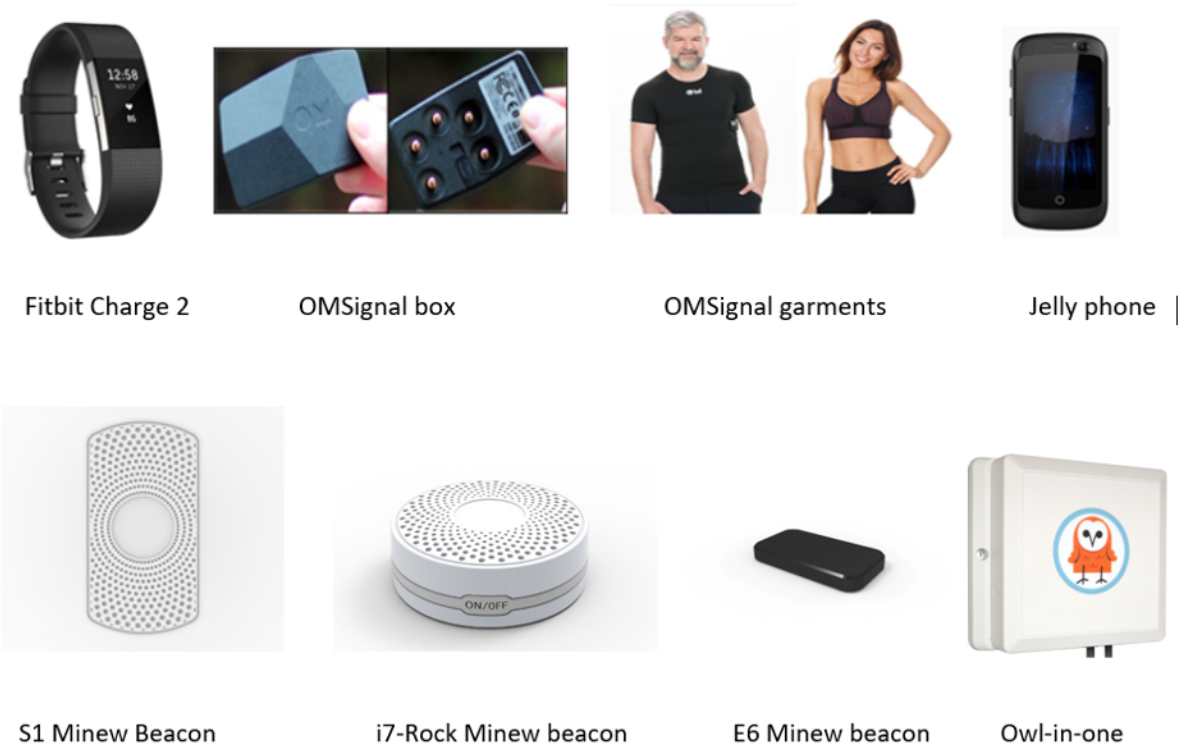
Participants were asked to wear multiple sensors over 10 weeks to collect physiological data including audio features, heart rate, respiratory rate, and sleep (see Table 2 for detailed descriptions; Figure 1 for images). Participants also completed survey batteries at the beginning and end of the study and daily surveys throughout the 10 weeks. Survey data were used in tandem with sensor-based data to provide a more comprehensive assessment of participant behavior.

The successful implementation of the TILES project required us to overcome a number of challenges, including identifying potential challenges with wearable sensors and making adjustments as needed, handling both the hospital's and potential participants' concerns regarding sensor-based data collection and privacy; garnering interest in a complex study with unfamiliar sensors; ensuring that participants were compliant with demanding, sensor-specific study procedures; and effectively implementing the study within budget and sensor manufacturer constraints. This study provides recommendations for successfully navigating such challenges based on what we learned before, during, and after study implementation.

Table 2. Wearable and nonwearable sensors used in the Tracking Individual Performance with Sensors (TILES) study.

Type of sensor	Description of sensor	Frequency of wear (if applicable)	Data collected from sensor
Wearable sensors			
Fitbit Charge 2 (FitBit)	Wrist-worn sensor; requires companion Fitbit phone app	24 hours a day for 10 weeks	Heart rate; step count; sleep duration
OM signal (OM signal)	Chest garment (bra for women, shirt for men) with clip-on box to collect data; requires companion OM signal app	Worn for the duration of a participant's work shift over the course of 10 weeks	Electrocardiogram; breathing rate, depth; body motion
Jelly phone (Unihertz)	Small Android phone that can be clipped onto participant's clothing as a badge	Worn for the duration of a participant's work shift over the course of 10 weeks	Audio features (eg, duration of speech and intonation)
Nonwearable sensors			
S1 Minew beacon (Minew)	Environmental sensor that can be placed in different rooms; communicates with Owl-in-One sensor	No participant interaction with this sensor is required	Temperature; humidity
i7-Rock Minew beacon (Minew)	Environmental sensor that can be attached to doors or placed in different rooms; communicates with Owl-in-One sensor	No participant interaction with this sensor is required	Motion
E6 Minew beacon (Minew)	Environmental sensor that can be placed in different rooms; communicates with Owl-in-One sensor	No participant interaction with this sensor is required	Light level
Owl-in-One (reelyActive)	Environmental sensor that plugs into an outlet; tracks participant proximity using Bluetooth pings from the Fitbit Charge 2	Plugged into different rooms within the hospital where participants spent most of their time; no participant interaction with these sensors was required	Participant proximity

Figure 1. Sensors used in the TILES project. This figure includes both wearable and nonwearable sensors that we used for this study. TILES: Tracking Individual Performance with Sensors.



Methods

Efficiency, security, and ecological validity were at the forefront when planning and implementing the TILES study and are reflected in the recommendations here. Recommendations were generated in the following ways: drawing on existing knowledge of and interaction with sensors, obtaining health care leadership's feedback on using sensors with their employees, resolving problems encountered with sensors during the study, or by post hoc reflections on what worked well and what did not. For purposes of this study, we limit our recommendations to those that relate specifically to the sensors; are believed to have the greatest impact on successful study implementation with regard to ecological validity, security, and efficiency; and can be generalized to other sensor-based research efforts.

We acknowledge that those recommendations that are grounded in the existing literature may be deemed more credible than those that are based on anecdotal challenges or post hoc reflections. Despite this, we remain confident that the strategies to overcome the challenges we encountered helped facilitate successful study implementation in a dynamic environment. The following sections outline the recommendations that were synthesized using the methods described above.

Results

Outcomes/Recommendations for Conducting Sensor-Based Research in a Health Care Setting

Recommendation 1: Pilot Test Sensors and Software Early and Often

Pilot studies are used to troubleshoot study design [19] and facilitate a smooth participant experience. In sensor-based research, conducting pilot studies can help identify whether proposed sensors are inappropriate or too complicated for participant use [19]. Sensor manufacturers may be willing to incorporate researchers' feedback to improve upon the sensors or their companion apps. Although conducting one pilot study may be suitable for some research projects, we recommend conducting several pilot studies early on to allow ample time to test, tweak, and retest sensor procedures. This recommendation is based on the existing sensor literature [20], which highlights some of the challenges of using sensors and the benefit of testing them before deploying them in naturalistic settings.

For our study, we conducted 2 internal pilot studies in which various research team members and student volunteers wore the sensors that we planned to use in the full-scale study. We aimed to assess the comfort of the sensors, gain insight into frustration that arose from participant interaction with the sensors or software, obtain feedback about the perceived invasiveness of the sensors, test the battery life, and assess incoming data quality from the sensors. Such considerations are unique to sensor-based data collection; failing to consider these aspects of the study may result in obtaining a smaller sample size than desired, participant attrition, data sparseness, or poor data quality.

In our first pilot study, participants (n=12) wore the 3 sensors for 2 weeks. Several men found the OM signal shirt to be tight, and all the women (n=3) experienced discomfort (ie, chafing, irritation, and blisters) with the bra. We provided this feedback to the manufacturers, who provided solutions to alleviate discomfort. Pilot participants also noticed that the Jelly phones that were hung around the neck would swing. This would likely frustrate hospital employees who leaned over patients' bedsides, so we attached a clip to the phone that could be secured onto clothing. A few participants felt uncomfortable about having their conversations potentially recorded, so we learned the importance of emphasizing that the phone was programmed to only record audio features and not content of conversations. Although this information was in our consent form, many participants expressed concerns before agreeing to participate in the pilot study and seeing the consent form. This underscores the importance of clearly articulating what data are being collected while recruiting participants, as participants will not see the consent form when they first hear about the study. The few battery issues that arose were overcome by adjusting the configuration parameters of the phone software. After making the necessary changes to sensor logistics and learning how to troubleshoot issues, we conducted another internal pilot 2 months later. Participants reported much less frustration, and the procedures were more seamless.

Before implementing the full-scale study, we conducted another pilot study within the hospital from where we would recruit study participants. A total of 6 hospital employees participated in this pilot study. During this study, we learned that participants' familiarity with technology varied dramatically, and when difficulties with syncing sensors to their mobile phones arose, participants were, understandably, frustrated. We changed our study procedures to sync participants' sensors to their mobile phones for them. Conducting an external pilot study allowed us to identify the challenges of conducting a sensor-based research study in a dynamic hospital environment and make the necessary adjustments to ensure a smooth participant experience.

Recommendation 2: Build/Maintain Trust With Those Wary of Sensors and Concerned About Privacy

Conducting a sensor-based research study in a health care setting requires researchers to effectively communicate study goals and details with both health care leadership and potential study participants. Researchers can use unique strategies tailored to each audience to build trust with those who are wary of introducing or using unfamiliar, complex sensors within the organization or concerned about the invasiveness of the sensors and how privacy may be compromised in a sensor-based study. This recommendation is grounded in the existing literature, which highlights that the acceptance of technology is an essential consideration for sensor-based data collection [21,22].

Build Trust Within the Organization

Building trust and developing buy-in with leadership is essential when conducting a large-scale research project in a dynamic environment [23]. The importance of doing so is heightened when employees are asked to use unfamiliar sensors that may be deemed as invasive. Developing relationships with multiple

leaders and incorporating their feedback into study implementation can build trust within the health care setting and assuage leadership's concerns about the use of sensors in an environment where privacy is of utmost importance.

Researchers are more likely to successfully conduct research in the workplace when they establish and maintain relationships with powerful organizational leaders [24]. For our study, the principal investigator leveraged an existing contact to learn about the organization's structure and key personnel. Forming additional relationships with the organization's leaders was crucial, as we had to ensure that the environmental sensors to be installed at the hospital would not adversely affect the hospital's technology infrastructure such as Wi-Fi networks, power outlets, and access. We effectively leveraged our existing relationship with hospital leadership to form and maintain relationships with those who gave us feedback and permission to use sensors in the hospital.

Physiological, behavioral, and contextual data from one's environment are sensitive in nature [25]. Leadership in a health care setting will likely have concerns about the sensors, so researchers will be tasked with assuaging such concerns [26]. Leadership will likely want to know how the research team will ensure anonymity of employee sensor data, and researchers should provide detailed information about how the sensors will be used, what data will be collected, and how those data will be used. In our study, engineers who were most knowledgeable about the sensors explained to leadership what data environmental sensors would collect (eg, temperature and humidity) and the purpose for collecting these data (eg, how does this context impact workers' affect, behavior, and cognition?). By meeting with leaders who were knowledgeable about their field and had decision-making authority, we were able to conduct research with our sensors of choice within a sensitive environment.

Build Trust With Potential Participants

Obtaining high-quality data hinges on participants feeling that they can trust the researchers [26]. Wearable sensors may be perceived as more invasive than self-report surveys, thus, building trust with potential participants is important in sensor-based research. Letting health care employees know that sensor data can provide insights into the demanding nature of their jobs can be an effective way to build trust with and motivate health care employees to take on the additional tasks of learning how to use unfamiliar sensors, charge sensors, wear sensors, and submit sensor data to the research team.

For the TILES study, we attempted to build trust with potential participants by being clear about the aim of our study and emphasizing that data were collected for research purposes only. Furthermore, we were sure to note the compensation that participants would receive for participating. We met with employees and emphasized that because we know that hospital employees tend to experience high levels of stress [27] and burnout, we were studying the factors that affected their ability to thrive at work and perform well. Afterward, we introduced the novel aspect of sensor-based research and pointed out that using sensors could unobtrusively collect data to provide insights into well-being and job performance. By emphasizing that we knew the challenges of their work and highlighting the advantages of using sensors and the greater impact of our study on participants and their patients, we built trust and found common ground via a shared goal.

Informed consent is also a way to reinforce trust with participants. In our study, we crafted a consent form that thoroughly explained what was expected of participants and what data were being collected. We also clearly specified how we planned to use these data. Although this information was clearly outlined in the consent form, we learned the importance of reiterating this information to potential participants during in-person interactions with them, as some individuals may not read a lengthy, detailed consent form in its entirety.

Eliciting and responding to potential participants' concerns regarding sensors can help build trust with potential participants, create excitement for the study, yield a larger sample size, increase compliance, and decrease attrition [28]. Researchers may ask participants to wear and interact with unfamiliar sensors that may be seen as complicated and invasive [29]. Potential participants may have questions or concerns about such sensors; therefore, researchers should be prepared to address them.

When introducing our study during our participant recruitment period, hospital employees expressed concern that the Jelly phone would record the content of conversations at work. A research team member who designed the Jelly phone software met with several employees to show them what the data from this sensor looked like and explained that the audio features in the study would be used to infer things such as stress and affect. Employees were also concerned that environmental sensor data could get back to their bosses who might ask them "What were you doing here at this time?" We clarified that location data would not be shared with hospital management and explained that environmental sensor data would allow us to estimate how much time participants spent at the bedside. We wanted to obtain an estimate of this as hospital leadership and employees said that patient care providers disliked going to medicine rooms, which decreased time spent with patients. It is important to note that this information was specified in the consent form that interested individuals gained access to after hearing about our study and expressing interest in participating. By engaging in open dialogue during the recruitment process, we appeased concerns regarding sensors, which led to participants trusting both our team and the sensors and expressing interest in participating.

Recommendation 3: Generate Excitement for Novel, New Technology During Recruitment

Previous attempts at longitudinal studies involving hospital employees have shown significant barriers with regard to participant recruitment and retention [30]. We were interested in learning about employee receptivity to the sensors; therefore, we took a participatory approach to recruitment. We asked directors and staff councils that comprised leaders for recommendations regarding how to initially engage and retain potential study participants, given that participants would have to learn how to use unfamiliar sensors and interact with them daily to collect sensitive data. The overarching recommendation that emerged from conversations with health care leadership

was to generate excitement for the sensors used in the study, as most of this technology would be new to participants.

Allow Potential Participants to Interact With Sensors

We generated excitement for the sensors by having them on display at study information tables in the hospital cafeteria and by allowing passersby to interact with them and ask questions. Having team members on-site can generate initial interest in the study [31]. Employees showed interest in the OM signal sensor and Jelly phone, which were the 2 sensors that we expected them to be unfamiliar with. Upon learning about the capability of the OM signal sensor to capture breathing and respiration rate, we received positive feedback about how *cool* the sensor and accompanying app were and how it could be nice to see these data displayed on the app when exercising. Once employees had the opportunity to interact with the sensors and learn about emerging technology and its use in daily life, they seemed more excited about the prospect of joining our study to try out the sensors.

Personally Reach Out to Those Who May Be Wary About Sensors

Translating initial interest in the sensors into study enrollment can be challenging. Although some employees may be enticed by novel sensors and incentives for participating [28], others may become warier about participating as they learn more about the sensors and their interaction with these devices. Researchers can decrease feelings of caution by personally reaching out to these individuals.

During recruitment, we identified a gap between the number of individuals who expressed initial interest and the number of individuals who signed up for an enrollment session. We contacted these individuals and learned that many of them had questions about the sensors, such as how often they would need to wear each device and if these devices would be safe to wear inside the hospital (eg, when visiting the magnetic resonance imaging room). Although much of this information was included on our study website, we personally reached out to these individuals and enrolled many of them after answering their questions. We encourage researchers to anticipate hesitance from potential participants in sensor-based research and personally reach out to effectively translate interest into study enrollment.

Consider the Level of Detail Needed When Introducing Sensors

The design and content of paper recruitment materials can affect an individual's decision to participate in a research study [32]. Researchers must balance informative yet brief content to engage the reader and must decide what sensor-related content, if any, to include. Although listing unfamiliar sensors may be meaningless or intimidating to potential participants, the enjoyment gained from learning about new products and technologies motivates people to participate in technology-related activities [33].

In our study recruitment materials, we emphasized the opportunity to be a part of a ground-breaking study that used cutting-edge technology. Paper materials contained a QR code that employees could easily scan with their mobile phone to be taken to our website to learn about the sensors. We encourage researchers to carefully consider how to best introduce the aspect of wearable sensors in their recruitment materials.

Recommendation 4: Monitor Incoming Sensor Data to Troubleshoot Sensor Issues

Monitoring incoming sensor data can help researchers identify technology glitches or participant issues with sensors that may otherwise go unnoticed or unreported. This recommendation is based on challenges we encountered during the study and the real-time solutions that we generated as well as post hoc reflections about what strategies worked. Not all sensors provide feedback regarding data quality and volume. In this case, researchers may have to rely on third-party quality metrics or a team member who is well versed on data analysis to determine what constitutes sensor data issues.

Monitoring Sensor Data to Identify Technology Glitches

Researchers should anticipate issues that may arise with wearable sensors, such as sensors breaking or companion apps crashing [34]. Participants may not be aware that a sensor or related software is malfunctioning; however, researchers may be able to identify when such an event occurs by monitoring the incoming sensor data.

The Jelly phone used in our study was initially programmed for voice activity detection to turn on the audio capture only after detecting vocal signals [35]; the battery lasted for approximately 10 to 12 hours for internal pilot testers in a university setting. While monitoring incoming Jelly data from hospital pilot participants, we noticed that we only received a couple of hours of Jelly data from a few participants. We learned that the Jelly battery died after a couple hours because of the high noise level in the hospital; thus, participants stopped wearing the Jelly. We reprogrammed the device to turn on only for a fixed duration of time (20 seconds) to save the battery. By monitoring incoming sensor data, we identified a technical glitch with our sensor and quickly developed an effective solution.

We also set up weekly calls with the manufacturers of OM signal, who provided us with usage statistics such as frequency of wear and data quality. We flagged 15.1% (32/212) of participants to check on and remeasured them to ensure appropriate sizing, checked that their companion app was collecting and sending data as expected, and reinstructed them on how to use this sensor. These issues may have otherwise gone unnoticed; therefore, we recommend that researchers monitor incoming data for both quality and volume and reach out in person to participants who fall below a predetermined threshold for both [36].

Monitor Sensor Data to Identify Participant Issues With Sensors

Identifying unique participant issues with sensors is important, as compliance with longitudinal research and sensors can be challenging [28,37], and researchers may ask participants to wear sensors that may be uncomfortable [38]. Monitoring incoming data allows researchers to identify sparse or low-quality sensor data, which may indicate a participant issue with the sensor, and take appropriate action [11,39].

By monitoring incoming OM signal data, we identified that several participants stopped wearing the garment, whereas several others only wore the garments for a couple of hours on each workday. We determined that 10% of participants experienced chafing, rashes, or general discomfort with the garment. For mild discomfort, we instructed participants to apply a bit of water or aloe vera to the electrodes that pressed against the skin. This solution decreased discomfort of several participants. Participants who experienced severe discomfort were instructed to refrain from wearing the garment for a few days; however, 5% continued to experience discomfort. We allowed these participants to participate without wearing the garment. We encourage researchers to anticipate that some participants may find certain wearable sensors uncomfortable and may fail to report instances of discomfort. Monitoring the incoming sensor data can help researchers identify how to accommodate participants and the unique issues they may experience with wearable sensors.

Monitor Sensor Data to Keep Abreast of Unresolved Issues

Depending on the sample size, number of sensors used in the study, and ease of participant sensor use, it may be challenging to identify and keep abreast of unresolved issues. Monitoring sensor data can help researchers ensure that the myriad of potential sensor issues that may arise do not go unnoticed or unresolved.

By monitoring incoming sensor data, we learned that data sparsity or low-quality data could be due to several things (eg, forgetting how to charge a sensor and not having a strong Wi-Fi connection when uploading data). We developed a shared logbook to note which participants had sparse or low-quality data so that we could follow up and assist them. We also developed a companion frequently asked questions/troubleshooting document that team members could reference to resolve sensor-related issues. Monitoring incoming data to manage and keep abreast of the different sensor issues allowed us to facilitate a smooth participant experience with the sensors.

Recommendation 5: Consider the Logistical Constraints of Sensor-Based Research

Researchers conducting sensor-based research will likely be constrained in their ability to purchase their most desired sensors or receive sensors in a timely manner. Although there is a seemingly endless number of constraints that may arise, budget and manufacturer constraints are 2 aspects of sensor-based research that can greatly impact the participant experience and study timeline. Thus, we urge researchers to consider the logistical constraints of sensor-based research, based on the challenges we encountered during the study, and on post hoc reflection about how we successfully navigated such challenges.

Consider Budget Constraints During Sensor Selection

The researcher's budget should be carefully considered when selecting sensors to be used in data collection. Although other criteria, such as battery life and ease of participant use should also be considered during sensor selection, the suite of sensors that a researcher chooses will ultimately be constrained by their budget [40]. Having a member of the research team who is well-versed in contractual obligations or financial resources is essential.

We tested several different sensors for our study and came to our final decision based on affordability, ease of participant use, and quality of data obtained. The cost of sensors (wearable and environmental) and sensor-related products was US $233,000 (US $1142 per participant; see Table 3). Each participant received 1 Fitbit Charge 2, 1 OM signal sensor, and 1 Jelly phone; however, we purchased extra sensors to accommodate sensors malfunctioning or breaking. In addition, we purchased a surplus of OM signal garments as we did not know the accurate size breakdown of our participants until after they enrolled in the study. Additional sensor-related costs included US $5000 for charging hubs and charging cables to make charging various sensors easy for participants and US $13,000 for 64 iPod touches to accommodate Android phone users who could not access the OM signal app without an Apple device. Our project manager who was well versed in our study budget was essential to the sensor selection and purchasing process.

Researchers should also consider the cost of incentivizing participants to wear sensors. The compensation for complying with sensor-related study procedures may be influenced by several factors including how often participants are asked to wear the sensor, how comfortable the sensor is, and ease of participant interaction with the sensor [36]. For this study, structuring these incentives to be motivating enough to encourage compliance, while staying within budget, was crucial.

Table 3. Cost of sensors.

Sensor	Quantity, n	Unit price (US $)	Total price (US $)	Total/person (n=204) (US $)
OM signal 250 boxes and 1250 garments	250	429	107,250	525.74
Jelly phones	250	138	34,500	169.12
Owl-in-one: Bluetooth data hub and Bluetooth device proximity sensor	261	155	40,455	198.31
Minew beacons (Bluetooth environment data sensors)	139	16	2224	10.90
Fitbit Charge 2	250	124	31,000	151.96
Total	__[a]	—	215,429	1056

[a]Not applicable.

We used recommendations from Lawler [41] to guide the development of our incentive system. We made it clear what behaviors were required to obtain the reward, and we made the incentives desirable to participants [41]. For example, participants received US $75 for completing an in-person enrollment session, during which they received their sensors and instructions for use. We also paid participants between US $10 to US $25 in a prepaid Visa gift card each week based on their level of weekly compliance in wearing sensors and answering surveys. To determine compensation for each week of participation, we generated a list of actions they could do for each day and allotted a specific number of points to each action. At the end of the study, we distributed grand prizes to the top 3 participants in each wave who earned the most points throughout the 10 weeks (ie, US $250, US $200, and US $150, respectively). Such incentives highlight the magnitude of investment that may be necessary to conduct a sensor-based study, so we encourage researchers to consider the cost of incentivizing participants to interact with sensors while planning a sensor-based study.

Consider Sensor Manufacturer Constraints

Researchers should anticipate constraints from sensor manufacturers when planning and conducting a sensor-based study. Manufacturers may run into production delays and may be unable to deliver sensors in accordance with the expected shipment schedule [42]. These events may delay the research team's timeline; therefore, constant communication and transparency with sensor manufacturers are crucial to the success of a sensor-based study.

Developing good rapport with sensor manufacturers helped us navigate the constraints of sensor-based research. The nature of our study required us to have all wearable sensors in hand before the study began. The OM signal garments were made to order and required 6 to 8 weeks of manufacturing and shipping time, which was challenging to navigate with a tight timeline and budget approval delays. In addition, batch ordering hundreds of sensors and receiving them on time was an obstacle, as the manufacturers were unable to process new orders during holidays. Finally, there were times when we had to exchange sensors for different sizes. When experiencing such issues, we expressed the urgency of timely delivery to the manufacturers, who accommodated us by sending sensors in partial shipments, so that we had enough sensors to cover our initial needs. We encourage researchers to delegate a few team members to establishing and maintaining relationships with sensor manufacturers, as doing so will likely help researchers navigate any manufacturer constraints that arise.

Discussion

Key Takeaways

Researchers will likely encounter several challenges when conducting sensor-based research. Pilot testing sensors and software early and often can help researchers iteratively improve the study procedures and user experience with sensors before the study begins. Building trust, both with key stakeholders and potential study participants who may be wary of sensor-based data collection and concerned about privacy, can aid in the recruitment of study participants and compliance with complex study procedures such as interacting with unfamiliar sensors. Researchers should aim to generate excitement for these novel sensors and technology during recruitment to spark initial interest in the study. Once the study is underway, monitoring incoming sensor data can help researchers troubleshoot issues that may have gone unnoticed. Finally, we urge researchers to consider the logistical constraints of sensor-based research. These general recommendations (summarized in Textbox 1) can be adapted to different apps of sensor technology.

Textbox 1. Summary of recommendations.

1. Conduct pilot tests of sensors and software early and do so several times to anticipate what works well and what challenges may arise with the sensors to be used.
2. Build trust with health care leadership and potential participants who may be wary of sensor-based data collection and concerned about privacy.
3. Generate excitement for novel, new technology during recruitment.
4. Monitor incoming sensor data to troubleshoot sensor issues.
5. Consider the logistical constraints of sensor-based research.

Limitations

We are confident that the recommendations put forth in this study will help facilitate a smoother investigator and participant experience with sensors. There are, however, several limitations of this study. The recommendations in this study were grounded in the existing literature, feedback from organizational leadership, challenges we encountered along the way and the solutions we generated, and post hoc reflections. This study would be strengthened with additional statistics demonstrating the effectiveness of each recommendation. These data, however, were either unavailable or challenging to quantify. For example, hospital leadership emphasized the importance of generating excitement for novel, new sensor technology rather than rattle off a list of unfamiliar sensors to participants. We recall seeing spikes in the number of participants who indicated interest on our website immediately after we spoke with them. Unfortunately, we did not keep track of the specific days and times that we spoke to participants and relayed information about "exciting" sensors; keeping such a log would have allowed us to provide descriptive statistics to speak to the effectiveness of this strategy. However, upon thoughtful reflection, feedback from participants, and seeing the number of interested individuals increase immediately after we conducted events (eg, having a study information booth in the hospital cafeteria), we

are confident that this (and other strategies) was effective in having the intended impact.

We limited the recommendations in this study to those that related specifically to the sensors; were believed to have had the greatest impact with regard to ecological validity, security, and efficiency; and could be generalized to other sensor-based research efforts. Therefore, researchers should consider the recommendations we provide within the broader scope of sensor-based research and not look to this study as the sole source of recommendations for implementing such a study. For example, other sensor-specific recommendations may relate to sensor selection (eg, learning how to prioritize ease of use, battery life, data quality, and cost) or implementing strategies to encourage participant compliance with sensors.

Another limitation is that we focus on using sensors to meet a specific goal in a specific population in a specific context. Although the primary aim of this study was to provide researchers with insight into assessing overworked, shift-based employees with sensors, we acknowledge that these recommendations may be applicable and helpful to those who wish to use sensors for other purposes. To overcome this limitation, the following section discusses how these recommendations can apply to an area that is already seeing significant growth from sensor-based data collection: telemedicine/remote patient monitoring.

Health Care Applications of Sensor Technology

Advancements in wearable sensor technology and wireless communications allow for real-time health care monitoring, which is a promising solution to treating those who live in remote areas far from medical centers or those who cannot afford inpatient care. Telemedicine, the integration of mobile communication with wearable sensors in patients, can be used to remotely monitor patients with diverse health issues including cardiac disease, diabetes, and autism spectrum disorder [43]. For example, electrodermal activity sensors, also known as skin conductance sensors, are effective for monitoring sympathetic nervous system arousal [44]. These sensors can provide those with autism spectrum disorder a means of measuring stress or anxiety when these feelings cannot be verbally or socially expressed [44]. The sensors can provide feedback to both caregivers and patients to help understand causes of stress or anxiety and help prevent negative behavior [44]. Moving toward technologies that enable continuous, outpatient monitoring is a cost-effective way to provide health care; however, health monitoring technology must be comfortable and unobtrusive, must be simple to use, and must provide privacy and security [44]. Thus, the recommendations provided in this study are applicable to the use of sensors in telemedicine.

Health care providers interested in using sensors to remotely monitor patients should pilot test sensors early and often. Pilot testing with the end user allows providers to ensure minimal discomfort to the patient and freedom of mobility, which allows the patient to carry out normal activities while under continuous monitoring. This may enhance the quality of data, as the patient's behavior may be closer to how it is when not wearing sensors [45]. Pilot tests can also be used to test the functional aspects of sensors, such as the battery life [44,46].

Building trust with patients and generating excitement for sensors to facilitate patient health is crucial to the success of using sensors for remote health care monitoring. A patient may not believe that the sensor will give good enough signals to health care providers [45], which may result in the individual deciding not to wear the sensor as frequently as asked. Health care providers should appease patient concerns that arise with sensors, as patients may be inclined to initially reject devices [47]. Previous research identifies such concerns, including being afraid that the sensor may fall off the skin if they move too much, being afraid of wearing the sensor, being afraid that it may suddenly stop working, and worrying that it may be visible to others [45]. In addition, patients may be concerned about the security and privacy of sensor-based data [47]. In this case, health care providers could discuss a patient-centric access policy with the patient, in which different privacy levels could be assigned to the sensor data, based on sensitivity. Different access privileges could then be assigned to health care providers based on their role [47]. Informing patients about the feasibility of using sensors in their daily lives and providing instructions for use can help patients trust the health care providers and the sensors themselves. Health care providers can also highlight the advantages of continuous, sensor-based patient monitoring, such as decreased cost and less time spent going to a clinic or to the patient. Highlighting these exciting advances in technology that result in decreased patient burden can help health care providers successfully implement wearable sensors into a patient's health care plan.

Monitoring incoming sensor data allows health care providers to help patients in real time. Sensors may have different interfaces, one for patients and one for health care providers, allowing for monitoring of incoming data. In one study, researchers employed a wearable sensor that extracted cardiac parameters such as heart rate and blood pressure to help in the early detection of diseases such as hypertension. The patient interface was the wearable sensor. This sensor extracted medical information from patients and used Bluetooth to transmit the data to an Android-based listening port, which transferred information to the Web server to show reports on the doctor interface [43]. This Web interface allowed doctors and medical centers to simultaneously view and diagnose patients' medical status. For example, the Web interface had a module containing alarming messages (eg, signaling arrhythmia) generated by the Android-based listening port. The interface also had location records so that the doctor could track the patient's location and send an ambulance to the patient, if needed [43]. Monitoring incoming sensor data may literally be the difference between life and death in remote patient monitoring.

Finally, we encourage health care providers to consider the logistical constraints of sensor-based patient monitoring. For example, health care providers may access sensor-based patient data for multiple patients at once [43]. In this case, the provider must consider the cost involved in obtaining sensors, teaching multiple patients how to use sensors, monitoring sensor data, and taking appropriate action (eg, calling an ambulance) during monitoring. Considering such constraints, in terms of cost and labor, will likely impact the patient experience. Sensor manufacturer constraints should also be considered when

selecting sensors for remote patient monitoring. Delays in manufacturing or shipment may have a more adverse impact on the end user, as health care providers who care for critically ill patients may have less flexibility in their timeline than researchers. Being mindful of the constraints of sensor-facilitated health care can help facilitate a smoother patient monitoring experience.

Conclusions

Advances in sensor-based technology allow for new ways of obtaining continuous physiological data from individuals in real time. Furthermore, these sensors can allow researchers to measure things that are difficult to measure with self-report and can potentially reduce the participant's burden. Although exciting, these new methods of data collections pose new challenges to researchers and require them to adapt new skills. Nevertheless, we are confident that implementing the recommendations in this study will facilitate successful study implementation and a smooth participant experience for investigators who are interested in conducting research with sensors and will provide patient care providers with insights into how to successfully use sensors for applications such as remote patient monitoring.

Acknowledgments

The research is based upon work supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via IARPA Contract No 2017-17042800005. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the US Government. The US Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

Conflicts of Interest

None declared.

References

1. Schmidt FL, Hunter JE. The validity and utility of selection methods in personnel psychology: practical and theoretical implications of 85 years of research findings. Psychol Bull 1998;124(2):262-274 [FREE Full text] [doi: 10.1037/0033-2909.124.2.262]
2. Wright TA, Bonett DG. The role of pleasantness and activation-based well-being in performance prediction. J Occup Health Psychol 1997;2(3):212-219. [doi: 10.1037/1076-8998.2.3.212]
3. Wright T, Cropanzano R. Psychological well-being and job satisfaction as predictors of job performance. J Occup Health Psychol 2000 Jan;5(1):84-94. [doi: 10.1037/1076-8998.5.1.84] [Medline: 10658888]
4. Judge TA, Thoresen CJ, Bono JE, Patton GK. The job satisfaction-job performance relationship: a qualitative and quantitative review. Psychol Bull 2001 May;127(3):376-407. [doi: 10.1037/0033-2909.127.3.376] [Medline: 11393302]
5. Barrick MR, Mount MK. The big five personality dimensions and job performance: a meta-analysis. Pers Psychol Durh 1991;44(1):1-26. [doi: 10.1111/j.1744-6570.1991.tb00688.x]
6. Donaldson S, Grant-Vallone E. Understanding self-report bias in organizational behavior research. J Bus Psychol 2002 Dec;17(2):245-260. [doi: 10.1023/A:1019637632584]
7. Bonato P. Wearable sensors/systems and their impact on biomedical engineering. IEEE Eng Med Biol Mag 2003;22(3):18-20. [doi: 10.1109/MEMB.2003.1213622]
8. Mukhopadhyay SC. Wearable sensors for human activity monitoring: a review. IEEE Sensors J 2015 Mar;15(3):1321-1330. [doi: 10.1109/JSEN.2014.2370945]
9. Baluja S, Pomerleau D. Non-Intrusive Gaze Tracking Using Artificial Neural Networks. In: Adv Neural Inf Process Syst. 1994 Presented at: Neural Information Processing Systems; 1994; Denver, CO, USA p. 460.
10. Berlin J, Storti K, Brach J. Using activity monitors to measure physical activity in free-living conditions. Phys Ther 2006 Aug;86(8):1137-1145. [doi: 10.1093/ptj/86.8.1137] [Medline: 16879047]
11. Chaffin D, Heidl R, Hollenbeck JR, Howe M, Yu A, Voorhees C, et al. The promise and perils of wearable sensors in organizational research. Organ Res Methods 2017;20(1):3-31. [doi: 10.1177/1094428115617004]
12. Patterson SM, Krantz DS, Montgomery LC, Deuster PA, Hedges SM, Nebel LE. Automated physical activity monitoring: validation and comparison with physiological and self-report measures. Psychophysiology 1993 May;30(3):296-305. [doi: 10.1111/j.1469-8986.1993.tb03356.x] [Medline: 8497559]
13. Kozey-Keadle S, Libertine A, Lyden K, Staudenmayer J, Freedson PS. Validation of wearable monitors for assessing sedentary behavior. Med Sci Sports Exerc 2011 Aug;43(8):1561-1567. [doi: 10.1249/MSS.0b013e31820ce174] [Medline: 21233777]
14. Olguin D, Gloor P, Pentland A. Wearable Sensors for Pervasive Healthcare Management. In: Proceedings of the 2009 3rd International Conference on Pervasive Computing Technologies for Healthcare. 2009 Presented at: PervasiveHealth'09; April 1-3, 2009; London, UK. [doi: 10.4108/icst.pervasivehealth2009.6033]
15. Levin WC. Age stereotyping: college student evaluations. Res Aging 1988;10(1):134-148. [doi: 10.1177/0164027588101007]

16. Zenonos A, Khan A, Kalogridis G, Vatsikas S, Lewis T, Sooriyabandara M. HealthyOffice: Mood Recognition at Work Using Smartphones and Wearable Sensors. In: Proceedings of the 2016 IEEE International Conference on Pervasive Computing and Communication Workshops. 2016 Presented at: PerCom Workshops'16; March 14-18, 2016; Sydney, NSW, Australia. [doi: 10.1109/percomw.2016.7457166]
17. Rachuri K, Musolesi M, Mascolo C, Rentfrow P, Longworth C, Aucinas A. EmotionSense: A Mobile Phones Based Adaptive Platform for Experimental Social Psychology Research. In: Proceedings of the 12th ACM international conference on Ubiquitous computing. 2010 Presented at: UbiComp'10; September 26-29, 2010; Copenhagen, Denmark p. 281-290. [doi: 10.1145/1864349.1864393]
18. LiKamWa R, Liu Y, Lane N, Zhong L. MoodScope: Building a Mood Sensor from Smartphone Usage Patterns. In: Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services. 2013 Presented at: MobiSys'13; June 25-28, 20132013; Taipei, Taiwan. [doi: 10.1145/2462456.2483967]
19. van Teijlingen E, Hundley V. The importance of pilot studies. Nurs Stand 2002;16(40):33-36. [doi: 10.7748/ns2002.06.16.40.33.c3214] [Medline: 12216297]
20. Bao L, Intille SS. Activity recognition from user-annotated acceleration data. In: Ferscha A, Mattern F, editors. Pervasive Computing. Berlin Heidelberg: Springer; 2004:1-17.
21. Davis FD, Bagozzi RP, Warshaw PR. User acceptance of computer technology: a comparison of two theoretical models. Manag Sci 1989;35(8):982-1003. [doi: 10.1287/mnsc.35.8.982]
22. Nasir S, Yurder Y. Consumers' and physicians' perceptions about high tech wearable health products. Procedia - Soc Behav Sci 2015;195:1261-1267. [doi: 10.1016/j.sbspro.2015.06.279]
23. Christopher S, Watts V, McCormick AK, Young S. Building and maintaining trust in a community-based participatory research partnership. Am J Public Health 2008 Aug;98(8):1398-1406. [doi: 10.2105/AJPH.2007.125757] [Medline: 18556605]
24. Harrison M, Shirom A. Organizational Diagnosis and Assessment: Bridging Theory and Practice. Thousand Oaks, California: Sage Publications; 1998.
25. Raij A, Ghosh A, Kumar S, Srivastava M. Privacy Risks Emerging from the Adoption of Innocuous Wearable Sensors in the Mobile Environment. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2011 Presented at: CHI'11; May 7-12, 2011; Vancouver, BC, Canada. [doi: 10.1145/1978942.1978945]
26. Cummings TG, Worley CG. Organizational Development & Change. Ninth Edition. Mason, OH, USA: Cengage Learning; 2009.
27. Sherman DW. Nurses' stress & burnout. How to care for yourself when caring for patients and their families experiencing life-threatening illness. Am J Nurs 2004 May;104(5):48-56; quiz 57. [doi: 10.1097/00000446-200405000-00020] [Medline: 15166716]
28. Fisher CD, To ML. Using experience sampling methodology in organizational behavior. J Organ Behav 2012;33(7):865-877. [doi: 10.1002/job.1803]
29. Pantelopoulos A, Bourbakis N. A survey on wearable biosensor systems for health monitoring. Conf Proc IEEE Eng Med Biol Soc 2008;2008:4887-4890. [doi: 10.1109/IEMBS.2008.4650309] [Medline: 19163812]
30. Broyles LM, Rodriguez KL, Price PA, Bayliss NK, Sevick MA. Overcoming barriers to the recruitment of nurses as participants in health care research. Qual Health Res 2011 Dec;21(12):1705-1718. [doi: 10.1177/1049732311417727] [Medline: 21844286]
31. Lindsay J. Getting the numbers: the unacknowledged work in recruiting for survey research. Field Methods 2005;17(1):119-128. [doi: 10.1177/1525822X04271028]
32. Peck LE, Sharpe PA, Burroughs EL, Granner ML. Recruitment strategies and costs for a community-based physical activity program. Health Promot Pract 2008 Apr;9(2):191-198. [doi: 10.1177/1524839906292819] [Medline: 17494948]
33. Eisma R, Dickinson A, Goodman-Deane J, Mival O, Syme A, Tiwari L. Mutual inspiration in the development of new technology for older people. In: Mutual inspiration in the development of new technology for older people Jan 1. 2003 Presented at: Proceedings of Include; 2003; London, UK p. 25-28.
34. Eaton C, Plaisant C, Drizd T. Visualizing missing data: graph interpretation user study. In: Human-Computer Interaction - INTERACT 2005. Berlin Heidelberg: Springer; 2005:861-872.
35. Feng T, Nadarajan A, Vaz C, Booth B, Narayanan S. TILES Audio Recorder: An Unobtrusive Wearable Solution to Track Audio Activity. In: Proceedings of the 4th ACM Workshop on Wearable Systems and Applications. 2018 Presented at: WearSys'18; June 10-10, 2018; Munich, Germany p. 33-38. [doi: 10.1145/3211960.3211975]
36. Musthag M, Raij A, Ganesan D, Kumar S, Shiffman S. Exploring Micro-Incentive Strategies for Participant Compensation in High-Burden Studies. In: Proceedings of the 13th International Conference on Ubiquitous Computing. 2011 Presented at: UbiComp'11; September 17-21, 2011; Beijing, China p. 435-444. [doi: 10.1145/2030112.2030170]
37. Ananthanarayan S, Sheh M, Chien A, Profita H, Siek K. PT Viz: Towards a Wearable Device for Visualizing Knee Rehabilitation Exercises. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2013 Presented at: CHI'13; April 27-May 2, 2013; Paris, France p. 1247-1250. [doi: 10.1145/2470654.2466161]
38. Intille SS, Lester J, Sallis JF, Duncan G. New horizons in sensor development. Med Sci Sports Exerc 2012 Jan;44(1 Suppl 1):S24-S31 [FREE Full text] [doi: 10.1249/MSS.0b013e3182399c7d] [Medline: 22157771]

39. Bonnici T, Orphanidou C, Vallance D, Darrell A, Tarassenko L. Testing of Wearable Monitors in a Real-World Hospital Environment: What Lessons Can Be Learnt? In: Proceedings of the 2012 Ninth International Conference on Wearable and Implantable Body Sensor Networks. 2012 Presented at: BSN'12; May 9-12, 2012; London, UK p. 79-84. [doi: 10.1109/bsn.2012.31]
40. Tait L. Dr Lynda Tait. 2014. How to Build a Strong Research Team: Tips on What Funders Are Looking for. URL: https://www.lyndatait.com/blog/2014/2/25/how-to-build-a-strong-research-team-tips-on-what-funders-are-looking-for [accessed 2019-07-15]
41. Lawler EE. The new pay: a strategic approach. Compens Benefits Rev 1995;27(4):14-22. [doi: 10.1177/088636879502700404]
42. Ruppert T, Abonyi J. Software sensor for activity-time monitoring and fault detection in production lines. Sensors (Basel) 2018 Jul 19;18(7):pii: E2346 [FREE Full text] [doi: 10.3390/s18072346] [Medline: 30029510]
43. Kakria P, Tripathi NK, Kitipawang P. A real-time health monitoring system for remote cardiac patients using smartphone and wearable sensors. Int J Telemed Appl 2015;2015:373474 [FREE Full text] [doi: 10.1155/2015/373474] [Medline: 26788055]
44. Fletcher R, Poh MZ, Eydgahi H. Wearable sensors: opportunities and challenges for low-cost health care. Conf Proc IEEE Eng Med Biol Soc 2010;2010:1763-1766. [doi: 10.1109/IEMBS.2010.5626734] [Medline: 21096416]
45. Fensli R, Pedersen PE, Gundersen T, Hejlesen O. Sensor acceptance model – measuring patient acceptance of wearable sensors. Methods Inf Med 2018 Jan 19;47(01):89-95. [doi: 10.3414/ME9106]
46. Maharatna K, Mazomenos E, Morgan J, Bonfiglio S. Towards the Development of Next-generation Remote Healthcare System: Some Practical Considerations. In: Proceedings of the 2012 IEEE International Symposium on Circuits and Systems. 2012 Presented at: ISCAS'12; May 20-23, 2012; Seoul, South Korea. [doi: 10.1109/iscas.2012.6270390]
47. Liang X, Barua M, Chen L, Lu R, Shen X, Li X, et al. Enabling pervasive healthcare through continuous remote health monitoring. IEEE Wireless Commun 2012;19(6):10-18. [doi: 10.1109/MWC.2012.6393513]

Abbreviations

IARPA: Intelligence Advanced Research Projects Activity
ODNI: Office of the Director of National Intelligence
TILES: Tracking Individual Performance with Sensors

Edited by G Eysenbach; submitted 10.01.19; peer-reviewed by C Fisher, B Johnson, K Huckvale; comments to author 27.04.19; revised version received 12.08.19; accepted 01.10.19; published 10.12.19.

Please cite as:
L'Hommedieu M, L'Hommedieu J, Begay C, Schenone A, Dimitropoulou L, Margolin G, Falk T, Ferrara E, Lerman K, Narayanan S Lessons Learned: Recommendations For Implementing a Longitudinal Study Using Wearable and Environmental Sensors in a Health Care Organization
JMIR Mhealth Uhealth 2019;7(12):e13305
URL: https://mhealth.jmir.org/2019/12/e13305
doi:10.2196/13305
PMID:31821155

Review

Digital Health Behavior Change Technology: Bibliometric and Scoping Review of Two Decades of Research

Fawad Taj[1]; Michel C A Klein[1], PhD; Aart van Halteren[1,2], PhD

[1]Vrije Universiteit Amsterdam, Amsterdam, Netherlands

[2]Philips Research, Eindhoven, Netherlands

Corresponding Author:
Fawad Taj
Vrije Universiteit Amsterdam
De Boelelaan 1105
Amsterdam, 1081 HV
Netherlands
Phone: 31 0205982310
Email: f.taj@vu.nl

Abstract

Background: Research on digital technology to change health behavior has increased enormously in recent decades. Due to the interdisciplinary nature of this topic, knowledge and technologies from different research areas are required. Up to now, it is not clear how the knowledge from those fields is combined in actual applications. A comprehensive analysis that systematically maps and explores the use of knowledge within this emerging interdisciplinary field is required.

Objective: This study aims to provide an overview of the research area around the design and development of digital technologies for health behavior change and to explore trends and patterns.

Methods: A bibliometric analysis is used to provide an overview of the field, and a scoping review is presented to identify the trends and possible gaps. The study is based on the publications related to persuasive technologies and health behavior change in the last 18 years, as indexed by the Web of Science and Scopus (317 and 314 articles, respectively). In the first part, regional and time-based publishing trends; research fields and keyword co-occurrence networks; influential journals; and collaboration network between influential authors, countries, and institutions are examined. In the second part, the behavioral domains, technological means and theoretical foundations are investigated via a scoping review.

Results: The literature reviewed shows a clear and emerging trend after 2001 in technology-based behavior change, which grew exponentially after the introduction of the smartphone around 2009. Authors from the United States, Europe, and Australia have the highest number of publications in the field. The three most active research areas are computer science, public and occupational health, and psychology. The keyword “mhealth” was the dominant term and predominantly used together with the term “physical activity” and “ehealth”. A total of three strong clusters of coauthors have been found. Nearly half of the total reported papers were published in three journals. The United States, the United Kingdom, and the Netherlands have the highest degree of author collaboration and a strong institutional network. Mobile phones were most often used as a technology platform, regardless of the targeted behavioral domain. Physical activity and healthy eating were the most frequently targeted behavioral domains. Most articles did not report about the behavior change techniques that were applied. Among the reported behavior change techniques, goal setting and self-management were the most frequently reported.

Conclusions: Closer cooperation and interaction between behavioral sciences and technological areas is needed, so that theoretical knowledge and new technological advancements are better connected in actual applications. Eventually, this could result in a larger societal impact, an increase of the effectiveness of digital technologies for health behavioral change, and more insight in the relationship between behavioral change strategies and persuasive technologies' effectiveness.

(JMIR Mhealth Uhealth 2019;7(12):e13311) doi:10.2196/13311

KEYWORDS

persuasive technology; digital health behavior; behavior change systems; behavior change support systems; bibliometric analysis; scoping review

Introduction

In the past two decades, researchers have spent a lot of effort to understand how digital technology can help people to make a positive change in their behavior. This research is often motivated by the increasing cost of our health care systems and the increasing demand for health care professionals. The field of influencing or changing human behavior through digital technologies started with the term *persuasive technology* (PT) around 1998. Fogg [1] states that persuasion is more than just computer-mediated communication but focuses on human-computer interaction. He defined PT as "how people can be persuaded when interacting with the technology and adjusting itself according to the actions, inputs, and context of persuaded party" [1]. Over time, many terms have emerged to describe technology-based behavior change interventions. For example, Oinas-Kukkonen and Harjumaa [2] define *persuasive systems* as "a computerized software or information system designed to reinforce, change, or shape attitude or behavior or both, without using coercion or deception". Later, from system perspective, Oinas-Kukkonen [3] defined the term *behavior change support systems* (BCSS): "a sociotechnical information system with psychological and behavioral outcomes designed to form, alter, or reinforce attitudes, behaviors or an act of complying without using coercion or deception". In this paper, we will use the terms "persuasive technology," "behavior change support systems," and "digital health interventions" intermixed.

Often, behavior change support systems are used to support individuals in making lifestyle changes that will lead to better health. When a sufficiently large portion of the population starts making positive changes to health-related behaviors, this will lead to lower utilization of health care and, eventually, to a significant reduction of health care expenditure. Although the potential and societal impact of digital health interventions is highly attractive, so far, its impact on health care utilization and expenditure has been minimal. This is partly because of the limited deployment and implementation, in terms of potential beneficiary users, of effective digital technology for behavior change.

The development of BCSS for health requires an interdisciplinary approach. For example, social psychology provides us a useful theory about the role of people's personalities, cognitions, and social environment, which system designers could use to better design and develop an effective system. The social science theories provide an accumulated understanding of what human behavior is and the contexts in which they occur, what are the mechanisms of action for change, and what are the ingredients required for change [4]. However, such theories are not often used when designing an intervention [4]. Note that the question *whether an intervention based on theory is more effective* is still under discussion; some reviews have found a positive association [5], whereas others found a negative association [6]. In this study, we tried, among other goals, to explore the adoption theoretical knowledge from related behavioral and psychological sciences, but we did not investigate the effectiveness itself.

In behavioral sciences, 92-item taxonomy of behavior change techniques (BCTs) has been developed to better report about behavior change interventions [7] and support their development. Moreover, some popular models and frameworks exist that provide a systematic procedure for understanding and changing behaviors. For example, Fogg [8] introduced a model called Fogg behavioral model that explains why a person does what he does and also defines functional triads of a persuasive system. Harjumaa extended Fogg's study and created a conceptual framework called *persuasive system design* (PSD) that can be directly applied for persuasive system development and evaluation [9,10]. Michie et al [11] combined 19 frameworks to propose a new framework called the *behavior change wheel* that comprises a layered structure that explains behavior with 3 components (capability, motivation, and opportunity) and links this to techniques known to change behavior and the concerned policy. There are some more latest models that combines the other well-established models and theories to translate clinical aims into behavioral strategies and endeavor to include the full scope of elements from behavioral principles to technological features [12,13].

To our knowledge, no bibliometric analysis of scientific literature on digital technologies for health behavior change has been published. However, a lot of research has been conducted on different aspects of PT and health domains, for example, the literature about the persuasive design principle in different technology domains [14-17] or systematic reviews about the effectiveness and prevalence of the persuasive systems [18-20]. Only a few studies evaluated the effectiveness of PT for health and well-being; it was hard to establish long-term effectiveness because of the lack of long-term evaluation methods.

Our study is related to the papers mentioned above, but in our study, the main focus is the application domains and the theoretical basis (eg, behavioral theories and BCTs) of actual BCSS. A better understanding of the main application domains, the usage of theories, and the gaps between them will help identify which areas of PT have sufficient evidence that implementation on a larger scale can be justified and which areas still require additional research.

In this study, we first conducted a bibliometric analysis of the literature to answer the questions about the quantitative trends in the literature and the geographical distribution of the researchers. With the help of a co-occurrence network of keywords and research fields, we tried to uncover meaningful insights based on the strength of links between the nodes in the network. We also studied the collaboration between scholars in the field and the collaboration between developers of digital interventions and health behavior change researchers. This is done at the author, institution, and country level.

Second, we have presented a scoping review that aims to critically evaluate the content of the published literature on digital health behavior change systems and answer the following questions: (1) *what are the trends in adopted technologies for different health domains*? and (2) *what are the theoretical foundations (ie, theories from behavioral sciences or systematic frameworks/models for the development of PT) of implemented systems?*

Finally, we concluded with a discussion on the identified pitfalls and possible future directions for research.

The information provided by this review will help analyze who are the relevant stakeholders that have contributed to the growing knowledge of digital health interventions and will help designers make more informed decisions regarding the development and design of PT for healthy behavior change.

Methods

This section describes the procedures that have been used in the different phases of our study. We discuss the query selection, the database selection, methodologies, eligibility criteria, data items, and the choice of the tools used for different types of analysis.

Data Collection

We chose to use the Web of Science (WoS) core collection and the Scopus database as basis for the study. The search query aimed to retrieve actually implemented behavior change systems with a technical component. Therefore, the following terms were used: ("persuas*" OR "ehealth" OR "mhealth") AND ("ICT" OR "prototype" OR "techno*" OR "system") AND "behavio* change." The search was limited to the articles publicated during January 2000 and December 2018. The process of merging the collected data is discussed below. The query search resulted in a set of 317 and 325 articles from WoS and Scopus, respectively. Among the 325 items from Scopus, 11 appeared empty or incomplete at further inspection, so 314 were left.

Study Design

For evaluating the existing scientific output through citations count, keywords, geographical data, authors collaborations, and discipline-wise interactions, the method of bibliometric analysis was used. The basic statistics included yearly publication output, publishing countries, the field of study, citation count, keyword co-occurrences, coauthorships, and collaboration networks between countries and institutions.

Several software packages are available to support a bibliometric analysis, each with different capabilities and limitations. Some of the most popular tools include HistCite [21], CiteSpace [22], BibExcel [23], and Science of Science (Sci2). In this study, the Sci2 tool was used, which is based on the Cyberinfrastructure Shell toolset, for the study of science [24].

In addition to tools for bibliometric analysis, we also used network visualization and analysis tools for the co-occurrence networks of keywords and research fields and the collaboration networks between countries and institutions. We used Gephi and VOSviewer that use a 3-dimensional render engine to render illustrations of large networks [25,26]. For the network analyses, the 394 extracted articles and their references were converted into graphs. For example, for the coauthor network, 394 articles resulted in 1777 nodes, where each node represented an author, and 5583 edges, where each edge represented coauthored studies.

A scoping review is a useful methodology to determine the coverage of a body of literature on a given topic and identify and analyze knowledge gaps [27]. A number of research questions in the introduction section were defined that will help investigate the use of behavior change theories and BCTs in actual BCSS.

Owing to the broad scope of the review, any conventional systematic review or meta-analysis framework was not followed strictly. However, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses for Scoping Reviews guidelines (PRISMA-Scr) were applied [28] for our scoping review.

Eligibility Criteria

To be considered for review, the following inclusion criteria were formulated: (1) publication in a peer-reviewed academic journal; (2) publication in the English language; (3) the research should have the primary purpose of changing behavior, either increasing or decreasing the behavior or stopping the behavior altogether; and (4) the article should discuss a technological solution that is used for behavior change.

Categorization

In this section, the key categories used in this study to classify the different interventions described in our dataset have been defined. The first category, *technology*, refers to the type of computer-based tool that is used to change a behavior, that is, a digital device, hardware, or a software solution. The *target domain* describes the behaviors that are targeted for change—for example, diet, sedentary behavior, mental health, or physical activity. The term *behavior change theory/model* is used to describe the theories and models about behavior change from behavioral sciences that are used as the basis for the persuasive systems. The category *development frameworks/models* covers a slightly different aspect; it describes the frameworks and models that are used to guide the design and development of the intervention. To further clarify the distinction between the two categories discussed above, the behavior change theory/model is used as the key behavior change principle for approaching the problem. For example, in a study by Lyons et al [29], the authors approached narrative transportation theory as a basis for increasing autonomous motivation about health behavior. In contrast, elements of the development frameworks category provide a systematic approach to select the behavioral principles and technological features during the design and development of the system. Finally, the category *BCT* describes the active ingredients of an intervention, for example, goal setting, reminder, and education. For this category, the taxonomy found in the study by Michie et al [7] was used.

Table 1 provides an overview of the categories and some examples of values. The categorization of the systems has been performed by the first author and finalized with the approval of other authors.

Table 1. Classification and coding scheme.

Category	Possible values
Technology	Web, mobile app, computer applications, mobile game, SMS, pedometers, virtual agent, and interactive voice response
Target domain	Physical activity, healthy eating, smoking cessation, carbon emission, and energy consumption
Theories/model employed for behavior change	Transtheoretical model, motivational interviewing, health belief model, and social cognitive theory
Development frameworks/model	Persuasive system design, behavior intervention technology, intervention mapping, and behavior change wheels
Behavior change techniques	Self-monitoring, motivation, goal setting, reward, punishment, and knowledge

Results

Data Source and Selection

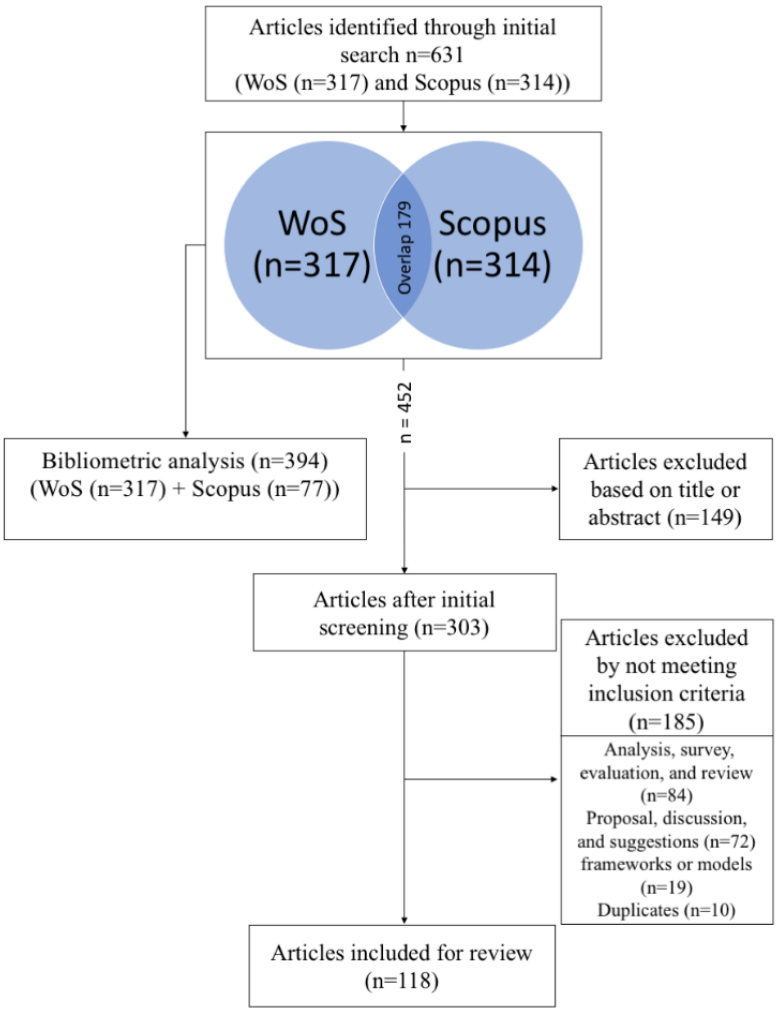
We applied our search query on 2 major literature databases, that is, WoS and Scopus. The WoS and Scopus databases returned 317 and 314 articles, respectively. There was an overlap of 179 articles between both datasets; thus, we identified 452 unique articles.

To perform our *bibliometric analysis*, we ideally would have merged the data from WoS with the data from Scopus. However, there are some technical constraints when performing a bibliometric analysis on different datasets: first, each database uses different sources for indexing the articles and, second, each database starts to index different journals at a different moment in time. For example, WoS started indexing the *Journal of Medical Internet Research* and its sister journal articles in 2015, whereas Scopus started in 2018 and is still indexing only a few of the journals published by JMIR Publications. Another issue is that the citation count for each article is different in the different databases, apparently, because different sources are used.

Owing to its quality and completeness of data, we decided to base our bibliometric analyses on data extracted from the WoS database [30,31]. To include as many relevant papers as possible, a manual search was performed in the WoS database for the 135 (314 minus 179 overlap) Scopus results that were not returned by the search query in WoS. Of those 135 articles, 77 were found in the WoS database, and 58 were not found. This resulted in a dataset of 394 items (317 as results to a query and 77 manually added items) for the bibliometric analysis.

Our *scoping review* started from the full set of 452 unique articles. A 2-phase screening was performed to determine their relevance for our scoping review. The first author initially screened the articles by reading the title and abstract and removing those that clearly did not match the inclusion criteria. As a result, 149 articles were excluded. The remaining 303 articles were thoroughly studied by the main author and compared with the inclusion criteria, which led to the exclusion of another 175 articles. The excluded articles were characterized in 3 main categories through group discussion. Furthermore, 10 articles were found as duplicates, as they were discussing the same intervention. Eventually, 118 articles were considered for the scoping review. The detailed flow of the selection process is presented in Figure 1.

Figure 1. Flowchart of study selection (n=the actual number of publications). WoS: Web of Science.



Bibliometric Analysis

Growth of the Literature and Distribution Over Countries

When looking at the number of publications in our selection, we can observe that it started in 2001, with 1 or 2 publications yearly, but rapidly increased since 2009 to 2010, as can be seen in Figure 2.

Table 2 shows the top 10 regions where publications originated. The United States is leading with 46.5% (183/394 records) of all contributions, followed by England and the Netherlands with 14.5% (57/394 records) and 8.4% (33/394 records), respectively. Furthermore, Australia and Canada are fourth and fifth in the ranking with 7.6% (30/394 records) and 5.8% (23/394 records), respectively.

Figure 2. Publication trend from 2000 to 2018.

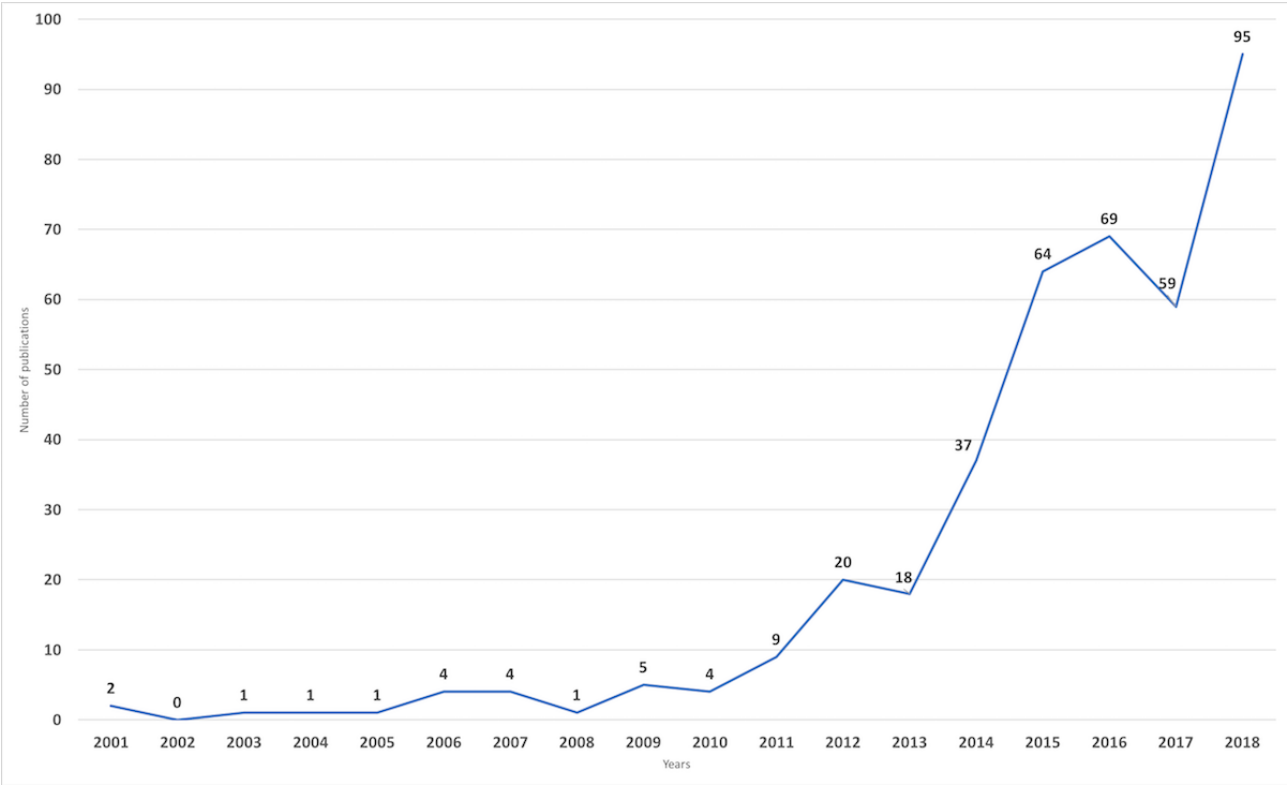


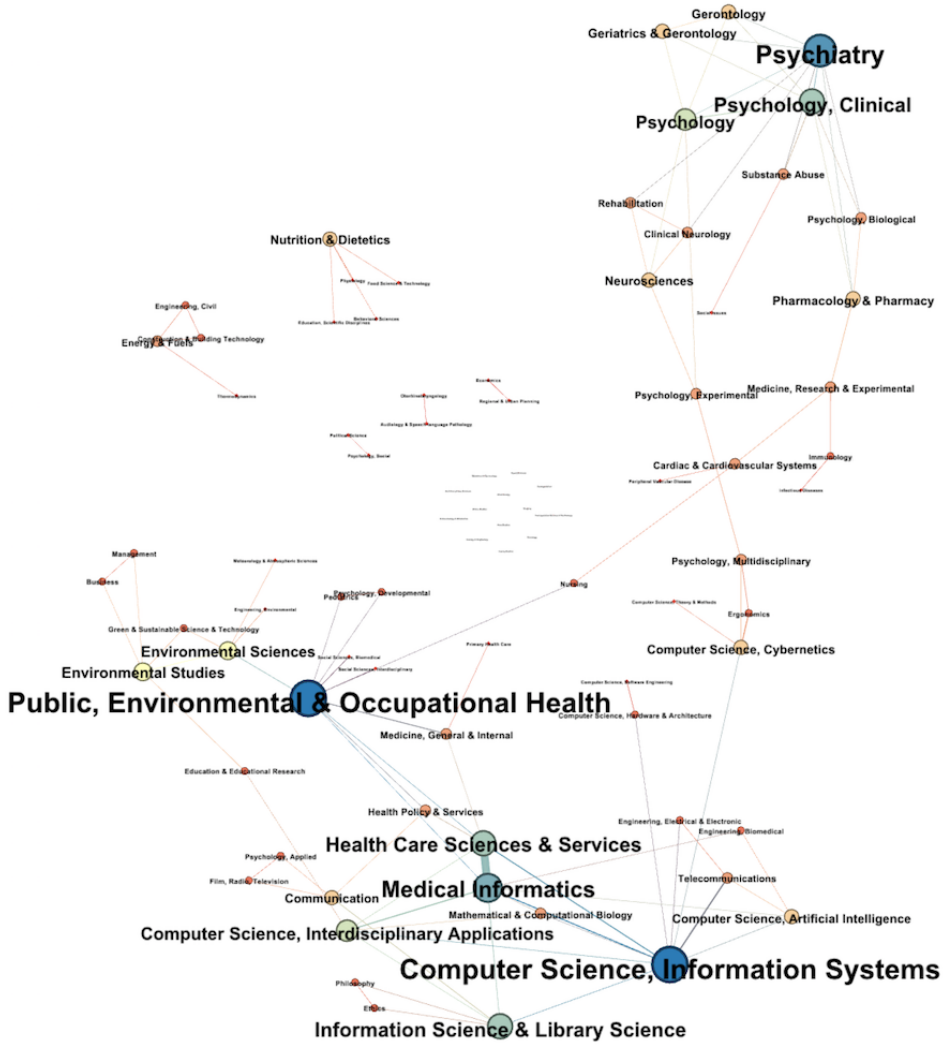
Table 2. Persuasive technology for behavior change, scholarly papers by region.

Country	Publications, n (%)
United States	183 (46)
England	57 (14)
The Netherlands	33 (8)
Australia	30 (7)
Canada	23 (6)
New Zealand	19 (5)
Finland	17 (4)
Italy	16 (4)
Belgium	13 (3)
Switzerland	11 (2)

Interdisciplinary Collaboration Network

From the *research field* column in the WoS database, a co-occurrence network of research areas has been created. The weights of the nodes in the graph were determined by the number of publications in each given category. The visualization of this network (see Figure 3) shows 3 clear clusters. The first cluster is the area of computer science and related fields, the second cluster is public health care and occupational health, and the third major cluster is psychology and subfields.

Figure 3. Disciplines involved in persuasive technology and health behavior change.



Citations

Tables 3 and 4 show the list of the 5 most cited articles, either globally or locally. The global citation count is based on the full WoS citation count, whereas the local citation count is the number of citations of an article within the network.

Table 3. Top 5 of the globally most citied articles.

Title	Reference	Global citation count
A behavior change model for internet interventions	[32]	235
New directions in electronic health communication: opportunities and challenges	[33]	205
Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis	[34]	162
Virtual self-modeling: the effects of vicarious reinforcement and identification on exercise behaviors	[35]	151
Online interventions for social marketing health behavior change campaigns: meta-analysis of psychological architectures and adherence factors	[36]	139

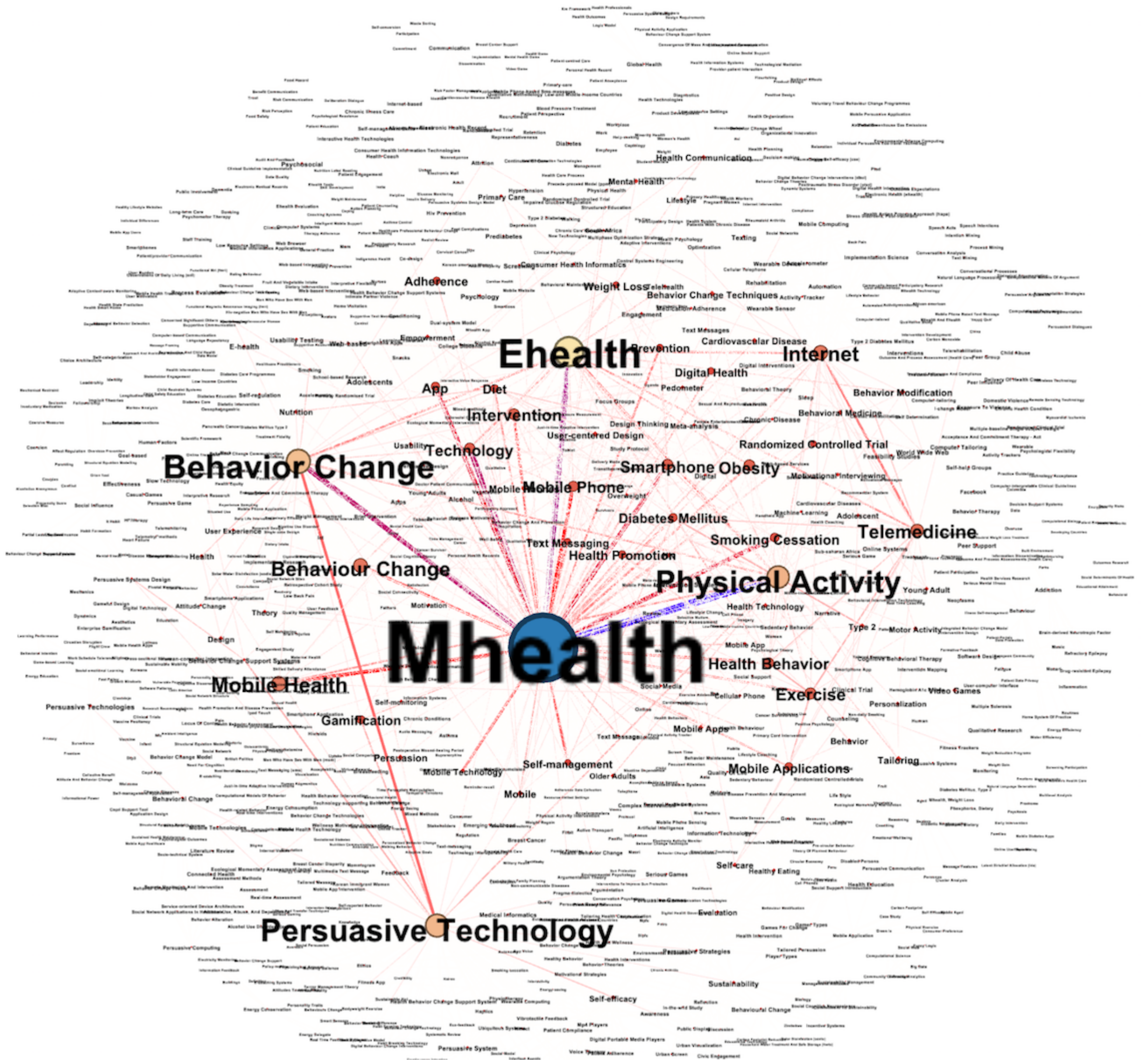
Table 4. Top 5 of the most citied articles within the network (locally).

Title	Reference	Local citation count
Health behavior models in the age of mobile interventions: are our theories up to the task?	[37]	45
Behavior change interventions delivered by mobile telephone short message service	[38]	44
Text messaging as a tool for behavior change in disease prevention and management	[39]	43
The theory of planned behavior	[40]	41
Persuasive technology: using computers to change what we think and do	[41]	41

Co-occurrence of Keywords

keywords are assumed to compose an adequate description of the content of a research article. The co-occurrence of keywords could provide an interesting structure of the research field, as it reveals the semantic relations in the scientific literature. The most frequently used keywords are "mhealth," "physical activity," "ehealth," "persuasive technology," "smart phone," and "behavior change" (see Figure 4). The size of the node and label reflects the co-occurrence count of a certain word. The higher the count, the larger the size of node and label.

Figure 4. The co-occurrence network of author keywords.

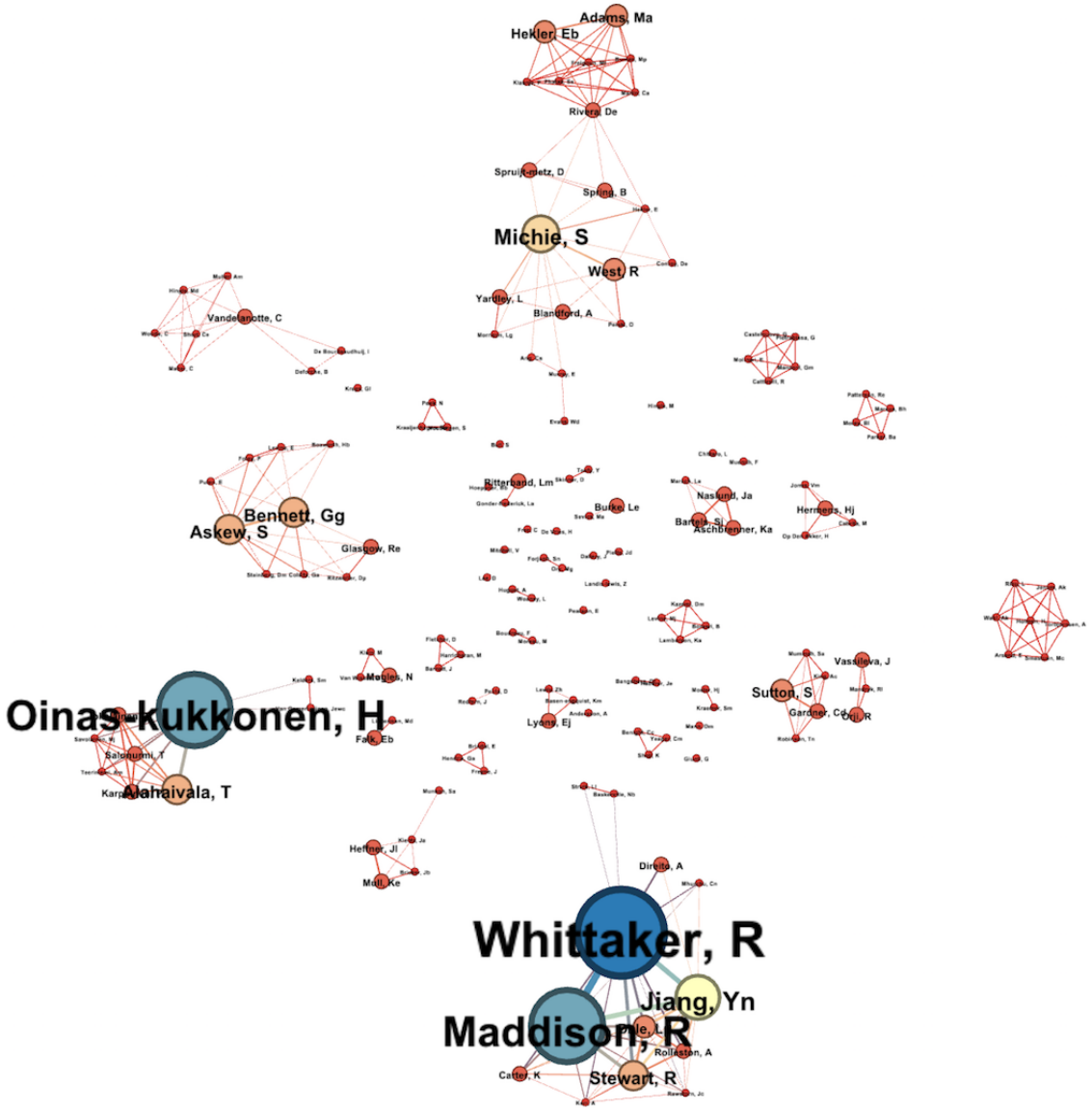


Coauthorship Network

A coauthorship network was extracted using the Sci2 tool. Each node represented an author, and a connection between 2 nodes represented a coauthorship. A total of 1777 authors were identified with 5583 connections. For better visualization, only authors who had at least two publications together were considered. Node sizes were based on the number of articles coauthored with other authors. There were 3 strong clusters among the network that were strongly intraconnected but not

interconnected (see Figure 5). In Figure 5, the highest coauthorship count (13 times) within the network was by R Whittaker. Her major coauthorship is with colleagues R Maddison and Y Jiang at the National Institute for Health Innovation, University of Auckland, Auckland, New Zealand. The other cluster is led by Oinas-Kukkonen, with a coauthorship count of 11 and most coauthorships with his colleague, T Alahäivälä, at the University of Oulu, Faculty of Information Technology and Electrical Engineering, Oulu Advanced Research on Service and Information Systems, Oulu, Finland. The last cluster is led by S Michie (Research Department of Clinical, Educational, and Health Psychology, University College London or UCL, London, United Kingdom), with a coauthorship count of 6 and, mostly, with R West (Health Behavior Research Center, UCL Epidemiology and Public Health, London, United Kingdom) and EB Hekler (School of Nutrition and Health Promotion, Arizona State University, Phoenix, Arizona, United States).

Figure 5. Coauthor graph.



Most Important Journals

The articles under review were published in 147 different journals. Table 5 shows the distribution of published articles in the top 6 journals. The leading journals are all related to JMIR Publications. The most often used subjournal is *JMIR mHealth and uHealth* with 45 publications out of 394 (11.4%). The main journal, *Journal of Medical Internet Research*, and the subjournal, *JMIR Research Protocols*, with 34 (8.6%) and 17 (4.3%) are in second and third places, respectively.

Table 5. List of top journal distribution.

Journal title	Publications (n)
JMIR mHealth and uHealth	45
Journal of Medical Internet Research	34
JMIR Research Protocols	17
BioMed Central Public Health	11
Personal and Ubiquitous Computing	10
International Journal of Medical Informatics	9

International Collaboration

To investigate the collaboration between countries and organizations/universities, the geographical distribution of the research was analyzed. Coauthorship networks between countries and organizations/universities were extracted. Figure 6 shows the network of 52 productive countries in 8 clusters; each cluster is represented by different colors. The size of the node represents the number of articles originating from a certain country and the thickness of the connection indicates the number of collaborations between the 2 countries. The largest cluster is the red one with around 18 nodes, led by England. The second biggest cluster is the blue one (12 nodes); the major country in this group is Australia.

Figure 6. Countries collaboration graph.

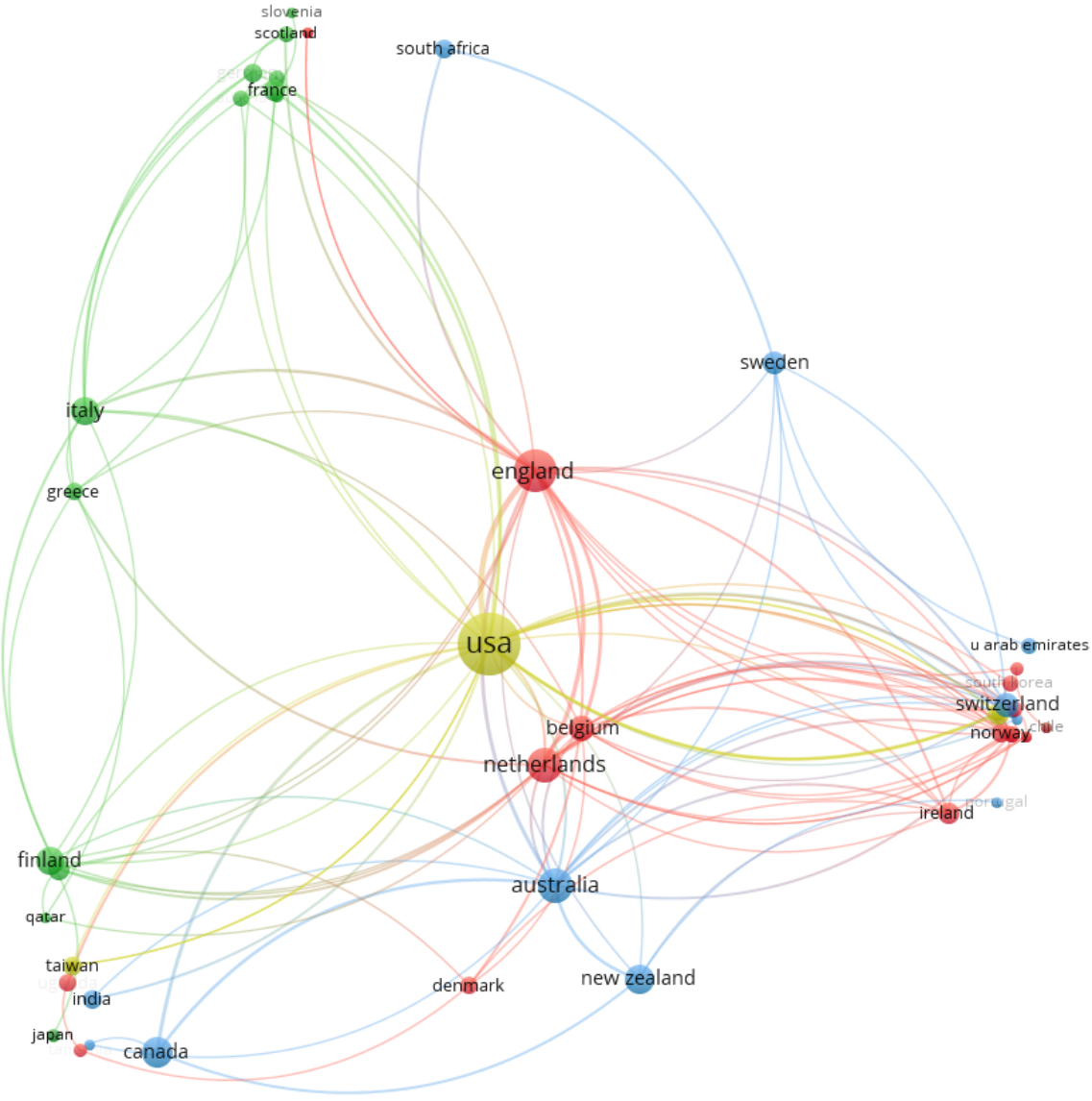
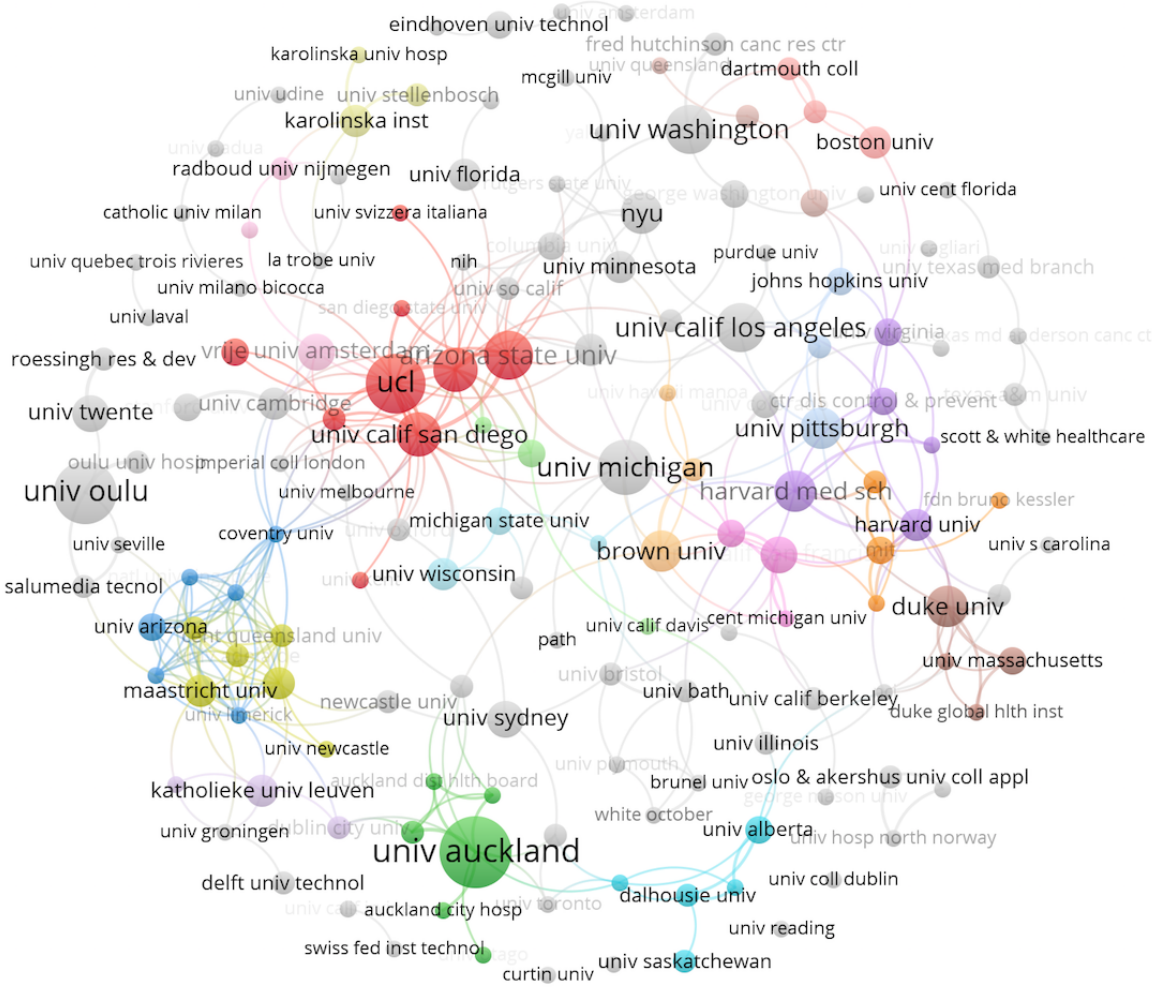


Figure 7 shows the productiveness of institutions and their collaboration with other institutions. There were about 511 organizations identified, but we only considered the organizations that had at least two or more articles published.

This filter left us with 90 institutions, which were divided into 10 clusters, each depicted with a different color. The largest and most productive cluster is the red one with 20 nodes. Most institutions in the cluster are from the United Kingdom and the United States, with UCL and Arizona State University as the leading ones. The second largest cluster is the blue one, with 12 nodes and University of Michigan and University of California as leading institutions. The size of nodes reflects the number of documents, and the thickness of the edge reflects the strength of collaboration.

Figure 7. Organization/institution collaboration graph. Coll: college; Hosp: hospital; Inst: Institute; NYU: New York University; Technol: technology; UCL: University College London; and Univ: University.



Scoping Review

The second objective of this study was a scoping review. The contents of the 118 articles that resulted from our selection process were thoroughly studied. This revealed some interesting insights and trends. The findings about the trends regarding technological choices and the theoretical basis of digital health interventions are presented in the following sections.

Technology Platforms Used

Table 6 summarizes the major technological platforms used for persuasive systems for behavior change. Apps on a mobile phone are the most frequently employed platform with 52.5% (62/118) of the studied systems, followed by SMS and the Web with 21.1% (25) and 19.4% (23), respectively. Less frequently adopted platforms are wearable sensors 16.1% (19), games 6% (8), desktop apps 3% (4), and social media 2% (3). Finally, 9.3% (11) use yet another platform. A complete list of technology usage with references is provided in Multimedia Appendix 1.

Table 6. Frequency of different technological platform used.

Digital technology	Usage count, n (%)
Mobile apps	62 (52)
SMS	25 (21)
Web	23 (19)
Wearable sensors	19 (16)
Others	11 (9)
Game	8 (6)
Desktop apps	4 (3)
Social media	3 (2)

Targeted (Health) Domains

The most often targeted behaviors are health-related behaviors. The top 8 is formed by physical activity, healthy eating, diabetes management, smoking cessation, weight control, AIDS/sexual behavior, cardiovascular diseases, and alcohol consumption. Physical activity makes up 28.8% (34/118) of all the reviewed studies, followed by healthy eating and diabetes with a total of 18.6% (22) and 11% (13), respectively (see Table 7). There are also some interventions targeting multiple health domains, for example, with physical activity for both diabetes control and cardiovascular diseases.

Table 7. Different targeted behavioral domains.

Targeted behavior	Count, n (%)
Physical activity	34 (28)
Healthy eating	22 (18)
Diabetes management	13 (11)
Smoking cessation	10 (8)
Weight control	10 (8)
AIDS or sexual behavior	6 (5)
Cardiovascular disease	5 (4)
Carbon dioxide emission	5 (4)
Energy saving	4 (3)
Alcohol consumption, medical adherence, lower back pain, mental illnesses	3 (2)
Overdose prediction, mammography adherence, asthma control, sedentary behavior, knee osteoarthritis, waste management, educational behavior	2 (1)
Psychotropic, multiple sclerosis, sleeping behavior, screen time	1 (0.8)

Behavior Change Theories

Most of the analyzed articles seem not to be based on proper theories. Only 33% (59/118) of articles reported at least one or more theories among the 21 theories identified during out review for designing the system. The social cognitive theory (SCT), transtheoretical model, self-determination theory, and motivational interviewing are most frequently reported theories. Table 8 shows the top 5 of most frequently used theories. The complete pie chart showing the usage percentages of all reported theories is provided in Multimedia Appendix 1.

Table 8. Percentage of reported theories (N=59).

Theory	Number reported, n (%)
Social cognitive theory	17 (29)
Transtheoretical model	6 (10)
Self-determination theory	4 (7)
Motivational interviewing	4 (7)
Theory of planned behavior	3 (5)

Framework/Model Adopted

Another important finding is about the usage of development frameworks and models. Such frameworks provide guidance on the development of a persuasive system, by suggesting a *coordinated set of activities* to help translate theory into practice [11]. Similar to the result of theories usage, we also see that the usage percentages are quite low; only 47 articles used any framework or model. We found a total of 15 frameworks, but they appeared to be rarely followed (see Table 9). However, it is also found that *gamification* is increasingly used as a paradigm for developing persuasive systems. Gamification is generally understood as the integration of specific features (eg, points, leaderboards, levels, competitions, rewards, and achievements) into the wider context of pursuing a goal [42]. These features are not only used in designing game-based interventions but also equally popular in other technology-based interventions. Among the list of 15 frameworks and models, PSD is a popular development framework with 20% (9/47) usage. The complete pie chart of reported frameworks/models is provided in Multimedia Appendix 1.

Table 9. Usage of different development framework/models (N=47).

Framework/model	Usage percentage, n (%)
Persuasive system design	9 (20)
Gamification	8 (17)
User-centered design	4 (9)
Intervention mapping	4 (9)
BJ Fogg persuasive principles and model	4 (9)
Theoretical domains framework	2 (4)

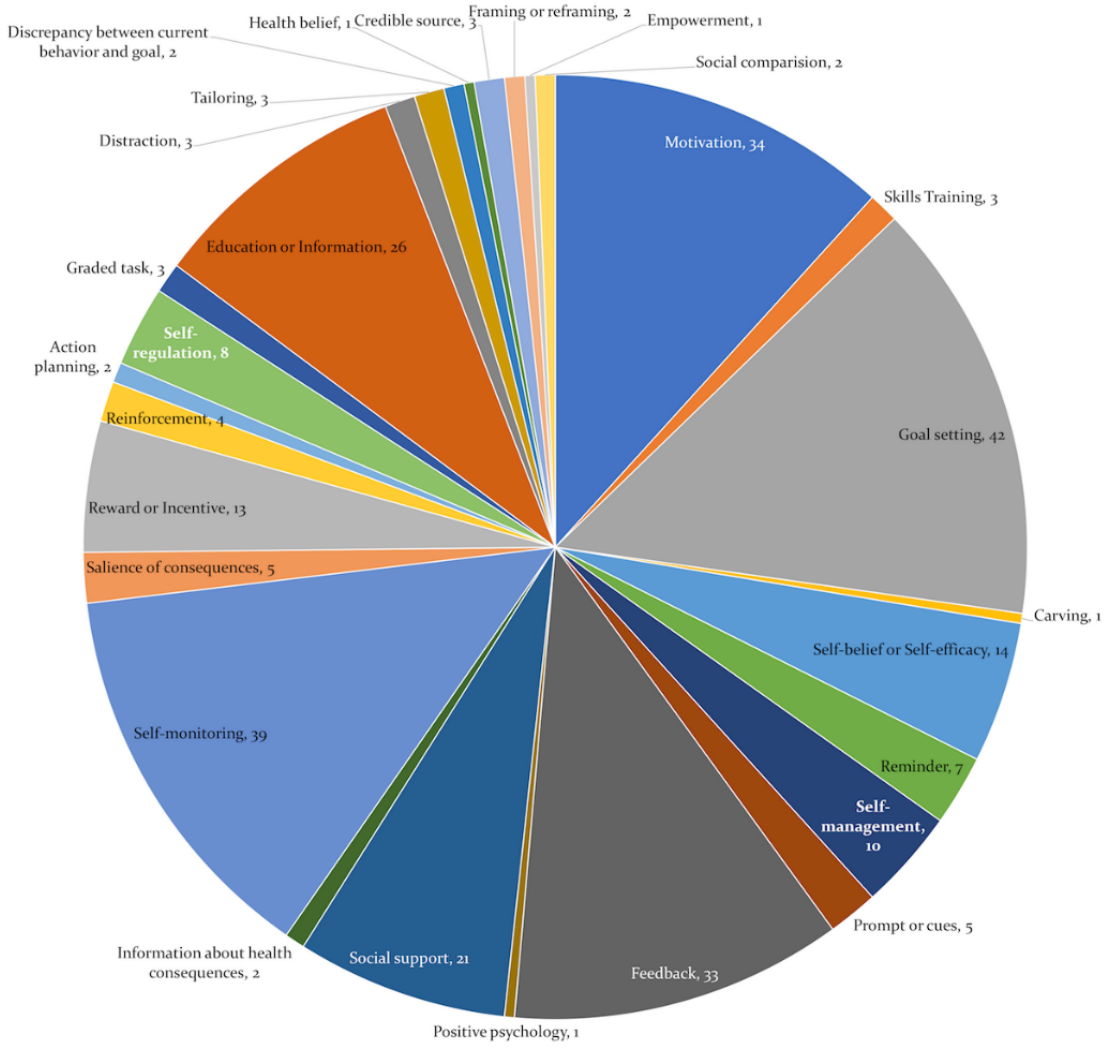
Behavior Change Technique Employed

There are many different BCTs that can be used to induce behavior change. Goal setting is the most frequently employed strategy with a total of 42 studies, followed by self-monitoring and motivation, by 39 and 34 times, respectively. Feedback is used for 33 times (see Figure 8).

It needs to be mentioned that there are no standard guidelines for reporting active components of the interventions, and people used different synonyms to report similar techniques [43]. The recently developed extensively agreed taxonomy of techniques used in behavior change interventions [7] is obviously not reported in papers before 2013 but also in newer papers, it is not often used (maybe partly because it is not widely known [44]). In our review, only 8 articles were identified that used this taxonomy for reporting the BCT adopted in the intervention.

It is relevant to mention that there is some overlap between the psychological constructs and BCTs [43]. Sometimes, articles do not explicitly mention the mechanism of change but do mention the construct they targeted. For example, a study [45] targeted the psychological construct *motivation* through text messaging but did not explicitly mention the type of motivation and what the mechanism of change was. In this part of our analysis, we focused on explicit descriptions of BCTs as mechanism of change. A list of BCTs with references is provided in Multimedia Appendix 1.

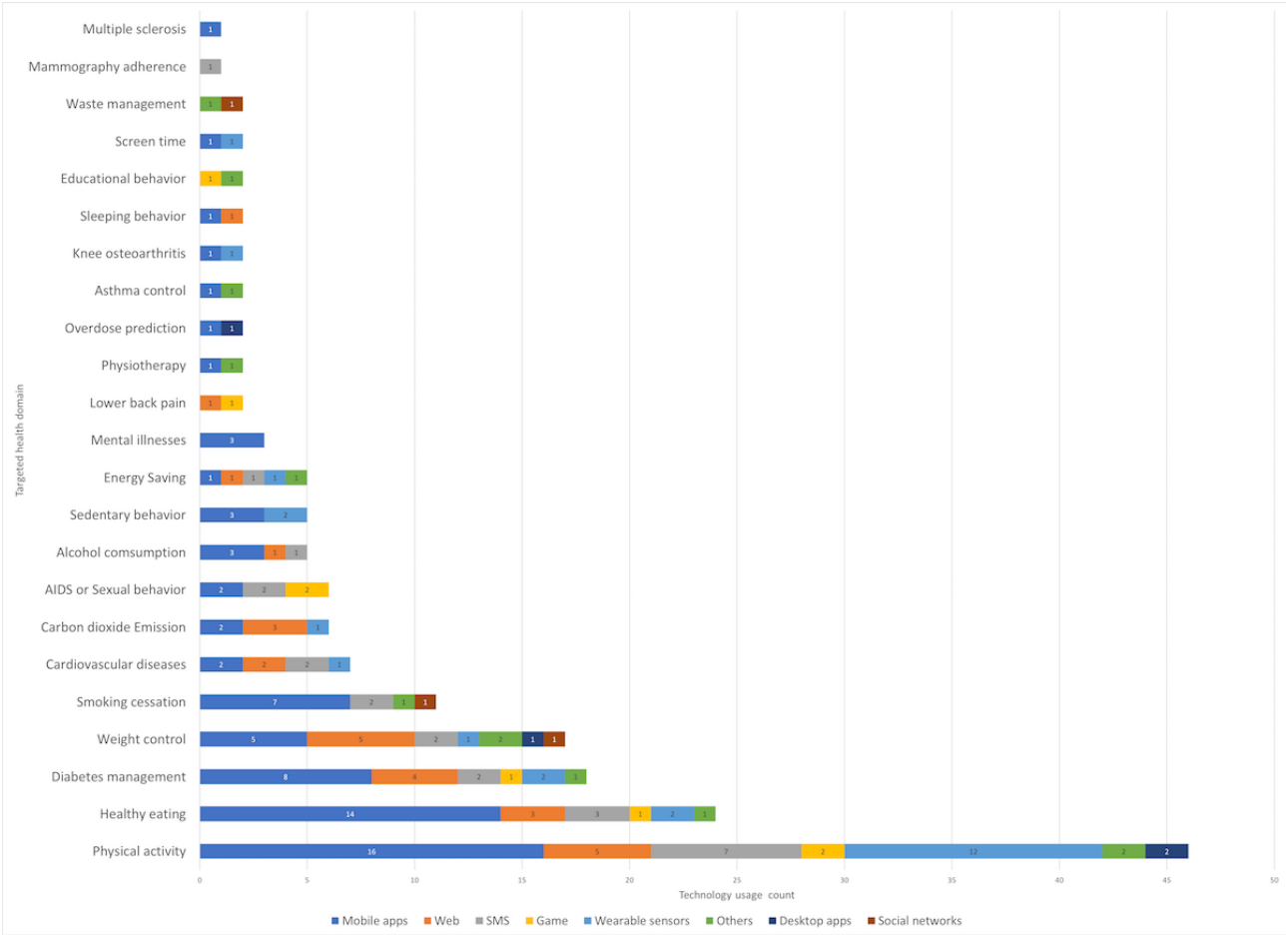
Figure 8. Frequency of different behavior change techniques adopted.



Relating Targeted Behavior With Technologies

After analyzing the usage of BCTs, theories, and technologies previously, we now analyze a combination of them. First, we compare the targeted behavior with the technological platforms that are used. We found that for increasing physical activity, almost all technological platforms were used, whereas for healthy eating, mobile apps were used in more than 80% of the cases (see Figure 9).

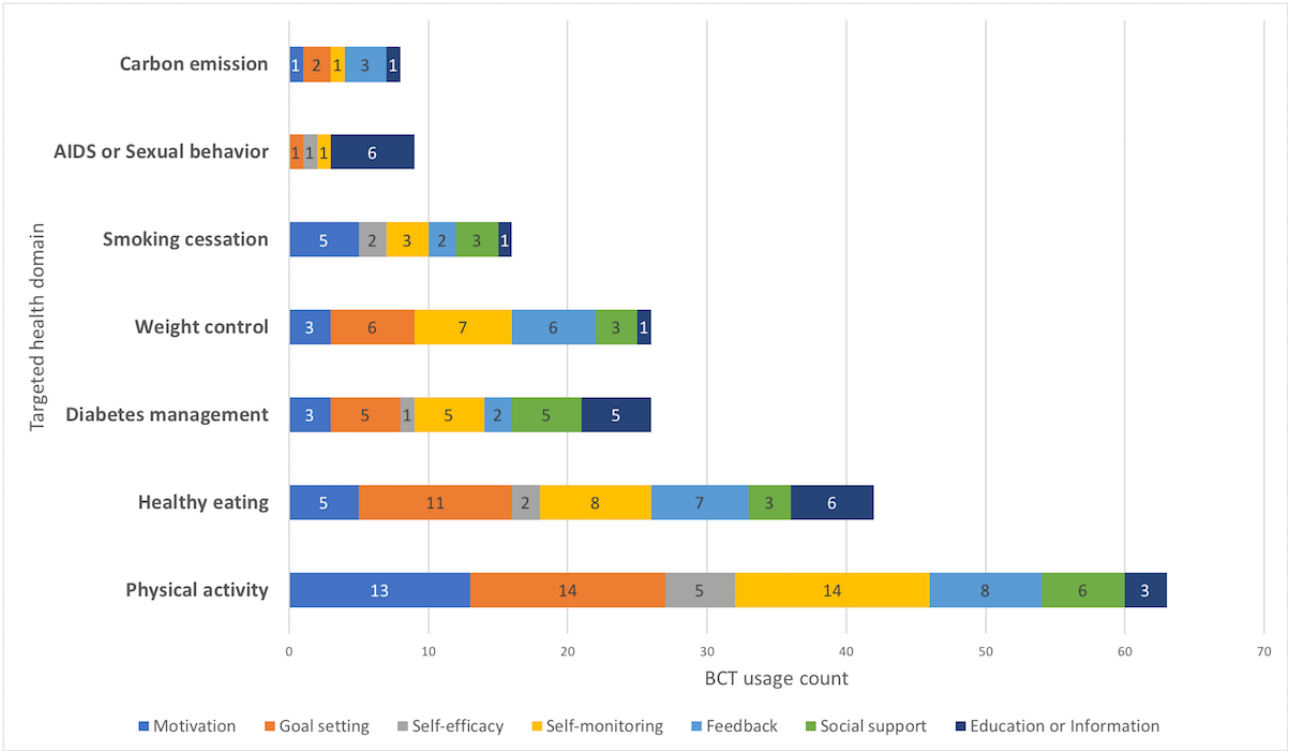
Figure 9. Bar graph representing the different targeted health domains using different technological platforms.



Relating Targeted Behavior With Behavior Change Techniques

Change in physical activity was targeted by a number of different BCTs, mostly by goal setting, self-monitoring, and motivation. Healthy eating was mostly targeted by self-monitoring, goal setting, and feedback (see Figure 10). From the findings, it seems quite hard to make any one-to-one link between behavior and BCT. The choices of BCTs for each intervention and system vary significantly, and each designer decides on his own accord.

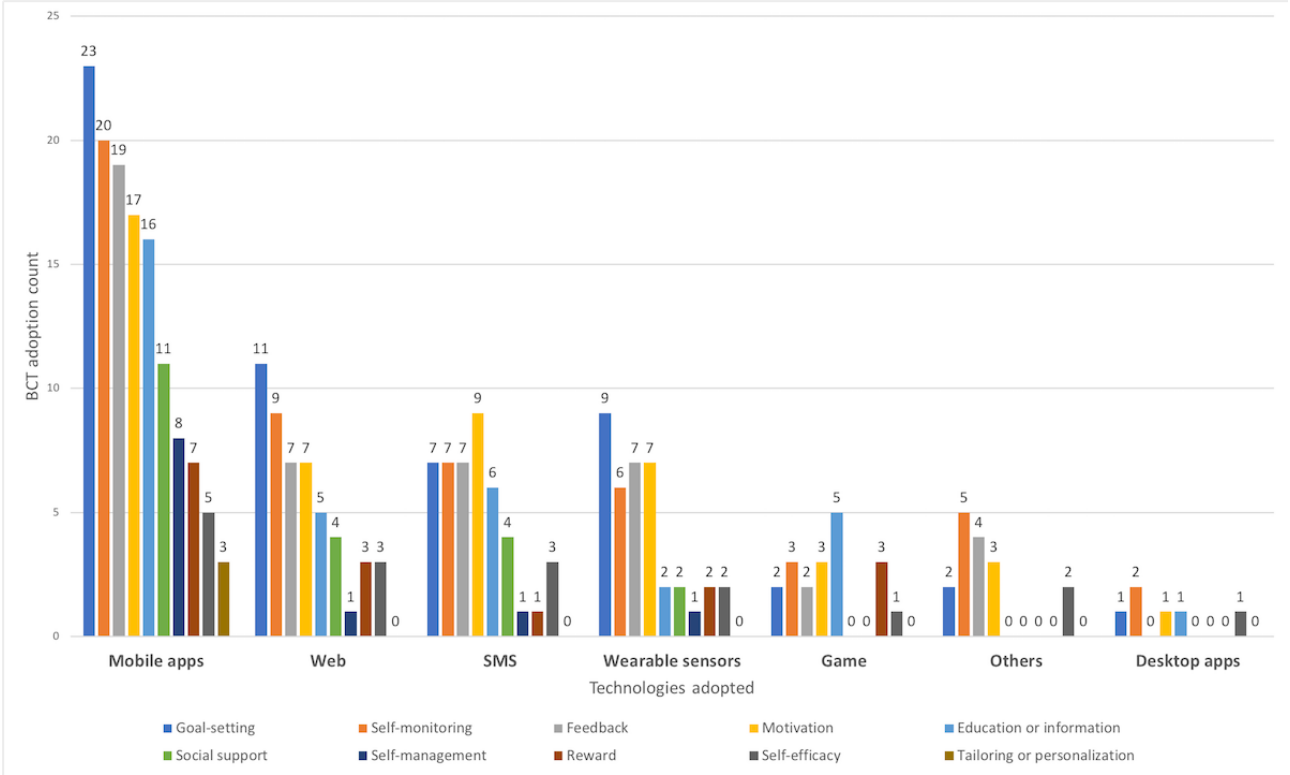
Figure 10. Bar graph represents the different target behavior using different behavior change techniques.



Relating Behavior Change Techniques With Technologies

Some techniques were more frequently applied in one technological platform than in another. In mobile apps, the most frequent strategies are goal setting (22 times), self-monitoring (20 times), and feedback (17 times; see Figure 11). Almost all the major BCTs are used on the mobile platform, probably because of its flexibility and accessibility. When using game, the most frequently used BCTs are education and reward, as those are important features of gamification. Goal setting is almost evenly used in all technological platforms.

Figure 11. Frequency of different behavior change techniques per technological platform.



Discussion

This study presents a comprehensive review of digital technologies and health behavior change. The review comprises 2 parts. First, a thorough bibliometric analysis has been conducted to present the scholarly networks and the global research trends. The bibliometric analysis identifies influential articles, authors, and collaboration networks among different stakeholders and shows where interdisciplinary collaboration is already strong and where further collaboration could strengthen the field. The bibliometric analysis is followed by a scoping review to map the collected literature and answer questions about the theoretical grounding of digital behavior change interventions and the use of technological platforms and targeted domains.

Bibliometric Analysis

The field of PT is still quite young, and the literature regarding persuasive technologies for health and well-being started to appear in early 2000 [46]. The literature has grown exponentially after 2009/2010, which seems to be because of the introduction of smartphones in this era. Given the development around ubiquitous technologies such as sensors and networks, it is reasonable to expect that the growth of the literature will continue. As the field begins to mature, we see the number of publications increasing and researchers from different fields (eg, psychology and behavior sciences) contributing to the field. The geographic distribution of the publications shows that the United States plays a leading role, with a decent contribution from Western European countries and Australia. Other regions such as Asia and Africa almost have no contributions. To be able to include different behaviors and cultures, it would be useful if the collaboration networks are extended to these regions.

Given the fact that technology for behavior change requires expertise from different scientific areas, we expected quite some collaborations between technological and behavioral scientists. This study shows that interdisciplinary collaboration was not as widespread as expected. As human-centered disciplines such as psychology and other behavior sciences are quite mature and can provide essential knowledge about human behavior, technological researches cannot develop effective digital behavior change interventions without their contribution. Similarly, behavioral scientists require knowledge and insights from technological areas to apply their knowledge with modern means.

The keyword network illustrates the most important knowledge structures and thematic evolution in the field of digital behavior change systems. The keyword “mhealth” was strongly connected with the word “physical activity” and “ehealth.” This finding was not so different from the findings of our scoping review. The author collaboration network can be useful for expanding the collaboration network. The network shows a very strong intracollaboration among 3 groups of authors, where one is specifically working on behavior change, another on mobile health interventions, and the last one on persuasive system or BCSS. There is an opportunity to increase the intergroup collaboration, which could add valuable knowledge to the field. For example, with the help of remote sensors and Internet of Things (IoT) devices, health practitioners can monitor and tune systems for better adherence for their target group.

Scoping Review

The translation of theories and theoretical frameworks/models into practice is essential for the development of any intervention. Davis et al [4] already concluded that only a few theories were used frequently. We also identified that the theoretical grounding for most of the systems was weak or, at best, not well described. An often-discussed reason for ignoring the majority of the theories is that intervention developers usually choose the most used theories and consider them as an easy and good choice. Another possible explanation is that the PT designer often lacks the skill to translate the theoretical determinants into technology design artifacts that are as effective as originally intended [4]. Conversely, a psychologist could study how digital technology enables new ways to personalize interventions and deliver just-in-time behavior change, with the potential to develop behavioral theories specific to the digital age.

The creation of a taxonomy of BCTs has been an important development in behavioral science, which is visible in the topmost cited articles in this review. However, these BCTs are still underreported in publications that describe persuasive systems [18]. Drawing on [47], the systematic review on mobile apps found that only 10 out of 93 BCTs were mentioned (mean of 2.42 BCTs were present in each app). Our finding in this review is not so different, only 6 articles explicitly used Michie coded taxonomy of BCTs. The most recent addition at the theoretical level is the link between the mechanism of actions and BCTs; this will probably pace up the development and evaluation of effective behavioral change interventions [48]. Owing to nonstandardized reporting, it is very hard to establish any relationship of effectiveness between 1 strategy and the success of persuasive systems. The result shows that each behavior domain is targeted with several techniques and strategies. Therefore, it is quite hard to determine the effectiveness of a certain technique or theory and the recommended process for design and evaluation of a behavior change support system.

Identified Pitfalls and Future Studies

On the basis our study, we can formulate a number of suggestions for future directions of the research in this domain.

Our study found a large gap in the process of designing digital technologies for health behavior change. They are, usually, weak in their theoretical grounding, and the papers describing them do not clearly report the different components, for example, persuasive strategies, theories, and BCTs. The main reason for this is the lack of design guidelines for these components. For better utilization and reporting of behavioral theories, the development frameworks also need to be updated to the most recent technological advances, for example, the IoT, that is, technologies capable of collecting a large amount of data, such as sensors and mobiles. Furthermore, computational models based on different theories can be designed that could be used by digital intervention developers [13].

Moreover, the relation between BCTs and behavior change theories/mechanisms requires more elaboration. For example, both for the SCT and theory of planned behavior, the use of self-efficacy construct is important. Self-efficacy is the strongest predictor of intention [49]. Furthermore, the correlations between determinants at different ecological levels for different behaviors need to be established. For example, the sitting time in cars has already been associated with urban design [50]. This requires more cross-sectional studies and controlled trials. These studies could help system developers make early decisions about the approach and techniques to follow for their digital interventions.

A clear framework or mechanism for reporting the components of health behavior change systems will not only advance the evaluation and its research design (eg, assess engagement, acceptability, and effectiveness) but also could enhance the costly process of development.

Limitations

There are 2 limitations to our study worth mentioning. First, owing to technical reasons, we only considered the WoS database for bibliometric analysis; 58 papers that were relevant according to Scopus were not included in our analysis. A second limitation is a possible subjectivity in our scoping review. The categorizing of the reviewed papers has been done in a thorough manner but might still be influenced by subjective interpretations. Unfortunately, this problem cannot be avoided in this type of studies.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Overview over behavior change theories, frameworks, models and techniques used in the included persuasive technology of health.
[PDF File (Adobe PDF File), 1081 KB - mhealth_v7i12e13311_app1.pdf]

References

1. Fogg BJ. Persuasive Computers: Perspectives and Research Directions. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 1998 Presented at: CHI'98; April 18-23, 1998; Los Angeles, CA, USA p. 225-232. [doi: 10.1145/274644.274677]
2. Oinas-kukkonen H, Harjumaa M. Towards Deeper Understanding of Persuasion in Software and Information Systems. In: Proceedings of the First International Conference on Advances in Computer-Human Interaction. 2008 Presented at: ACHI'08; February 10-15, 2008; Sainte Luce, Martinique p. 200-205. [doi: 10.1109/ACHI.2008.31]
3. Oinas-Kukkonen H. A foundation for the study of behavior change support systems. Pers Ubiquit Comput 2012;17(6):1223-1235. [doi: 10.1007/s00779-012-0591-5]
4. Davis R, Campbell R, Hildon Z, Hobbs L, Michie S. Theories of behaviour and behaviour change across the social and behavioural sciences: a scoping review. Health Psychol Rev 2015;9(3):323-344 [FREE Full text] [doi: 10.1080/17437199.2014.941722] [Medline: 25104107]
5. Routhe AS, Jones RE, Feldman DL. Using theory to understand public support for collective actions that impact the environment: alleviating water supply problems in a Nonarid Biome*. Soc Sci Q 2005;86(4):874-897. [doi: 10.1111/j.0038-4941.2005.00361.x]
6. Stephenson JM, Imrie J, Sutton SR. Rigorous trials of sexual behaviour interventions in STD/HIV prevention: what can we learn from them? AIDS 2000;14(Suppl 3):S115-S124. [Medline: 11086854]
7. 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]
8. Fogg BJ. A Behavior Model for Persuasive Design. In: Proceedings of the 4th International Conference on Persuasive Technology. 2009 Presented at: Persuasive'09; April 26-29, 2009; Claremont, California, USA. [doi: 10.1145/1541948.1541999]
9. Oinas-Kukkonen H, Harjumaa M. A Systematic Framework for Designing and Evaluating Persuasive Systems. In: Proceedings of the 3rd international conference on Persuasive Technology. 2008 Presented at: PERSUASIVE'08; June 4-6, 2008; Oulu, Finland p. 164-176. [doi: 10.1007/978-3-540-68504-3_15]
10. Torning K, Oinas-Kukkonen H. Persuasive System Design: State of the Art and Future Directions. In: Proceedings of the 4th International Conference on Persuasive Technology. 2009 Presented at: Persuasive'09; April 26-29, 2009; Claremont, California, USA. [doi: 10.1145/1541948.1541989]
11. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]

12. Mohr DC, Schueller SM, Montague E, Burns MN, Rashidi P. The behavioral intervention technology model: an integrated conceptual and technological framework for eHealth and mHealth interventions. J Med Internet Res 2014 Jun 5;16(6):e146 [FREE Full text] [doi: 10.2196/jmir.3077] [Medline: 24905070]
13. Klein M, Mogles N, van Wissen A. Why Won't You Do What's Good for You? Using Intelligent Support for Behavior Change. In: Proceedings of the International Workshop on Human Behavior Understanding. 2011 Presented at: HBU'11; November 16, 2011; Amsterdam, The Netherlands p. 104-115. [doi: 10.1007/978-3-642-25446-8_12]
14. 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]
15. Hamari J, Koivisto J, Pakkanen T. Do Persuasive Technologies Persuade? - A Review of Empirical Studies. In: Proceedings of the 9th International Conference on Persuasive Technology. 2014 Presented at: PERSUASIVE'14; May 21-23, 2014; Padua, Italy p. 118-136. [doi: 10.1007/978-3-319-07127-5_11]
16. Shafin NA, Abdullah NH. Persuasive system practices in mobile application development. J Phys Conf Ser 2018;1049:012025. [doi: 10.1088/1742-6596/1049/1/012025]
17. Win KT, Roberts MR, Oinas-Kukkonen H. Persuasive system features in computer-mediated lifestyle modification interventions for physical activity. Inform Health Soc Care 2018 Oct 23:1-29. [doi: 10.1080/17538157.2018.1511565] [Medline: 30351975]
18. Direito A, Carraça E, Rawstorn J, Whittaker R, Maddison R. mHealth technologies to influence physical activity and sedentary behaviors: behavior change techniques, systematic review and meta-analysis of randomized controlled trials. Ann Behav Med 2017 Apr;51(2):226-239. [doi: 10.1007/s12160-016-9846-0] [Medline: 27757789]
19. Iribarren SJ, Cato K, Falzon L, Stone PW. What is the economic evidence for mHealth? A systematic review of economic evaluations of mHealth solutions. PLoS One 2017;12(2):e0170581 [FREE Full text] [doi: 10.1371/journal.pone.0170581] [Medline: 28152012]
20. Kazemi DM, Borsari B, Levine MJ, Li S, Lamberson KA, Matta LA. A systematic review of the mHealth interventions to prevent alcohol and substance abuse. J Health Commun 2017 May;22(5):413-432 [FREE Full text] [doi: 10.1080/10810730.2017.1303556] [Medline: 28394729]
21. Garfield E. From the science of science to Scientometrics visualizing the history of science with HistCite software. J Inform 2009 Jul;3(3):173-179. [doi: 10.1016/j.joi.2009.03.009]
22. CiteSpace. Visualizing Patterns and Trends in Scientific Literature URL: http://cluster.cis.drexel.edu/~cchen/citespace/ [accessed 2019-10-24]
23. Olle P, Rickard D, Wiborg JS. How to use Bibexcel for various types of bibliometric analysis. In: Fredrik A, Rickard D, Birger L, Wiborg JS, editors. Celebrating Scholarly Communication Studies: A Festschrift for Olle Persson at His 60th Birthday. Berlin: International Society forScientometrics and Informetrics; 2009:9-24.
24. Sci2 Tool - Indiana University. 2009. Science of Science (Sci2) Tool URL: https://sci2.cns.iu.edu [accessed 2009-01-01]
25. Bastian M, Heymann S, Jacomy M. Gephi: An Open Source Software for Exploring and Manipulating Networks. In: Proceedings of the Third International AAAI Conference on Weblogs and Social Media. 2009 Presented at: ICWSM'09; May 17-20, 2009; San Jose, California USA. [doi: 10.1007/978-1-4614-7163-9_299-1]
26. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 2010 Aug;84(2):523-538 [FREE Full text] [doi: 10.1007/s11192-009-0146-3] [Medline: 20585380]
27. Munn Z, Peters MD, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Nov 19;18(1):143 [FREE Full text] [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
28. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 2;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
29. Lyons EJ, Baranowski T, Basen-Engquist KM, Lewis ZH, Swartz MC, Jennings K, et al. Testing the effects of narrative and play on physical activity among breast cancer survivors using mobile apps: study protocol for a randomized controlled trial. BMC Cancer 2016 Mar 9;16:202 [FREE Full text] [doi: 10.1186/s12885-016-2244-y] [Medline: 26960972]
30. Chadegani AA, Salehi H, Yunus MM, Farhadi H, Fooladi M, Farhadi M, et al. A comparison between two main academic literature collections: web of science and scopus databases. Asian Soc Sci 2013;9(5):18-26. [doi: 10.5539/ass.v9n5p18]
31. Aguillo IF. Is Google Scholar useful for bibliometrics? A webometric analysis. Scientometrics 2012;91(2):343-351. [doi: 10.1007/s11192-011-0582-8]
32. Ritterband LM, Thorndike FP, Cox DJ, Kovatchev BP, Gonder-Frederick LA. A behavior change model for internet interventions. Ann Behav Med 2009 Aug;38(1):18-27 [FREE Full text] [doi: 10.1007/s12160-009-9133-4] [Medline: 19802647]
33. Kreps GL, Neuhauser L. New directions in eHealth communication: opportunities and challenges. Patient Educ Couns 2010 Mar;78(3):329-336. [doi: 10.1016/j.pec.2010.01.013] [Medline: 20202779]

34. Lyons EJ, Lewis ZH, Mayrsohn BG, Rowland JL. Behavior change techniques implemented in electronic lifestyle activity monitors: a systematic content analysis. J Med Internet Res 2014 Aug 15;16(8):e192 [FREE Full text] [doi: 10.2196/jmir.3469] [Medline: 25131661]
35. Fox J, Bailenson JN. Virtual self-modeling: the effects of vicarious reinforcement and identification on exercise behaviors. Media Psychol 2009;12(1):1-25. [doi: 10.1080/15213260802669474]
36. Cugelman B, Thelwall M, Dawes P. Online interventions for social marketing health behavior change campaigns: a meta-analysis of psychological architectures and adherence factors. J Med Internet Res 2011 Feb 14;13(1):e17 [FREE Full text] [doi: 10.2196/jmir.1367] [Medline: 21320854]
37. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
38. Fjeldsoe BS, Marshall AL, Miller YD. Behavior change interventions delivered by mobile telephone short-message service. Am J Prev Med 2009 Feb;36(2):165-173. [doi: 10.1016/j.amepre.2008.09.040] [Medline: 19135907]
39. Cole-Lewis H, Kershaw T. Text messaging as a tool for behavior change in disease prevention and management. Epidemiol Rev 2010;32:56-69 [FREE Full text] [doi: 10.1093/epirev/mxq004] [Medline: 20354039]
40. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-t]
41. Fogg BJ. Persuasive Technology: Using Computers to Change What We Think and Do. Burlington, Massachusetts: Morgan Kaufmann; 2002.
42. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013 Dec 12;1(1):e3 [FREE Full text] [doi: 10.2196/games.3139] [Medline: 25658754]
43. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
44. Conroy DE, Yang C, Maher JP. Behavior change techniques in top-ranked mobile apps for physical activity. Am J Prev Med 2014 Jun;46(6):649-652. [doi: 10.1016/j.amepre.2014.01.010] [Medline: 24842742]
45. Hors-Fraile S, Schneider F, Fernandez-Luque L, Luna-Perejon F, Civit A, Spachos D, et al. Tailoring motivational health messages for smoking cessation using an mHealth recommender system integrated with an electronic health record: a study protocol. BMC Public Health 2018 Jun 5;18(1):698 [FREE Full text] [doi: 10.1186/s12889-018-5612-5] [Medline: 29871595]
46. Orji R, Moffatt K. Persuasive technology for health and wellness: state-of-the-art and emerging trends. Health Informatics J 2018 Mar;24(1):66-91. [doi: 10.1177/1460458216650979] [Medline: 27245673]
47. Dunn EE. Semantic Scholar. 2017. Sedentary Behaviour, Physical Activity, and Mobile Apps Among University Students URL: https://pdfs.semanticscholar.org/2812/f92745cf313ca8cd44cff168382cfe2f4014.pdf [accessed 2019-10-24]
48. Connell LE, Carey RN, de Bruin M, Rothman AJ, Johnston M, Kelly MP, et al. Links between behavior change techniques and mechanisms of action: an expert consensus study. Ann Behav Med 2019 Jul 17;53(8):708-720 [FREE Full text] [doi: 10.1093/abm/kay082] [Medline: 30452535]
49. Tolma EL, Reininger BM, Evans A, Ureda J. Examining the theory of planned behavior and the construct of self-efficacy to predict mammography intention. Health Educ Behav 2006 Apr;33(2):233-251. [doi: 10.1177/1090198105277393] [Medline: 16531515]
50. Owen N, Sugiyama T, Eakin EE, Gardiner PA, Tremblay MS, Sallis JF. Adults' sedentary behavior determinants and interventions. Am J Prev Med 2011 Aug;41(2):189-196. [doi: 10.1016/j.amepre.2011.05.013] [Medline: 21767727]

Abbreviations

BCSS: behavior change support system
BCT: behavior change technique
ISI: Institute for Scientific Information
IoT: Internet of Things
PSD: persuasive system design
PT: persuasive technology
Sci2: Science of Science
SCT: social cognitive theory
UCL: University College London
WoS: Web of Science

Edited by G Eysenbach; submitted 04.01.19; peer-reviewed by J Lacroix, S Jalil, N Ribeiro; comments to author 28.04.19; revised version received 20.07.19; accepted 26.09.19; published 13.12.19.

Please cite as:
Taj F, Klein MCA, van Halteren A
Digital Health Behavior Change Technology: Bibliometric and Scoping Review of Two Decades of Research
JMIR Mhealth Uhealth 2019;7(12):e13311
URL: https://mhealth.jmir.org/2019/12/e13311
doi:10.2196/13311
PMID:31833836

Original Paper

Clustering Insomnia Patterns by Data From Wearable Devices: Algorithm Development and Validation Study

Sungkyu Park[1*], MSc; Sang Won Lee[2*], MD, PhD; Sungwon Han[3], BSc; Meeyoung Cha[3,4], PhD

[1]Graduate School of Culture Technology, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea

[2]Kyungpook National University Chilgok Hospital, Daegu, Republic of Korea

[3]School of Computing, Korea Advanced Institute of Science and Technology, Daejeon, Republic of Korea

[4]Data Science Group, Center for Mathematical and Computational Sciences, Institute for Basic Science, Daejeon, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Meeyoung Cha, PhD
Data Science Group
Center for Mathematical and Computational Sciences
Institute for Basic Science
55 EXPO-ro, Doryong-dong, Yuseong-gu
Daejeon
Republic of Korea
Phone: 82 42 878 9300
Email: meeyoung.cha@gmail.com

Abstract

Background: As societies become more complex, larger populations suffer from insomnia. In 2014, the US Centers for Disease Control and Prevention declared that sleep disorders should be dealt with as a public health epidemic. However, it is hard to provide adequate treatment for each insomnia sufferer, since various behavioral characteristics influence symptoms of insomnia collectively.

Objective: We aim to develop a neural-net based unsupervised user clustering method towards insomnia sufferers in order to clarify the unique traits for each derived groups. Unlike the current diagnosis of insomnia that requires qualitative analysis from interview results, the classification of individuals with insomnia by using various information modalities from smart bands and neural-nets can provide better insight into insomnia treatments.

Methods: This study, as part of the precision psychiatry initiative, is based on a smart band experiment conducted over 6 weeks on individuals with insomnia. During the experiment period, a total of 42 participants (19 male; average age 22.00 [SD 2.79]) from a large university wore smart bands 24/7, and 3 modalities were collected and examined: sleep patterns, daily activities, and personal demographics. We considered the consecutive daily information as a form of images, learned the latent variables of the images via a convolutional autoencoder (CAE), and clustered and labeled the input images based on the derived features. We then converted consecutive daily information into a sequence of the labels for each subject and finally clustered the people with insomnia based on their predominant labels.

Results: Our method identified 5 new insomnia-activity clusters of participants that conventional methods have not recognized, and significant differences in sleep and behavioral characteristics were shown among groups (analysis of variance on rank: $F_{4,37}$=2.36, P=.07 for the sleep_min feature; $F_{4,37}$=9.05, P<.001 for sleep_efficiency; $F_{4,37}$=8.16, P<.001 for active_calorie; $F_{4,37}$=6.53, P<.001 for walks; and $F_{4,37}$=3.51, P=.02 for stairs). Analyzing the consecutive data through a CAE and clustering could reveal intricate connections between insomnia and various everyday activity markers.

Conclusions: Our research suggests that unsupervised learning allows health practitioners to devise precise and tailored interventions at the level of data-guided user clusters (ie, precision psychiatry), which could be a novel solution to treating insomnia and other mental disorders.

(JMIR Mhealth Uhealth 2019;7(12):e14473) doi:10.2196/14473

KEYWORDS

insomnia; precision psychiatry; cluster analysis; time-series data; unsupervised learning; convolutional autoencoder

Introduction

Approximately 30% of contemporary people have one or more symptoms of insomnia, and insomnia sufferers encounter difficulty falling or staying asleep [1,2]. Work schedule; sleep irregularity; naps; and nicotine, alcohol, and caffeine consumption can have significant effects on insomnia symptoms [3,4]. While an individual's intermixed behavioral characteristics might affect sleep behaviors, these are not considered in current diagnostic systems such as the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* or the International Classification of Diseases.

With concepts of precision psychiatry emerging, individual characteristics including genetic or neuroimaging, behavioral characteristics, and individual symptoms of illness are being used to make better decisions for diagnosis or treatment. Advances in machine learning and deep learning techniques can make precision psychiatry possible in clinical situations. For instance, unsupervised learning has been applied to distinguish traits of patients with various psychiatric disorders from those of healthy subjects [5,6]. When considering the heterogeneous characteristics of insomnia patients, an approach using precision psychiatry concepts can help develop better treatment methods for insomnia.

There has been considerable research using unsupervised machine learning methodologies in medical sciences not limited to psychiatry. One study reviewed the literature on detecting various diseases via computer-aided diagnosis and identified the best machine learning methodology for each disease [7]. Specific to using images as inputs, researchers have developed a single-layer sparse autoencoder to automatically classify tissue types from dynamic contrast-enhanced magnetic resonance imaging [8]. Another study discovered ground-truth networks in the brain through unsupervised learning of functional magnetic resonance imaging data [9]. As these research projects demonstrate, unsupervised learning has shown its potential to produce highly accurate models in the real world.

Several approaches have been proposed to handle the unstructured data type common in medical research. The support vector machine algorithm has been used on brain images for detecting Alzheimer disease in patients as well as for finding related brain parts with signs of Alzheimer disease [10]. One study built a convolutional neural network model to learn relevant features from unstructured raw data automatically and then made a convolutional neural network–based multimodal disease risk prediction model [11]. Another work collected data from an online social network (Twitter), constructed symptom weighting vectors by exploiting sentiment analysis, and tried to detect latent infectious diseases [12]. Some researchers have adopted wearable devices as inputs and built a neural network based on a multilayered perceptron to detect cardiovascular disease [13].

Despite the potential of a precision approach to insomnia treatment, little effort has been made to classify patients with insomnia using sleep and behavioral patterns. We hypothesized that certain symptom clusters representing specific sleep and/or behavioral patterns exist among individuals with insomnia. Sleep and behavioral characteristics that can affect insomnia symptoms are easily collected through smart bands. In this study, we tried to find certain clusters representing different sleep or behavioral characteristics using smart band data. Our unsupervised learning classification approach can be a cost-effective alternative that can bring positive insights for developing everyday interventions to assist in insomnia recovery (see Multimedia Appendix 1 for more information about implementation details including codes and datasets).

Methods

Recruitment

We applied unsupervised learning to gathered time-series data from an experiment. Participants were recruited using an online board of a large university community. To recruit subjects with insomnia symptoms, we used the Insomnia Severity Index (ISI) [14], Korean version [15]. The ISI is calculated based on responses to 7 questions that ask about sleep problems individuals have experienced in the most recent 2 weeks. For example, the questionnaire asks "How satisfied/dissatisfied are you with your current sleep pattern?" and the respondents answer using a 5-point Likert scale (1=very satisfied, 2=satisfied, 3=moderately satisfied, 4=dissatisfied, 5=very dissatisfied). This index is used as a standard metric of treatment response in clinical research.

A total of 50 participants whose ISI scores were 15 or above (ie, an indication of a mild to severe level of insomnia) participated (26 male; average age 22.63 [SD 3.02] years). Height and weight information were also gathered to calculate body mass index (BMI). In the study, we recruited young, healthy participants. Applicants being treated for medical or psychiatric illnesses were excluded from the experiment. The experiment was approved by the Korea Advanced Institute of Science and Technology institutional review board (approval number KH2018-40). Subjects' behavioral and sleep characteristics were collected via a smart band, the Charge 2 (Fitbit Inc) [16].

Intervention

The experiment was conducted for 6 weeks from April 23 to June 3, 2018, and the subjects' data were sent to a server. The first author gave weekly reminders to encourage subjects to wear the device continuously. Eight subjects were excluded from the study since they failed to wear the device for longer than 2 consecutive days. The remaining 42 participants finished our experiment, and we analyzed data from the subjects (19 male; average age 22.00 [SD 2.79]). The data features gathered and analyzed in our study are shown in Textbox 1. All values were normalized before analysis (see Figure 1 for the normalization method).

Textbox 1. List of data gathered from the smart band–wearing experiment.

Modality 1: Sleep pattern (source: Fitbit)

- sleep_start_time: time when a user goes to bed in units of seconds
- sleep_end_time: time when a user gets out of bed
- onbed_min: total time duration for staying in bed
- sleep_min: total time duration for actual sleeping
- sleep_efficiency: sleep min/onbed min
- awaken_min: total time duration of waking while sleeping
- awaken_moments: total frequency of waking up while sleeping

Modality 2: Daily activity (Fitbit)

- calories_consumed: total number of calories consumed per day
- active_calories: total number of calories consumed for activities per day
- Walks: total frequency of steps per day
- distance: total distance a user moves per day
- Stairs: total frequency at which a user climbs stairs per day
- active_ratio: daily moving time over the total time wearing the device

Modality 3: Personal demographic (Survey)

- Age: age of the participant
- Gender: gender of the participant
- Body mass index: kg/m2
- Insomnia Severity Index: result from the survey

Figure 1. Min-max scaler for normalizing each feature, where all values for each feature are normalized between 0 and 1.

$$normalized\_val = \frac{val - min}{max - min}$$

Fitbit provides a rich set of information about the wearer's sleep. For example, it tracks the total time that a person is in bed as well as the predicted total sleep time. Its autodetection and prediction are based on various behavioral or biological patterns such as information on heart rate and movement. There is one paper confirming the validity of the data from the Charge 2 [17]; it reports that sleep sensitivity (ie, predicting falling asleep) of Fitbit is remarkable, about 0.96, and the measured wake after sleep onset is similar to that of the existing polysomnography (PSG) method. Another study also demonstrated that the retrieved data from actigraphy corresponded acceptably well to that of PSG [18]. However, at the same time, we need to be cautious of using the data retrieved from wearables, since there might be a specificity (relatively low precision) issue with the wearables [19]. While it is possible for smart bands to give false sleep reports, we chose smart bands over PSG because they are easier to use and can obtain sleep measures regularly not in the clinical setting but in the wild. In addition to the sleep log, the Fitbit also reports a variety of information about the subjects' activity, including the calories consumed, steps walked, distance moved, and stairs climbed.

Data Analysis

Among the Fitbit-provided features in Textbox 1, we considered the reliability of information about sleep stages to be relatively low; therefore, 3 sleep features (rapid eye movement sleep, narrow sleep, deep sleep) were excluded from the analysis. Instead, sleep efficiency was added, which is calculated as the fraction of time dedicated to actual sleep out of the time spent lying in bed. Sleep efficiency is an essential factor in determining sleep quality [20]. Additionally, we introduced a new variable, the active ratio, which examines the total fraction of time spent on any activity.

Due to the nature of sleep, we needed to examine time-series data and consider the entire sequence of activities leading to each sleep event [21]. Today's sleep can be affected collectively by today's activity, yesterday's sleep, and so on. Therefore, it may be acceptable to jointly consider logs like this across multiple sleep and behavioral features. Our goal was to automatically identify clusters of sleep and behavioral features in hopes of identifying meaningful groups of individuals with insomnia who exhibit similar sleep-related dysfunctions.

Results

Approach Overview of Cluster Analysis

Cluster analysis has been a popular research topic in computer science [22]. In this study, we have tried two different approaches to clustering: synchronic and diachronic. In the synchronic approach, we would consider the dataset as one whole snapshot and cluster features within a snapshot all together. There were two issues with the synchronic approach (see Multimedia Appendix 1 for more detailed research procedures and results of this approach): (1) the clustering performance was not good enough in general and (2) this approach does not consider the consecutive patterns within and between features. Thus, we developed a novel diachronic approach, presented in the following section, in which we cluster the chunk consecutive daily logs first and then classify users based on their dominant clusters of chunk logs. This method represents insomnia-related patterns as a sequence of images.

Diachronic Unsupervised Learning of Insomnia Patterns

Supervised learning of clinical data may develop an overfitting problem because data are unstructured, sparse, noisy, or irregular [23]. In such conditions, unsupervised learning can be a good alternative. In particular, a convolutional autoencoder (CAE) that learns the abstract latent representations of data is known to be appropriate for the task [24]. We implemented a CAE and prepared a collage of daily features as an input image. This approach allowed us to capture the entire set of features simultaneously within the assigned time window.

When clustering high-dimensional data, it is crucial to reduce dimensions to avoid expensive computational costs and memory loss [25]. In handling high-dimensional data, the target clusters often lie in subspaces of the full space [26]. Hence, reducing dimensions via conventional methods such as singular value decomposition or principal component analysis may not yield correct clustering results, mainly when the target clusters are not in the same subspace. To avoid this issue, we implemented a CAE to effectively reduce the dimensions in the data since the latent variables of the CAE may include subspace information via filters. We also chose to implement CAE before clustering because a previous study argued that subsequence time-series clustering based on a sliding window might not be meaningful [27]. By treating the data with CAE first and then applying to cluster, the outcome patterns of the clusters are affected not by the subsequence time-series data but by the derived latent variables per image. Moreover, we can also expect that by presuming the data into images, the latent variables of CAE are more robust to the noisy daily changes of the Fitbit logs.

Clustering Steps

The neural net–based clustering approach operates in 5 steps.

Step 1: Preprocessing of Time-Series Data

We combine modality 1 (sleep pattern) and modality 2 (daily activity) in Textbox 1 to construct a multiplex vector of 12 dimensions (cf, modality 3, personal demographics, was not included since its data did not frequently change over the 6-week period, so it was solely employed in the comprehensive qualitative analysis process). In this step, with respect to modality 1, the *onbed_min* feature was excluded from the source dataset because it showed high correlations with other features, similar to the synchronic approach (see Multimedia Appendix 2 for the cross-correlations among 12 features of modalities 1 and 2). Let sleep vector $\mathbf{v_s}$=*{x*1, *x*2, ..., *x*8*}* and activity vector $\mathbf{v_a}$=*{y*1, *y*2, ..., *y*6*}*; then, the daily multiplex vector of subject $\alpha\mathbf{v}_\alpha$=*{x*1_α, ..., *x*8_α, *y*1_α, ..., *y*6_α*}*.

Step 2: Composing Sequential Images From Data

Sequentially connected multiplex vectors can be treated as an image chunk. We then composed the consecutive multiplex matrix to learn latent variables from images that represent individuals' daily log series. We modeled the daily sleep patterns as a Markov decision process because, according to previous sleep research, sleep deprivation (ie, sleep debt) has cumulative effects on waking function, and tonight's sleep debt will mostly be affected by recent sleep patterns (ie, sleep 1 to 2 days before) [28]. As a result, we connected each daily multiplex vector with a sliding window over 8 days (ie, the total 8-day logs are combined as one window always to include a weekend, where a window slides to the next vector, which is the next day). We applied a discount factor (γ) such that on the first window, the consecutive multiplex vector of the subject α per day can be formulated as in Figure 2.

We then finalized the sequential multiplex matrix, as shown in Figure 3. Fixing 8 days as a window size leaves a total of 42 days worth of data (ie, 42 data points). This result means that each user is represented by 35 sets of images (42–8+1=35), which is the batch size. The total number of input image chunks for the CAE is 1470 (42 users×35 images=1470). Let the sequential multiplex matrix of the first image chunk for the subject α be defined as in Figure 4. As a consequence, the sequential multiplex matrix of all subjects for all 35 images can be framed as in Figure 5 (see Multimedia Appendix 1 for more information about the mathematical notations).

Figure 2. Consecutive multiplex vector of the subject α per day on the 1st image chunk, where t_1 is defined as the 9th of v^α (ie, the 9th day from the starting date) and m as the corresponding closeness rank number (m^{th} of v^α) to t_1 within the image chunk (1≤m≤8).

$$\mathbf{v}^{\alpha}_{\mathbf{t_1-m}}{}' = \gamma^{\mathbf{m-1}}\mathbf{v}^{\alpha}_{\mathbf{t_1-m}}$$

Figure 3. Converting the consecutive sleep and activity data to images: sliding windows of 8-day chunks (# of Features=12, # of Days [ie, win-size]=8).

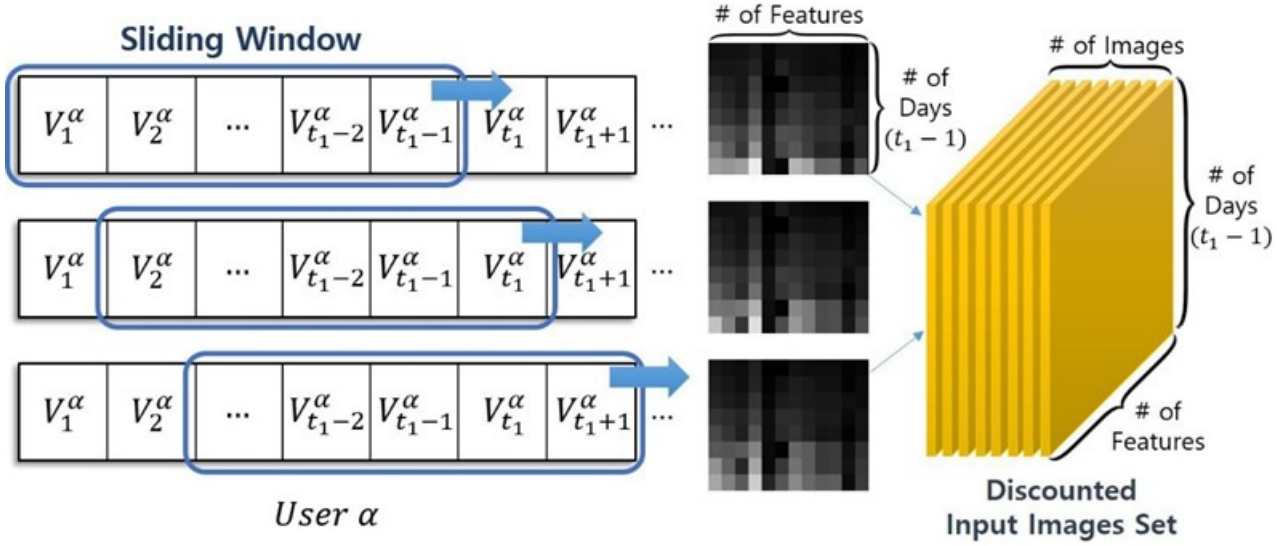


Figure 4. Sequential multiplex matrix of the subject α on the 1st image chunk, where t_1=9.

$$\mathbf{M}_{\mathbf{t_1}}^{\alpha} = \begin{pmatrix} \mathbf{v}_{\mathbf{t_1}-8}^{\alpha\ \prime} \\ \cdots \\ \mathbf{v}_{\mathbf{t_1}-2}^{\alpha\ \prime} \\ \mathbf{v}_{\mathbf{t_1}-1}^{\alpha\ \prime} \end{pmatrix} = \begin{pmatrix} \gamma^7 \mathbf{v}_{\mathbf{t_1}-8}^{\alpha} \\ \cdots \\ \gamma^1 \mathbf{v}_{\mathbf{t_1}-2}^{\alpha} \\ \gamma^0 \mathbf{v}_{\mathbf{t_1}-1}^{\alpha} \end{pmatrix}$$

Figure 5. Sequential multiplex matrix of all subjects for all 35 images, where β indicates the 2nd participant, $t_n \leq 43$, and $t_{n+1}=t_n+1$.

$$\mathbf{M_t} = \begin{pmatrix} \mathbf{M}_{\mathbf{t_1}}^{\alpha} \\ \cdots \\ \mathbf{M}_{\mathbf{t_{35}}}^{\alpha} \\ \mathbf{M}_{\mathbf{t_1}}^{\beta} \\ \cdots \\ \mathbf{M}_{\mathbf{t_{35}}}^{\beta} \\ \cdots \end{pmatrix}$$

Step 3: Learning Representations With Reducing Dimensions via Convolutional Autoencoder

In the proposed model depicted in Figure 6 for the designed CAE, the pooling layer has been omitted due to the small size of the input images (8×12). Additionally, we set up only one convolutional layer due to the relatively small amount of data (141,120 pixels=1470×8×12). Discount factor γ was set between 0.60 and 0.99. Then we iterated the model until finding the optimal learning rate, optimal γ value, optimal vertical size (cf, the horizontal size was fixed to 12, ie, the number of features), number of convolutional filters, and number of latent variables (ie, *encoded_size* in Figure 6).

We tested all combinations among 5 hyperparameters and found that the optimal values were $1e^{-4}$ with the AdamOptimizer (optimal learning rate); 0.75 (optimal γ value); 3 (optimal vertical size); 30 (number of convolutional filters); and 15 (number of latent variables), for each of these hyperparameters with the lowest L2-norm regularized reconstruction loss value of 1.09 (see Multimedia Appendix 1 for more information about the optimization and overfitting issue). To confirm the learning results, Figure 7 depicts the 35 input images and 35 reconstructed outputs of one random participant (*UserId* 26). This visual coherence ensures that the CAE efficiently reduced the input size dimensions (ie, 8×12→15) by learning the vital latent representations of data.

Figure 6. Convolutional autoencoder to find the latent variables of insomnia-related patterns per 8 days.

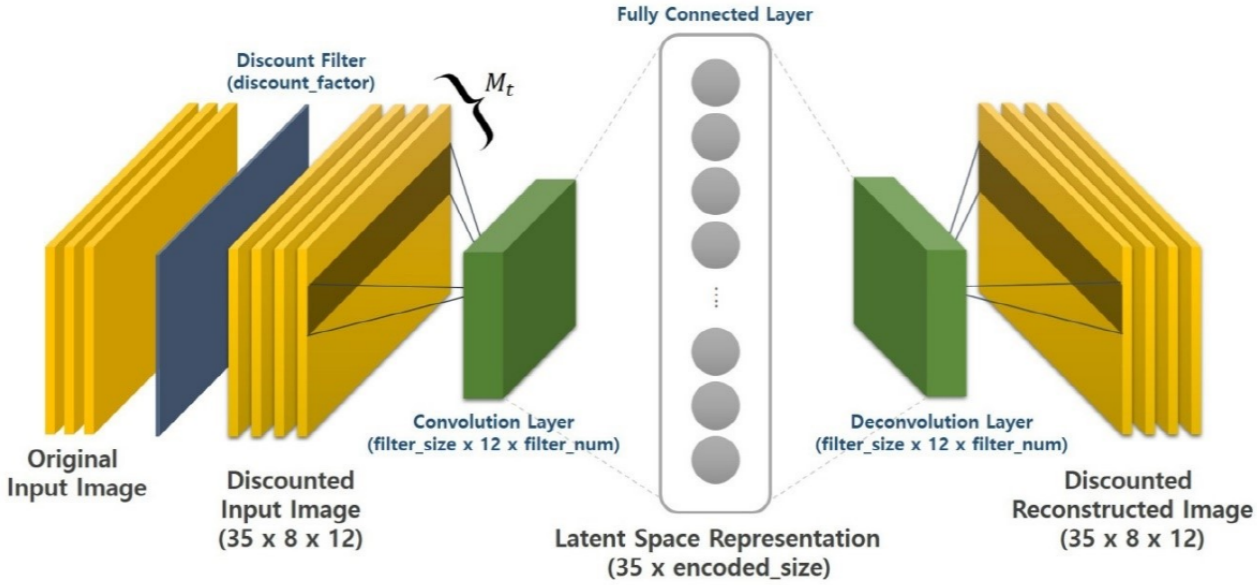
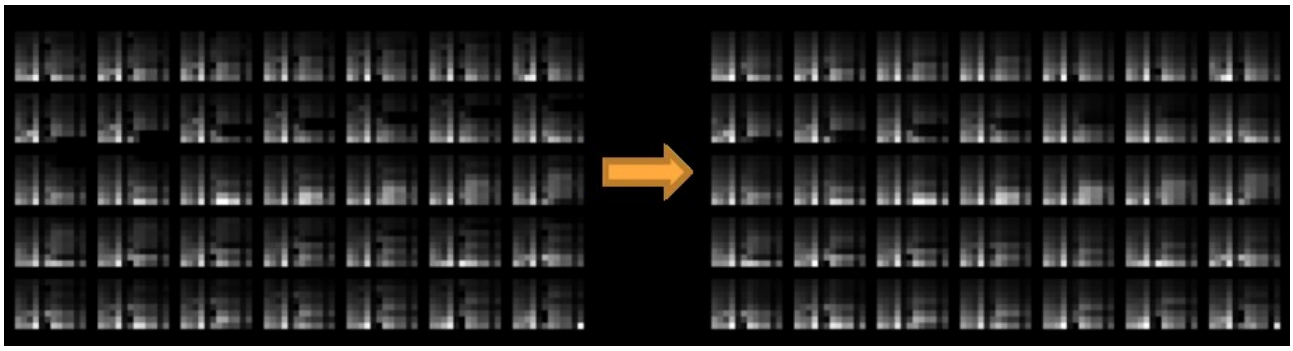


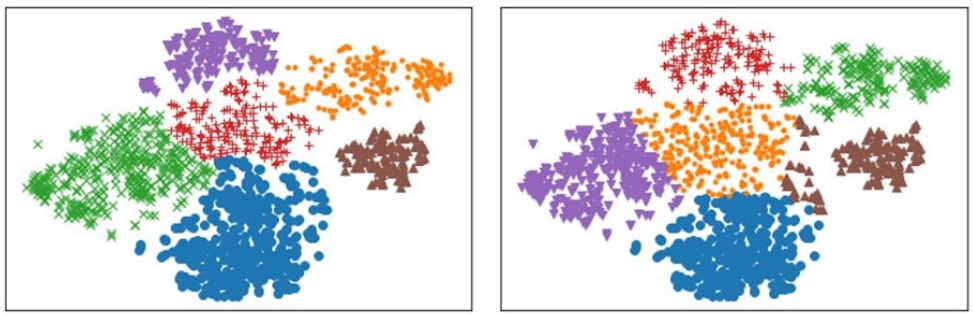
Figure 7. Convolutional autoencoder (CAE) reconstruction results of the learned CAE for one random subject.



Step 4: Clustering Images via Latent Variables

We clustered the latent variables as follows. As an initial step, we used t-SNE (ie, stochastic neighbor embedding with t-distribution) to reduce the number of dimensions from 15 to 2 additionally (ie, major hyperparameters for t-SNE were perplexity=30.0 [29,30], see Multimedia Appendix 1 for more information related to t-SNE; metric=euclidean; method=barnes_hut; the clustering result from the largest value of average silhouettes [AS] was chosen out of 100 trials). We then applied k-means and hierarchical clustering algorithms to form clusters of the 8-day chunk images. The results are displayed in Figure 8; both clustering results indicate the same cluster number of 6, and they were found reasonable by the AS values. However, the results of k-means clustering showed better results following the AS as well as the sum of squared errors (SSE). Therefore, we chose k-means to be the choice of the final user clustering step.

Figure 8. Clusters of 8-day chunk images based on sequential sleep/activity patterns. D: dimensions; C: clusters; AS: average silhouettes; SSE: sum of squared errors.



Step 5: Clustering Individuals With Insomnia

Each cluster of the 8-day images was labeled as cluster A, B, C, D, E, and F. We then composed a sequence of codes for each user (eg, the best-matched cluster code for each of the 35 images representing the user's logs, such as "*A-A-D-C-C-E-B-...*"). Finally, to determine one dominant cluster type per user, the PageRank algorithm was implemented and used among the stochastic relations between those 6 extracted labels. This intuition is feasible since participants' sequential codes, which are the series of the latent space representations per 8-day image for both sleep pattern and daily activity modalities, became the subjects' series of the comprehensive insomnia-related patterns; therefore, the codes can be modeled as a hidden Markov model, meaning that one's next state is dependent on the current state [28]. We then set the largest stochastic value of the labels per participant to be the dominant cluster, which is identical to finding a cluster with the highest PageRank value (ie, the probability of arriving at the cluster after many steps).

Evaluation Outcomes

Table 1 shows the clustering evaluation results among the two approaches concerning the diachronic clustering of daily vectors (**v**). The first approach is without CAE, meaning the original dataset is converted into chunk images (one image is composed of the 8 consecutive daily vectors with 96 dimensions) and then filtered via t-SNE (96→2 dimensions) and clustered by 3 clustering methods. The second approach is our model, and as introduced above, the original dataset is converted into chunk images and then inputted into CAE, and the latent variables of CAE are clustered by 3 clustering methods after the variables are filtered through t-SNE (96→15→2 dimensions).

We calculated two metrics to measure the clustering performance: the AS and the SSE. With AS, if the value is greater or equal to 0.4, then the clustering result is guaranteed to be significant [31]. By contemplating consecutive pattern information together via composing 35 images from the weekly based consecutive **v** per user, we could confirm that k-means clustering with our process showed the best performance (ie, the highest value of AS and the lowest of SSE) in clustering **v**. With respect to the first approach, the clustering performances were worse, and the possible reason for this result is that although the number of dimensions becomes large due to the chunked **v**, the approach could not successfully extract the latent features of the chunks while dramatically reducing the number of dimensions from 96 to 2.

Table 2 presents how the 42 subjects were eventually divided across the identified clusters of **v** by using our process with k-means clustering. Cluster F did not contain any members. The table shows the PageRank value, which represents the strength of the association between subjects and clusters. Cluster IDs are sorted by this PageRank value. Both cluster A and cluster D have the largest subject count of 12 people. The average BMI and ISI values are also given for each of the clusters. The BMI and ISI values were similar across clusters.

A psychiatrist conducted both qualitative and quantitative analyses of the 5 clusters. The number of subjects for each cluster is small and does not meet normality, and therefore, we performed an analysis of variance (ANOVA) on rank and Kruskal-Wallis test. There were several features that showed statistically significant differences among clusters (ANOVA on rank: $F_{4,37}$=2.36, *P*=.07 for *sleep_min*; $F_{4,37}$=9.05, *P*<.001 for *sleep_efficiency*; $F_{4,37}$=8.16, *P*<.001 for *active_calorie*; $F_{4,37}$=6.53, *P*<.001 for *walks*; $F_{4,37}$=3.51, *P*=.02 for *stairs*). Detailed results were described in Figure 9 as well as Tables MA3-a and MA3-b in Multimedia Appendix 3 (we also performed the ANOVA and post hoc Tukey honest significant difference test and included the results in the same supplementary material for reference).

Table 1. Evaluation of user clustering results for two diachronic approaches.

Approach	NC[a]	AS[b]	SSE[c]	NEO[d]	Remark
First with DBSCAN[e]	1	0.2075	119.4467	14	Chunk images used without a CAE[f]
First with hierarchical clustering	4	0.3394	35.0190	—[g]	Chunk images used without a CAE
First with k-means clustering	5	0.3941	22.6103	—	Chunk images used without a CAE
Ours with DBSCAN	4	0.1275	64.7179	30	Chunk images used with a CAE
Ours with hierarchical clustering	6	0.4485	24.3988	—	Chunk images used with a CAE
Ours with k-means clustering	6	0.4653	21.5532	—	Chunk images used with a CAE

[a]NC: number of clusters.

[b]AS: average silhouettes.

[c]SSE: sum of squared errors.

[d]NEO: number of excluded outliers.

[e]DBSCAN: density-based spatial clustering of applications with noise.

[f]CAE: convolutional autoencoder.

[g]Not applicable.

Table 2. Diachronic clustering results of 42 participants suffering from insomnia.

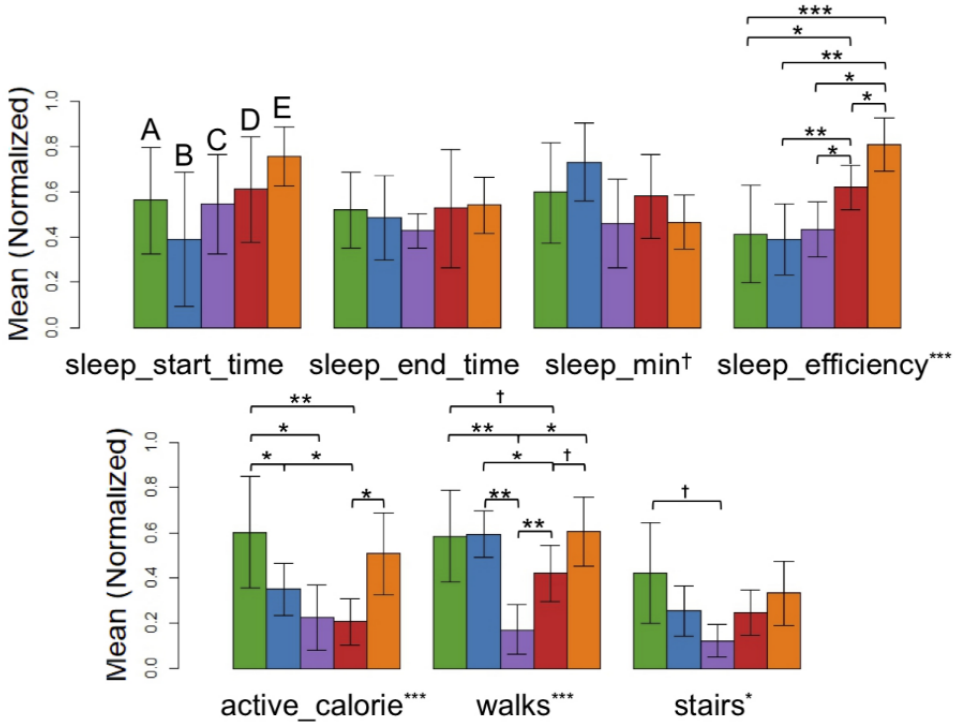
Cluster ID	A	B	C	D	E	F
Subjects, n (%)	12 (29)	9 (21)	4 (10)	12 (29)	5 (12)	0 (0)
PageRank	0.57	0.54	0.45	0.42	0.36	—[a]
BMI[b] (kg/m^2), mean (SD)	22.8 (3.3)	20.8 (2.3)	26.0 (2.7)	20.2 (3.0)	22.9 (2.3)	—
ISI[c], mean (SD)	19.1 (1.5)	19.1 (2.4)	18.2 (1.6)	16.9 (1.8)	18.8 (2.4)	—

[a]Not applicable.

[b]BMI: body mass index.

[c]ISI: Insomnia Severity Index.

Figure 9. Bar plots of the major smartband features per cluster (†P<.10, *P<.05, **P<.01, ***P<.001 in analysis of variance on rank and Kruskal-Wallis test results for 5 groups, also find Tables MA3-a and MA3-b in Multimedia Appendix 3).



Discussion

Principal Findings

This research conducted a 6-week experiment to collect smart band data and analyzed it to identify relationships between insomnia and daily activities. Our analysis finds that participants who seemingly face similar levels of insomnia based on the ISI score belonged to different clusters based on unsupervised learning. This finding means that our neural net–based clustering method could identify, beyond conventional diagnosis, new meaningful sleep-activity relationships that could be used to devise tailored interventions. Our method could determine which cluster an individual belongs to via data indicators of sleep and behavior acquired in the study.

Among the derived clusters of people with insomnia (see Figure 9), we found clusters *B*, *C*, and *E* to be more similar to one another and clusters *A* and *D* to contain similar sleep-activity profiles. Within the former group, subjects in cluster *B* were relatively long sleepers (even with insomnia) with low sleep efficiency. These subjects were in bed long hours, yet they could not easily fall asleep. Sleep restriction was a potentially useful intervention for these subjects [32]. In contrast, the subjects in cluster *E* exhibited a short sleep time with high sleep efficiency. These subjects also burned more calories per day. Chronic restriction of sleep may have affected symptoms of insomnia in this group [33]. While subjects in cluster *C* seemed to be morning persons, as indicated by the awake time patterns, their activity levels were relatively low, as shown in the active calorie, walk, and stair data. These subjects were treated with different interventions from the other similar clusters. In the latter group, clusters *A* and *D* had similar sleep patterns, such as total sleep time and sleep efficiency. However, subjects in cluster *A* indicated higher activity levels, such as walks and stairs, compared with cluster *D*. Cluster *D* also showed the lowest ranges for both BMI and ISI compared with other clusters, where a smaller BMI value represents nonobese status and a smaller

ISI value represents a milder insomnia level. Such subtle differences would have been difficult to notice with conventional methods or supervised learning. These results suggest that cluster-based intervention for individuals with insomnia could be an accurate method for alleviating symptoms of insomnia.

This study demonstrates that unsupervised learning of insomnia activity data gathered from wearable fitness devices could identify meaningful clusters. The 5 derived clusters each had distinctive characteristics, meaning that they can be used to derive different therapies and diagnoses. In accordance with our finding, one study also classified thousands of participants who answered online surveys by using latent class analysis so that they could find insomnia disorder subtypes [34], yet they were more focused not on past behavioral and sleeping patterns but on individual personalities as well as stressful surroundings in revealing the clusters. One more difference is that they solely used a conventional statistical method. Our study highlights that merely looking at sleep logs can reveal a limited aspect of insomnia; one needs to look at the full spectrum of both daily activities and sleep logs. Finally, our findings align with those of a previous study that revealed the average sleep quality is better related to many behavioral aspects of the person than the average sleep quantity [35].

Limitations

This work has several limitations. First, participants were from the same university, which may induce sampling biases. A larger scale study can draw conclusions that apply to the general public. In the future, we plan to conduct experiments targeting generic patients who visit hospitals regularly due to insomnia to ensure the generality of the clustering result. Second, a small sample size can reduce the statistical power, and it can affect the reducibility for future study. Third, various behaviors, such as consuming alcohol, tobacco, or caffeine, can affect patterns of sleep and activity. However, these factors were not considered in the current clustering analyses, mainly due to the small sample size and limited information on these behaviors (eg, consuming frequency per day or week). In that sense, a future study is needed to verify our preliminary insights. Fourth, when constructing the CAE, we prepared the input and output images irrespective of subjects. As a consequence, the learned latent variables cannot capture subtle dissimilarities that may exist across participants. To reduce this potential bias, a learning model may diminish the latent factors in composing images among participants while training. Last, we used t-SNE when reducing the dimensionality of the latent features of CAE images, and there could be type 1 errors when clustering dimension-reduced data because t-SNE forcibly fits original data into a t-distribution. We further plan to adopt and test other dimensionality reduction methods toward the extracted latent variables of CAE images to relax the possible type 1 error issue.

Conclusion

This study, although preliminary, gives new insights for future studies in the field of mobile health. Motivated by the results of this study, in the future we hope to develop tailored intervention strategies that can be matched to each cluster for relieving insomnia symptoms, which will be a meaningful step in precision psychiatry. In the meantime, we plan to consider qualitative factors such as the level of sleep as well as quantitative aspects of sleep-related factors.

Acknowledgments

The authors would like to thank Donghyun Ahn for his contribution on managing the experiment and collecting data. This research was supported by the Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Science and Information and Communications Technology (#NRF-2017R1E1A1A01076400).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Description of the datasets and the codes utilized at the current research and detailed procedure of user clustering based on the synchronic approach and its evaluation outcomes with additional descriptions on the clustering steps of the diachronic unsupervised learning approach.
[PDF File (Adobe PDF File), 113 KB - mhealth_v7i12e14473_app1.pdf]

Multimedia Appendix 2
Cross rank correlation matrix among 12 features of modality 1 and 2 used in clustering, regardless of the derived cluster ID (N=42) and with regard to each derived cluster ID (# of clusters=5).
[PDF File (Adobe PDF File), 114 KB - mhealth_v7i12e14473_app2.pdf]

Multimedia Appendix 3
Test results with respect to sleep pattern-related (modality 1) and daily activity-related (modality 2) features among 5 derived clusters of insomnia sufferers: (1) analysis of variance on rank and Kruskal-Wallis and (2) analysis of variance and post hoc Turkey honest significant difference.
[PDF File (Adobe PDF File), 275 KB - mhealth_v7i12e14473_app3.pdf]

References

1. Ohayon MM. Epidemiology of insomnia: what we know and what we still need to learn. Sleep Med Rev 2002 Apr;6(2):97-111. [doi: 10.1053/smrv.2002.0186] [Medline: 12531146]
2. Morphy H, Dunn K, Lewis M, Boardman H, Croft P. Epidemiology of insomnia: a longitudinal study in a UK population. Sleep 2007 Mar;30(3):274-280. [Medline: 17425223]
3. Yang Y, Shin JC, Li D, An R. Sedentary behavior and sleep problems: a systematic review and meta-analysis. Int J Behav Med 2017 Aug;24(4):481-492. [doi: 10.1007/s12529-016-9609-0] [Medline: 27830446]
4. National Sleep Foundation. What causes insomnia? URL: https://www.sleepfoundation.org/insomnia/what-causes-insomnia [accessed 2019-04-23] [WebCite Cache ID 77q3rvC3u]
5. Silver H, Shmoish M. Analysis of cognitive performance in schizophrenia patients and healthy individuals with unsupervised clustering models. Psychiatry Res 2008 May 30;159(1-2):167-179. [doi: 10.1016/j.psychres.2007.06.009] [Medline: 18440074]
6. Zeng L, Shen H, Liu L, Hu D. Unsupervised classification of major depression using functional connectivity MRI. Hum Brain Mapp 2014 Apr;35(4):1630-1641. [doi: 10.1002/hbm.22278] [Medline: 23616377]
7. Fatima M, Pasha M. Survey of machine learning algorithms for disease diagnostic. JILSA 2017;09(01):1-16. [doi: 10.4236/jilsa.2017.91001]
8. Shin H, Orton M, Collins D, Doran S, Leach M. Autoencoder in time-series analysis for unsupervised tissues characterisation in a large unlabelled medical image dataset. 2011 Presented at: Proceedings of the ICMLA; 2011; Honolulu. [doi: 10.1109/icmla.2011.38]
9. Bai Z, Walker P, Tschiffely A, Wang F, Davidson I. Unsupervised network discovery for brain imaging data. 2017 Presented at: Proceedings of the SIGKDD; 2017; Halifax. [doi: 10.1145/3097983.3098023]
10. Zhang Y, Dong Z, Phillips P, Wang S, Ji G, Yang J, et al. Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning. Front Comput Neurosci 2015;9:66 [FREE Full text] [doi: 10.3389/fncom.2015.00066] [Medline: 26082713]
11. Chen M, Hao Y, Hwang K, Wang L, Wang L. Disease prediction by machine learning over big data from healthcare communities. IEEE Access 2017;5:8869-8879. [doi: 10.1109/access.2017.2694446]
12. Lim S, Tucker CS, Kumara S. An unsupervised machine learning model for discovering latent infectious diseases using social media data. J Biomed Inform 2017 Feb;66:82-94 [FREE Full text] [doi: 10.1016/j.jbi.2016.12.007] [Medline: 28034788]
13. Oresko JJ, Duschl H, Cheng AC. A wearable smartphone-based platform for real-time cardiovascular disease detection via electrocardiogram processing. IEEE Trans Inf Technol Biomed 2010 May;14(3):734-740. [doi: 10.1109/TITB.2010.2047865] [Medline: 20388600]
14. Bastien C, Vallières A, Morin C. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. Sleep Med 2001 Jul;2(4):297-307. [doi: 10.1016/s1389-9457(00)00065-4] [Medline: 11438246]
15. Cho YW, Song ML, Morin CM. Validation of a Korean version of the insomnia severity index. J Clin Neurol 2014 Jul;10(3):210-215 [FREE Full text] [doi: 10.3988/jcn.2014.10.3.210] [Medline: 25045373]
16. Fitbit. Fitbit Charge 2 URL: https://www.fitbit.com/uk/charge2 [accessed 2019-04-23] [WebCite Cache ID 77q540vXN]
17. de Zambotti M, Goldstone A, Claudatos S, Colrain IM, Baker FC. A validation study of Fitbit Charge 2™ compared with polysomnography in adults. Chronobiol Int 2018 Apr;35(4):465-476. [doi: 10.1080/07420528.2017.1413578] [Medline: 29235907]
18. Blackwell T, Redline S, Ancoli-Israel S, Schneider J, Surovec S, Johnson N, Study of Osteoporotic Fractures Research Group. Comparison of sleep parameters from actigraphy and polysomnography in older women: the SOF study. Sleep 2008 Feb;31(2):283-291 [FREE Full text] [doi: 10.1093/sleep/31.2.283] [Medline: 18274276]
19. Marino M, Li Y, Rueschman MN, Winkelman JW, Ellenbogen JM, Solet JM, et al. Measuring sleep: accuracy, sensitivity, and specificity of wrist actigraphy compared to polysomnography. Sleep 2013 Nov 01;36(11):1747-1755 [FREE Full text] [doi: 10.5665/sleep.3142] [Medline: 24179309]
20. 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]
21. Ding H, Trajcevski G, Scheuermann P, Wang X, Keogh E. Querying and mining of time series data. 2008 Aug 01 Presented at: Proceedings of the VLDB; 2008; Auckland p. 1542-1552. [doi: 10.14778/1454159.1454226]
22. Saxena A, Prasad M, Gupta A, Bharill N, Patel OP, Tiwari A, et al. A review of clustering techniques and developments. Neurocomputing 2017 Dec;267:664-681. [doi: 10.1016/j.neucom.2017.06.053]
23. Lasko TA, Denny JC, Levy MA. Computational phenotype discovery using unsupervised feature learning over noisy, sparse, and irregular clinical data. PLoS One 2013;8(6):e66341 [FREE Full text] [doi: 10.1371/journal.pone.0066341] [Medline: 23826094]
24. Masci J, Meier U, Cirean D, Schmidhuber J. Stacked convolutional auto-encoders for hierarchical feature extraction. 2011 Presented at: Proceedings of the International Conference on Artificial Neural Networks; 2011; Espoo. [doi: 10.1007/978-3-642-21735-7_7]

25. Aghabozorgi S, Seyed Shirkhorshidi A, Ying Wah T. Time-series clustering: a decade review. Inf Syst 2015 Oct;53:16-38. [doi: 10.1016/j.is.2015.04.007]
26. Steinbach M, Ertöz L, Kumar V. The challenges of clustering high dimensional data. New Directions Statistical Physics 2004;273:309. [doi: 10.1007/978-3-662-08968-2_16]
27. Keogh E, Lin J. Clustering of time-series subsequences is meaningless: implications for previous and future research. Knowl Inf Syst 2005 Aug 1;8(2):154-177. [doi: 10.1007/s10115-004-0172-7]
28. Dinges D, Pack F, Williams K, Gillen K, Powell J, Ott G, et al. Cumulative sleepiness, mood disturbance, and psychomotor vigilance performance decrements during a week of sleep restricted to 4-5 hours per night. Sleep 1997 Apr;20(4):267-277. [Medline: 9231952]
29. Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res 2008 Nov:2579-2605.
30. Wattenberg M, Viégas F, Johnson I. How to use t-SNE effectively. Distill 2016 Oct;1(10). [doi: 10.23915/distill.00002]
31. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Computat Appl Math 1987 Nov;20:53-65. [doi: 10.1016/0377-0427(87)90125-7]
32. Falloon K, Elley CR, Fernando A, Lee AC, Arroll B. Simplified sleep restriction for insomnia in general practice: a randomised controlled trial. Br J Gen Pract 2015 Aug;65(637):e508-e515 [FREE Full text] [doi: 10.3399/bjgp15X686137] [Medline: 26212846]
33. Levine B, Lumley M, Roehrs T, Zorick F, Roth T. The effects of acute sleep restriction and extension on sleep efficiency. Int J Neurosci 1988 Dec;43(3-4):139-143. [doi: 10.3109/00207458808986162] [Medline: 3243670]
34. Blanken TF, Benjamins JS, Borsboom D, Vermunt JK, Paquola C, Ramautar J, et al. Insomnia disorder subtypes derived from life history and traits of affect and personality. Lancet Psychiatry 2019 Feb;6(2):151-163. [doi: 10.1016/s2215-0366(18)30464-4]
35. Pilcher JJ, Ginter DR, Sadowsky B. Sleep quality versus sleep quantity: relationships between sleep and measures of health, well-being and sleepiness in college students. J Psychosom Res 1997 Jun;42(6):583-596. [doi: 10.1016/s0022-3999(97)00004-4] [Medline: 9226606]

Abbreviations

ANOVA: analysis of variance
AS: average silhouettes
BMI: body mass index
CAE: convolutional autoencoder
ISI: Insomnia Severity Index
PSG: polysomnography
SSE: sum of squared errors
t-SNE: t-distributed stochastic neighbor embedding

Edited by G Eysenbach; submitted 23.04.19; peer-reviewed by B Jeong; comments to author 03.06.19; revised version received 30.09.19; accepted 19.10.19; published 05.12.19.

Please cite as:
Park S, Lee SW, Han S, Cha M
Clustering Insomnia Patterns by Data From Wearable Devices: Algorithm Development and Validation Study
JMIR Mhealth Uhealth 2019;7(12):e14473
URL: https://mhealth.jmir.org/2019/12/e14473
doi:10.2196/14473
PMID:31804187

Original Paper

Objectively Monitoring Amyotrophic Lateral Sclerosis Patient Symptoms During Clinical Trials With Sensors: Observational Study

Luis Garcia-Gancedo[1], PhD; Madeline L Kelly[2], PhD; Arseniy Lavrov[3], MD; Jim Parr[4], PhD; Rob Hart[4], MEng; Rachael Marsden[5], BSc; Martin R Turner[5], PhD; Kevin Talbot[5], PhD; Theresa Chiwera[6], BSc; Christopher E Shaw[6], FRACP; Ammar Al-Chalabi[6], PhD

[1]Advanced Biostatistics & Data Analytics Centre of Excellence, R&D Projects Clinical Platforms & Sciences, GlaxoSmithKline, Stevenage, United Kingdom

[2]Translational Medicine, Future Pipeline Discovery, GlaxoSmithKline, Stevenage, United Kingdom

[3]Clinical Development, AveXis, Bannockburn, IL, United States

[4]McLaren Technology Centre, McLaren Applied Technologies, Woking, United Kingdom

[5]Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom

[6]Maurice Wohl Clinical Neuroscience Institute, Department of Basic and Clinical Neuroscience, King's College London, London, United Kingdom

Corresponding Author:
Luis Garcia-Gancedo, PhD
Advanced Biostatistics & Data Analytics Centre of Excellence
R&D Projects Clinical Platforms & Sciences
GlaxoSmithKline
Gunnels Wood Road
Stevenage, SG1 2NY
United Kingdom
Phone: 44 1438 762129
Email: luis.x.garcia-gancedo@gsk.com

Abstract

Background: Objective symptom monitoring of patients with Amyotrophic Lateral Sclerosis (ALS) has the potential to provide an important source of information to evaluate the impact of the disease on aspects of real-world functional capacity and activities of daily living in the home setting, providing useful objective outcome measures for clinical trials.

Objective: This study aimed to investigate the feasibility of a novel digital platform for remote data collection of multiple symptoms—physical activity, heart rate variability (HRV), and digital speech characteristics—in 25 patients with ALS in an observational clinical trial setting to explore the impact of the devices on patients' everyday life and to record tolerability related to the devices and study procedures over 48 weeks.

Methods: In this exploratory, noncontrolled, nondrug study, patients attended a clinical site visit every 3 months to perform activity reference tasks while wearing a sensor, to conduct digital speech tests and for conventional ALS monitoring. In addition, patients wore the sensor in their daily life for approximately 3 days every month for the duration of the study.

Results: The amount and quality of digital speech data captured at the clinical sites were as intended, and there were no significant issues. All the home monitoring sensor data available were propagated through the system and were received as expected. However, the amount and quality of physical activity home monitoring data were lower than anticipated. A total of 3 or more days (or partial days) of data were recorded for 65% of protocol time points, with no data collected for 24% of time points. At baseline, 24 of 25 patients provided data, reduced to 13 of 18 patients at Week 48. Lower-than-expected quality HRV data were obtained, likely because of poor contact between the sensor and the skin. In total, 6 of 25 patients had mild or moderate adverse events (AEs) in the skin and subcutaneous tissue disorders category because of skin irritation caused by the electrode patch. There were no reports of serious AEs or deaths. Most patients found the sensor comfortable, with no or minimal impact on daily activities.

Conclusions: The platform can measure physical activity in patients with ALS in their home environment; patients used the equipment successfully, and it was generally well tolerated. The quantity of home monitoring physical activity data was lower than expected, although it was sufficient to allow investigation of novel physical activity end points. Good-quality in-clinic speech

data were successfully captured for analysis. Future studies using objective patient monitoring approaches, combined with the most current technological advances, may be useful to elucidate novel digital biomarkers of disease progression.

(JMIR Mhealth Uhealth 2019;7(12):e13433) doi:10.2196/13433

KEYWORDS

amyotrophic lateral sclerosis; objective symptom monitoring; clinical trial; physical activity; digital phenotyping; digital biomarker; heart rate; speech; accelerometer; wearable

Introduction

Background

Amyotrophic Lateral Sclerosis (ALS) is a neurodegenerative disorder affecting motor neurons, characterized by progressive weakness, leading to increased disability and eventually death from neuromuscular respiratory failure, typically within 5 years [1,2].

There is no known cure for ALS; the existing licensed disease-modifying medications riluzone (trade names: Rilutek, Teglutik), and edaravone (trade names: Radicut, Radicava) are only modestly effective in impacting the disease course or improving survival [3-5]. Thus, there remains a significant unmet medical need in ALS for therapies to slow progression of functional decline and improve survival. Numerous therapies and experimental agents have been tested in ALS and have failed, sometimes after positive results were achieved in early efficacy studies, suggesting a need to improve outcome measures of disease progression and overall study design [6,7]. Existing disease progression measures have limitations in terms of sensitivity, requiring long trials with large sample sizes. The ALS Functional Rating Score (Revised) (ALSFRS-R) [8] is the most commonly used instrument to monitor the progression of disability in patients with ALS; however, it relies on patients' recollection of their clinical function rather than direct objective assessment. Testing burden in ALS clinical trials is heavy and often involves a significant number of in-clinic assessments for overall clinical function, respiratory function, and muscle strength, which are mainly conducted at clinical visits. This can be tiring for patients, particularly as the disease progresses, leading to patient dropout and, consequently, missing data. As a result, the ability to draw conclusions from the data is often impacted. In that regard, objective monitoring of patients has the potential to (1) enable the development of novel *digital* biomarkers to accurately quantify changes in function and disease progression with greater sensitivity than current clinical methods, thereby enabling smaller and shorter trials, and (2) provide additional, important information to assess the impact of the disease and treatment on clinical function and activities of daily living in a real-life setting, while reducing the testing burden.

For these reasons, there has been an increasing interest in developing new technologies for remote and objective clinical assessment of ALS, which may be useful outcome measures in clinical trials [9-11].

Amyotrophic Lateral Sclerosis Digital Biomarker Candidates

Although the presentation of ALS varies among patients, its main characteristics are (1) upper and lower motor neuron symptoms and signs, resulting in skeletal muscle weakness and spasticity (compromising mobility and activities of daily living), (2) speech and swallowing difficulties, and (3) respiratory problems. In addition, other symptoms such as impaired cardiac autonomic control, weight loss, cramps and fasciculations, emotional lability, and frontal lobe-type cognitive dysfunction are not uncommon [10,12-16]. Technology advances over the last few years have provided an opportunity to objectively quantify some of these manifestations.

Movement sensors have been previously used to quantify mobility and activities of daily living in patients with rheumatoid arthritis (RA), chronic obstructive pulmonary disease (COPD), or Parkinson disease (PD). Hashimoto et al [17] found that the mean daily activity level of patients with RA was significantly lower than that in healthy controls, and the number of sedentary periods was significantly higher and moderately correlated with the Health Assessment Questionnaire Disability Index. Van Buul et al [18] found that patients with more symptomatic COPD do fewer steps a day and spend less time in moderate-and-vigorous physical activity than those with less symptomatic disease. Lipsmeier et al [19] found multiple aspects of significantly reduced everyday motor behavior in patients with PD compared with controls. Objective activity data have not been reported for patients with ALS thus far, although ongoing studies (such as the AT HOME study [9]) aim to investigate the use of armbands for evaluating disease progression through physical activity measures. Patients are anticipated to be increasingly less mobile as the disease progresses, which would be reflected in multiple activity measures, including overall time spent active, average daily activity levels, or activity fragmentation. In this study, we hypothesized that these measures could be objectively assessed with a standard accelerometer, as long as patients wear the device for a period of time long enough to be representative of their typical behavior (continuous in-home monitoring).

Portable electrocardiogram (ECG) systems have been used to evaluate cardiac function and perform analysis of heart rate variability (HRV) data. An increase in the mean heart rate at rest, a decrease in standard deviation of interbeat interval (tRR), as well as in proportion of number of pairs of successive beat-to-beat intervals that differ more than 50 ms divided by the total number of beat-to-beat intervals, and an increase in the low-frequency/high-frequency (LF/HF) component ratio were found in patients with ALS, indicating a vagal-sympathetic imbalance [10,20,21].

Different acoustic recording systems have previously been used to demonstrate associations between acoustic measures and speech intelligibility in patients with ALS. Patients were found to have significantly slower syllabic and speaking rate and lower speech intelligibility, with speaking rate typically declining much earlier in the disease stage than speech intelligibility [11,22,23]. In this study, we evaluate speech characteristics at the clinical sites to allow for high-fidelity data capture equipment and a controlled quiet environment that may allow more accurate quantification of changes in speech formats.

All previously reported efforts to objectively monitor patients with ALS evaluate individual symptoms only; however, the heterogeneity of symptom presentation (among patients and within a patient over time) makes it likely that a combination of measurements would be needed to accurately quantify changes in disease progression with high sensitivity for a diverse cohort of patients. Thus, platforms incorporating multiple objective symptom assessment technologies present the best opportunity to investigate multiple digital biomarkers—and ultimately composite digital biomarkers—that would be useful measures of disease progression for a broad patient population and disease stages.

The objective of this clinical study was 3-fold: (1) to investigate the feasibility of a novel platform for objective data collection of multiple ALS manifestations (physical activity, HRV, and speech characteristics) in a clinical study; (2) to explore the impact of the devices on patients' everyday life and to record safety and tolerability related to the devices and study procedures over 48 weeks; and (3) to explore digital disease progression markers in ALS, which may be useful outcome measures in clinical trials. The results for the latter objective will be reported separately.

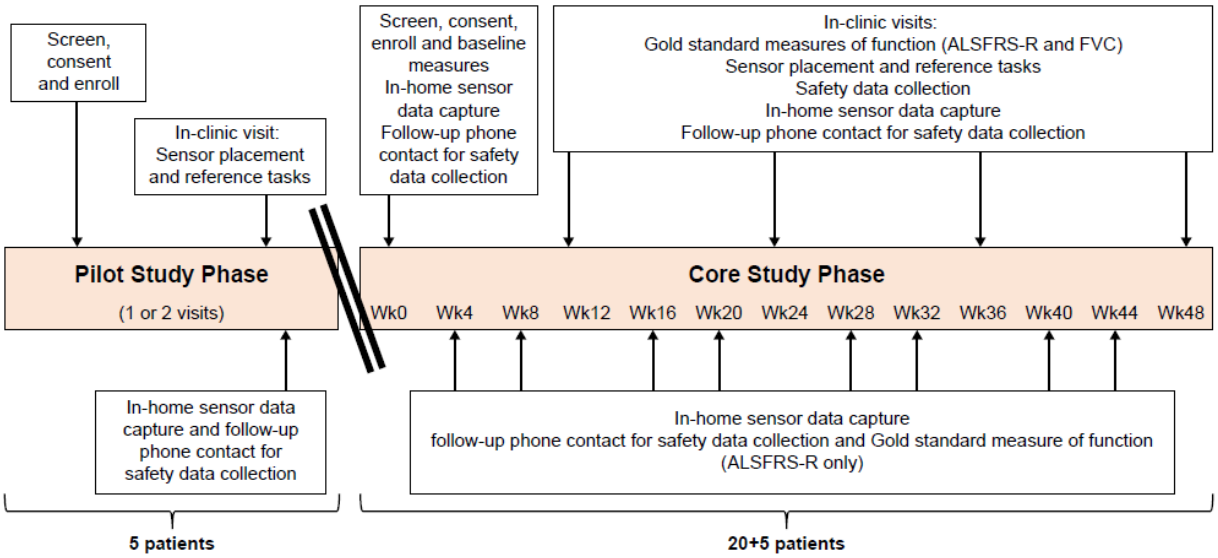
Methods

Study Overview

An exploratory, noncontrolled, nondrug study in ALS (NCT02447952) was sponsored by GlaxoSmithKline (GSK) and conducted in collaboration with McLaren Applied Technologies (MAT). A total of 2 clinical sites, both in the United Kingdom, enrolled patients. The study was initiated on June 30, 2015 (first patient screened), and completed on June 1, 2017 (last patient, last visit). The study comprised 2 phases, a variable length Pilot Study Phase (5 patients enrolled) and a 48-week Core Study Phase (25 patients enrolled, including the 5 patients who progressed from Pilot Study Phase to Core Study Phase), as shown schematically in Figure 1. The 2-phase approach was utilized to allow for refinement of the equipment and data transmission processes during the Pilot Study Phase and enable adaptation of components of the digital platform, which were performing suboptimally before embarking on the Core Study Phase. In such an exploratory setting, the 2-part design helped mitigate taking forward equipment, algorithms, data capture methods, and data transmission processes that may require adaptation and lessened the risk that major changes would be needed during the Core Study Phase. For the duration of the study, patients attended a clinical site visit every 12 weeks to perform various assessments and tasks; in addition, they wore a sensor in their daily life for approximately 3 consecutive days every month (home monitoring periods), allocating 2 hours per 24-shour period to recharge the sensor. The study did not include any specific patient retention strategy. The home monitoring period duration was selected to minimize patient burden while capturing sufficient data to enable extraction of clinically relevant data. At all study time points, patients were provided with a diary as a tool to record details about their experience with the sensor and their activity type and level while wearing the sensor.

The study was designed to enable exploration of novel disease progression markers; for the purposes of the sample size justification, the correlation between the ALSFRS-R and the novel end points was considered. With 20 patients, it was assumed that there would be an 80% chance of detecting a within-patient correlation of greater than or equal to 0.6 (considered a moderate to strong correlation), if the true correlation was 0.7. A total of 25 patients were enrolled to allow 20 evaluable patients (anticipating 20% dropout rate).

Figure 1. Study design overview: All 5 patients from the Pilot Study Phase progressed to the Core Study Phase. ALSFRS-R: ALS Functional Rating Score (Revised); FVC: forced vital capacity; Wk: week.



Patient Population

The intended study population comprised patients with a diagnosis of ALS, who were ambulant and had a relatively high level of clinical function at baseline. The diagnosis was required to have been made by a neurologist with ALS expertise within 18 months of symptom onset. Eligible participants were 18 to 80 years of age, capable of giving signed (or verbal) informed consent and were capable of, and willing to follow the study protocol. Patients were excluded from study participation if they met any of the following criteria: had neurological (other than ALS) or nonneurological comorbidities; presented with clinically significant cognitive impairment; had a regionally restricted form of ALS or other atypical variant; required mechanical ventilation; and had an active implantable cardiac medical device or were at a high risk for needing external defibrillation or had a history of skin hypersensitivity to adhesives. There were no prohibited medications, but enrollment into an investigational drug trial (in addition to participation in this trial) could be prohibited if, in the opinion of the investigator, the investigational drug might have impacted the objectives of this study. Table 1 summarizes the baseline population characteristics of all patients and the completers' population. As expected, the most common concomitant medications were from the nervous system class, with the most common being riluzole (64%), zopiclone (24%), and citalopram (20%; Table 2).

Table 1. Patient demographics and baseline disease characteristics.

Characteristics	Overall population (N=25)	Completers population (n=18)
Gender, n (%)		
Male	21 (84)	16 (89)
Female	4 (16)	2 (11)
Age (years), mean (SD)	53.1 (9.93)	52.9 (11.3)
Race, n (%)		
White	23 (92)	17 (94)
Asian	2 (8)	1 (6)
Duration of Amyotrophic Lateral Sclerosis, n (%)		
<18 months	22 (88)	16 (89)
Missing	3 (12)	2 (11)
Phenotype at onset, n (%)		
Upper limbs	15 (60)	10 (56)
Lower limbs	6 (24)	5 (28)
Upper and lower limbs	2 (8)	1 (6)
Bulbar	2 (8)	2 (11)
ALSFRS-R[a] total score, mean (SD)	41.6 (4.98)	43.0 (2.71)
FVC[b], mean (SD)	3.927 (1.432)	4.262 (1.238)

[a]ALSFRS-R: Amyotrophic Lateral Sclerosis Functional Rating Score (Revised).

[b]FVC: forced vital capacity.

Table 2. Most common (≥2 patients) concomitant medications.

Anatomical Therapeutic Chemical medication grouping	Core Study Phase (N=25), n (%)
Any medication	22 (88)
Nervous system	20 (80)
Musculoskeletal system	4 (16)
Alimentary tract and metabolism	3 (12)
Cardiovascular system	3 (12)
Sensory organs	3 (12)
Genitourinary system and sex hormones	2 (8)
Respiratory system	2 (8)

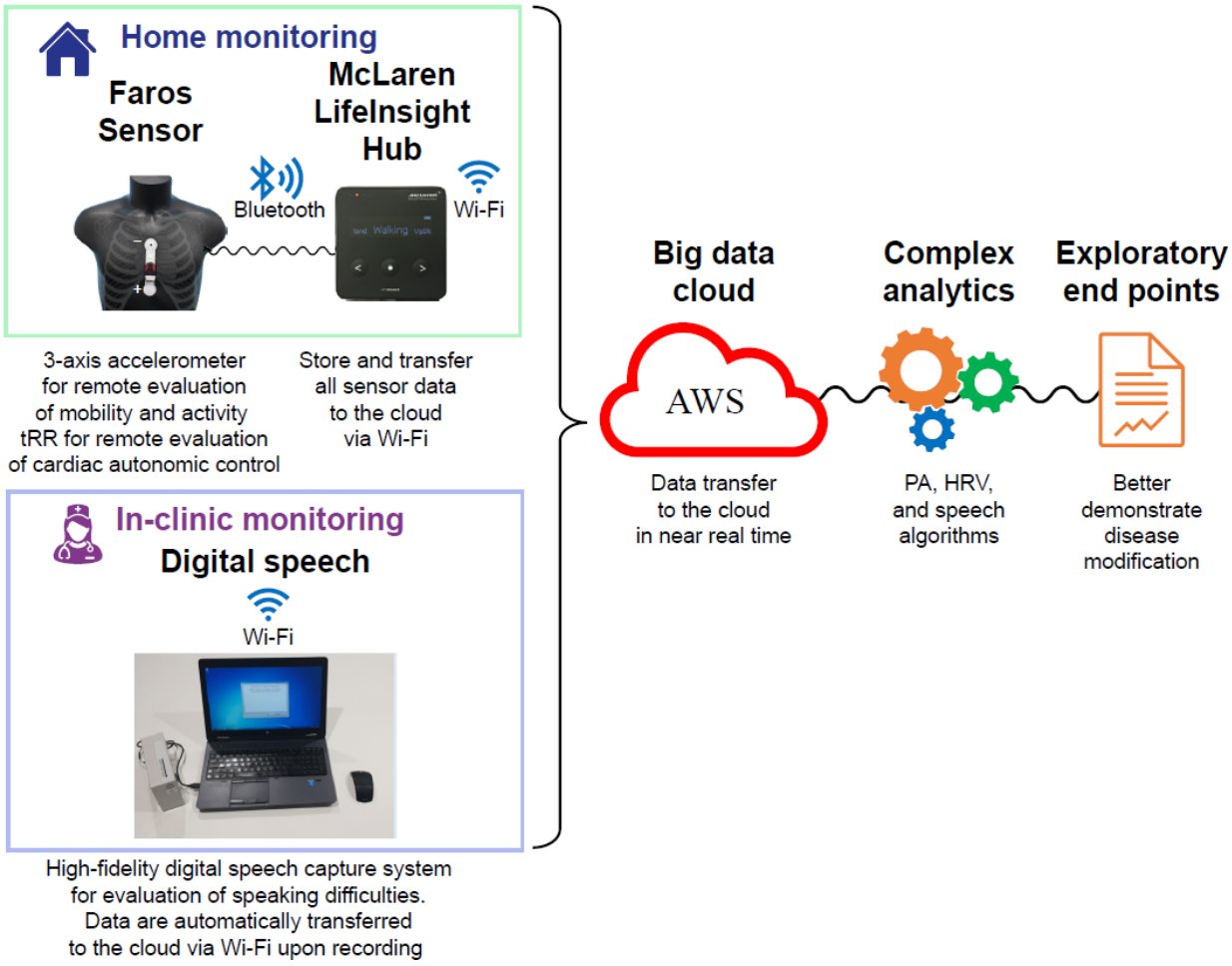
Technology Description

A monitoring system (Figure 2) was developed to allow evaluation of key ALS symptoms. It comprised the following 3 main components:

1. The commercially available Mega Faros 180 accelerometer (3 axes, 50 Hz) and 2-lead ECG sensor (Mega Electronics Ltd, Finland). The selected heartbeat sensing electrode was the Mega Fast Fix electrode disposable patch. The accelerometer and the electrode were attached to the chest (Figure 2). It was expected that physical activity and tRR data would be continuously captured by this sensor (extracted from 1 kHz ECG data, enough for HRV analysis [24]; full ECG data are not available with the sensor configuration utilized). This device is in conformity with the provisions of the Council Directive 93/42/EEC of June 14, 1993, concerning medical devices, and it was compliant with the standards of measurement from the European Society of Cardiology and the North American Society of Pacing and Electrophysiology [25].
2. A "LifeInsight Hub (v.2.0.6.)," developed by MAT, received data from the sensor via a secure Bluetooth wireless signal every 2 min. The hub then automatically uploaded the data in near real time (every 10 min) to secure cloud servers (Amazon Web Services EC2) via a secure connection on a third-generation mobile phone network. The hub allowed secure connection to additional devices, and it may be utilized to receive and transmit data from different data capture technologies, depending on future study needs.

3. A digital speech capture system comprising a high-fidelity microphone connected to a computer, with bespoke software that instructed the patients to say a series of vowels, words, and paragraphs, which were then recorded and immediately automatically transferred to a secure server via mobile connectivity.

Figure 2. Schematic representation of the monitoring platform. AWS: Amazon Web Service; HRV: heart rate variability; PA: physical activity; tRR: interbeat interval.



Physical Activity

A total of 2 types of activity algorithms (summarized in the Multimedia Appendix 1) were developed to interpret the raw data from the accelerometer: "Activity score" algorithms to evaluate "how much" activity the patients performed and "Activity classification" algorithms to evaluate "what" activities the patients performed.

The activity score is used as a measure of physical activity and is based on the intensity of movement. To develop the activity classification algorithms, a series of reference tasks were performed by the patients at each clinic visit in both the Pilot and Core Study Phases. A group of healthy volunteers outside the clinical study also performed the same tasks before the Core Study Phase; this can enhance algorithm accuracy when the number of patients is small [26]. All the reference tasks served as a "blueprint" for specific movements measured by the accelerometer. The data generated from these tasks helped in developing the algorithms and evaluating their performance using a leave-one-patient-out validation procedure. The reference tasks (which included sitting, standing, lying down, walking, climbing stairs, and transitions such as sit to stand, stand to sit, stand to lying, lying to stand, and nine-hole peg test of manual dexterity) were predefined to enable the development of clinically meaningful digital markers of disease progression for investigation. This process aligns with the recommendations from the Clinical Trials Transformation Initiative [27].

Heart Rate Variability

There are a number of widely accepted HRV metrics [17]; 2 such metrics include the Root Mean Square of the Successive Differences (RMSSD; time domain) and LF/HF ratio (frequency domain).

It is important to note that measures of HRV are derived from tRR data and are impacted by the duration of the time series (number of data points), body orientation, time of day, and activity being performed. Where possible, these factors have been considered by using the recommended 5-min duration of tRR and providing values specific to each activity.

From our analysis (summarized in the Multimedia Appendix 1), the RMSSD metric is less sensitive to the amount of data points missing. As per our analysis (Multimedia Appendix 1), around 88% of the data was needed to reliably compute RMSSD for a 5-min window, whereas almost all the data (99%) were

needed to reliably compute LF/HF. Therefore, the RMSSD may be considered a metric that is more robust for patients with lower data quality. This is a similar finding to other studies that have found time-based metrics to be less sensitive to missing data and poor data quality than frequency-based metrics [28].

Speech

The speech data collection was performed during the clinical site visits for simplicity and to allow for a more sensitive and structured assessment at this early exploratory stage. Speech algorithms to extract acoustic, quantitative, and linguistic audio features were developed based on classical speech processing techniques (eg, formants and fundamental frequencies) and used, where possible, open-source components to implement these techniques.

The patients were asked to perform 4 speech tests that were identified as clinically relevant. The first 2 tests used the phonation of the "Ah" sound, one being shortly sustained and repeated 7 times and the other being a long-sustained sound for 10 seconds [22]. The subsequent test was to pronounce the word "doily" 3 times [11]. Finally, the patients were asked to read a short 100-word paragraph ("Bamboo" passage) [29]. This particular speech data collection protocol comprised tests that had individually been shown to demonstrate associations between acoustic measures and speech intelligibility in patients with ALS but who had not been jointly investigated within the same patient group.

In contrast to physical activity and HRV, the data and processing of speech were done offline using a MATLAB script to analyze the speech data and generate a comma-separated value file with the desired end points. This process is summarized in the Multimedia Appendix 1.

Results

Pilot Study Phase: Feasibility of Equipment and Data Transmission

A total of 5 patients with ALS were recruited for the Pilot Study Phase and attended 2 clinical visits. Patients also wore the sensor in their routine home-life setting for approximately 3 days after each clinical visit.

At the time of reviewing the data from the Pilot Study Phase, clinic reference task data were available for 4 of the 5 patients and home monitoring was available for 3 out of 5 patients. The missing data did not propagate through the system within the expected timeframe, but data were subsequently fully recovered. For every patient, the number of days having least 18 hours' data available varied from 2 to 3. The device recharging routine (2 hours each day) was successfully adopted by 3 of the 5 patients.

HRV data were available for 4 of the 5 patients and the percentage of success of good-quality HRV windows of data varied from 74.8% to 100% for RMSSD analyses and 53.2% to 100% for LF/HF analyses. This was deemed to be acceptable and no adaptations for the Core Study Phase were considered necessary. The amount and quality of the speech data captured were as planned.

Most (4; 80%) patients were able to attach and remove the sensor without assistance. A patient reported that the sensor fell off on a single occasion. Most patients (4; 80%) found the device comfortable to wear. All patients reported that wearing the sensor did not have an impact on the ability to perform daily activities. Overall, the sensor was well tolerated, although 1 of the 5 patients experienced mild skin irritation.

Satisfactory user acceptance for the new technology, combined with the successful capture of study data and absence of adverse events (AEs), enabled the study to continue to the Core Study Phase with minimal tweaks to the monitoring platform. No protocol amendments based on the Pilot Phase results were necessary.

Core Study Phase

A total of 25 patients (including the 5 patients who completed the Pilot Study Phase) were enrolled in the Core Study Phase. A total of 18 of the 25 patients completed the study; 4 patients discontinued from the study because of withdrawal of consent (3 of whom had become too unwell to continue with the study because of progression of ALS, the fourth patient did not provide a specific reason), 2 patients discontinued because of AEs, and 1 patient was discontinued at the investigator's discretion. The mean (standard deviation) ALSFRS-R score at baseline was 41.6 (4.98). The mean (standard error) monthly rate of change for ALSFRS-R total score was 0.9 (0.23) points/month.

Physical Activity and Heart Rate Variability: Impact, Data Quantity, Quality, and Algorithm Accuracy

At the start of the Core Phase, a majority of patients (16/25, 64% patients) were able to attach and remove the sensor without assistance; however, the proportion of patients who did not require assistance gradually decreased over the course of the study, and by Week 48, only 6 of 15 patients (40.0%) were able to attach the sensor without some assistance. A majority of patients found the device comfortable to wear; the majority of patients who reported that the device was uncomfortable reported symptoms of skin irritation (skin itching and local skin reactions likely because of an allergy to the adhesive). A majority of patients reported that wearing the sensor had no or minimal impact on the ability to perform daily activities during the Core Phase; only 1 patient reported a moderate impact on daily activities at Week 12.

A low number of patients reported a minimal impact of the sensor on sleep: the reasons given included difficulty in getting comfortable, local itching, and impact of the flashing light at night. A patient reported a moderate negative impact on sleep.

Most patients adhered to protocol requirements on sensor wear: 65% of visits with data across the subjects had at least three days of data. However, only 6 of 25 (24%) patients had data captured for all 13 home monitoring time points.

Figure 3 illustrates the number of days of raw accelerometer and tRR data captured for each patient at each time point. There are examples of good data coverage where the patient has been able to follow the protocol (eg, patients A, B, and C) and many examples of deviations from the ideal data coverage. Patients O, P, T, V, W, X, and Y ended the study early.

There are significant gaps in the data coverage for patients Q, R, S, and U, as indicated by the zeros in Figure 3. Others have smaller gaps in the data coverage. Across all 25 patients, no data were collected for 24% of time points expected.

There are several instances where more than 3 days, or partial days, of data were collected (partial days represent those days where any data were captured). Patient O used the sensor for 8 days during 1 home monitoring period, and there are several occurrences of 5 or 6 days of wear time across different patients.

At baseline, 23 of 25 (92%) patients provided data for 3 or more days (or partial days) during the home monitoring; however, at Week 48, this reduced to 10 of 18 (56%) patients, indicating that either an increasing number of patients were unable to meet the protocol requirements as the study progressed or they were less willing to comply with study procedures. Over the course of the study, there was deterioration in the ALSFRS-R total score, indicating worsening of ALS (loss of physical function), and a decreasing number of patients provided any home monitoring sensor data (Figure 4).

If a patient wore the sensor for any period during the day which was significantly lower than the requested 22-hour period, then the derivation of any activity type end point, even if normalized to wear time, may be biased because of the heterogeneity of activity patterns throughout the day. The total wear times over each 24-hour period showed that patients did not wear the device for long enough. Average total wear time over the 24-hour recording period showed mean total wear time was 1106 min (approximately 18.4 hours) per day at baseline. The mean total wear times did not decrease significantly throughout the study to Week 48 (1084 min; 18.1 hours) per day and did not decrease significantly from the first day of monitoring, “Day 1” (1086 min; 18.1 hours) to the third, “Day 3” (1051 min; 17.5 hours). Therefore, although fewer patients provided data as the study progressed, the quality of the daily data provided by participating patients did not deteriorate significantly.

Figure 3. Home monitoring data coverage: number of days (or partial days) of data captured for each patient and time point. Blank entries indicate early withdrawals. Patients had either a Pilot Study Phase or Core Study Phase baseline home monitoring period but not both. For clarity, in 5 instances, data were excluded from the analysis because of predefined data quality deviation rules unrelated to the monitoring platform.

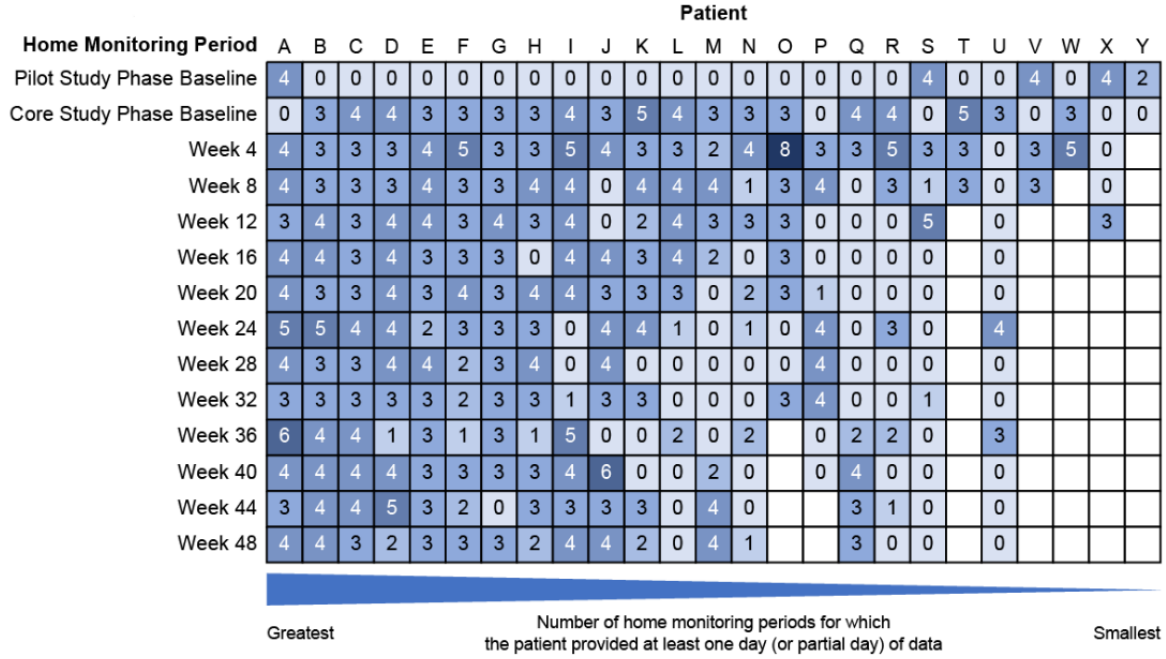
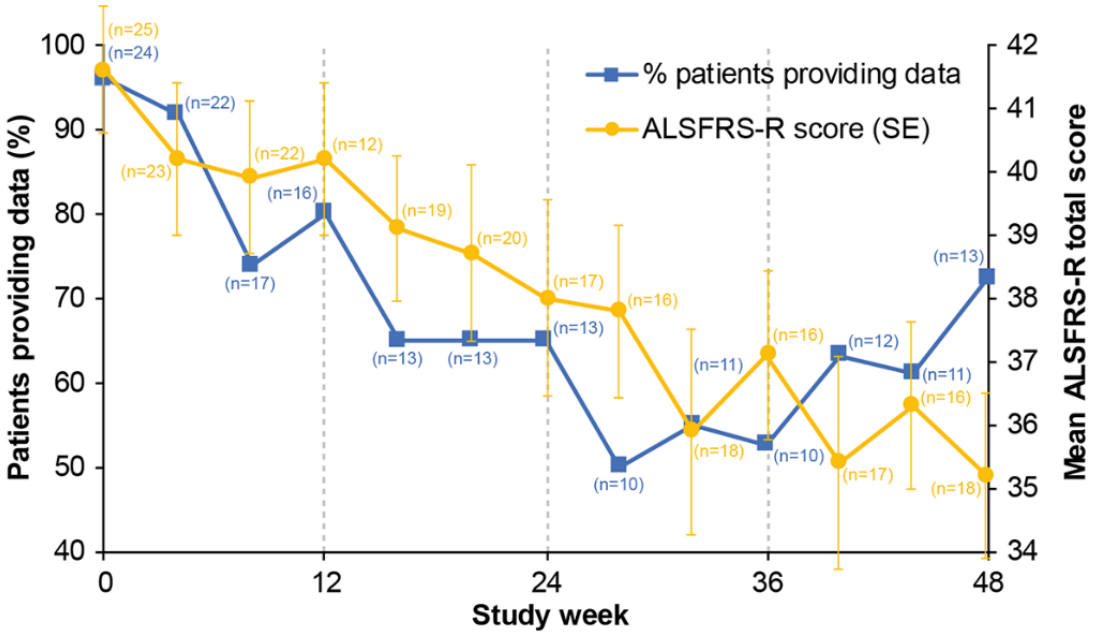


Figure 4. Percentage of patients providing home monitoring data throughout the study; mean Amyotrophic Lateral Sclerosis Functional Rating Score (Revised) total score of all patients remaining in the study. The number in brackets next to each time point represents the number of patients in the study. ALSFRS-R: ALS Functional Rating Score (Revised).



Physical activity algorithms were developed to classify "active," "sedentary but not lying," and "lying," as these 3 activity classes were expected to enable describing patients' level of physical function. For clarity, "sedentary but not lying" refers to any activity requiring low levels of intensity (but excluding "lying down"), such as standing, sitting, or performing low-intensity movements. The algorithms were tested using the reference tasks data. Their overall performance is illustrated by the confusion matrix in Figure 5, where it can be seen that a few "active" labels are being predicted as "sedentary but not lying"; nonetheless, the overall performance is considered high, as accuracy, sensitivity, and specificity are greater than 97% for all 3 activity classes.

Overall, the amount and quality of the data collected and the accuracy of the algorithms developed were sufficient to evaluate changes over time in patients' activities of daily living. Changes from baseline at Week 48 in the main physical activity end points investigated (normalized to wear time) are shown in Table 3. A reduction in the patients' ability to perform activities of daily living over time can be observed across all end points.

Figure 6 provides a high-level summary of the HRV data quality for each patient across the study. The effect of muscle activation can be considered small enough, which was supported by the tRR data. Some patients had poor data quality throughout the study, whereas others had periods of both good and poor data. Very few patients had good data quality throughout the study. The percentage of windows with enough good-quality data to compute the RMSSD and LF/HF metrics varied across the patients. Some patients had very poor data quality, with very few windows having sufficient data to compute the HRV metrics (eg, patient P and patient W). It is possible that those patients had a poor contact between the sensor and the skin using the Fast Fix.

Figure 5. The confusion matrix for the physical activity algorithms. The percentages represent the predictions of "Actual classes." NL: number of data labels from reference tasks, collected by 24 Amyotrophic Lateral Sclerosis patients (1 patient did not provide reference task data). Each label contains 1 min of accelerometer data.

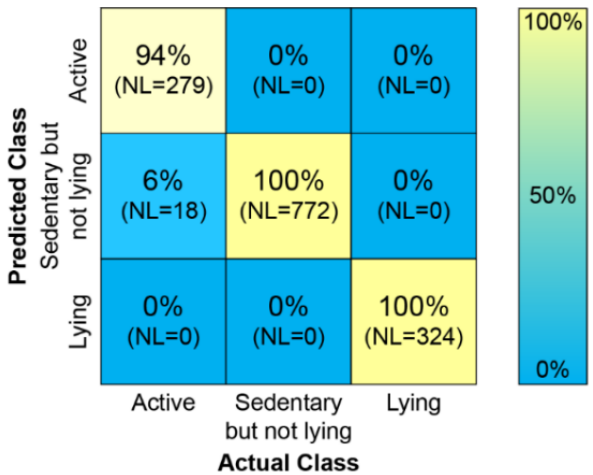
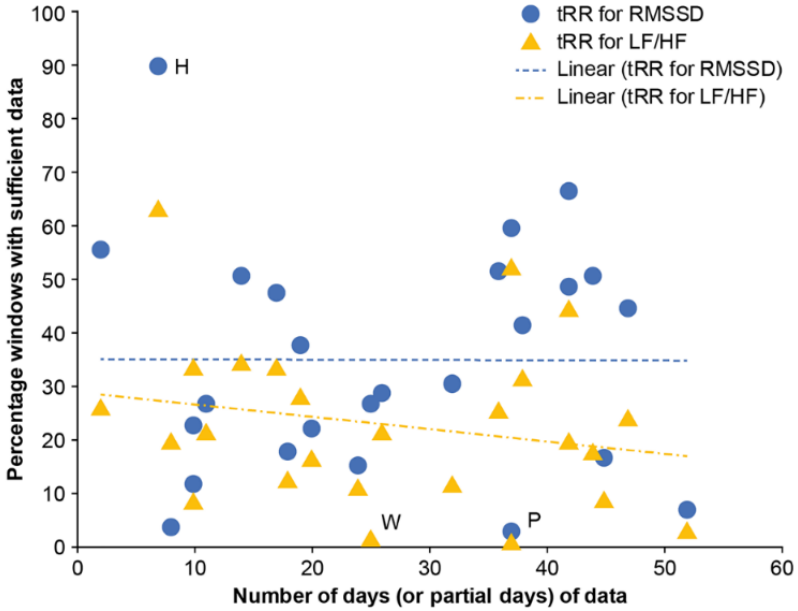


Table 3. Changes in physical activity end points from baseline to Week 48.

End point	Baseline (n=24), mean (SE)	Week 48 (n=13), mean (SE)
Average daytime active (minutes)	34.42 (6.00)	23.58 (8.05)
Percentage of daytime active (%)	5.36 (0.85)	3.59 (1.16)
Average daytime sedentary (minutes)	602.22 (26.89)	651.11 (43.36)
Percentage of daytime sedentary (%)	94.64 (0.85)	96.41 (1.16)
Total daytime activity score per hour (counts)	3336.55 (541.83)	1972.51 (612.89)
Total 24-hour activity score per hour (counts)	2275.65 (370.22)	1430.35 (464.89)
Maximum daytime activity score per hour (counts)	2803.91 (575.50)	1647.04 (456.40)
Mean maximum daytime activity score per hour (counts)	1618.40 (263.01)	1002.06 (280.53)
Daytime number of active periods per hour (minutes)		
>1 to ≤2	0.28 (0.06)	0.25 (0.10)
>2 to ≤5	0.11 (0.03)	0.10 (0.04)
>5 to ≤15	0.04 (0.02)	0.02 (0.01)
Average duration of active periods >1 min (minutes)	2.60 (0.24)[a]	2.08 (0.13)[b]

[a]n=21.
[b]n=7.

Figure 6. Percentage of windows with sufficient data to compute heart rate variability metrics for each patient. LF/HF: low-frequency/high-frequency; RMSSD: Root Mean Square of the Successive Differences; tRR: interbeat interval.



Some patients had much better data quality. In particular, nearly 90% of the windows for patient H had sufficient data to compute RMSSD; however, this patient only wore the sensor for 2 home monitoring periods (7 days of data in total); therefore, this high percentage is biased by the limited wear time (Figure 3).

For all data across the whole study, only 33% of 5-min windows of tRR data were sufficient to compute RMSSD metrics, and only 19% were sufficient to compute the LF/HF metrics. Overall, the percentage of RMSSD and LF/HF quality data did not correlate with the number of days of data; there was only a slight decrease in LF/HF quality data, with increasing number of days of data.

A slight decrease in the RMSSD HRV mean and variance data was observed over time (data not shown), but this finding should be treated with caution, given the small amount of data available. Unsurprisingly (because of the limited amount of LF/HF data available), no discernible trends were observed over time in the LF/HF data.

Speech: Data Quantity and Quality

The quantity and quality of the digital speech data files captured during the Core Study Phase of the study was as intended, with no significant issues related to the methods or equipment. Both at baseline and at Week 48, all (100%) patients captured the digital speech data successfully. However, patients J, P, and U did not perform the speech test at Weeks 12, 24, or 36. All data

files captured were transferred as anticipated, and no data were lost. All speech data were successfully analyzed. Change from baseline values at Week 48 were low for all speech end points studied, with no obvious pattern of change over time.

Safety Evaluation: Adverse Events and Serious Adverse Events

Safety evaluation comprised reported AEs and serious AEs (SAEs) related to the study equipment, devices, or procedures only. A total of 6 (24.0%) patients reported AEs assessed as related to the devices (including the sensor and Fast Fix adhesive patches); there were no reports of SAEs and no deaths, and only 2 (8.0%) patients had AEs that led to withdrawal, both because of contact dermatitis. Table 4 summarizes AEs in the study.

All the AEs reported in the study were in the skin and subcutaneous tissue disorders system organ class (SOC) and all were skin reactions related to the use of the Fast Fix adhesive patch. The most commonly reported AE was contact dermatitis.

A total of 21 AEs were reported by 6 patients. Events were mild or moderate in intensity in 5 of 6 (83.3%) patients and all events were recovered/resolved at the end of study (1 patient had an AE of contact dermatitis, which was recovered/resolved with sequelae). A patient had 2 events of pruritic rash of severe intensity: the first event lasted for 4 days; the second event lasted for 5 days.

Table 4. Summary of adverse events in the Core Study Phase (N=25).

Adverse event	Adverse events reported	Patients reporting adverse events
Adverse events	21	C, F, J, U, W, and Y
Severity of events		
Mild	14	Y, F, J, and C
Moderate	5	Y, U, and W
Severe	2	U
Severe adverse events	0	—[a]
System organ class		
Skin and subcutaneous tissue	21	C, F, J, U, W, and Y
Disorders		
Dermatitis contact	15	Y, F, U, and W
Rash	3	J and C
Rash pruritic	2	U
Skin irritation	1	Y
Adverse events leading to withdrawal	2	W and Y
Deaths	0	—

[a]No patient reported an adverse event.

Medical Device Incidents, Near-Miss Incidents, and Malfunctions

A near-miss incident was reported in the Core Study Phase: it was discovered that, despite being CE-marked, in some circumstances, it was possible for the mains chargers' casing to come apart, exposing the internal wires and components, resulting in a risk of an electrical shock. The clinical sites were informed immediately, and all the chargers were recalled from all the patients and replaced. A patient reported a replacement charger made crackling noises and smelled of burning when it was plugged into the mains. The issue was reported to the manufacturer, who confirmed there were no other instances of failure in over 15,000 units sold over a 5-year period before receiving this report. The patient was issued with a replacement charger. Before receiving the new charger, the patient had been able to use a mobile phone charger; therefore, no data were lost.

No incidents or malfunctions were reported with the use of Mega Faros sensor, Fast Fix electrode, or LifeInsight hub in this study.

Discussion

Principal Findings

Objective monitoring of disease manifestations in clinical trials using sensors is a rapidly advancing area, with over 100 actively recruiting trials using technology-based outcomes across a range of diseases (as of July 2019), according to clinical trial registers [30]. Interest is building across multiple conditions, including other neurodegenerative diseases, such as PD and multiple sclerosis [31]; RA [26,32], and COPD [33]. This study successfully developed and explored the feasibility of a novel integrated platform to objectively monitor multiple ALS symptoms across several domains (physical activity, HRV, and speech). This is a key advantage compared with other existing approaches that monitor symptoms individually [10,34].

Moreover, the platform is flexible and could be tailored to the needs of future studies, as components for measuring different symptoms may be added or removed as required. The system allowed for real-world and in-clinic data collection, both of which could be reviewed remotely in near real time.

The dual-phase study design proved to be useful, as the Pilot Study Phase enabled adaptation of components and algorithms before embarking on the Core Study Phase, thereby lessening the risk that major changes would be needed during the Core Study Phase. From an ethical perspective, this is particularly important in ALS, as patients have a limited life expectancy and are making a significant effort donating their time to the study; therefore, clinical trial sponsors have a duty to maximize the value of their time and data collected. It is also important to note that the study participants continued to receive ALS standard of care (eg, riluzole and multidisciplinary care).

Core Study Phase results demonstrate that the monitoring platform can measure physical activity and digital speech for patients with ALS over the course of the 48-week study, although the specific sensor and analysis methodology used were not successful for continuously measuring tRR data, and, hence, HRV.

Importantly, this novel study showed that it was possible to assess the physical activity of patients with ALS in their daily life (real-world monitoring): the data propagated through the system and were received as expected, although in some instances, the data experienced delay in being transmitted because of poor network connectivity, which prevented the data to be available for review in true real time. At baseline, 23 of 25 patients collected at least three days of home monitoring data; however, this decreased to 10 of 18 patients at Week 48. The decreasing amount of data collected with study progression may be partly explained because of loss of physical function with disease progression; patients had increased reliance on carers to attach or remove the device and were less able to follow the study procedures. It might also be possible that patients were simply willing to provide less data as the “novelty” effect faded away over time. This highlights that it is important to achieve a balance between patient burden and the ease of use vs the amount of the data being collected. If technology is too intrusive, awkward, or requires frequent interaction with patients, then it is likely that compliance will be adversely affected, reducing the amount of valuable and useful data collected. Wrist-worn devices have shown excellent patient acceptance in previous studies [35]; however, they are currently unable to monitor HRV continuously, and from a biomechanical perspective, the wrist is a less-than-ideal wear location for accurate activity classification.

HRV data quality appeared to be affected by a poor connection between the Fast Fix electrode patch and the chest, yielding insufficient HRV data on which one may draw any robust clinical conclusions for both the RMSSD and LF/HF analyses. The lower than anticipated quality of HRV data could have been prevented if we had identified this as an issue at the end of the Pilot Study Phase. However, the Pilot Study Phase data showed a much greater percentage of good-quality HRV windows; this was likely biased because of the low number of patients and monitoring periods. A greater number of patients and/or monitoring periods would have been needed at the Pilot Study Phase to identify this issue. In future studies, data quality could be improved by optimizing study design and using newer, less invasive sensors with a more reliable skin attachment and not needing to be recharged daily. Owing to the rapid technological evolution that has occurred in the wearable device market, wearable devices are now available and are expected to continue improving reliability of data collection over time, but they were not available at the time this study was designed. Moreover, the analysis of HRV data in this study was performed using the typical 5-min duration of tRRs, but the literature is unclear as to how long HRV analysis windows should be to offer the best compromise between the quality/accuracy of the metrics extracted and the quantity of data resulting from the analysis. The longer the analysis window, the more reliable the HRV parameters; the shorter the analysis window, the lower the amount of data that needs to be excluded from analysis because of data points missing. Recent studies suggest that 2-min windows and 3-to-4 min windows would not negatively affect the quality of RMSSD and LF/HF analyses, respectively [36]; therefore, these might be a better choice to maximize data availability.

The amount and quality of the speech data captured were as intended, and there were no significant issues with the methods, equipment, or the data analysis. However, speech analysis was performed only during in-clinic visits. Future studies may employ different technologies, such as mobile phones, to collect digital speech data at home and more frequently. This approach is currently being followed in a number of ongoing studies, such as the AT Home study in ALS [9] or the mPower study in PD [37]. From a statistical perspective, more frequent data collected at home would be useful to investigate novel end points, particularly if within-patient and between-patient end point variability is high.

Patients were able to use all the study equipment successfully, with the majority of patients reporting that the sensor was comfortable to wear and had no or minimal impact on their ability to perform daily activities, nor did it affect their sleep quality. The technology was generally well tolerated with AEs in the skin and subcutaneous tissue disorders SOC reported by 6 of 25 (24.0%) patients, with the events being mild or moderate in intensity in 5 of 6 (83.3%) patients. This is a typical drawback of using sensors that are directly affixed to the skin; wearing armbands or wrist-worn devices often lessens the risk of skin irritation, but these sensors are less appropriate for continuous HRV monitoring.

The near-miss incident with the study technology, which occurred during the study, highlighted the importance of patient safety considerations that need to be taken into account before deciding upon which technologies are acceptable to use. Thorough risk assessments need to be carried out and risk management measures put in place before study starts, in addition to the use of CE-marked (or equivalent) equipment.

Conclusions

In conclusion, the novel monitoring platform tested in this exploratory study was successful in collecting ALS patient data

remotely, which may be useful in identifying digital markers of disease progression; the home monitoring of physical activity was a particular accomplishment that may generate insights into the real lives of patients with ALS not previously appreciated, thus providing the clinicians with a valuable new understanding of their patients' well-being. The limitations of the study included a large amount of missing data, partly because of a number of patients withdrawing from the study early, which limited the ability to draw robust clinical conclusions; however, the platform was able to capture objective data to monitor patients outside of clinical visits, in near real time, and further improvements are expected with the ever-rapid advancements in digital sensor technologies. The relationship between the exploratory biotelemetry end points and the clinical gold standard ALS measures will be reported separately (manuscript under preparation). Ultimately, the goal should be to develop a platform that seamlessly integrates the data from various sources and standardizes ALS clinically relevant measures, procedures, and data reporting and validation and, ultimately, regulatory acceptance of the digital clinical trial end points.

Acknowledgments

The authors are very grateful to all the patients who participated in this study and their families. The authors wish to thank GSK colleagues: Theodora Adamou for involvement in the data analysis, Jane Temple for her contribution to the study design, Jo Barnard and Alison Peacock for their clinical operational input, and Michelle Crouthamel for her support in managing the project. The authors also wish to thank Nemanja Tiosavljevic, MAT, for developing the speech algorithms, and Jennie McLean, PhD, Fishawack Indicia Ltd, for language editing and redrawing figures (funded by GSK). This study (201283) was funded by GSK. This study represents independent research, part-funded by the National Institute for Health Research Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London, United Kingdom.

Authors' Contributions

MLK, AL, MRT, KT, CS, and AA-C contributed to study conception or design. LG-G, AL, JP, RH, RM, MRT, KT, TC, CES, and AA-C contributed to data acquisition. LG-G, MK, AL, JP, RH, MRT, KT, CES, and AA-C helped in data analysis or interpretation.

Conflicts of Interest

LG-G and MLK are employees of GSK and hold stocks/shares. At the time of the study, AL was an employee of GSK; currently, he holds stocks/shares. CES has previously consulted for GSK (>5 years ago), has received research grants from Vertex and Chronos Therapeutics in the past, and has an active grant with Eli Lily. AA-C reports consultancies for GSK, Cytokinetics, Biogen Idec, Treeway Inc, Chronos Therapeutics, OrionPharma, and Mitsubishi-Tanabe Pharma and was chief investigator for commercial clinical trials run by OrionPharma and Cytokinetics. JP, RH, RM, MRT, KT, and TC have no conflicts of interest to declare.

Multimedia Appendix 1
Additional methodological information on the algorithms used for physical activity, heart rate variability, and speech.
[DOCX File , 278 KB - mhealth_v7i12e13433_app1.docx]

References

1. Kiernan MC, Vucic S, Cheah BC, Turner MR, Eisen A, Hardiman O, et al. Amyotrophic lateral sclerosis. Lancet 2011;377(9769):942-955. [doi: 10.1016/s0140-6736(10)61156-7]
2. Talbot K. Motor neuron disease: the bare essentials. Pract Neurol 2009 Oct;9(5):303-309. [doi: 10.1136/jnnp.2009.188151] [Medline: 19762894]
3. Bensimon G, Lacomblez L, Meininger V. A controlled trial of riluzole in amyotrophic lateral sclerosis. ALS/Riluzole Study Group. N Engl J Med 1994 Mar 3;330(9):585-591. [doi: 10.1056/NEJM199403033300901] [Medline: 8302340]
4. Fang T, Al Khleifat A, Meurgey J, Jones A, Leigh PN, Bensimon G, et al. Stage at which riluzole treatment prolongs survival in patients with amyotrophic lateral sclerosis: a retrospective analysis of data from a dose-ranging study. Lancet Neurol 2018;17(5):416-422. [doi: 10.1016/s1474-4422(18)30054-1]
5. Hardiman O, van den Berg LH. Edaravone: a new treatment for ALS on the horizon? Lancet Neurol 2017 Jul;16(7):490-491. [doi: 10.1016/S1474-4422(17)30163-1] [Medline: 28522180]
6. Petrov D, Mansfield C, Moussy A, Hermine O. ALS clinical trials review: 20 years of failure. Are we any closer to registering a new treatment? Front Aging Neurosci 2017;9:68 [FREE Full text] [doi: 10.3389/fnagi.2017.00068] [Medline: 28382000]
7. Katyal N, Govindarajan R. Shortcomings in the current amyotrophic lateral sclerosis trials and potential solutions for improvement. Front Neurol 2017;8:521 [FREE Full text] [doi: 10.3389/fneur.2017.00521] [Medline: 29033893]
8. Cedarbaum JM, Stambler N, Malta E, Fuller C, Hilt D, Thurmond B, et al. The ALSFRS-R: a revised ALS functional rating scale that incorporates assessments of respiratory function. BDNF ALS Study Group (Phase III). J Neurol Sci 1999 Oct 31;169(1-2):13-21. [doi: 10.1016/s0022-510x(99)00210-5] [Medline: 10540002]

9. ClinicalTrials. ALS Testing Through Home-based Outcome Measures (ALS AT HOME) URL: https://www.clinicaltrials.gov/ct2/show/NCT03016897 [accessed 2018-11-08] [WebCite Cache ID 73mFSuvKD]
10. Prell T, Ringer T, Witte O, Grosskreutz J. Heart rate variability is decreased in patients with amyotrophic lateral sclerosis. Clin Neurophysiol 2015;126(8):e117. [doi: 10.1016/j.clinph.2015.04.180]
11. Yunusova Y, Green JR, Greenwood L, Wang J, Pattee GL, Zinman L. Tongue movements and their acoustic consequences in amyotrophic lateral sclerosis. Folia Phoniatr Logop 2012;64(2):94-102 [FREE Full text] [doi: 10.1159/000336890] [Medline: 22555651]
12. Rowland LP, Shneider NA. Amyotrophic lateral sclerosis. N Engl J Med 2001;344(22):1688-1700. [doi: 10.1056/nejm200105313442207]
13. Brooks BR, Miller RG, Swash M, Munsat TL, World Federation of Neurology Research Group on Motor Neuron Diseases. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. Amyotroph Lateral Scler Other Motor Neuron Disord 2000 Dec;1(5):293-299. [doi: 10.1080/146608200300079536] [Medline: 11464847]
14. Lomen-Hoerth C, Anderson T, Miller B. The overlap of amyotrophic lateral sclerosis and frontotemporal dementia. Neurology 2002 Oct 8;59(7):1077-1079. [doi: 10.1212/wnl.59.7.1077] [Medline: 12370467]
15. Fang T, Jozsa F, Al-Chalabi A. Nonmotor symptoms in amyotrophic lateral sclerosis: a systematic review. Int Rev Neurobiol 2017;134:1409-1441. [doi: 10.1016/bs.irn.2017.04.009] [Medline: 28805578]
16. Chida K, Sakamaki S, Takasu T. Alteration in autonomic function and cardiovascular regulation in amyotrophic lateral sclerosis. J Neurol 1989 Mar;236(3):127-130. [doi: 10.1007/bf00314326] [Medline: 2709061]
17. Hashimoto T, Yoshiuchi K, Inada S, Shirakura K, Wada N, Takeuchi K, et al. Physical activity of elderly patients with rheumatoid arthritis and healthy individuals: an actigraphy study. Biopsychosoc Med 2015;9:19 [FREE Full text] [doi: 10.1186/s13030-015-0046-0] [Medline: 26442128]
18. van Buul AR, Kasteleyn MJ, Chavannes NH, Taube C. Physical activity in the morning and afternoon is lower in patients with chronic obstructive pulmonary disease with morning symptoms. Respir Res 2018 Mar 27;19(1):49 [FREE Full text] [doi: 10.1186/s12931-018-0749-4] [Medline: 29587841]
19. Lipsmeier F, Taylor KI, Kilchenmann T, Wolf D, Scotland A, Schjodt-Eriksen J, et al. Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 Parkinson's disease clinical trial. Mov Disord 2018 Aug;33(8):1287-1297 [FREE Full text] [doi: 10.1002/mds.27376] [Medline: 29701258]
20. Linden D, Diehl RR, Berlit P. Reduced baroreflex sensitivity and cardiorespiratory transfer in amyotrophic lateral sclerosis. Electroencephalogr Clin Neurophysiol 1998 Oct;109(5):387-390. [doi: 10.1016/s0924-980x(98)00035-6] [Medline: 9851294]
21. Pinto S, Pinto A, De Carvalho M. Decreased heart rate variability predicts death in amyotrophic lateral sclerosis. Muscle Nerve 2012 Sep;46(3):341-345. [doi: 10.1002/mus.23313] [Medline: 22907223]
22. Lundy DS, Roy S, Xue JW, Casiano RR, Jassir D. Spastic/spasmodic vs tremulous vocal quality: motor speech profile analysis. J Voice 2004 Mar;18(1):146-152. [doi: 10.1016/j.jvoice.2003.12.001] [Medline: 15070235]
23. Rong P, Yunusova Y, Wang J, Zinman L, Pattee GL, Berry JD, et al. Predicting speech intelligibility decline in amyotrophic lateral sclerosis based on the deterioration of individual speech subsystems. PLoS One 2016;11(5):e0154971 [FREE Full text] [doi: 10.1371/journal.pone.0154971] [Medline: 27148967]
24. Ellis RJ, Zhu B, Koenig J, Thayer JF, Wang Y. A careful look at ECG sampling frequency and R-peak interpolation on short-term measures of heart rate variability. Physiol Meas 2015 Sep;36(9):1827-1852. [doi: 10.1088/0967-3334/36/9/1827] [Medline: 26234196]
25. Task Force of the European Society of Cardiology and the North American Society of Pacing Electrophysiology. Heart rate variability: standards of measurement, physiological interpretation and clinical use. Circulation 1996 Mar 1;93(5):1043-1065. [doi: 10.1161/01.cir.93.5.1043] [Medline: 8598068]
26. Andreu-Perez J, Garcia-Gancedo L, McKinnell J, van der Drift A, Powell A, Hamy V, et al. Developing fine-grained actigraphies for rheumatoid arthritis patients from a single accelerometer using machine learning. Sensors (Basel) 2017 Sep 14;17(9):pii: E2113 [FREE Full text] [doi: 10.3390/s17092113] [Medline: 28906437]
27. Clinical Trials Transformation Initiative. Durham, North Carolina; 2017 Jun. Developing Novel Endpoints Generated by Mobile Technology for Use in Clinical Trials URL: https://www.ctti-clinicaltrials.org/briefing-room/recommendations/developing-novel-endpoints-generated-mobile-technology-use-clinical [accessed 2018-11-08] [WebCite Cache ID 73m2G39gY]
28. Kuss O, Schumann B, Kluttig A, Greiser KH, Haerting J. Time domain parameters can be estimated with less statistical error than frequency domain parameters in the analysis of heart rate variability. J Electrocardiol 2008;41(4):287-291. [doi: 10.1016/j.jelectrocard.2008.02.014] [Medline: 18367200]
29. Yunusova Y, Graham NL, Shellikeri S, Phuong K, Kulkarni M, Rochon E, et al. Profiling Speech and Pausing in Amyotrophic Lateral Sclerosis (ALS) and Frontotemporal Dementia (FTD). PLoS One 2016;11(1):e0147573 [FREE Full text] [doi: 10.1371/journal.pone.0147573] [Medline: 26789001]
30. WebCite. Explore 289,090 Research Studies in All 50 States and in 205 Countries URL: https://www.webcitation.org/73m1rJyNU [accessed 2019-11-27] [WebCite Cache ID 73m1rJyNU]
31. DasMahapatra P, Chiauzzi E, Bhalerao R, Rhodes J. Free-living physical activity monitoring in adult US patients with multiple sclerosis using a consumer wearable device. Digit Biomark 2018;2(1):47-63. [doi: 10.1159/000488040]

32. Crouthamel M, Quattrocchi E, Watts S, Wang S, Berry P, Garcia-Gancedo L, et al. Using a ResearchKit Smartphone App to Collect Rheumatoid Arthritis Symptoms From Real-World Participants: Feasibility Study. JMIR Mhealth Uhealth 2018 Sep 13;6(9):e177 [FREE Full text] [doi: 10.2196/mhealth.9656] [Medline: 30213779]
33. Gimeno-Santos E, Raste Y, Demeyer H, Louvaris Z, de Jong C, Rabinovich RA, PROactive consortium. The PROactive instruments to measure physical activity in patients with chronic obstructive pulmonary disease. Eur Respir J 2015 Oct;46(4):988-1000 [FREE Full text] [doi: 10.1183/09031936.00183014] [Medline: 26022965]
34. Andres PL, Skerry LM, Munsat TL, Thornell BJ, Szymonifka J, Schoenfeld DA, et al. Validation of a new strength measurement device for amyotrophic lateral sclerosis clinical trials. Muscle Nerve 2012 Jan;45(1):81-85. [doi: 10.1002/mus.22253] [Medline: 22190312]
35. Doherty A, Jackson D, Hammerla N, Plötz T, Olivier P, Granat MH, et al. Large scale population assessment of physical activity using wrist worn accelerometers: The UK Biobank study. PLoS One 2017;12(2):e0169649 [FREE Full text] [doi: 10.1371/journal.pone.0169649] [Medline: 28146576]
36. Bourdillon N, Schmitt L, Yazdani S, Vesin JM, Millet GP. Minimal window duration for accurate HRV recording in athletes. Front Neurosci 2017;11:456 [FREE Full text] [doi: 10.3389/fnins.2017.00456] [Medline: 28848382]
37. Bot BM, Suver C, Neto EC, Kellen M, Klein A, Bare C, et al. The mPower study, Parkinson disease mobile data collected using ResearchKit. Sci Data 2016 Mar 3;3:160011 [FREE Full text] [doi: 10.1038/sdata.2016.11] [Medline: 26938265]

Abbreviations

AE: adverse event
ALS: Amyotrophic Lateral Sclerosis
ALSFRS-R: ALS Functional Rating Score (Revised)
COPD: chronic obstructive pulmonary disease
ECG: electrocardiogram
GSK: GlaxoSmithKline
HRV: heart rate variability
LF/HF: low-frequency/high-frequency
MAT: McLaren Applied Technologies
PD: Parkinson disease
RA: rheumatoid arthritis
RMSSD: Root Mean Square of the Successive Differences
SAE: serious adverse event
SOC: system organ class
tRR: interbeat interval

Edited by G Eysenbach; submitted 18.01.19; peer-reviewed by J Salisbury, B Johnson, K Okamoto, S Helou; comments to author 27.04.19; revised version received 11.07.19; accepted 26.09.19; published 20.12.19.

Please cite as:
Garcia-Gancedo L, Kelly ML, Lavrov A, Parr J, Hart R, Marsden R, Turner MR, Talbot K, Chiwera T, Shaw CE, Al-Chalabi A
Objectively Monitoring Amyotrophic Lateral Sclerosis Patient Symptoms During Clinical Trials With Sensors: Observational Study
JMIR Mhealth Uhealth 2019;7(12):e13433
URL: https://mhealth.jmir.org/2019/12/e13433
doi:10.2196/13433
PMID:31859676

Original Paper

Adults' Preferences for Behavior Change Techniques and Engagement Features in a Mobile App to Promote 24-Hour Movement Behaviors: Cross-Sectional Survey Study

Ann DeSmet[1,2,3], PhD; Ilse De Bourdeaudhuij[3], PhD; Sebastien Chastin[3,4], PhD; Geert Crombez[5], PhD; Ralph Maddison[6], PhD; Greet Cardon[3], PhD

[1]Clinical and Health Psychology, Université Libre de Bruxelles, Brussels, Belgium
[2]Research Foundation – Flanders, Brussels, Belgium
[3]Department of Movement and Sport Sciences, Ghent University, Ghent, Belgium
[4]Glasgow Caledonian University, Glasgow, United Kingdom
[5]Department of Experimental-Clinical and Health Psychology, Ghent University, Ghent, Belgium
[6]Institute for Physical Activity and Nutrition, Deakin University, Geelong, Australia

Corresponding Author:
Ann DeSmet, PhD
Clinical and Health Psychology
Université Libre de Bruxelles
Franklin Rooseveltlaan 50
Brussels, 1050
Belgium
Phone: 32 2 650 32 82
Email: Ann.DeSmet@ulb.be

Abstract

Background: There is a limited understanding of components that should be included in digital interventions for 24-hour movement behaviors (physical activity [PA], sleep, and sedentary behavior [SB]). For intervention effectiveness, user engagement is important. This can be enhanced by a user-centered design to, for example, explore and integrate user preferences for intervention techniques and features.

Objective: This study aimed to examine adult users' preferences for techniques and features in mobile apps for 24-hour movement behaviors.

Methods: A total of 86 participants (mean age 37.4 years [SD 9.2]; 49/86, 57% female) completed a Web-based survey. Behavior change techniques (BCTs) were based on a validated taxonomy v2 by Abraham and Michie, and engagement features were based on a list extracted from the literature. Behavioral data were collected using Fitbit trackers. Correlations, (repeated measures) analysis of variance, and independent sample *t* tests were used to examine associations and differences between and within users by the type of health domain and users' behavioral intention and adoption.

Results: Preferences were generally the highest for information on the health consequences of movement behavior self-monitoring, behavioral feedback, insight into healthy lifestyles, and tips and instructions. Although the same ranking was found for techniques across behaviors, preferences were stronger for all but one BCT for PA in comparison to the other two health behaviors. Although techniques fit user preferences for addressing PA well, supplemental techniques may be able to address preferences for sleep and SB in a better manner. In addition to what is commonly included in apps, sleep apps should consider providing tips for sleep. SB apps may wish to include more self-regulation and goal-setting techniques. Few differences were found by users' intentions or adoption to change a particular behavior. Apps should provide more self-monitoring (P=.03), information on behavior health outcome (P=.048), and feedback (P=.04) and incorporate social support (P=.048) to help those who are further removed from healthy sleep. A virtual coach (P<.001) and video modeling (P=.004) may provide appreciated support to those who are physically less active. PA self-monitoring appealed more to those with an intention to change PA (P=.03). Social comparison and support features are not high on users' agenda and may not be needed from an engagement point of view. Engagement features may not be very relevant for user engagement but should be examined in future research with a less reflective method.

Conclusions: The findings of this study provide guidance for the design of digital 24-hour movement behavior interventions. As 24-hour movement guidelines are increasingly being adopted in several countries, our study findings are timely to support the design of interventions to meet these guidelines.

(JMIR Mhealth Uhealth 2019;7(12):e15707) doi:10.2196/15707

KEYWORDS

physical activity; sleep; sedentary behavior; 24-hour movement; mobile health; mobile apps; behavior change technique; engagement; adult

Introduction

Movement Behaviors

Physical activity (PA), sleep, and sedentary behavior (SB) are modifiable determinants of several negative health outcomes among adults. More time spent being physically active (more light-intensity PA [LPA] and a minimum of 30-min moderate-to-vigorous PA [MVPA] per day), a sleep duration of 7-9 hours per night, and less time spent on SB are associated with beneficial health outcomes, including lower risk of weight gain and obesity [1,2], lower risk of type 2 diabetes [3-5], lower risk of cardiovascular diseases [3,6], and higher health-related quality of life [1-7]. In Europe, 61% of the adult population meets the guidelines for MVPA [8], and 73% of adults without diseases meet the recommendations for a healthy sleep duration (7-8 hours per night) [9]. Between 10% and 42% of adults in Europe have been reported sitting for more than 7.5 hours per day [10], whereas less than 7 hours of SB per day (mainly measured via self-reports) has been suggested as beneficial for health [11]. Despite the benefits of PA, reducing SB, and getting enough sleep, most studies to date have examined the duration of these movement behaviors in isolation [12]. This is problematic, as these behaviors are interrelated; over the course of 24 hours, a change in any of these given behaviors impacts the duration of (at least one) movement behavior(s). Consequently, these behaviors need to be considered together in a multibehavior program, as targeting one behavior will cause a time displacement in another behavior, and not all displacements are equally favorable to health [13,14].

User-Centered Design of Multibehavior Programs

To date, very few multibehavior programs have been designed taking into account 24-hour movement behaviors, with no effectiveness data available thus far [15,16]. As a result, there is a limited understanding of which components should be included and how behavior recommendations are best combined in such multibehavioral interventions [17,18]. Behavior change techniques (BCTs) are uniquely identifiable components of an intervention that can be considered active ingredients of behavior change [19]. To ensure effectiveness, exposure to and active elaboration of intervention content are also required, referred to as user engagement [20]. User-centered design, in which user preferences for BCTs and other intervention features are taken into account, can help increase user engagement [20-22]. This is especially important in digital multibehavior programs that face additional challenges to user engagement because of a lack of in-person support [20]. Although BCTs are an important feature of behavior change interventions, apps may also include specific features to enhance user engagement, such as the use of celebrities [23,24], narratives [25], gamification, challenge and competition elements [25,26], interactive features such as a chat function and virtual coaches [27,28], or a social media connection [28]. In line with the Elaboration Likelihood Model, when users are not able or motivated to process the message in a rational way, the message is processed via such visual and contextual cues, here referred to as engagement features [29,30].

BCTs are best selected to fit specific determinants of behavior [31]. This may imply that different BCTs are needed and preferred by users for different behaviors. Users' needs and preferences for techniques in an intervention may change as users proceed to adopting and maintaining the behavior change [32] and may also differ between users with a higher or lower motivation to change behavior, where those with low motivation may, for example, be more interested in peripheral cues than those with high motivation for the health behavior [32,33]. We do not expect preferences for peripheral engagement features to differ by specific behavior because these peripheral cues do not rely on content matching with specific behaviors. This is therefore not included in our study.

Study Aims and Research Questions

To the best of our knowledge, no study has assessed adult users' preferences for BCTs and engagement features in relation to combined PA, sleep, and SB intervention and investigated whether these differ between movement behaviors and by the user's intention and current adoption for each health behavior domain. When designing interventions for 24-hour movement behaviors, these insights are important to ensure that users will engage with the intervention and continue to use the app for as long as it is necessary to change behavior. This study aimed to assess the following research questions: (1) What are the preferences for BCTs in mobile apps aiming to improve PA, sleep, and SB? (2) Do differences exist between participants in terms of their preferences for specific movement behavior (PA, sleep, and SB)? (3) What are their preferences for engagement features? (4) Do differences exist in participants' preferences for BCTs or features by their intention to change the health behavior and current behavioral adoption (PA, sleep, and SB)? The results of this study may inform the evidence-based design of mobile health interventions, promoting 24-hour movement behaviors in a general adult population.

Methods

Study Design

The data used for this study were part of the Healthy Worker study, a 2-week intensive measurement study that assessed PA,

sleep, SB, and their behavioral determinants. Incentives for study participation consisted of individual feedback and the possibility to win a folding bike via a raffle. On study completion, users completed a voluntary Web-based process evaluation survey using QuestionPro (Austin, Texas, US). The data reported here were derived from the process evaluation survey. Methods and results were reported in accordance with the Checklist for Reporting Results of Internet E-Surveys guidelines [34]. Surveys were pseudonymized to ensure confidentiality and avoid social desirability bias. The survey did not use a randomized order of items to maintain a logical flow. Adaptive questioning (branching and skips) was used to reduce respondent burden. Surveys were sent to individual email addresses, disabling multiple logins. Only completed surveys were included. Most items required a response to continue. Participants could save responses and continue later. It took, on average, 12 min to complete the survey, which consisted of 40 questions. Data were collected between February 2017 and May 2017. Participants received a personal feedback report at study completion and, as an incentive, they could win a folding bike via a raffle. The ethics committee of Ghent University Hospital approved this study (reference 2016/1231).

Participants and Sampling Procedure

A random sample of working-age adults (aged 22-55 years) was drawn from the civil registry in Ghent, a city of roughly 250,000 inhabitants in Flanders, Belgium. A total of 2453 letters were sent out, of which 47 were returned undeliverable. Participants were included if they were aged 22-55 years and owned a smartphone with access to the internet. Participants were requested to notify the researchers of their interest to participate by email and provide their phone number and postal address. Home visits were made to each participant to set up the Fitbit app (San Francisco, California, US) and wearable device and to retrieve the Fitbit app after study completion.

Measurements

Sociodemographic Information

Sociodemographic information including gender (male/or female), age (continuous), and highest educational degree (primary school, secondary school, postsecondary nonacademic education, or academic education) was assessed in a Web-based survey 1 week before the start of the study.

Movement Behaviors

Participants were provided a Fitbit Charge 2, a commercial wrist-worn activity tracker that includes a triaxial accelerometer and a heart rate monitor. Fitbit devices have been shown to be valid and reliable commercial devices to measure the time spent on these movement behaviors [35,36]. The Fitbit Charge 2 measured minute-by-minute activities classified as LPA, moderate PA (MPA), vigorous PA (VPA), sleep, or SB. The usability of the Fitbit trackers was assessed with Likert scale items (1-5) based on the Short Usability Scale [37] and was rated as high, with average scores between agree and completely agree (≥4 or 5) for ease of use, self-confidence in using it, quickly learning how to use it, good integration of functionality, and interest in using it in the future. Full descriptive results are provided in Multimedia Appendix 1.

Preferences for Behavior Change Techniques and Engagement Features in a Mobile App on Physical Activity, Sedentary Behavior, and Sleep

Preferences for BCTs and engagement features were rated per behavior on a 5-point rating scale. A short BCT list, consisting of 22 BCTs (Multimedia Appendix 2) [38] rather than the more extensive list of 93 BCTs [19], was used to reduce respondent burden. The BCT list was inspired by commonly used techniques in health apps [28,39,40] when scored based on BCT taxonomy of Abraham and Michie [38]. Participants rated their preferences for specific engagement app features on a 5-point rating scale including social media connection [28], gamification [25,26], competition with others [25], a virtual coach to give instructions [27,28], being able to ask questions via a chat function [27], having a narrative (eg, fictional drama) and being a character in this narrative [25], and having celebrity endorsement of the app [24].

Intention to Change Movement Behaviors

Validated scales to measure intention-to-change behavior were used for each behavior, where possible. As a result, the operational definitions of intentional phases differed between each health behavior. Intention was measured before the start of the study in a Web-based survey for PA and SB and in diaries during the 14-day follow-up period for sleep. For PA, the intention phase was determined by one item from the Belgian Environmental Physical Activity Study survey reflecting their stages of change: (1) premotivational phase: “I am not sufficiently active and have no intention to change in the next 6 months,” (2) action phase: “I am not sufficiently active and have the intention to change in the next 6 months,” and (3) maintenance phase: “I am sufficiently physically active” [41]. The action and maintenance phases were grouped into the intentional phase. For sleep, no validated behavioral determinant questionnaires existed then. A questionnaire was developed in the framework of this study and validated elsewhere (DeSmet et al, unpublished, 2019). The intention to improve sleep was defined as users’ intention to go to bed on time (measured in evening diaries, averaged over 14 days, and dichotomized by no intention ≤3 [*completely disagree to neutral*] and >3 [*rather to completely agree*] on a 1- to 5-point rating scale). For SB, users were asked for their intention to reduce their time spent sitting based on a previous scale for occupational sitting time [42], assessed here across five different domains (ie, while watching TV, using computers in leisure time, during other leisure activities, during transport, and at work). If users indicated they wanted to reduce their sitting in at least one of these areas in the next 6 months, they were considered to have an intention to reduce their SB.

Analysis

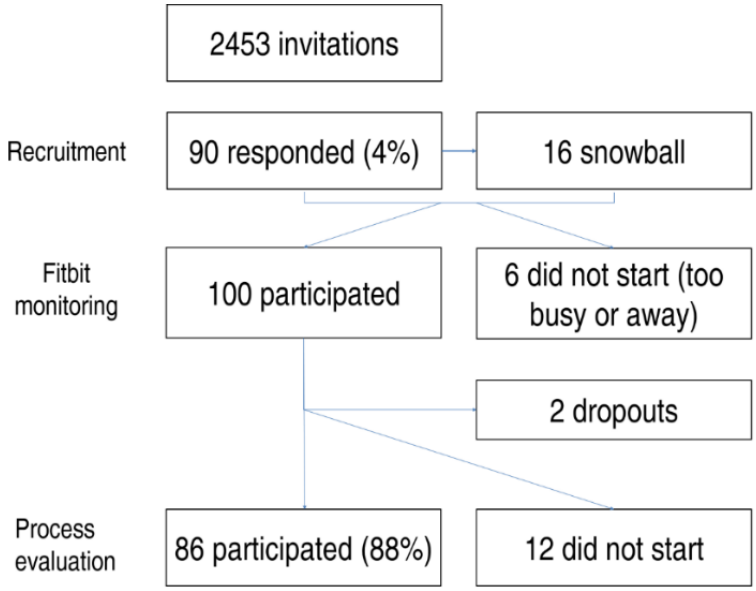
Descriptive statistics assessed participants’ average preferences for BCT by each movement behavior and for app features (research question 1). Associations of preferences with age were assessed with Pearson correlations, and differences in preferences by gender were assessed with analysis of variance (ANOVA) tests. Repeated measures ANOVA analyses were conducted to assess differences among users in their preferences for BCTs in each of the three movement behaviors (PA, sleep,

and SB). No between-factor test was used. Significant results were subsequently tested via independent sample *t* tests to detect differences between pairs (research question 2). ANOVA tests were used to assess differences among participants by their intention to change and current behavior in relation to PA, sleep, and SB. Homogeneity of variances was assessed, and significance tests were only performed for item sets with nonsignificant tests on homogeneity of variances (research question 3). All statistics were performed in SPSS, version 25 (IBM corporation, Armonk, NY).

Results

A total of 98 participants completed the main study, with 86 participants completing the process evaluation (response rate 88%; Figure 1). The analyzed sample (n=86) mostly consisted of people living in an urban area (65/86, 76%), who were well educated (71/86, 83% completed postsecondary education), living together with a partner (58/86, 67%), and working full time (66/86, 77%). The mean age was 37.4 years (SD 9.2), and 57% (49/86) of the participants were female. Fitbit data were, on average, available for 11.9 of 14 days (median 14, SD 4.3). Participants spent, on average, 4.6 hours (SD 1.19) on LPA, 25.3 min (SD 15.01) on MPA, 29.4 min (SD 20.7) on VPA, 7.6 hours (SD 0.80) asleep, and 10.3 hours (SD 1.31) on SB per 24 hours. There were no significant correlations between the number of days Fitbit data were available for and users' activity profile. Of the 74 preference items, there were significant correlations with one technique and one feature (ie, having more measured days was positively correlated with a preference for social comparison with others on PA: r=0.22, *P*=.04; having more measured days was negatively correlated with a preference for a narrative: r=−0.22, *P*=.04).

Figure 1. Study flow.



Across movement behaviors, the same BCTs were the most preferred, that is, information on the health consequences of movement behaviors, self-monitoring of behavior, feedback on how well they do, obtaining insight into their healthy lifestyles, and receiving tips and instructions on how to do the behavior (Multimedia Appendix 2). Moreover, the standard deviations for preferences on the specific techniques of *information on the link between behavior health outcome* and *behavioral self-monitoring* were small, indicating a strong agreement among users for these BCTs. Although the same BCTs were ranked highest across behaviors, there were significant differences in participants' preferences for the BCT items between the health domains of PA, sleep, and SB on all but one BCT (ie, time management skills; Multimedia Appendix 2). There was a higher preference for several BCTs to promote PA than for sleep or SB. This was the case for several BCTs associated with goal setting (ie, setting and adjusting personally desired outcomes, setting and adjusting personally relevant goals, building up toward more difficult goals, and receiving a reminder). This was also the case for several BCTs in relation to positive reinforcement (ie, receiving rewards of incentives, providing encouragement, and obtaining social support) and BCTs related to (role) modeling (ie, receiving videos that model the desired behavior and being a role model for others).

Specific BCTs were less preferred for SB than for sleep and PA. This was noted for several BCTs in relation to instructions to perform the behavior, self-monitoring, and feedback (ie, information on link between behavior and health outcome, instructions on how to perform the behavior, self-monitoring of behavior, feedback on behavior, and receiving regular feedback on how the behavior contributes to the health outcome). Identifying barriers for the behavior was less preferred for SB than for sleep and PA. For two BCTs— receiving tips tailored to their profile and to compare themselves with others with a similar profile—there was a lower preference for SB than for PA, but there was no difference with sleep. For obtaining insights into the differences between what the person does and what is needed to meet the health outcome, the preference was significantly higher for PA than for the other behaviors, and this BCT preference for sleep was significantly higher than for SB. In sum, the preference was lowest for SB on all BCTs but one (no significant difference in time management skills). Goal setting, role modeling, and reinforcement BCTs were generally

preferred more for PA than for either sleep or SB. Instructions, behavioral monitoring, and feedback were less preferred for SB than for either sleep or PA.

The preference for engagement features was low to very low (Multimedia Appendix 3).

Very few BCTs showed a difference in preference between participants with a low or high intention to change the specific movement behavior (Multimedia Appendices 4-6.) For PA only, participants with a higher intention to increase this behavior had a stronger preference for monitoring their behavior than participants with a lower intention. There were no significant differences by intention to change behavior on any other BCT preference (Multimedia Appendices 4-6). Participants who had fewer hours of sleep per night showed a higher preference for information on behavior health outcome, behavioral monitoring, and feedback (Multimedia Appendix 5). For PA, lower levels of MVPA were associated with a higher preference for a virtual coach that gives instructions and for video modeling (Multimedia Appendix 4).

Discussion

Principal Findings

When designing interventions for 24-hour movement behaviors, it is important to assess whether users would prefer a different approach for each behavior. This study examined adult users' preferences for different BCTs in a mobile app to improve 24-hour movement behaviors of PA, sleep, and SB and to assess whether these preferences differed by behavior and users' behavioral intention and adoption. The study also examined preferences for engagement features and differences in preferences by users' behavioral intention and adoption.

In general, the same BCTs were top-ranked across 24-hour movement behaviors, although the strength of these preferences was higher for PA than for SB or sleep. This suggests that the same BCTs can be used universally for combined intervention on the 24-hour movement behavior rather than behavior-specific BCTs. For all behaviors, information on behavior health outcomes, obtaining insights into their healthy lifestyles, self-monitoring, feedback, instructions, or tips were ranked the highest, and social support and comparison were ranked lowest in user preferences. The high preference for self-monitoring techniques is in line with earlier research on user preferences for apps to promote PA [28] and underscores the upcoming interest among users in commercial wearable activity trackers that monitor PA, sleep, and SB. In Belgium, 19% of adults had a wearable activity tracker in 2017 compared with only 5% owning one in 2016 [43]. Some BCTs, such as self-regulation techniques, were less preferred for both sleep and SB than for PA. A lower preference for self-regulation techniques, such as action and coping planning, to improve sleep in comparison with PA, aligns with previous research showing that users felt that self-regulation techniques were not always easy to apply, given limited control over wake time and bedtimes [44], or where integrating self-regulation techniques in mental imagery was found more useful to promote sleep than the traditional form of creating implementation intentions, given the habitual nature of bedtime routines [45].

This study moreover examined preferences and differences in user preferences for BCTs and engagement features by behavioral intention and adoption. The interest in the engagement features was generally low. A previous study on engagement features in digital health interventions to reduce alcohol consumption showed that personalization, control features (being able to make choices), and interactive features (allowing you to enter information and take a game or quiz) were most appreciated. Action plans and challenge features (competition against others) were ranked as least important by some, but as more important by others in the study. Narrative features (storyline, in which the user can be a character) were rated as not very important for engagement with the app [25]. This largely fits with our findings. This earlier study found little consistency in user preferences [25]. For example, it showed that users only valued social comparison features if they expected that they would not be outperformed by others; however, our study did not find a higher user preference for such features among users with a behavior that was in a healthier range compared with those who scored worse on these behaviors. We also found no difference in preferences for engagement features by intentional phase, disconfirming our expectation that these features would appeal more via a peripheral route among those with a lower motivation. However, our self-reported method of assessment may not have been appropriate to capture any interest in features via a peripheral route. For example, in entertainment education, where a narrative and characters are used in an entertainment format to convey health messages, a crucial condition for its effectiveness is that the message is provided unobtrusively and that the audience is unaware of the intent to change their opinions or behavior [46]. An experimental method or a conjoint analysis method where examples of apps differing in these features are shown and rated by users, rather than asking them to reflect on the importance of these features, may be better suited to assess user preferences for such engagement features in apps to promote PA, sleep, and SB.

A stronger preference for BCTs of self-monitoring, reminders or cues, role modeling, and rewards for PA than for SB and sleep may be unexpected, as these BCTs are mainly assumed to change automatic, habitual behavior [31,47,48] and would hence be expected as more preferred for SB and sleep than for PA. Possibly, a higher intention to change PA and familiarity with PA may have resulted in a stronger preference for BCTs to change this movement behavior than for SB or sleep. This is, however, not supported by our findings for SB, as no difference was noted in BCT preferences between participants with an intention to change SB and those without such an intention nor were there any significant associations found between the time spent on SB and any of the BCT preferences. For sleep, however, those with a longer sleep duration were indeed less interested in using self-monitoring for sleep. The lower preferences for SB and sleep on these BCTs than for PA may also be a consequence of the taxonomy that was used. A relatively brief set of BCTs was selected to reduce respondent burden when scoring these for three behaviors. The taxonomy

used was inspired by previous scorings of mobile health apps [28,39,40] but was designed specifically for PA and diet interventions [38,49]. Possibly for sleep and SB, other relevant techniques may need to be included, such as sleep hygiene practices and cognitive behavioral therapy for sleep, or environmental changes, emotional persuasion, and education for SB [50-52], which were not examined here. A selection of BCTs that is relevant for SB and sleep, for example, based on a larger taxonomy of 93 BCTs has been constructed with expert input from several behavioral areas [19], may be appropriate for future studies on user ratings of techniques to address 24-hour movement behaviors. Finally, we may expect that features differentially implemented across behaviors in the Fitbit that the participants had worn before filling out this survey may have affected their preference ratings. This does not appear to be the case, as preferences were higher for PA both for BCTs more strongly integrated for PA than for other behaviors (eg, detailed and real-time self-monitoring), as well as for features more strongly implemented for SB than for PA (ie, reminders and buzzes).

Some recommendations can be made when comparing user preferences with what is currently commonly used in apps. PA tracker apps often include information, self-monitoring, instructions and feedback, goal-setting techniques, social support, modeling, and social comparison [53]. Our findings show that all BCTs, except social comparison techniques, that showed the lowest preference are important to include in PA apps to ensure user engagement. Sleep tracker apps include fewer BCTs and mainly consist of self-monitoring and feedback [53], which align well with preferred BCT factors for sleep observed in our study. Our findings suggest that adding *giving tips* as BCT in sleep apps may further increase user engagement. In addition, for SB, activity tracker apps mostly include self-monitoring and feedback [53]. A previous study suggested including more social support features in SB apps [53]. On the basis of our findings, we expect that this would not increase user engagement, as preferences were lowest for social comparison and support. Self-regulation and goal-setting techniques, however, were highly preferred and may be a more useful addition to SB apps. In addition, two mobile apps have been designed to improve multibehavior patterns of PA, sleep, and SB in adults: *Balanced* and *BeWell24*. *Balanced* included goal setting, information on behavior health outcome, self-monitoring, and feedback as BCTs [16]. *BeWell24* included goals and planning, feedback, and monitoring, shaping knowledge, and associations for all behaviors, complemented with shaping knowledge, natural consequences, and repetition and substitution for some behaviors [15]. It appears that these fit well with user preferences for these behaviors, with the exception of goal setting, which is included but not highly preferred for all behaviors and tips and instructions, which are highly preferred but not included.

In addition, for BCTs, we did not find any differences according to users' intentions to improve their behaviors. Those with a lower intention to change their behavior may be more interested in BCTs that increase risk perception and positive outcome expectancies, such as feedback and information on the behavior health link [31], and those with a higher intention may have a higher preference for self-regulation techniques [32]. This was not consistent with our results; however, we did observe a higher interest in these features, not by intention, but by behavioral adoption of sleep: Participants who had fewer hours of sleep per night showed a higher preference for information on behavior health outcome, behavioral monitoring, and feedback. Thus, these techniques may be important to create awareness and initiation of health behavior change [54]. Other specific BCTs and engagement features also appeared to be preferred by people who showed lower levels of healthy behavior, indicating that these may be especially useful to support initial behavior change. Those with already higher levels of sleep wished to receive more social support for their sleep. Those with a high intention to change PA were more interested in self-monitoring their behavior. For PA, lower levels of MVPA were associated with a higher preference for a virtual coach that gives instructions and for video modeling. This suggests that more instructions and support are needed on specific exercises via which to actually achieve more MVPA. Such virtual coaches have been integrated into several research-grade mobile apps or digital interventions for PA [55-57], but in commercial settings, apps are often provided at a premium (eg, *Fitbit Coach*), although some trackers have also started providing free audio guidance during sports activities (eg, *Polar Beat* and Moov). From a public health perspective, it is unfortunate that a virtual coach would be offered at a premium because it would then be available only to those who can afford it. On the other hand, this is where an opportunity may lie for public health and academic organizations to develop such an app that incorporates a virtual coach and can be provided free as an add on to the service provided by commercial apps and wearable trackers.

Limitations and Strengths

The study had some limitations. Despite using a random sampling method to recruit a sample representative of the population, the final sample was highly educated. This is a common problem in health promotion research. The Belgian National Health Survey, a nationally representative survey conducted among 12,038 Belgian individuals aged 15 years or older, with 3191 individuals from the Flemish region, shows that 48% of Flemish individuals have attained postsecondary education compared with 83% in our sample of individuals aged 22-55 years [58]. Although some differences may result from the inclusion of a younger age group that could not have attained postsecondary education yet (ie, aged 15-21 years, approximately 9%), it is clear that our sample overrepresents people who are highly educated. This may have had an influence on behavior outcomes, as both SB and PA were higher among well-educated individuals than among less-educated people in the Belgian National Health Survey [59]. It is unclear how this may have affected the preferences for BCTs, as we are not aware of any studies assessing the differences in preferences for BCTs by educational background. We could, however, expect a lower preference for self-quantification among less educated individuals based on existing research [60]. Our findings may thus not be generalizable to users who are less educated. Our list of chosen BCTs was based on previous research on mobile health apps. A more extensive list of BCTs may have uncovered more unmet needs of users in current mobile apps. Preferences

were elicited via self-report measures in a hypothetical way. Future research should validate these findings by using experimental studies with experiences from an actual app. Eliciting user preferences is the first step in a user-centered design of a digital health intervention. On the basis of these user preferences, a prototype can be developed that can be optimized based on subsequent rounds of user feedback [61].

This study also had several strengths. It examined BCT preferences for each behavior separately, indicating important differences that should be taken into account in 24-hour activity interventions. The study examined not only BCTs but also other features that may increase engagement. As engagement and effectiveness are closely entwined, examining engagement features is important in intervention design considerations. Finally, examining differences between users in their preferences provides a useful basis for tailoring 24-hour activity interventions.

Conclusions

Across behaviors, preferences were generally highest for information on the link between behavior health outcomes, self-monitoring, behavioral feedback, insight into healthy lifestyles, and tips and instructions. In general, very few differences were found by users' intentions or adoption to change a particular movement behavior. The same ranking of preferred BCTs was found for all 24-hour movement behaviors, indicating that the same selection for all three behaviors could fit user preferences. However, some techniques may be added to more closely meet preferences for sleep and SB. In addition to what is commonly included in apps, sleep apps should consider giving tips for sleep and provide more information on behavior health outcome and feedback to support those who are further removed from healthy sleep. A virtual coach and video modeling may provide appreciated support to those who are less physically active. Social comparison and support features are not high on users' agenda and may not be needed from an engagement point of view. Engagement features may not be very relevant for user preferences but should be examined in future research with a less reflective method.

Acknowledgments

AD (FWO/16/PDO/060) is supported by the Research Foundation Flanders (Fonds Wetenschappelijk Onderzoek).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Perceived usability of the Fitbit tracker.
[DOCX File , 13 KB - mhealth_v7i12e15707_app1.docx]

Multimedia Appendix 2
Repeated measured analyses assessing within-user differences in behavior change technique preferences between health domains.
[DOCX File , 16 KB - mhealth_v7i12e15707_app2.docx]

Multimedia Appendix 3
Descriptive statistics for engagement features.
[DOCX File , 12 KB - mhealth_v7i12e15707_app3.docx]

Multimedia Appendix 4
Differences between participants in behavior change technique preferences for physical activity by users' intention to change behavior and behavioral adoption.
[DOCX File , 17 KB - mhealth_v7i12e15707_app4.docx]

Multimedia Appendix 5
Differences between participants in behavior change technique preferences for sleep by users' intention to change behavior and behavioral adoption.
[DOCX File , 16 KB - mhealth_v7i12e15707_app5.docx]

Multimedia Appendix 6
Differences between participants in behavior change technique preferences for sedentary behavior by users' intention to change behavior and behavioral adoption.
[DOCX File , 16 KB - mhealth_v7i12e15707_app6.docx]

References

1. Magee L, Hale L. Longitudinal associations between sleep duration and subsequent weight gain: a systematic review. Sleep Med Rev 2012 Jun;16(3):231-241 [FREE Full text] [doi: 10.1016/j.smrv.2011.05.005] [Medline: 21784678]
2. Maher CA, Mire E, Harrington DM, Staiano AE, Katzmarzyk PT. The independent and combined associations of physical activity and sedentary behavior with obesity in adults: NHANES 2003-06. Obesity (Silver Spring) 2013 Dec;21(12):E730-E737 [FREE Full text] [doi: 10.1002/oby.20430] [Medline: 23512825]
3. Bassuk SS, Manson JE. Epidemiological evidence for the role of physical activity in reducing risk of type 2 diabetes and cardiovascular disease. J Appl Physiol (1985) 2005 Sep;99(3):1193-1204 [FREE Full text] [doi: 10.1152/japplphysiol.00160.2005] [Medline: 16103522]
4. Biswas A, Oh PI, Faulkner GE, Bajaj RR, Silver MA, Mitchell MS, et al. Sedentary time and its association with risk for disease incidence, mortality, and hospitalization in adults: a systematic review and meta-analysis. Ann Intern Med 2015 Jan 20;162(2):123-132. [doi: 10.7326/M14-1651] [Medline: 25599350]
5. Cappuccio FP, D'Elia L, Strazzullo P, Miller MA. Quantity and quality of sleep and incidence of type 2 diabetes: a systematic review and meta-analysis. Diabetes Care 2010 Feb;33(2):414-420 [FREE Full text] [doi: 10.2337/dc09-1124] [Medline: 19910503]
6. Cappuccio F, Cooper D, D'Elia L, Strazzullo P, Miller MA. Sleep duration predicts cardiovascular outcomes: a systematic review and meta-analysis of prospective studies. Eur Heart J 2011 Jun;32(12):1484-1492. [doi: 10.1093/eurheartj/ehr007] [Medline: 21300732]
7. Guallar-Castillón P, Bayán-Bravo A, León-Muñoz LM, Balboa-Castillo T, López-García E, Gutierrez-Fisac JL, et al. The association of major patterns of physical activity, sedentary behavior and sleep with health-related quality of life: a cohort study. Prev Med 2014 Oct;67:248-254. [doi: 10.1016/j.ypmed.2014.08.015] [Medline: 25138382]
8. Marques A, Sarmento H, Martins J, Saboga Nunes L. Prevalence of physical activity in European adults - Compliance with the World Health Organization's physical activity guidelines. Prev Med 2015 Dec;81:333-338. [doi: 10.1016/j.ypmed.2015.09.018] [Medline: 26449407]
9. Cassidy S, Chau JY, Catt M, Bauman A, Trenell MI. Cross-sectional study of diet, physical activity, television viewing and sleep duration in 233,110 adults from the UK Biobank; the behavioural phenotype of cardiovascular disease and type 2 diabetes. BMJ Open 2016 Mar 15;6(3):e010038 [FREE Full text] [doi: 10.1136/bmjopen-2015-010038] [Medline: 27008686]
10. Loyen A, Verloigne M, Van Hecke L, Hendriksen I, Lakerveld J, Steene-Johannessen J, DEDIPAC consortium. Variation in population levels of sedentary time in European adults according to cross-European studies: a systematic literature review within DEDIPAC. Int J Behav Nutr Phys Act 2016 Jun 28;13:71 [FREE Full text] [doi: 10.1186/s12966-016-0397-3] [Medline: 27350251]
11. Chau JY, Grunseit AC, Chey T, Stamatakis E, Brown WJ, Matthews CE, et al. Daily sitting time and all-cause mortality: a meta-analysis. PLoS One 2013;8(11):e80000 [FREE Full text] [doi: 10.1371/journal.pone.0080000] [Medline: 24236168]
12. Pedišić Z. Measurement issues and poor adjustments for physical activity and sleep undermine sedentary behaviour research—the focus should shift to the balance between sleep, sedentary behaviour, standing and activity. Kinesiology 2014;46(1):135-146 [FREE Full text]
13. Chastin SF, Palarea-Albaladejo J, Dontje ML, Skelton DA. Combined effects of time spent in physical activity, sedentary behaviors and sleep on obesity and cardio-metabolic health markers: a novel compositional data analysis approach. PLoS One 2015;10(10):e0139984 [FREE Full text] [doi: 10.1371/journal.pone.0139984] [Medline: 26461112]
14. Chaput J, Carson V, Gray C, Tremblay M. Importance of all movement behaviors in a 24 hour period for overall health. Int J Environ Res Public Health 2014 Dec 4;11(12):12575-12581 [FREE Full text] [doi: 10.3390/ijerph111212575] [Medline: 25485978]
15. Buman MP, Epstein DR, Gutierrez M, Herb C, Hollingshead K, Huberty JL, et al. BeWell24: development and process evaluation of a smartphone 'app' to improve sleep, sedentary, and active behaviors in US Veterans with increased metabolic risk. Transl Behav Med 2016 Sep;6(3):438-448 [FREE Full text] [doi: 10.1007/s13142-015-0359-3] [Medline: 27528532]
16. Duncan MJ, Vandelanotte C, Trost SG, Rebar AL, Rogers N, Burton NW, et al. Balanced: a randomised trial examining the efficacy of two self-monitoring methods for an app-based multi-behaviour intervention to improve physical activity, sitting and sleep in adults. BMC Public Health 2016 Jul 30;16:670 [FREE Full text] [doi: 10.1186/s12889-016-3256-x] [Medline: 27473327]
17. Albarracín D, Wilson K, Chan M, Durantini M, Sanchez F. Action and inaction in multi-behaviour recommendations: a meta-analysis of lifestyle interventions. Health Psychol Rev 2018 Mar;12(1):1-24. [doi: 10.1080/17437199.2017.1369140] [Medline: 28831848]
18. Meader N, King K, Wright K, Graham HM, Petticrew M, Power C, et al. Multiple risk behavior interventions: meta-analyses of RCTs. Am J Prev Med 2017 Jul;53(1):e19-e30 [FREE Full text] [doi: 10.1016/j.amepre.2017.01.032] [Medline: 28258777]
19. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]

20. Yardley L, Spring B, Riper H, Morrison L, Crane D, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
21. Kelders S, Pots W, Oskam M, Bohlmeijer E, van Gemert-Pijnen JE. Development of a web-based intervention for the indicated prevention of depression. BMC Med Inform Decis Mak 2013 Feb 20;13:26 [FREE Full text] [doi: 10.1186/1472-6947-13-26] [Medline: 23425322]
22. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
23. Hoffman S, Tan C. Biological, psychological and social processes that explain celebrities' influence on patients' health-related behaviors. Arch Public Health 2015;73(1):3 [FREE Full text] [doi: 10.1186/2049-3258-73-3] [Medline: 25973193]
24. Chan A, Kow R, Cheng JK. Adolescents' perceptions on smartphone applications (Apps) for health management. J Mob Technol Med 2017 Aug;6(2):47-55. [doi: 10.7309/jmtm.6.2.6]
25. Perski O, Baretta D, Blandford A, West R, Michie S. Engagement features judged by excessive drinkers as most important to include in smartphone applications for alcohol reduction: A mixed-methods study. Digit Health 2018;4:2055207618785841 [FREE Full text] [doi: 10.1177/2055207618785841] [Medline: 31463077]
26. Edwards E, Lumsden J, Rivas C, Steed L, Edwards L, Thiyagarajan A, et al. Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. BMJ Open 2016 Oct 4;6(10):e012447 [FREE Full text] [doi: 10.1136/bmjopen-2016-012447] [Medline: 27707829]
27. Morrison L, Yardley L, Powell J, Michie S. What design features are used in effective e-health interventions? A review using techniques from Critical Interpretive Synthesis. Telemed J E Health 2012 Mar;18(2):137-144. [doi: 10.1089/tmj.2011.0062] [Medline: 22381060]
28. Middelweerd A, van der Laan DM, van Stralen MM, Mollee J, Stuij M, te Velde S, et al. What features do Dutch university students prefer in a smartphone application for promotion of physical activity? A qualitative approach. Int J Behav Nutr Phys Act 2015 Mar 1;12:31 [FREE Full text] [doi: 10.1186/s12966-015-0189-1] [Medline: 25889577]
29. Petty RE, Cacioppo JT. Communication and persuasion: Central and peripheral routes to attitude change. New York: Springer Science & Business Media; 2012.
30. Ahmed R, Beard F, Yoon D. Examining and extending advertising's dual mediation hypothesis to a branded mobile phone app. J Interact Advert 2016;16(2):133-144. [doi: 10.1080/15252019.2016.1237315]
31. Eldredge LK, Markham CM, Ruiter RA, Kok G, Fernandez ME, Parcel GS. Planning Health Promotion Programs: An Intervention Mapping Approach. San Francisco: John Wiley & Sons; 2016.
32. Schwarzer R, Luszczynska A. How to overcome health-compromising behaviors: The health action process approach. Eur Psychol 2008 Jan;13(2):141-151. [doi: 10.1027/1016-9040.13.2.141]
33. Hagger M, Trost N, Keech J, Chan D, Hamilton K. Predicting sugar consumption: application of an integrated dual-process, dual-phase model. Appetite 2017 Sep 1;116:147-156. [doi: 10.1016/j.appet.2017.04.032] [Medline: 28461198]
34. Eysenbach G. Improving the quality of Web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
35. Evenson K, Goto M, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12:159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
36. Rosenberger M, Buman M, Haskell W, McConnell M, Carstensen LL. Twenty-four hours of sleep, sedentary behavior, and physical activity with nine wearable devices. Med Sci Sports Exerc 2016 Mar;48(3):457-465 [FREE Full text] [doi: 10.1249/MSS.0000000000000778] [Medline: 26484953]
37. Brooke J. UNIGE - Université de Genève. SUS-A Quick and Dirty Usability Scale URL: https://cui.unige.ch/isi/icle-wiki/_media/ipm:test-suschapt.pdf [accessed 2019-11-01]
38. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
39. Direito A, Dale L, Shields E, Dobson R, Whittaker R, Maddison R. Do physical activity and dietary smartphone applications incorporate evidence-based behaviour change techniques? BMC Public Health 2014 Jun 25;14:646 [FREE Full text] [doi: 10.1186/1471-2458-14-646] [Medline: 24965805]
40. Middelweerd A, Mollee J, van der Wal CN, Brug J, Te Velde S. Apps to promote physical activity among adults: a review and content analysis. Int J Behav Nutr Phys Act 2014 Jul 25;11:97 [FREE Full text] [doi: 10.1186/s12966-014-0097-9] [Medline: 25059981]
41. van Dyck D, Cardon G, Deforche B, Giles-Corti B, Sallis J, Owen N, et al. Environmental and psychosocial correlates of accelerometer-assessed and self-reported physical activity in Belgian adults. Int J Behav Med 2011 Sep;18(3):235-245. [doi: 10.1007/s12529-010-9127-4] [Medline: 21038103]
42. de Cocker K, Duncan M, Short C, van Uffelen JG, Vandelanotte C. Understanding occupational sitting: prevalence, correlates and moderating effects in Australian employees. Prev Med 2014 Oct;67:288-294. [doi: 10.1016/j.ypmed.2014.07.031] [Medline: 25117522]
43. Vanhaelewyn B, De Marez L. imec. Digimeter 2017. Measuring digital media trends in Flanders. Leuven: imec; 2017.

44. Berger AM, Kuhn BR, Farr LA, Lynch JC, Agrawal S, Chamberlain J, et al. Behavioral therapy intervention trial to improve sleep quality and cancer-related fatigue. Psychooncology 2009 Jun;18(6):634-646. [doi: 10.1002/pon.1438] [Medline: 19090531]
45. Loft M, Cameron LD. Using mental imagery to deliver self-regulation techniques to improve sleep behaviors. Ann Behav Med 2013 Dec;46(3):260-272. [doi: 10.1007/s12160-013-9503-9] [Medline: 23640130]
46. Slater MD, Rouner D. Entertainment - education and elaboration likelihood: understanding the processing of narrative persuasion. Commun Theory 2002;12(2):173-191. [doi: 10.1111/j.1468-2885.2002.tb00265.x]
47. Lally P, Gardner B. Promoting habit formation. Health Psychol Rev 2013;7(Suppl 1):S137-S158. [doi: 10.1080/17437199.2011.603640]
48. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
49. Michie S, Ashford S, Sniehotta F, Dombrowski S, Bishop A, French DP. A refined taxonomy of behaviour change techniques to help people change their physical activity and healthy eating behaviours: the CALO-RE taxonomy. Psychol Health 2011 Nov;26(11):1479-1498. [doi: 10.1080/08870446.2010.540664] [Medline: 21678185]
50. Gardner B, Smith L, Lorencatto F, Hamer M, Biddle SJ. How to reduce sitting time? A review of behaviour change strategies used in sedentary behaviour reduction interventions among adults. Health Psychol Rev 2016;10(1):89-112 [FREE Full text] [doi: 10.1080/17437199.2015.1082146] [Medline: 26315814]
51. Irish LA, Kline CE, Gunn HE, Buysse DJ, Hall MH. The role of sleep hygiene in promoting public health: a review of empirical evidence. Sleep Med Rev 2015 Aug;22:23-36 [FREE Full text] [doi: 10.1016/j.smrv.2014.10.001] [Medline: 25454674]
52. Knowlden A, Sharma M, Bernard AL. A Theory of Planned Behavior research model for predicting the sleep intentions and behaviors of undergraduate college students. J Prim Prev 2012 Feb;33(1):19-31. [doi: 10.1007/s10935-012-0263-2] [Medline: 22293980]
53. Duncan M, Murawski B, Short C, Rebar A, Schoeppe S, Alley S, et al. Activity trackers implement different behavior change techniques for activity, sleep, and sedentary behaviors. Interact J Med Res 2017 Aug 14;6(2):e13 [FREE Full text] [doi: 10.2196/ijmr.6685] [Medline: 28807889]
54. Rothman A, Sheeran P, Wood W. Reflective and automatic processes in the initiation and maintenance of dietary change. Ann Behav Med 2009 Dec;38(Suppl 1):S4-17. [doi: 10.1007/s12160-009-9118-3] [Medline: 19787308]
55. Watson A, Bickmore T, Cange A, Kulshreshtha A, Kvedar J. An internet-based virtual coach to promote physical activity adherence in overweight adults: randomized controlled trial. J Med Internet Res 2012 Jan 26;14(1):e1 [FREE Full text] [doi: 10.2196/jmir.1629] [Medline: 22281837]
56. Eyck A, Geerlings K, Karimova D, Meerbeek B, Wang L, IJsselsteijn W, et al. Effect of a Virtual Coach on Athletes' Motivation. In: Proceedings of the International Conference on Persuasive Technology. 2006 Presented at: PERSUASIVE'06; May 18-19, 2006; Eindhoven, The Netherlands p. 158-161.
57. Ellis T, Latham N, DeAngelis T, Thomas C, Saint-Hilaire M, Bickmore TW. Feasibility of a virtual exercise coach to promote walking in community-dwelling persons with Parkinson disease. Am J Phys Med Rehabil 2013 Jun;92(6):472-81; quiz 482 [FREE Full text] [doi: 10.1097/PHM.0b013e31828cd466] [Medline: 23552335]
58. Health Interview Survey. Full reports: Health Survey Belgium 2018 URL: https://his.wiv-isp.be/nl/SitePages/Volledige_rapporten_2018_backup.aspx [accessed 2019-10-13]
59. Drieskens S, Nguyen D, Gisle L. Sciensano Rapportnummer: D/2019/14.440/54. Brussel, België; 2019. Gezondheidsenquête 2018: Lichaamsbeweging. (National Health Survey 2018: Physical Activity) URL: https://his.wiv-isp.be/nl/SitePages/Volledige_rapporten_2018_backup.aspx [accessed 2019-10-13]
60. Régnier F, Chauvel L. Digital inequalities in the use of self-tracking diet and fitness apps: interview study on the influence of social, economic, and cultural factors. JMIR Mhealth Uhealth 2018 Apr 20;6(4):e101 [FREE Full text] [doi: 10.2196/mhealth.9189] [Medline: 29678807]
61. Morrison L, Muller I, Yardley L, Bradbury K. The person-based approach to planning, optimising, evaluating and implementing behavioural health interventions. Eur Health Psychol 2018;20(3):464-469 [FREE Full text]

Abbreviations

BCT: behavior change technique
LPA: light physical activity
MPA: moderate physical activity
MVPA: moderate-to-vigorous physical activity
PA: physical activity
SB: sedentary behavior

Edited by G Eysenbach; submitted 31.07.19; peer-reviewed by E Ortega, D Wolff-Hughes; comments to author 25.09.19; revised version received 13.10.19; accepted 20.10.19; published 20.12.19.

Please cite as:
DeSmet A, De Bourdeaudhuij I, Chastin S, Crombez G, Maddison R, Cardon G
Adults' Preferences for Behavior Change Techniques and Engagement Features in a Mobile App to Promote 24-Hour Movement Behaviors: Cross-Sectional Survey Study
JMIR Mhealth Uhealth 2019;7(12):e15707
URL: http://mhealth.jmir.org/2019/12/e15707/
doi:10.2196/15707
PMID:31859680

Original Paper

Medical Videography Using a Mobile App: Retrospective Analysis

Julia C Cambron[1], BS; Kirk D Wyatt[2], MD; Christine M Lohse[3], MS; Page Y Underwood[4], JD; Thomas R Hellmich[5], MD

[1]Mayo Clinic Alix School of Medicine, Mayo Clinic, Rochester, MN, United States

[2]Division of Pediatric Hematology/Oncology, Department of Pediatric and Adolescent Medicine, Mayo Clinic, Rochester, MN, United States

[3]Division of Biomedical Statistics and Informatics, Mayo Clinic, Rochester, MN, United States

[4]Legal Department, Mayo Clinic, Scottsdale, AZ, United States

[5]Department of Emergency Medicine, Mayo Clinic, Rochester, MN, United States

Corresponding Author:
Thomas R Hellmich, MD
Department of Emergency Medicine
Mayo Clinic
200 First Street Southwest
Rochester, MN, 55905
United States
Phone: 1 507 255 2216
Email: Hellmich.Thomas@mayo.edu

Abstract

Background: As mobile devices and apps grow in popularity, they are increasingly being used by health care providers to aid clinical care. At our institution, we developed and implemented a point-of-care clinical photography app that also permitted the capture of video recordings; however, the clinical findings it was used to capture and the outcomes that resulted following video recording were unclear.

Objective: The study aimed to assess the use of a mobile clinical video recording app at our institution and its impact on clinical care.

Methods: A single reviewer retrospectively reviewed video recordings captured between April 2016 and July 2017, associated metadata, and patient records.

Results: We identified 362 video recordings that were eligible for inclusion. Most video recordings (54.1%; 190/351) were captured by attending physicians. Specialties recording a high number of video recordings included orthopedic surgery (33.7%; 122/362), neurology (21.3%; 77/362), and ophthalmology (15.2%; 55/362). Consent was clearly documented in the medical record in less than one-third (31.8%; 115/362) of the records. People other than the patient were incidentally captured in 29.6% (107/362) of video recordings. Although video recordings were infrequently referenced in notes corresponding to the clinical encounter (12.2%; 44/362), 7.7% (22/286) of patients were video recorded in subsequent clinical encounters, with 82% (18/22) of these corresponding to the same finding seen in the index video. Store-and-forward telemedicine was documented in clinical notes in only 2 cases (0.5%; 2/362). Videos appeared to be of acceptable quality for clinical purposes.

Conclusions: Video recordings were captured in a variety of clinical settings. Documentation of consent was inconsistent, and other individuals were incidentally included in videos. Although clinical impact was not always clearly evident through retrospective review because of limited documentation, potential uses include documentation for future reference and store-and-forward telemedicine. Repeat video recordings of the same finding provide evidence of use to track the findings over time. Clinical video recordings have the potential to support clinical care; however, documentation of consent requires standardization.

(JMIR Mhealth Uhealth 2019;7(12):e14919) doi:10.2196/14919

KEYWORDS

photography; video recording; telemedicine; medical informatics applications

Introduction

Background

Mayo Clinic's institutional archives reference the use of video cameras to record surgeries as early as the 1930s using equipment that would be considered bulky by today's standards. Historical photographs show videographers perched from balconies above operating theaters to record surgeries (Figure 1). As the process of video recording patients has become easier, the use of rich media for clinical documentation and diagnostic purposes has evolved.

At present, video cameras remain in use to capture medical procedures for the purposes of quality improvement, training, and research [1,2]. A recent systematic review on video recording open surgeries included 110 articles that discussed camera use to capture open surgery with miniature cameras, such as GoPro (GoPro, Inc.) and GoogleGlass (Alphabet, Inc) [3]. Outside of surgical specialties, video recordings are widely used to correlate clinical findings with electroencephalogram findings in patients with suspected seizures [4-6] and may also be used to study team dynamics and technical skills in emergency cardiopulmonary resuscitation [7]. In addition to asynchronous review of video recordings after they have been captured, live video can be used to conduct synchronous telemedicine encounters. For example, provider-to-provider telemedicine is used to guide neonatal resuscitation [8] and provider-to-consumer telemedicine is used for urgent care visits for minor illnesses [9].

At present, cameras available in smartphones arguably capture photographs and video recordings with better quality than most consumer-grade digital cameras that were available less than a decade ago. In a time when at least 77% of Americans own smartphones and more than 86% of physicians report using an electronic health record (EHR), merging both technologies presents an attractive and convenient opportunity to enhance point-of-care documentation [10,11]. Indeed, leading EHR vendors have seized this opportunity to integrate point-of-care image capture into their mobile apps, and institutions have developed home-grown apps for this purpose [12-14].

Figure 1. Videographer recording a surgical procedure at Mayo Clinic in Rochester, Minnesota, in March 1937. Copyright 1937 Mayo Foundation for Medical Education and Research. Used with the permission of Mayo Foundation for Medical Education and Research, all rights reserved.

Prior Work

We previously described the implementation of a mobile app (PhotoExam) for point-of-care clinical photography at Mayo Clinic [12], including its integration in the primary care setting [15]. Later evaluations revealed that the app's primary use was not for teleconsultation but rather for improved documentation of physical examination findings within the EHR. Our initial assessment observed that residents and fellows represented the primary user base, with surgical specialties making up the largest number of users and dermatology accounting for most of the photographs [12]. Photographs were observed to be of acceptable quality, and the app's release did not appear to be associated with a decrease in the use of traditional medical photography services. Our ongoing efforts are focused on assessing how point-of-care medical photography affects patient care–related outcomes.

The PhotoExam app has predominantly been used for its photography function, which allows authorized users to securely photograph a clinical finding and tag it with metadata (eg, anatomical site captured and description of finding). However, more recent versions of the PhotoExam app include capability to capture and upload video recordings.

Goals

As it was unclear how the video recording function was being used in clinical practice and we were unable to identify literature about the use of point-of-care mobile clinical video recording

technology elsewhere, we aimed to assess the following attributes of video recordings captured using the PhotoExam app: (1) who was capturing video recordings (in terms of specialty and work role), (2) what clinical findings were being captured on video recordings, and (3) was there a measurable impact of video recordings on patient care. Therefore, we retrospectively reviewed video recordings, their associated metadata, and patient records to assess the use and impact of the PhotoExam video recording feature.

Methods

The PhotoExam App

The PhotoExam app is available to health care providers at all Mayo Clinic sites via an internal *App Store* and is compatible with recent versions of iOS. Its release and updates were announced in internal communications (ie, staff newsletter). Any health care staff with access to the EHR can use the app while their device is securely connected behind the institutional firewall. Use of the app was not mandated by the institution. Although we are not aware of specific, formal departmental initiatives promoting use, we cannot rule out the possibility that departmental quality improvement initiatives may have encouraged use within specific specialties. Our anecdotal experience has been that clinical *champions* within a department often encouraged colleagues to use the app. The features and functionality of the video recording function are identical to those previously described for the photography function [12] except that only 1 video recording can be uploaded for each anatomic site (compared with multiple photographs per anatomic site), and the newest version automatically launches from within our EHR vendor's proprietary mobile EHR app in a manner that hands off patient context, thereby obviating the need to manually search for the patient record within the PhotoExam app. Once the patient record is opened within the PhotoExam app, a hard stop verifies that the appropriate consent for photography or videography has been obtained according to departmental policies. The user is then permitted to capture a video recording using the device's camera. The video recording is securely uploaded to a Digital Imaging and Communications in Medicine standard-compliant Digital Clinical Asset Management System. The captured video recordings are stored on the local device only temporarily until they are successfully uploaded or the user closes the app—whichever comes first—after which, they are permanently deleted from the user's device in a manner that is compliant with the Health Insurance Portability and Accountability Act (HIPAA). Video recordings are not accessible by other apps on the mobile device. The app was not designed for patient use, and patients were not able to utilize the app to upload self-shot video recordings.

The users did not receive any formal training on practical aspects such as use case scenarios or how to capture clinically relevant video recordings with high fidelity. However, an internally accessible website provided technical support and instructions on the app's use. We are unaware of formal integration of the app into medical training programs; however, users included residents and fellows in training.

Human Subjects Protection

The study procedures were reviewed and approved by the Mayo Clinic's institutional review board.

Patient Selection and Data Collection

We queried the clinical asset management database to identify all video recordings taken using the PhotoExam app between April 15, 2016, (when the video recording feature first became available) and July 17, 2017. We excluded records corresponding to patients who refused the use of their medical records for research purposes, known test patients (ie, fictitious medical records within the production environment used for training and testing) and incidental video recordings (ie, user had not intended to record a video).

Employee work role and department were identified by cross-referencing a human resources database. These data were missing in the data source for a small number of users who were no longer employed by Mayo Clinic at the time of the query. Data available by review of the video recording and the patient medical records were extracted by 1 reviewer (JCC). Data were extracted into a data extraction form in REDCap [16] and exported for analysis.

Quality Assessment

Lacking a standardized tool to assess the quality of provider-captured clinical video recordings, we adapted the quality assessment rubric used previously for assessing photographs taken using PhotoExam [12]. By consensus, we arrived at the following quality assessment items that were included in our rubric:

1. Does image quality or blurriness limit ability to see the area of focus?
2. Does the video recording objectively portray size using a ruler?
3. Does the video recording zoom in or out or move around to optimally characterize the finding as needed?
4. Is there sufficient lighting and color differentiation to see the area of focus?
5. Is the audio clear and audible (if present)?
6. Is the video recording image stable?
7. Is the video recording right side up?

A favorable score was awarded for an answer of *yes* on questions 2, 3, 4, 5, 6, and 7, and a favorable score was awarded for an answer of *no* on question 1. A quality score was calculated as a percentage of applicable favorable scores awarded on rubric items. For video recordings that included sound, the total score was calculated as a percentage of items out of 7. For video recordings that did not include sound, the total score was calculated as a percentage of items out of 6 because item 5 did not apply.

Data Analysis

Continuous features were summarized with means and standard deviations when approximately normally distributed and with medians and interquartile ranges otherwise. Categorical features were summarized with frequency counts and percentages. For differences in quality score, referral generation, and consent documentation between specialties, analysis of variance and

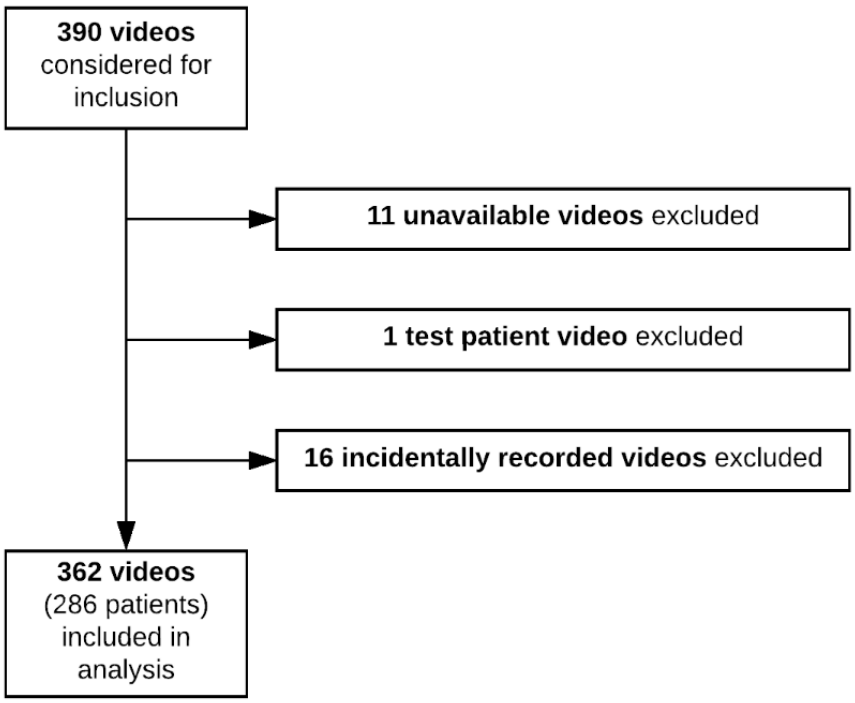
chi-square tests were performed across the most common user specialties (ie, orthopedic surgery, neurology, ophthalmology, and emergency medicine) with all other specialties grouped together as *other*. Throughout the paper, sample sizes for features with missing data or for subsets of interest are indicated in italics in parentheses. Results are reported at the video recording level—rather than patient level—unless otherwise specified. Statistical analyses were performed using SAS version 9.4 (SAS Institute). All tests were 2 sided, and *P* values less than .05 were considered statistically significant.

Results

Included Records

We identified 390 video recordings that were potentially eligible for inclusion. We observed that 11 video recordings were unavailable for review, 1 video recording corresponded to a *test patient* record that had not been identified in the initial screening process, and 16 video recordings appeared to be incidentally recorded. These video recordings were all excluded, yielding 362 video recordings of 286 patients that were included in the study (Figure 2).

Figure 2. Patient record selection.



Demographics

App use in terms of the number of clinical video recordings captured using the PhotoExam app over time is shown in Figure 3.

In terms of patient demographics, there was a slight predominance of males (58.0%, 210/362), and patients were largely white (88.6% [321/362]; Table 1). The most common site where video recordings took place was the Mayo Clinic Rochester campus, corresponding to nearly 4 out of 5 video recordings (79.0% [286/362]; Table 1). Most (70.7%, 256/362) video recordings took place in the outpatient clinic setting, followed by the inpatient hospital setting (20.4%, 74/362), emergency department (8.3%, 30/362), and, unexpectedly, patients' homes (1.1% [4/362]; Table 1). The mean video length was 21 seconds (SD 12 seconds; Figure 4).

Figure 3. Videorecordings over time.

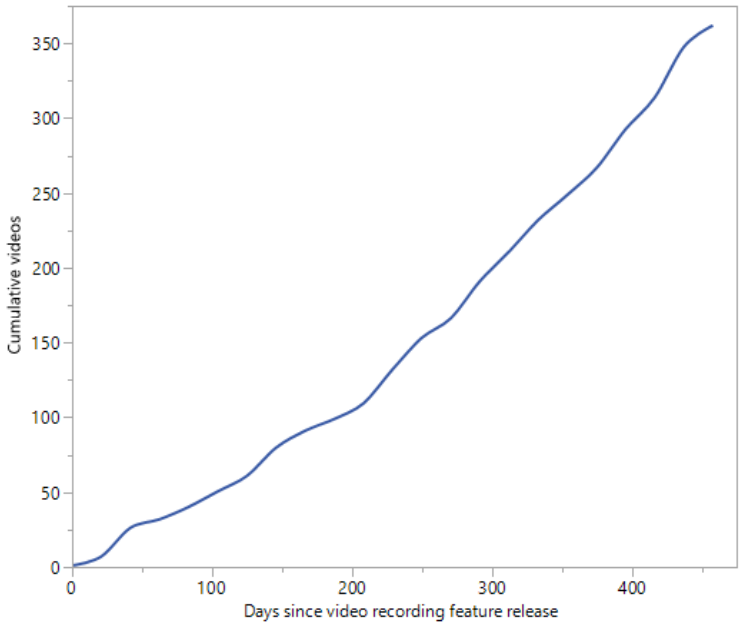
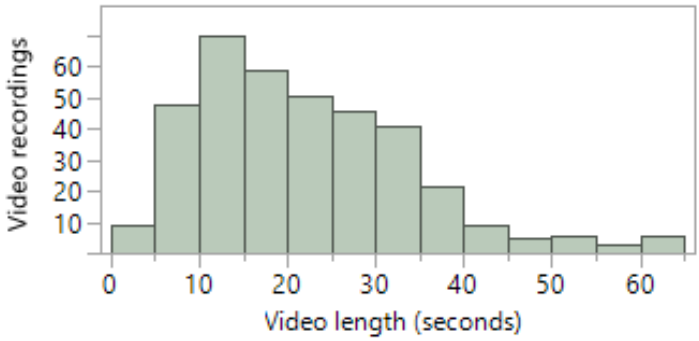


Table 1. Patient demographics and location of video recording, including work site and clinical setting.

Patient demographics (N=362)	Value
Age (years), mean (SD)	45.0 (24.6)
Sex, n (%)	
Male	210 (58.0)
Female	152 (41.9)
Race, n (%)	
White	321 (88.6)
Black/African American	10 (2.7)
Asian	4 (0.2)
Other	18 (5.0)
Unknown	9 (2.5)
Location of video recording	
Work site, n (%)	
Mayo Clinic Rochester	286 (79.0)
Mayo Clinic Arizona	43 (12.0)
Mayo Clinic Florida	10 (2.8)
Mayo Clinic Health System	23 (6.4)
Clinical setting, n (%)	
Hospital	74 (20.4)
Clinic	256 (70.7)
Emergency department	30 (8.3)
Patient's home	2 (0.5)

Figure 4. Distribution of videorecording lengths.



User Demographics

The work role and primary clinical department of the provider who captured each video recording were assessed. The majority of video recordings were captured by attending physicians (54.1%; 190/351), followed by nurses (19.7%; 69/351) and residents/fellows (18.2%; 64/351). When assessed according to clinical department, approximately one-third (34.8%; 122/351) of video recordings were taken within orthopedic surgery, 21.9% (77/351) were taken within neurology, and 15.7% (55/351) of video recordings were taken within ophthalmology.

Video Recordings and Photographs Within the Medical Record

To assess the extent of utilization of the app, we assessed the number of video recordings or photographs taken of the patient during each clinical encounter. To assess the provider's experience level, we measured the cumulative number of video recordings and photographs the recording provider had previously captured using the app as of the time of each video recording. In most cases, only 1 video recording was recorded during the encounter and no photographs were captured (Table 2). Furthermore, providers had previously captured a median of 46.5 photographs and 3.5 video recordings using the PhotoExam app as of the time the video recording was recorded (Table 2). The distribution of provider experience level with videography and photography is graphically demonstrated in a histogram form in Figures 5 and 6, respectively.

Table 2. Photographs and video recordings associated with each clinical encounter and provider experience level at the time of use.

Associated photographs and video recordings	Value, median (IQR)
Extent of app use during encounter	
Number of video recordings captured at visit	1 (1-2)
Number of photographs captured at visit	0 (0-2)
Provider experience level	
Cumulative number of video recordings captured by provider	3.5 (2-8)
Cumulative number of photographs captured by provider	46.5 (14-111)

Figure 5. Distribution of provider experience level with videography. A single outlier who captured nearly 50 videorecordings is not included on the histogram.

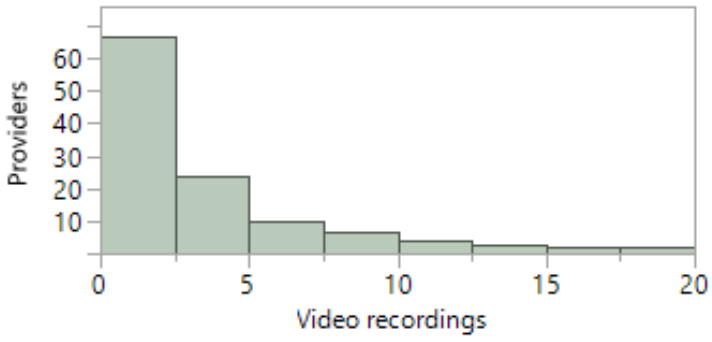
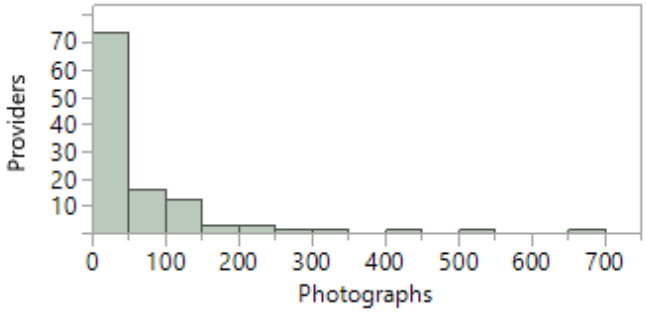


Figure 6. Distribution of provider experience with photography. A single outlier who captured over 1000 photographs is not included on the histogram.



Patients included in 22 video recordings (7.7% of patients; N=286) were video recorded during subsequent clinical encounters, with 82% (18/22) of video recordings of these patients demonstrating the same finding seen in the index video.

Similarly, 19.2% (55/286) of patients' video recordings were associated with additional photographs taken using the PhotoExam app. In 71% (39/55) of these cases, the photographed finding was the same as the video-recorded finding.

Clinical Findings Recorded

Independent of the recording provider's identified specialty, clinical findings were classified according to finding specialty categories. Table 3 summarizes the types of specific findings captured, stratified by finding category for the 4 most common specialty uses of orthopedic/musculoskeletal/physiatrics, neurologic, ophthalmologic, and dental, which accounted for more than three-quarters of all video recordings.

Table 4 includes additional details about the findings documented in the reviewed video recordings. Most of the video recordings (68.8%; 249/362) captured volitional movements by patients, and 29.6% (107/362) of them captured either passive movements initiated by the examiner or static findings. The vast majority (92.3%; 334/362) of video recordings captured physical examination findings; however, less common uses included demonstration of cares (eg, wound care), capture of externally captured media (eg, video recording originally captured on patient's cellphone or portable ultrasound), and objects external to the patient (eg, foreign body). A variety of anatomic sites were captured, with the most common sites being the upper extremities (31.5%; 114/362) and the eyes (18.0%; 65/362).

Table 3. Specialty-specific findings included in the video recordings.

Specialty findings (N=362)	Value, n (%)
Orthopedic/musculoskeletal/physiatrics	130 (35.9)
Range of motion	102 (28.2)
Muscle strength test	4 (1.1)
Other	38 (10.5)
Neurologic	80 (22.1)
Motor function and balance	19 (5.2)
Coordination test	17 (4.7)
Cranial nerve test	7 (1.9)
Mental status	5 (1.4)
Reflexes	4 (1.1)
Other	40 (11.0)
Ophthalmologic	58 (16.0)
Extraocular movement	26 (7.2)
Slit lamp examination	5 (1.4)
Pupillary response	2 (0.6)
Other	31 (8.6)
Dental	13 (3.6)
Surgical prosthesis or device	11 (3.0)
Other	2 (0.6)
Other findings	83 (22.9)

Table 4. General characteristics of the findings captured in video.

General finding characteristics (N=362)	Value, n (%)
Volition	
Nonvolitional	107 (29.6)
Volitional	249 (68.8)
Both	6 (1.7)
Finding type	
Object outside patient	3 (0.8)
Capture of video recording on another device	5 (1.4)
Capture of imaging findings	11 (3.0)
Physical exam	334 (92.3)
Demonstration of care	9 (2.5)
Anatomic site captured	
Hand	75 (20.7)
Eye	65 (18.0)
Arm	39 (10.8)
Entire body/unspecified	44 (12.2)
Leg	29 (8.0)
Face	23 (6.4)
Mouth	15 (4.1)
Chest	14 (3.9)
Head	12 (3.3)
Other	46 (12.7)

In 107 (29.6%; 107/362) video recordings, we noted that people other than the patient were incidentally captured. In 80 (74.8%; 80/107) of these, medical personnel were included, and in 31 (29.0%; 31/107) video recordings, other people who had accompanied the patient to the visit (ie, family or friends) were included. As both medical personnel and other people who had accompanied a patient to a visit could have been captured in the same video recording, these percentages add up to greater than 100%.

Captured Audio

As the PhotoExam app allows the user to include or omit the recording of sound, we were interested in the inclusion and content of the recorded audio. Audio was only recorded in 118 (32.6%; 118/362) video recordings. In 36.4% (43/118) of cases, only background noise was captured, suggesting that audio did not need to be recorded in those cases. In 55.1% (65/118) of cases, the audio included provider instructions (such as commands to the patient to complete physical examination maneuvers), and in 11.0% (13/118) of cases, providers recorded commentary (eg, description) about the observed findings. In 28.0% (33/118) of cases, sounds made by the patient, which may be relevant to an observer, were captured.

Upon manual review of these 118 video recordings that included audio, the audio recorded findings were deemed to be the primary finding in 2.5% (3/118) of the video recordings, audio recorded and video recorded findings were considered equally important in 27.1% (32/118) of cases, and, in the remaining 70.3% (83/118), the video recorded finding was deemed the primary finding.

Video Recording Quality

Our quality assessment revealed that most video recordings were acceptably in focus with adequate lighting, stability, and sound. However, video recordings frequently did not adequately demonstrate scale (ie, by using a ruler) or perspective (ie, by rotating the camera around clinical findings). The mean quality score across video recordings was 67.8 (SD 7.7)%. There was no significant difference in quality score when compared across specialties. Average quality scores ranged between 65.8% in emergency medicine and 68.5% in *other* specialties, with quality scores of ophthalmology, neurology, and orthopedic surgery falling within this range (P=.59).

Consent

In 115 (31.8%; 115/362) video recordings, patient consent was clearly documented in the medical record: for 113 of these video recordings (98.2%), a signed media consent form was scanned into the medical record, and for 2 video recordings (1.7%), the provider documented in the clinical note that verbal consent was obtained. There was a significant difference between specialties in terms of the percentage of video recordings that were associated with explicitly documented consent within clinical notes or a signed media consent form (P<.001). Exploratory comparisons between individual specialties and

the others revealed driving factors to be high rates in orthopedic surgery (53.3%, 65/122; *P*<.001 vs all others) compared with low rates in emergency medicine (13.3%, 4/30; *P*=.02 vs all others) and *other* (ie, not orthopedic surgery, neurology, ophthalmology, or emergency medicine) specialties (11.5%, 9/78; *P*<.001 vs all others).

Clinical Impact

We first evaluated clinical notes to identify evidence of the clinical impact of video recordings. Disappointingly, only 44 (12.2%; 44/362) video recordings were referenced in the clinical notes corresponding to the visit, and in only 2 cases (0.6%; 2/362) the use of telemedicine with remote, asynchronous review of the video recording by another provider was explicitly documented in the clinical notes.

In 30 cases (8.3%; 30/362), an outcome of the visit was generation of a referral to a specialist. In the majority of these cases (20/30; 67%), the referral occurred on the same day. Of the 20 same-day referrals, 14 (70%) were specialist referrals made for patients seen in the emergency department. Overall, 16 total referrals were made in the emergency department, meaning that only 2 patients for whom a referral was generated in the emergency department were not seen by the specialist on the same day the referral was made. Neurology and plastic surgery accounted for the majority of the specialists consulted, contributing 10 and 6 referrals, respectively. There was a significant difference between departments in the proportion of video recordings associated with a referral (*P*<.001), with the main driver being a high rate of referrals made from the emergency department, where half of the video recordings were associated with a referral compared with 7% or fewer encounters for all other specialties.

Qualitative Aspects and Real-World Use Cases

In this study, we categorized and summarized the myriad ways that health care providers have used video recordings to document clinical findings. One unfortunate consequence of aggregating data is that the qualitative richness of the video content becomes lost. To capture these uses, we therefore include a few qualitative observations to highlight innovative or interesting use cases. One coauthor (TRH) utilized the app to capture the work of the breathing of an infant with bronchiolitis who required hospital admission. Another coauthor (KDW) who was covering the pediatric ward was able to view the video from elsewhere in the hospital, and this was noted to facilitate the appropriate disposition of the patient to the intensive care unit.

Overall, 19 video recordings demonstrated *motor function and balance* in a variety of ways. Common use cases included capturing of abnormal gait (eg, ataxia), tremors, and dystonia. Face, letter, and number recognition exercises were recorded and used to document mental status. Ophthalmologic video recordings captured slit lamp examinations, which were generally conducted without the use of a magnifying objective.

Innovative uses included capturing radiographic videos (eg, fluoroscopy, echocardiography and other ultrasonography captured on portable machines and at outreach locations that utilize a different EHR), dressing of a wound and application of vacuum-assisted closure device (ie, intended to be instructive to other providers), recording brief procedures (eg, plantar corn removal) and capturing patient-provided video recordings (eg, capturing a video recording of a video playing on the patient's phone in order to integrate it into the EHR).

Discussion

To our knowledge, this is the first report in the literature to summarize the use of a point-of-care mobile app to capture and upload clinical video recordings to the EHR in a manner that is secure and HIPAA compliant.

Principal Findings and Comparison With Prior Work

In our initial report on the use of the photography feature of the PhotoExam app, attending physicians captured more photographs than other user groups (29.5%; 6725/22,784), with nurses taking a similar number of photographs (28.3%; 6446/22,784). In this study, attending physicians and nurses were still the leading users, though there was a greater discrepancy between the percentage of video recordings captured by the top user groups when compared with photographs, with attending physicians and nurses capturing 54.1% (190/351) and 19.7% (69/351) of video recordings, respectively. With respect to specialty, our initial report on PhotoExam revealed that 54.1% (12,315/22,784) of photographs were captured within dermatology and 25.6% (5825/22,784) were captured within surgery. In contrast, leading user departments in this study were orthopedic surgery (33.7%; 122/362), neurology (21.3%; 77/362), and ophthalmology (15.2%; 55/362). This is not surprising as dermatology and surgery may be able to rely on static clinical findings, whereas orthopedic surgery, neurology, and ophthalmology commonly refer to active physical examination findings.

Analysis of usage patterns indicated steady growth in use over time, though the number of photographs taken using the app dwarf the number of video recordings. For example, in the first 8 months after the launch of the PhotoExam app, 22,784 photographs were captured compared with 148 video recordings in the 8 months following the addition of video recording functionality to the app [12]. There are several possible explanations for this discrepancy. One possibility is that users are not as aware of or familiar with the video recording feature. Another possibility is that many users may find the photography feature to be sufficient to capture clinical findings, thereby obviating the need to capture a video recording.

Most video recordings were recorded at the Mayo Clinic Rochester campus. This likely—at least partly—reflects differences in patient visit volumes between sites. The majority of video recordings were taken by attending physicians. However, nurses and residents/fellows also captured a significant number of video recordings. In comparison, in our previous study on the use of the app to capture photographs, attending physicians captured more photographs than other user groups (29.5%; 6725/22,784) but nurses took a similar number of photographs (28.3%; 6446/22,784) [12]. On the basis of review of the video recordings and medical records, as well as practical experience with the app's use, nurses and

residents/fellows may take video recordings on their own initiative or may be delegated by an attending physician to capture a video recording.

In addition to permitting moving images to be captured, video recordings also permit sound to be recorded. Common inclusions on audio tracks included provider commands (which allow the viewer to follow what is being asked of a patient) and a provider's verbal description (to indicate what is being captured). Furthermore, sounds made by patients, which could be relevant to the clinical finding, were also captured.

In general, video recordings were of high quality, though many did not extensively demonstrate perspective by rotating about the clinical finding, and many did not explicitly demonstrate size/scale using an objective measuring device (ie, ruler). In retrospect, perspective may be less important for some findings (ie, seizure) than for others (ie, raised subcutaneous abscess) and therefore may be less relevant to consider when assessing these video recordings for quality. In addition, our rubric may have been excessively stringent by requiring the use of a ruler to be deemed as adequately demonstrating size/scale. Our previous use of this rubric for assessment of photography quality [12] did not strictly require size to be portrayed using a ruler—photographs could be considered to demonstrate size if other anatomical landmarks that infer size were included. In that study, 12.0% (12/100) of photographs did not adequately demonstrate size [12] compared with 99.4% (360/362) of video recordings in this study. In retrospect, many video recordings that we reviewed for this study were able to reasonably demonstrate size/scale by including an object of reference within the frame or by starting the video recording showing the patient's entire body or recognizable landmarks and then slowly panning and zooming to the area of interest to capture size. Therefore, the use of overly stringent criteria to judge demonstration of size represents a limitation of the quality adjudication criteria used in this study. Furthermore, assessment of quality by only 1 reviewer was another limitation. In future studies, we will adjust the quality rating scale based on our observations, and we will consider reviewing at least a subset of video recordings in duplicate to ensure consistency in the quality assessment process.

Patient consent was not always clearly documented in the EHR. However, we are confident that patient consent was obtained for all video recordings because the app includes a hard stop (toggle switch and popup dialog box) that requires the user to attest that consent was obtained from the patient before being permitted to proceed with photography. One reason verbal consent may not have been clearly documented in clinical notes could be that providers may have thought that attesting that consent was obtained within the app was sufficient for documentation purposes. One possible improvement to the app would be to have the user select the method of consent (ie, verbal or written) and indicate the person providing consent (ie, patient, parent, guardian, and legal representative), which would then be documented with the video recordings. During the early stages of the app's development, the idea of having patients sign the screen of the device to provide written consent was considered. However, this option was not pursued because providers' personal mobile devices were often used and would ideally need be sanitized before having patients handle them. Furthermore, logistical issues relating to internal form processing and practical considerations of handing a patient an unlocked mobile device that contains protected health information and other private, personal information limited our ability to pursue this mechanism for consent documentation.

In many cases, people other than the patient were incidentally captured in video recordings. Most of the individuals incidentally captured on the video recordings appeared to be staff members, but others appeared to be family or friends who accompanied the patient. It is unclear whether these individuals also separately provided consent to be included in the video recordings or whether their consent was implied by facilitating video recording (ie, parent holding a child while the provider captures the video recording). From a legal standpoint, patients have a fundamental right to privacy that includes the right to consent to be photographed or video recorded. Staff, family, and friends do not have the same privacy rights. Courts have held that when individuals are in public areas, such as common areas within a hospital, they do not have reasonable expectations of privacy and can be photographed and, in most states, video recorded without the need for their prior consent. In a minority of states, a recording of a conversation requires the consent of every person in the conversation. If the video recording occurs in one of these *2-party* consent states, consent is required of each staff member, family member, or friend who is a party to the recorded conversation. However, consent can be expressed or implied based on the circumstances. As in the case of the parent holding a child, if the individual is aware of the video recording and continues to participate, consent may be implied.

We observed evidence that video recordings favorably affected clinical care. For example, we observed that video recorded findings were either video recorded or photographed at subsequent visits, suggesting that clinical findings were being tracked over time using the app and presumably referenced at those visits for comparison. In 8% of cases, a referral to a specialist was generated, and in two-thirds of these cases, referrals occurred the same day. Furthermore, video recordings captured in the emergency department were more likely to be associated with a referral than those captured within other specialties. Although it is tempting to speculate that the existence of a video recording within the EHR facilitated or expedited referrals, we do not have additional evidence to support this hypothesis. The authors' anecdotal experience has been that consulting specialists find video recordings helpful because they allow specialists to gather important information and provide tentative management advice before seeing the patient in-person. Review of video recordings may also help specialists prioritize consultations or expedite care when appropriate (eg, notify the operating room staff of an anticipated procedure).

We were disappointed that providers only infrequently mentioned video recordings within clinical notes. In fact, when video recordings were mentioned in clinical notes, they were only mentioned in notes corresponding to the clinical encounter where the video recording was recorded and were not mentioned in subsequent follow-up visits. We had anticipated that other providers who see the patient in follow-up might reference

previously taken video recordings. As noted above, we speculate that providers who captured repeat video recordings or photographs at follow-up visits viewed those video recordings, but we do not have evidence of this. It is unclear whether the absence of further references in follow-up and referred consultation notes reflects an incomplete search of the patients' medical records, a lack of viewership by subsequent physicians, a simple omission on the part of the provider in documentation, or a lack of impact of the video recording on future medical care. Video recording access log audits would help demonstrate how often other providers view the video recordings.

Strengths and Limitations

Strengths include that we captured data from a variety of sources, including photograph metadata, human resources records, manual review of video recordings, and review of clinical documentation. We also reviewed longitudinal data to identify when additional photographs or video recordings of the captured findings were obtained.

The major limitation was the use of a single reviewer to abstract data and review video recordings. Although we included patients seen at multiple sites within various settings, another limitation that limits the generalizability of our findings is that our patient population is not representative. Furthermore, we were unable to clearly associate video recordings with objective clinical outcomes. We systematically assessed the technical quality of video recordings; however, in this descriptive study, we did not include an assessment of suitability for clinical use. Conducting such an assessment in the course of our review presents several challenges. For one, the reviewer may not be familiar with the clinical requirements of a given subspecialty and may therefore not be appropriate to assess the suitability for clinical purposes. Second, the clinical utility may not be apparent or evident to those who are not directly involved in the patient's care during a retrospective review (ie, third-party reviewer). In part, to address this aspect of the app's use and assess the clinical impact of the app, we separately conducted Web-based user surveys (unpublished) that included questions about use cases and outcomes of photography. To keep the survey brief and maximize the response rate, the questions asked about the app in general and did not ask the same questions separately about the photography and videography features. Most users utilized the app exclusively to capture photographs—rather than video recordings—and, therefore, responses in general were in reference to photography. Therefore, the clinical utility of the video recording feature remains an area needing study.

Areas of Future Research

Through the authors' use of the app in clinical practice, a number of legal questions have arisen. Although it is out of the scope of this study to review them all, several of these questions are worth raising. For example, in video recordings where people other than the patient are included, have consent laws been violated? In some cases, consent for video recording may be implied—for example, if a parent holds a child in their lap and poses them to the video camera for photography, they are providing implied consent. However, if another patient is captured on a video recording while walking by, they have not provided implied consent. Even in cases where others may be incidentally included in video recordings, the HIPAA privacy rule permits certain incidental disclosures as long as reasonable safeguards are put in place to limit these disclosures. For example, policies, procedures, and training may be reasonable safeguards to limit incidental disclosures. In cases where others are incidentally included in recordings, other privacy rights beyond HIPAA may limit reuse of the video recording without the express consent of all parties included in the video recording.

If a recording is to be used in court proceedings, state laws may dictate consent requirements. For example, some states have *two-party consent* laws that require consent of parties involved to be used as evidence in court. Another legal consideration if the video recording is to be used for legal proceedings is maintenance of the chain of custody. The use of electronic log files that document the creation and any modifications to the video recording can help maintain the chain of custody.

Conclusions

In conclusion, the video recording feature of the PhotoExam app has proven to be a versatile tool that has uses within many different specialties at Mayo Clinic. Clinical video recordings offer the potential to augment clinical documentation, support assessments at follow-up visits, and facilitate telemedicine. In particular, the ability to closely track patient examination findings over time offers the potential to more accurately document chronic and progressive conditions. This technology may be used to capture objective assessments of the efficacy of treatments and medical interventions over time, and it can facilitate collaboration among multiple members of a multidisciplinary patient care team.

Acknowledgments

REDCap database resources were funded by grant UL1TR002377. Funds to publish in an open-access journal are provided by the Mayo Clinic Department of Emergency Medicine.

Conflicts of Interest

All coauthors are students (JCC and KDW) and/or employees (KDW, TRH, and CML) at Mayo Clinic. Mayo Clinic developed and owns intellectual property rights for the PhotoExam app; however, the app is currently only being used internally and is not being licensed or sold.

References

1. Thia BC, Wong NJ, Sheth SJ. Video recording in ophthalmic surgery. Surv Ophthalmol 2019;64(4):570-578. [doi: 10.1016/j.survophthal.2019.01.005] [Medline: 30703406]
2. Kimmich N, Zimmermann R, Kreft M. Video analysis for the evaluation of vaginal births: a prospective observational study. Swiss Med Wkly 2018;148:w14634 [FREE Full text] [doi: 10.4414/smw.2018.14634] [Medline: 30044472]
3. Saun TJ, Zuo KJ, Grantcharov TP. Video technologies for recording open surgery: a systematic review. Surg Innov 2019 Oct;26(5):599-612. [doi: 10.1177/1553350619853099] [Medline: 31165687]
4. Wyatt KD, Mandrekar J, Wong-Kisiel L, Nickels K, Wirrell E. Predictors of recording an event during prolonged inpatient video electroencephalogram monitoring in children. Pediatr Neurol 2014 May;50(5):458-463. [doi: 10.1016/j.pediatrneurol.2014.01.038] [Medline: 24731843]
5. Jain SV, Dye T, Kedia P. Value of combined video EEG and polysomnography in clinical management of children with epilepsy and daytime or nocturnal spells. Seizure 2019 Feb;65:1-5. [doi: 10.1016/j.seizure.2018.12.009] [Medline: 30590283]
6. Cho YW, Motamedi GK, Kim KT. The clinical utility of non-invasive video-electroencephalographic monitoring has been diversifying. Neurol Sci 2019 Aug 8. [doi: 10.1007/s10072-019-04022-y] [Medline: 31392643]
7. Lowe DJ, Dewar A, Lloyd A, Edgar S, Clegg GR. Optimising clinical performance during resuscitation using video evaluation. Postgrad Med J 2017 Aug;93(1102):449-453. [doi: 10.1136/postgradmedj-2016-134357] [Medline: 27986970]
8. Fang JL, Campbell MS, Weaver AL, Mara KC, Schuning VS, Carey WA, et al. The impact of telemedicine on the quality of newborn resuscitation: a retrospective study. Resuscitation 2018 Apr;125:48-55. [doi: 10.1016/j.resuscitation.2018.01.045] [Medline: 29408329]
9. Foster CB, Martinez KA, Sabella C, Weaver GP, Rothberg MB. Patient satisfaction and antibiotic prescribing for respiratory infections by telemedicine. Pediatrics 2019 Sep;144(3). [doi: 10.1542/peds.2019-0844] [Medline: 31371464]
10. Jamoom E, Yang N. Centers for Disease Control and Prevention. 2016. National Electronic Health Records Survey: 2015 State and National Electronic Health Record Adoption Summary Tables URL: https://www.cdc.gov/nchs/data/ahcd/nehrs/2015_nehrs_web_table.pdf [accessed 2019-05-31]
11. Smith A. Pew Research Center. 2017. Record shares of Americans now own smartphones, have home broadband URL: https://www.pewresearch.org/fact-tank/2017/01/12/evolution-of-technology [accessed 2019-05-30]
12. Wyatt KD, Willaert BN, Pallagi PJ, Uribe RA, Yiannias JA, Hellmich TR. PhotoExam: adoption of an iOS-based clinical image capture application at Mayo Clinic. Int J Dermatol 2017 Dec;56(12):1359-1365. [doi: 10.1111/ijd.13648] [Medline: 28497467]
13. Landman A, Emani S, Carlile N, Rosenthal DI, Semakov S, Pallin DJ, et al. A mobile app for securely capturing and transferring clinical images to the electronic health record: description and preliminary usability study. JMIR Mhealth Uhealth 2015 Jan 2;3(1):e1 [FREE Full text] [doi: 10.2196/mhealth.3481] [Medline: 25565678]
14. Li MK, Howard DP, King R. 'A picture tells a thousand words' smartphone-based secure clinical image transfer improves compliance in open fracture management. Injury 2019 Jul;50(7):1284-1287. [doi: 10.1016/j.injury.2019.05.010] [Medline: 31147186]
15. Pecina JL, Wyatt KD, Comfere NI, Bernard ME, North F. Uses of mobile device digital photography of dermatologic conditions in primary care. JMIR Mhealth Uhealth 2017 Nov 8;5(11):e165 [FREE Full text] [doi: 10.2196/mhealth.8257] [Medline: 29117934]
16. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]

Abbreviations

EHR: electronic health record
HIPAA: Health Insurance Portability and Accountability Act

Edited by G Eysenbach; submitted 03.06.19; peer-reviewed by T Schopf, J Pecina, F North; comments to author 31.08.19; revised version received 06.09.19; accepted 24.09.19; published 03.12.19.

Please cite as:
Cambron JC, Wyatt KD, Lohse CM, Underwood PY, Hellmich TR
Medical Videography Using a Mobile App: Retrospective Analysis
JMIR Mhealth Uhealth 2019;7(12):e14919
URL: https://mhealth.jmir.org/2019/12/e14919
doi:10.2196/14919
PMID:31793894

Original Paper

A Hospital-Community-Family–Based Telehealth Program for Patients With Chronic Heart Failure: Single-Arm, Prospective Feasibility Study

Xiaorong Guo[1,2,3*], MA; Xiang Gu[1,2,3*], MD, PhD; Jiang Jiang[1,2,3*], MA; Hongxiao Li[1,3], MD; Ruoyu Duan[2,3], MA; Yi Zhang[3], MA; Lei Sun[3], MA; Zhengyu Bao[3], MD; Jianhua Shen[3], MA; Fukun Chen[3], MD

[1]Clinical Medical College, Yangzhou University, Yangzhou, Jiangsu, China

[2]Dalian Medical University, Dalian, Liaoning, China

[3]Department of Cardiology, Subei People's Hospital, Yangzhou, Jiangsu, China

[*]these authors contributed equally

Corresponding Author:
Xiang Gu, MD, PhD
Clinical Medical College
Yangzhou University
No 98, Nantong West Road
Yangzhou, Jiangsu, 225009
China
Phone: 86 0514 87373366
Email: guxiang@yzu.edu.cn

Abstract

Background: An increasing number of patients with chronic heart failure (CHF) are demanding more convenient and efficient modern health care systems, especially in remote areas away from central cities. Telehealth is receiving increasing attention, which may be useful to patients with CHF.

Objective: This study aimed to evaluate the feasibility of a hospital-community-family (HCF)–based telehealth program, which was designed to implement remote hierarchical management in patients with CHF.

Methods: This was a single-arm prospective study in which 70 patients with CHF participated in the HCF-based telehealth program for remote intervention for at least 4 months. The participants were recruited from the clinic and educated on the use of smart health tracking devices and mobile apps to collect and manually upload comprehensive data elements related to the risk of CHF self-care management. They were also instructed on how to use the remote platform and mobile app to send text messages, check notifications, and open video channels. The general practitioners viewed the index of each participant on the mobile app and provided primary care periodically, and cardiologists in the regional central hospital offered remote guidance, if necessary. The assessed outcomes included accomplishments of the program, usability and satisfaction, engagement with the intervention, and changes of heart failure–related health behaviors.

Results: As of February 2018, a total of 66 individuals, aged 40-79 years, completed the 4-month study. Throughout the study period, 294 electronic medical records were formed on the remote monitoring service platform. In addition, a total of 89 remote consultations and 196 remote ward rounds were conducted. Participants indicated that they were generally satisfied with the intervention for its ease of use and usefulness. More than 91% (21/23) of physicians believed the program was effective, and 87% (20/23) of physicians stated that their professional knowledge could always be refreshed and enhanced through a library hosted on the platform and remote consultation. More than 60% (40/66) of participants showed good adherence to the care plan in the study period, and 79% (52/66) of patients maintained a consistent pattern of reporting and viewing their data over the course of the 4-month follow-up period. The program showed a positive effect on self-management for patients (healthy diet: P=.046, more fruit and vegetable intake: P=.02, weight monitoring: P=.002, blood pressure: P<.001, correct time: P=.049, and daily dosages of medicine taken: P=.006).

Conclusions: The HCF-based telehealth program is feasible and provided researchers with evidence of remote hierarchical management for patients with CHF, which can enhance participants' and their families' access and motivation to engage in self-management. Further prospective studies with a larger sample size are necessary to confirm the program's effectiveness.

(JMIR Mhealth Uhealth 2019;7(12):e13229) doi:10.2196/13229

KEYWORDS

telehealth; chronic heart failure; feasibility studies; precise follow-up; self-management

Introduction

Background

Chronic heart failure (CHF) is a major public health issue, affecting nearly 1 in every 100 people aged above than 65 years, and its prevalence is increasing with the aging population [1]. Patients with CHF have a poor quality of life and a very low 5-year survival rate [2]. In addition, they are particularly vulnerable to readmission. Previous studies indicated that the 30-day readmission rate was 5.6% for CHF, and more than half of them were readmitted a year later [3,4]. In addition to the morbidity of repeat hospitalizations, the cost of hospitalizations is high [5,6]. It undoubtedly increases the burden on the health care system and results in unnecessary wastage of medical resources.

Considering the demographic changes of patients with CHF, there is a need for a solution that can facilitate more convenient and effective access to medical service, especially in remote areas. With ubiquitous penetration of wireless internet, mobile phone, and portable personal health tracking devices, telehealth has become possible. On account of its potential to provide a much larger number of patients across a much greater geography with specialist care, it is increasingly being valued as a promising long-term management strategy for patients with chronic diseases. It promises the opportunity to *remotely* provide patients with consistent education, motivation to become engaged in their own self-care, and assistance in monitoring on a regular basis [7].

To date, many experiments have confirmed the remarkable achievements of telehealth in decreasing patients' risk of CHF exacerbations and hospital readmissions [8-11], which addresses the burdens associated with disease management and reduces CHF-associated health care costs [12]. However, a number of prior research studies have selected general practitioners or cardiologists to be the only management lead [13,14] or adopted a single management strategy to support self-care and home telemonitoring, such as structured telephone programs [10,15], mobile apps [11], or Web platforms [16]. To our knowledge, no existing system has incorporated a Web platform, mobile app, and smart health tracking devices, together with collaborative work between general practitioners and cardiologists, to engage and empower patients in disease self-management.

Objectives

We developed a hospital-community-family (HCF)–based telehealth program and aimed to explore its feasibility for the implementation of remote hierarchical management in patients with CHF.

Methods

Innovative Digital Devices in the Hospital-Community-Family–Based Telehealth Program

Overall, the program consisted of three innovative digital devices: (1) a remote monitoring service platform (Physio-Gate PG 1000, GETEMED Medizin- und Informationstechnik AG) that collects and integrates patients' data, (2) a personal health tracking mobile app (King OPTO-Electronic) for each patient, and (3) a few smart health tracking devices.

The Integrated Remote Monitoring Service Platform

The integrated remote monitoring service platform is a cloud-based, tablet computer–accessed, secure Web platform that was designed by the Information Academy of Yangzhou University to collect and integrate data. Its domain name is iccvd.com. It is available for both doctors and participants. Each doctor has his/her own specific log-in account. Participants can browse the Web directly without logging in. In the HCF-based telehealth program, this platform acted as a medium for information transmission and sharing of resources among regional central hospitals, community hospitals, and patients. The main functions of this platform are as follows:

- The electronic medical record of each participant, including medical history, physical examinations, laboratory and imaging findings, clinical diagnosis, general treatment, medication, and individualized reminder, was recorded on the remote monitoring service platform. There are three different important data sources: (1) data registered at the clinic, (2) data exported from the hospital information system, and (3) data recorded by participants themselves.
- In the electronic medical record, previous data about laboratory and imaging findings were cumulated and arranged in a chronological order. Each value was color marked to indicate the health status of participants: green for improvement, orange for medium risk, red for high risk, and no color for normal or close to normal range. Thus, doctors can assess a patient's condition more quickly and comprehensively from high-to-moderate risk to low risk, as all medical information is available at a glance.
- What makes the platform different is the function of analyzing the cumulative data of each individual. A trend diagram of quantitative data such as blood pressure (BP), heart rate, and 6-minute walk distance can be generated automatically by the platform, which can help visualize the overall dynamic change of data.
- The platform, equipped with a dedicated audio-video system, has high confidentiality and anti-interference ability, which is different from public video-audio platforms such as QQ or WeChat (Tencent, Shenzhen, China).
- The platform was designed as a library about CHF, including the current research state, latest developments,

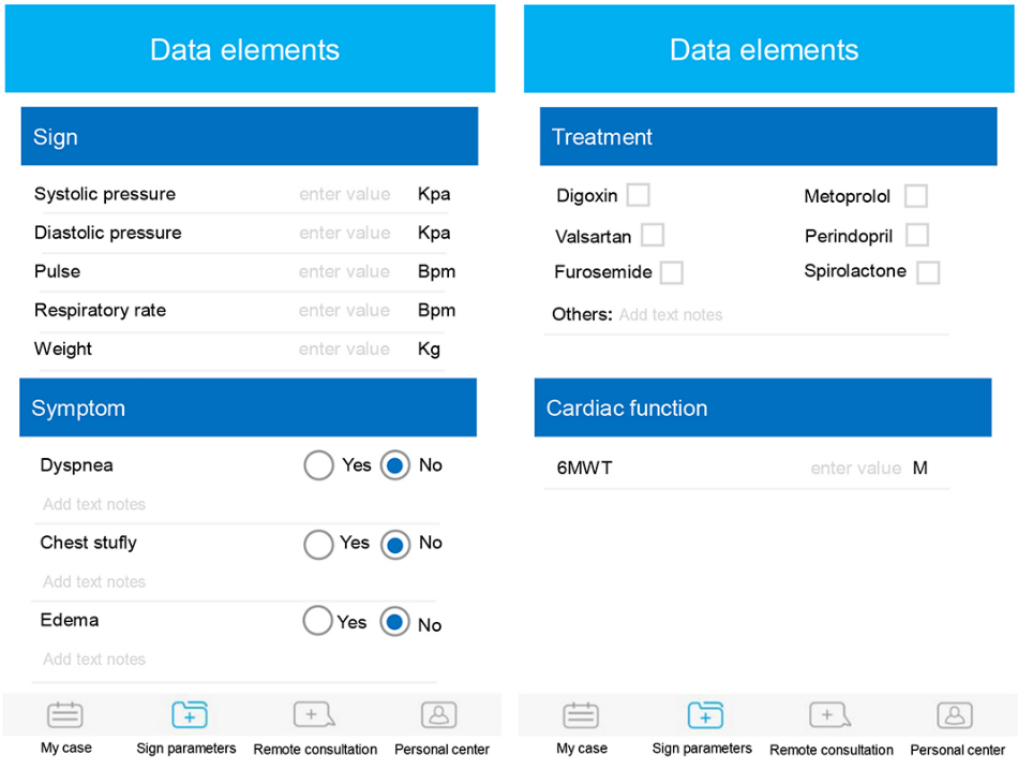
future trends, and most influential technologies and theories, with more advanced medical knowledge browsing links to the MedSci medical website.

Mobile App

A Web-based app, created on both Android and iOS platforms, was developed for reporting health information, monitoring clinical signs, and enabling direct supervision and instruction for patients. The app was designed with a link to the remote monitoring service platform, and two versions were available: one for physicians and another for patients. Both participants and physicians have the right to use the mobile app free of charge. It enables participant management from a distance and allows the physicians to observe and follow the health status of participants at any time.

The patient's mobile app is secured by their own specific password. It has the following main functions: data uploading, remote consultations, electronic medical record viewing, and medical appointments (Figure 1). Participants can use the app at home to record and upload comprehensive data elements related to the risk of CHF self-care management, including daily recording of symptom and sign changes and medication adherence. Physicians will analyze these incoming values. When data are outside an acceptable range, participants will receive video calls or text messages via the app. In addition, participants can communicate with physicians more conveniently through their handheld mobile phone with the assistance of the remote consultations function of our app. They can send text messages to their general practitioners round the clock for help and receive personalized guidance on long-term self-care management of CHF via the mobile app. Moreover, participants and their family members can view the participant's cumulative data, including examination report and clinical diagnosis and medication. Finally, reminders for tracking medical appointments and next visits are visible on the mobile app to keep track of their schedule.

Figure 1. Screenshots from the mobile app of patient.

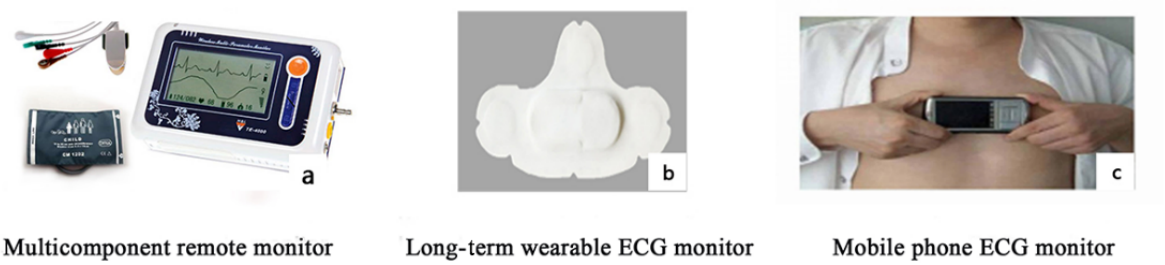


A mobile app for physicians was also designed. When a participant uploads data or clicks on the remote consultation icon, general practitioners receive text messages that alert them. They can then view the electronic medical record of the participant under their jurisdiction at any time by logging into the mobile app. If necessary, they can send text or video messages to participants and provide them with individualized advice through mobile apps.

Smart Health Tracking Devices

The basic devices including a weight scale and a BP monitor with a cuff were prepared by participants themselves. Some smart health tracking devices are available in our program, if necessary: multicomponent remote monitor (for an electrocardiogram [ECG], peripheral capillary oxygen saturation, and BP; hlTE-4000X, TE-4000Y, Beijing Haileying Medical Technology Co, Ltd, Beijing, China), long-term wearable ECG monitor (BECG1200-A, Thoth Medical Technology Co, Ltd, Suzhou, China), and mobile phone ECG monitor (Zhongwei Laikang Technology Development Co, Ltd, Beijing, China; Figure 2). These health tracking devices were invented to allow remote monitoring, recording, transmitting, and analysis of health parameters during daily activities.

Figure 2. Smart health tracking devices. ECG: electrocardiogram.



Study Design

Population

This single-arm, prospective, quasi-experimental study was conducted in the midland of Jiangsu province. The study received approval from the institutional review board of Subei People's Hospital. Patients were recruited consecutively from the outpatient clinics of our medical center between June 2017 and September 2017 and were followed up for at least 4 months.

Inclusion Criteria

The inclusion criteria were as follows: (1) age ≥18 years; (2) left ventricular ejection fraction of ≤45% as assessed by echocardiogram, but left ventricular ejection fraction can be >45% for patients with cardiac insufficiency caused by atrial fibrillation, valvular heart disease, and hypertrophic cardiomyopathy; (3) history of CHF of ≥3 months with New York Heart Association functional class I to IV, of which class IV was without strict bed rest; (4) the general intervention rules are based on the current guidelines [5,17]; and (5) ability to understand the requirements of the study and will to provide written informed consent.

Exclusion Criteria

The exclusion criteria were as follows: (1) some secondary cardiomyopathy (eg, hyperthyroid heart disease, anemic heart disease); (2) a history of malignancy and life expectancy <1 year; (3) severe primary hepatic and renal insufficiency (alanine aminotransferase level ≥ 100 U/L, serum creatinine level > 3.0 mg/dL, and serum albumin level < 2.5g/L); (4) refusal to participate; (5) inability to visit outpatient clinics periodically; (6) ambulatory status; and (7) judged to be inappropriate for the study by the researchers.

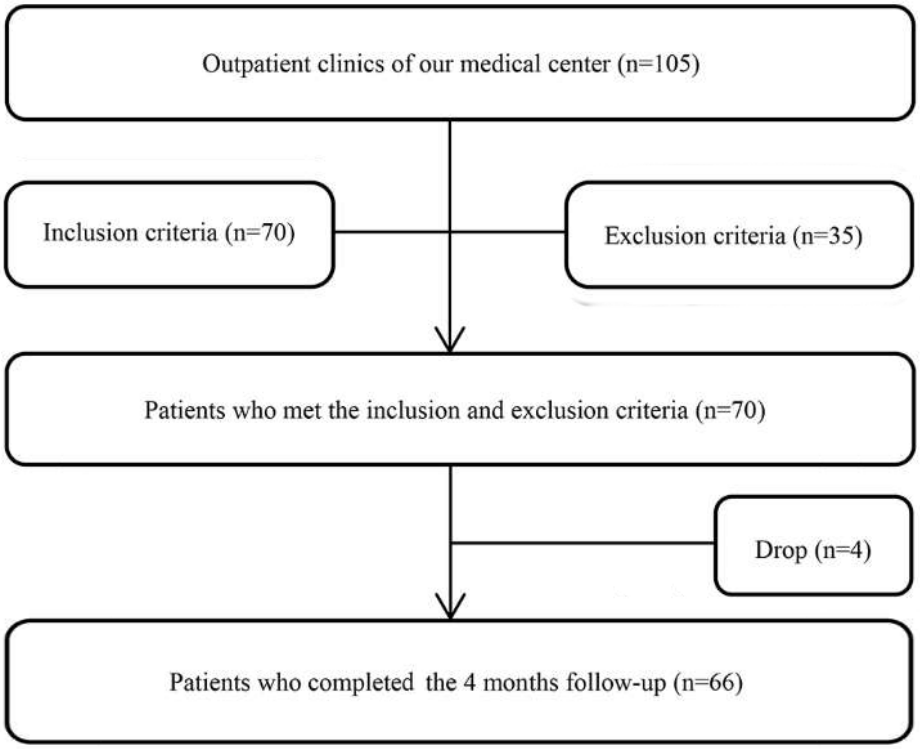
Study Procedure

General practitioners in 12 peripheral community hospitals who agreed to participate and signed the contract were organized and trained about the study procedures, the latest standardized management, and the treatment of CHF before the trial began.

On the patient's first outpatient visit, the investigator introduced the study plan to the patient and his or her caregiver (if present) who met the inclusion and exclusion criteria. The mobile app was downloaded to patient's iOS or Android mobile phone or tablet device after the procedures were explained, and written informed consent of the patient was obtained. Trained research assistants then instructed the patient on participation of the HCF-based telehealth program (described below) and the use of digital devices. At study enrollment, information on participant demographics was collected, and the first electronic medical record was established on the remote monitoring service platform. Thereafter, participants were assigned to the nearest participating community hospitals for further follow-up and management. Participants then took their health tracking devices and mobile apps home to join our HCF-based telehealth program, in which they were reminded to use the app or browse the Web platform more than once a week. Telephone follow-up was scheduled at weeks 1, 4, 8, and 16 to evaluate and provide technical support regarding the use of digital devices.

At the end of the study, the accomplishments of the HCF-based telehealth program were summarized and engagement in the intervention was assessed. Participants, their family members, and professional health care professionals were also interviewed for their experiences and opinions about the telehealth program. Participants were encouraged to provide their honest and candid feedback about the program. The study flow is illustrated in Figure 3.

Figure 3. Study flow of HCF-based telehealth program. HCF: hospital-community-family.



The Hospital-Community-Family–Based Telehealth Program

The HCF-based telehealth program consisted of multidisciplinary team members, including experienced cardiologists, general practitioners, participants, and their family members (Figure 4). Mobile app, as the main communication tool between different parts, is at the core of the telehealth structure. The HCF-based telehealth program was not a purely Web-based trial, and there were face-to-face components included in the intervention and assessment.

The program was partly financed by the project of the Jiangsu Provincial Science and Technology Department (project code: BL2013022) for items such as remote equipment, software development, labor subsidies, and academic expenses. The primary care physicians and cardiologists who joined the program were compensated or remunerated according to their workload and contribution to the project.

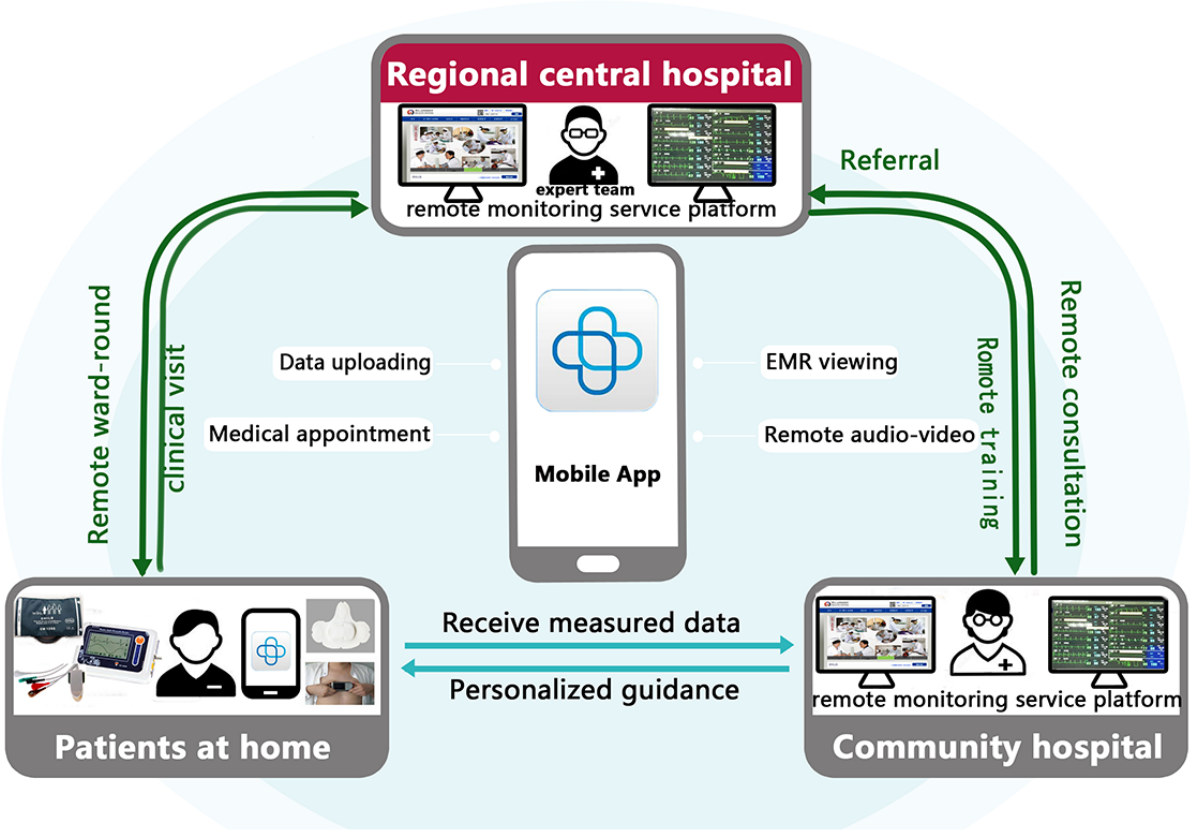
During the study, participants were required to visit the cardiologist clinic at least once every 2-4 months, and examination of indices related to CHF was arranged every 3-6 months or on the basis of the participant's condition. During regular clinical visits, participants could provide their health information via the mobile app at home and browse the Web-based platform at any time. As for participants with CHF who were at high risk of arrhythmia, a smart health tracking device was provided to record and send health parameters at any moment. Owing to the remote consultation function of the mobile app, patients could contact with the general practitioner 24 hours a day if they felt uncomfortable. Subsequently, they would receive feedback about their measurements, education about their symptoms and their diseases, reminders, and encouragement to follow care plans [18].

The general practitioners in community hospitals examined the data submitted by the participants and artificially identified abnormal or worsening health data of participants. Depending on the risk level of the participant, the general practitioners would send out a phone confirmation or perform other interventions, such as consulting experienced cardiologists at our regional central hospital. In some cases, the result may be a referral patient to a superior hospital through a *green passage*. Usually, general practitioners would initiate weekly interactive voice calls to assess the health status and provide technical support on the use of digital devices. Changes in the health status of each participant would be reported on the remote monitoring service platform in the electronic medical records form monthly by general practitioners. The comorbidities and medication would be kept up to date, as they were the reasons for adjusting the treatment regimen.

Experienced cardiology specialists in regional central hospitals, through the app's audio and video system, conducted weekly ward rounds for patients reported to be critically ill by general practitioners, tailored CHF management, and adjusted outpatient visit schedules according to the clinical situation of each participant. In addition, experts regularly held remote video lectures on the treatment and management of CHF and provided free training for general practitioners throughout the study.

A unique feature of this project was the collaborative work between general practitioners and cardiologists using a Web-based platform, a variety of intelligent health tracking devices, and mobile apps to achieve a comprehensive and personalized intervention for patients with CHF.

Figure 4. The hospital-community-family–based telehealth program in patients with chronic heart failure. EMR: electronic medical record.



Outcome Measures

Usability of the HCF intervention for patients was assessed based on the 12-item Perceived Health Web Site Usability Questionnaire [19], which is a widely used tool to assess usability for a variety of technologies. It consists of three separate components that ask about a patient's satisfaction with the intervention, the ease of use, and the effectiveness. All items were rated on a 1-7 scale. Responses were averaged for each component and across all items. In this study, the Cronbach alpha for satisfaction, ease of use, usefulness, and overall usability were .81, .74, .82, and .83, respectively. The physician's satisfaction with the intervention was assessed through self-reported questionnaires.

At the beginning and end of the study, lifestyle and health behaviors of participants were collected by interview, which would be used to assess changes in self-management for participants. Engagement with the intervention was assessed objectively via daily Web portal log-ins and use of the mobile app, including data uploading, remote consultations, and electronic medical record viewing. During the interview, a qualitative method was used to examine perceptions of the intervention components for participants.

Data Analyses

Descriptive statistics were computed for participants' characteristics and all outcome variables. Data were presented as numbers and percentages for categorical variables and mean and standard deviations for continuous variables. The Wilcoxon signed-rank tests for nonparametric data were used to assess satisfaction with the HCF intervention. Differences between categorical variables were analyzed by the Chi-square test when $P<.05$ was considered statistically significant.

Results

Baseline Characteristics of Participants

As of February 2018, of the 105 patients with CHF assessed for eligibility, 70 subjects met the inclusion criteria and agreed to participate in this study. Of the 70 participants, 4 (6%) dropped out from the study: 2 of them were lost to follow-up and 2 others withdrew voluntarily because they had migrated abroad and were unable to visit outpatient clinics periodically. The final analysis was performed in 66 (94%) participants who completed the baseline and 4-month measurements. All demographic information for participants who completed the study is shown in Table 1.

Table 1. Baseline demographic characteristics of the study participants.

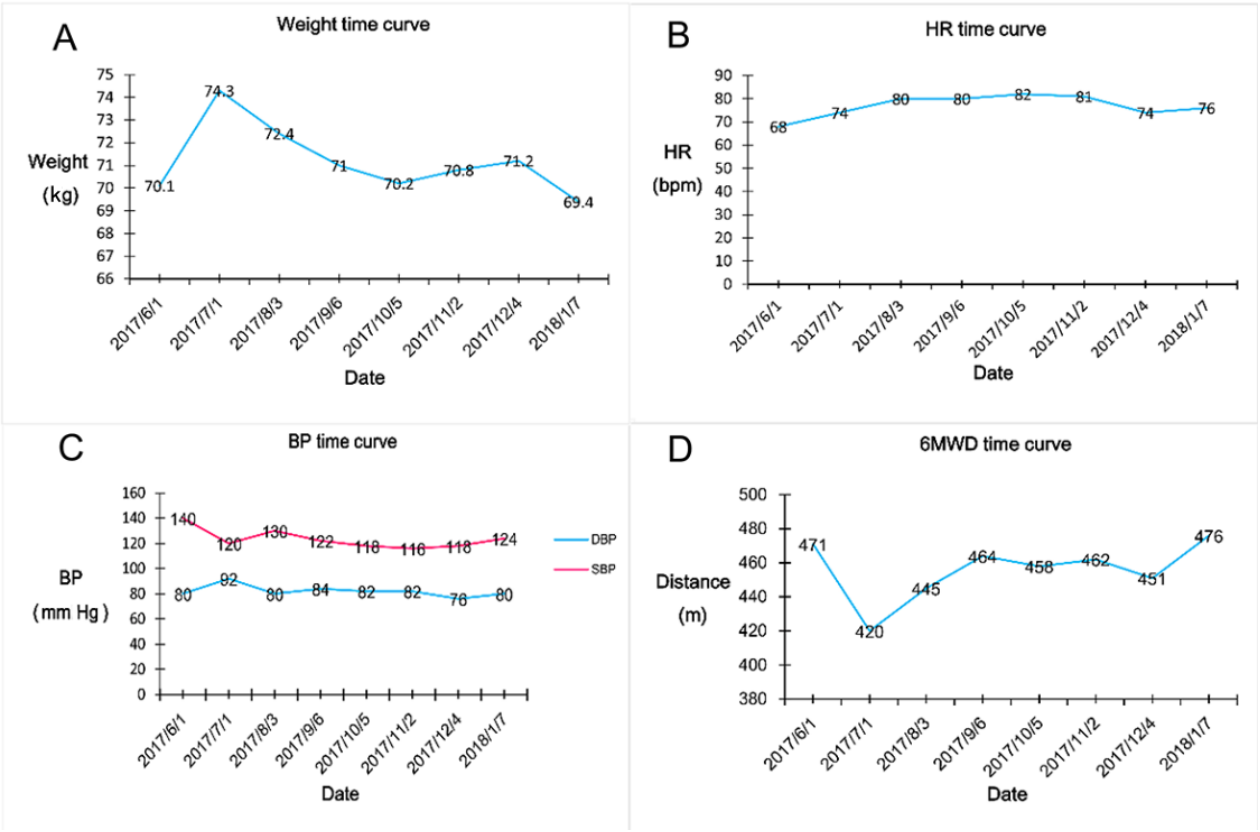
Characteristics	Values
Age (years), mean (SD)	69.35 (11.15)
Sex, n	
Male	34
Female	32
Body mass index (kg/m^2), mean (SE)	22.17 (1.69)
Clinical history, n (%)	
Hypertension	37 (56)
Coronary heart disease	11 (17)
Valvular heart disease	11 (17)
Atrial fibrillation	32 (48)
Cardiomyopathy	23 (35)
Diabetes	22 (33)
Medications, n (%)	
Angiotensin-converting enzyme inhibitors/angiotensin receptor blocker	36 (55)
Beta blocker	40 (61)
Diuretic	32 (48)
Digitalis	21 (32)
Ivabradine	0 (0)
Device, n (%)	
Cardiac resynchronization therapy	3 (5)
New York Heart Association, n (%)	
I	11 (17)
II	32 (48)
III	23 (35)
IV	0 (0)
Education, n (%)	
Less than high school	51 (77)
High school degree or more	15 (23)
Home caregiver, n (%)	
Spouse	34 (52)
Relative	21 (32)
Other	11 (17)
Monthly income (CNY $), n (%)	
≤400	31 (47)
400-800	21 (32)
≥800	14 (21)

Accomplishments of the Hospital-Community-Family–Based Telehealth Program

Participants uploaded data elements recorded by themselves or their caregivers weekly via the mobile app, submitting a total of 1096 reports. Data and the trend diagram of health parameters from one of the participants, analyzed and generated automatically by the platform, is shown in Figure 5. In addition, 294 electronic medical records were recorded on the remote monitoring service platform. Throughout the study period, general practitioners consulted the experienced cardiologists in the regional central hospital remotely 89 times by the mobile app, and one participant was transferred to a superior hospital.

Cardiologists completed 196 remote ward rounds and gave advice on adjustments of medications and lifestyle to participants. A total of eight remote interventions failed because of wireless network problems and operational errors by patients.

Figure 5. Data trends of a person on the remote monitoring service platform. HR: heart rate; BP: blood pressure; DBP: diastolic blood pressure; SBP: systolic blood pressure; 6MWD: 6-min walk distance.



Participants and Physicians' Experience of the Hospital-Community-Family Intervention

As shown in Table 2, the overall usability of the program was slightly above the midpoint of the 1-7 scale. Participants rated their satisfaction with and usefulness of the HCF intervention higher than they rated the ease of use of the HCF intervention.

At the end of the study, 23 physicians, including 18 general practitioners and 5 experienced cardiologists, were surveyed. Results from the survey data showed that 21 (91%) physicians believed that this remote hierarchical management program improves management efficiency, with 20 (87%) physicians stating their professional knowledge can always be refreshed and enhanced through the library hosted on the platform and remote consultation, and all 23 (100%) physicians firmly believed that the program should be explored and promoted on a larger scale.

Table 2. Participant-reported usability of hospital-community-family intervention (average scores; possible range=1-7; higher scores indicate greater overall usability, higher degree of satisfaction, easier use, and better effectiveness of the hospital-community-family intervention).

Usability of intervention	Mean (SD)	Range
Overall usability	4.79 (1.03)	2.00-7.00
Satisfaction	5.04 (0.89)	3.00-7.00
Ease of use	4.40 (1.08)	2.00-7.00
Usefulness	4.69 (1.09)	2.00-7.00

Feedback on Intervention Components

Most participants indicated that they found the mobile app useful as a tool to track CHF-related information, contact clinicians easily, and be reminded to take their medication, and they were also able to share their logged information with the people whom they trust most, such as family members or close friends who could help them during their medical appointments. Participants also responded positively to the knowledge propaganda feature of the remote monitoring service platform, which provided them with useful information, particularly relating to sustaining health. The participants were more interested in smart health tracking devices, which could help them keep track of health conditions anywhere, so that they felt more secure and involved in their own care. Generally, most of them were somewhat or very willing to recommend the telehealth program to patients with

CHF around them. We also heard a few complaints about the components that negatively affected their experience. These complaints were mainly focused on the complex operation steps and the consumption of mobile data traffic. Moreover, one patient complained that voice or video call is the largest segment of mobile data traffic unless they can directly connect to the Wi-Fi. Another older patient said that operating this complicated system has more or less obstacles for him.

Engagement With the Intervention

The overall engagement with the two main components of the intervention (ie, the mobile app and the Web platform) was assessed by participants or their caregivers' usage logging. More than 60% (40/66) of the study participants showed great adherence to the HCF-based telehealth program, which was defined as the use of the app or the access of Web platform more than once a week and the visit of specialist clinics as scheduled during the study period. Although most of them engaged through the mobile app, 12 participants (18%) who displayed high adherence to the program were found to have used the mobile app and the Web platform simultaneously. In addition, 79% (52/66) of the patients maintained a consistent pattern of reporting and viewing their data over the course of the 4-month follow-up period. We found that overall engagement decreased following the first 8 weeks, but 40 of the 66 participants continued to engage with the care plan throughout each week of the study. Moreover, nearly one-third (21/66) of the participants used a mobile app by themselves, and the rest were mostly family members or caregivers.

Changes in Lifestyle and Health Behaviors

As shown in Table 3, the program may also have some value in improving health behaviors (increasing fruit and vegetable intake, controlling BP and weight, and reducing salt intake) and drug compliance in participants.

Table 3. Changes in lifestyle and health behaviors.

Lifestyle and health behaviors	Baseline, n (%)	At 4 months, n (%)	*P* value
Diet			
Low salt, low fat, and low sugar	37 (56)	48 (73)	.046
More fruits or vegetables	22 (33)	35 (53)	.02
Self-monitoring			
Blood pressure	15 (23)	41 (62)	<.001
Weight	5 (8)	19 (29)	.002
Medicine adherence			
Correct daily dosages	41 (62)	55 (83)	.006
Correct time	35 (53)	46 (70)	.049

Discussion

Principal Findings

This is a prospective experimental study to investigate the feasibility of the HCF-based telehealth program in patients with CHF. To the best of our knowledge, this is the first study evaluating the use of such a telehealth program incorporating remote monitoring service platforms, mobile apps, and smart health tracking devices to manage patients with CHF at their own homes. The initial results have been achieved in the program. The results indicate that satisfaction and participation of doctors and patients are relatively higher; meanwhile, the patients' lifestyle has been effectively improved. Our research is expected to lay a foundation for further large-scale randomized controlled clinical trials related to remote hierarchical medical care.

Comparison With Prior Work

The main contribution of this program lies in the development of a remote hierarchical management system centered on patients with CHF. This system has multiple intelligent tracking devices that can support patients' self-care at home and strengthen communications among patients, medical service providers, and home caregivers for better care transition and coordination. To our knowledge, this is one of few systems with collaborative work between general practitioners and cardiologists for HF remote hierarchical management through remote monitoring service platforms, mobile apps, and smart health tracking devices.

This paper presented the design and use of the HCF-based telehealth program for patients with CHF. Similar to other pilot studies [20], we focused on user perceptions and experience because user perspective was the most important dimension in the development phase of telehealth projects [21,22]. According to the interviews of participants, their family members, and physicians, they were generally satisfied with the service. The CHF management team was key in engaging patients to participate the program; the benefits of the program also include support in recording and tracking health status, encouragement and reassurance received from medical staff, timely detection, recognition and management of subtle changes in the condition, and more convenient and faster communication among all participants. Previous investigation showed that physicians did not seem very enthusiastic about telemedicine. The main causes of such opposition were found to be the lack of a medical services delivery system and a professional management team [23]. The HCF-based telehealth program can serve as a platform for providing more continuous care, linking primary and

specialty care to support management of patients with severe and chronic diseases. Some physicians perceived the platform to be effective as a communication tool to share data in a timely, accurate, and visual manner, so that they can be armed with all relevant health information contained in one system, especially in an emergency or unfamiliar health care setting, for care planning.

The poor compliance of patients is a major obstacle to the management of most chronic diseases [24,25]. An earlier large trial showed that 14% of the intervention patients never used their equipment, and only 55% were still using the system at the end of the study (6 months from the baseline) [26]. Our most significant finding was that 79% of patients maintained a consistent pattern of reporting and viewing their data over the course of the 4-month follow-up period. In contrast, we found higher engagement levels over the course of the study. A possible explanation for this findings is the user friendliness of digital devices and that family-targeted self-care interventions [27] may play an important role in promoting patient compliance. The caregiver app will show the patient's real-time data (with the patient's consent) and receive messages from the patient's health care providers, so that he/she can help monitor the patient's health. In fact, telehealth apps have been used to enhance patient-caregiver engagement [28]. Inviting family members or caregivers in training sessions was effective for older adults to continue to adopt the telehealth tools in their daily life. To date, educational interventions intentionally including family involvement in the care of patients with CHF are few, although such family involvement is explicitly recommended in the existing CHF management guidelines [29].

In accordance with reports from the World Health Organization, nearly two-thirds of people's quality of life and health status lies with their lifestyle and health behaviors, and 53% of the death causes were also associated with lifestyle and personal behaviors. Therefore, we can say that elderly individuals who adopt improved health behaviors would experience a healthy old age [30]. Emerging evidence suggests that telehealth interventions may improve self-care behaviors and disease management for elderly patients [11,31,32]. The Web-based social networks acting as an active communication framework can be an effective means of promoting healthy lifestyles [33]. In the study, patients were encouraged to document their conditions and medicine use and then received personalized health coaching tailored for their individual situation through *face-to-face* communications, which can intensify the patient's health behaviors, medication adherence, and self-management ability. One possible reason is that the use of mobile health via an app with real-time representation of data trends would strengthen patient empowerment and decision making in self-management. They felt more empowered and confident to perform their self-care activities at home with the help of the program.

Limitations

This study had some limitations. First, our sample size was small, and the duration of intervention was short. In addition, as this was only a feasibility pilot study, we did not conduct a formal sample size calculation. Furthermore, this was a single-arm, experimental, prospective design study rather than a randomized evaluation and may be not powered to detect the effects of the intervention on clinical outcomes. Large-scale randomized prospective controlled studies will be necessary to test the program. In addition, those who participated in our study were predominantly elderly patients with CHF having varying socioeconomic and educational backgrounds; this limits the generalizability of our findings and precludes us from extrapolating the findings from this study to other populations. Finally, because of technical restriction, semiautomatic input of data in this program greatly wastes human resources and possibly increases the error rate.

Conclusions

Given the serious and complex condition of patients with CHF, a more convenient and effective access to medical services is urgently needed, especially in remote areas. The HCF-based telehealth program offers them a glimmer of light. On the basis of remote monitoring service platform, mobile app, intelligent health tracking device, and professional management team, this study realized remote hierarchical management of patients with CHF. This study provided evidence on the feasibility of HCF-based telehealth program, potentially enhancing the opportunities and incentives for patients with CHF and their families to participate in self-management. In addition, it may lay the foundation for further large-scale randomized controlled studies. In the near future, we would like to expand the HCF-based telehealth program to other cardiovascular diseases such as atrial fibrillation, coronary heart disease, and stroke.

Acknowledgments

This trial was financed by the project of the Science and Technology Department of Jiangsu Province (funding code: BL2013022). We would like to express our sincere thanks to Shu-Hang Miao, Qing-Qing Shi, Meng-Yuan Dai, Min Gu, Suo-Ya Pan, Mei-Qian Wu, Hai-Yan Li, You-Jin Chen, and other cardiovascular clinic staff and patients involved and the research team at the Network Information Center for making this study possible (including Yi-Jun Ju and Lu Chen of the Information Department of Subei People's Hospital, Xiao-Chao Ding of Information Academy of Yangzhou University, and Yue Shen of King OPTO-Electronic).

Conflicts of Interest

None declared.

References

1. Mozaffarian D, Benjamin EJ, Go AS, Arnett DK, Blaha MJ, Cushman M, American Heart Association Statistics Committee and Stroke Statistics Subcommittee. Heart disease and stroke statistics--2015 update: a report from the American Heart Association. Circulation 2015 Jan 27;131(4):e29-322. [doi: 10.1161/CIR.0000000000000152] [Medline: 25520374]
2. Shah KS, Xu H, Matsouaka RA, Bhatt DL, Heidenreich PA, Hernandez AF, et al. Heart failure with preserved, borderline, and reduced ejection fraction: 5-year outcomes. J Am Coll Cardiol 2017 Nov 14;70(20):2476-2486 [FREE Full text] [doi: 10.1016/j.jacc.2017.08.074] [Medline: 29141781]
3. Tuppin P, Cuerq A, de Peretti C, Fagot-Campagna A, Danchin N, Juillière Y, et al. Two-year outcome of patients after a first hospitalization for heart failure: a national observational study. Arch Cardiovasc Dis 2014 Mar;107(3):158-168 [FREE Full text] [doi: 10.1016/j.acvd.2014.01.012] [Medline: 24662470]
4. Khan H, Greene SJ, Fonarow GC, Kalogeropoulos AP, Ambrosy AP, Maggioni AP, EVEREST Trial Investigators. Length of hospital stay and 30-day readmission following heart failure hospitalization: insights from the EVEREST trial. Eur J Heart Fail 2015 Oct;17(10):1022-1031 [FREE Full text] [doi: 10.1002/ejhf.282] [Medline: 25960401]
5. Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE, Drazner MH, et al. 2013 ACCF/AHA guideline for the management of heart failure: executive summary: a report of the American College of Cardiology Foundation/American Heart Association Task Force on practice guidelines. Circulation 2013 Oct 15;128(16):1810-1852. [doi: 10.1161/CIR.0b013e31829e8807] [Medline: 23741057]
6. Ziaeian B, Sharma PP, Yu T, Johnson KW, Fonarow GC. Factors associated with variations in hospital expenditures for acute heart failure in the United States. Am Heart J 2015 Feb;169(2):282-9.e15 [FREE Full text] [doi: 10.1016/j.ahj.2014.11.007] [Medline: 25641538]
7. Nissen L, Lindhardt T. A qualitative study of COPD-patients' experience of a telemedicine intervention. Int J Med Inform 2017 Nov;107:11-17. [doi: 10.1016/j.ijmedinf.2017.08.004] [Medline: 29029687]
8. Gu X, Zhu Y, Zhang Y, Sun L, Bao Z, Shen J, et al. Effect of telehealth interventions on major cardiovascular outcomes: a meta-analysis of randomized controlled trials. J Geriatr Cardiol 2017 Aug;14(8):501-508 [FREE Full text] [doi: 10.11909/j.issn.1671-5411.2017.08.013] [Medline: 29089966]
9. Bernocchi P, Vitacca M, La Rovere MT, Volterrani M, Galli T, Baratti D, et al. Home-based telerehabilitation in older patients with chronic obstructive pulmonary disease and heart failure: a randomised controlled trial. Age Ageing 2018 Jan 1;47(1):82-88. [doi: 10.1093/ageing/afx146] [Medline: 28985325]
10. Inglis SC, Clark RA, Dierckx R, Prieto-Merino D, Cleland JG. Structured telephone support or non-invasive telemonitoring for patients with heart failure. Heart 2017 Feb 15;103(4):255-257. [doi: 10.1136/heartjnl-2015-309191] [Medline: 27864319]
11. Piette JD, List J, Rana GK, Townsend W, Striplin D, Heisler M. Mobile health devices as tools for worldwide cardiovascular risk reduction and disease management. Circulation 2015 Nov 24;132(21):2012-2027 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.114.008723] [Medline: 26596977]
12. Jiang X, Ming W, You JH. The cost-effectiveness of digital health interventions on the management of cardiovascular diseases: systematic review. J Med Internet Res 2019 Jun 17;21(6):e13166 [FREE Full text] [doi: 10.2196/13166] [Medline: 31210136]
13. Bussières S, Tanguay A, Hébert D, Fleet R. Unité de Coordination Clinique des Services Préhospitaliers d'Urgence: A clinical telemedicine platform that improves prehospital and community health care for rural citizens. J Telemed Telecare 2017 Jan;23(1):188-194. [doi: 10.1177/1357633X15627234] [Medline: 27072126]
14. Haas K, Hayoz S, Maurer-Wiesner S. Effectiveness and feasibility of a remote lifestyle intervention by dietitians for overweight and obese adults: pilot study. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e12289 [FREE Full text] [doi: 10.2196/12289] [Medline: 30973338]
15. Grustam AS, Severens JL, de Massari D, Buyukkaramikli N, Koymans R, Vrijhoef HJ. Cost-effectiveness analysis in telehealth: a comparison between home telemonitoring, nurse telephone support, and usual care in chronic heart failure management. Value Health 2018 Jul;21(7):772-782 [FREE Full text] [doi: 10.1016/j.jval.2017.11.011] [Medline: 30005749]
16. Marcelle ET, Nolting L, Hinshaw SP, Aguilera A. Effectiveness of a multimodal digital psychotherapy platform for adult depression: a naturalistic feasibility study. JMIR Mhealth Uhealth 2019 Jan 23;7(1):e10948 [FREE Full text] [doi: 10.2196/10948] [Medline: 30674448]
17. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, Authors/Task Force Members, Document Reviewers. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur J Heart Fail 2016 Aug;18(8):891-975 [FREE Full text] [doi: 10.1002/ejhf.592] [Medline: 27207191]
18. Schmidt C, Öner A, Mann M, Krockenberger K, Abbondanzieri M, Brandewiede B, et al. A novel integrated care concept (NICC) versus standard care in the treatment of chronic cardiovascular diseases: protocol for the randomized controlled trial CardioCare MV. Trials 2018 Feb 20;19(1):120 [FREE Full text] [doi: 10.1186/s13063-018-2502-1] [Medline: 29458404]
19. Nahm E, Resnick B, Mills ME. Development and pilot-testing of the perceived health Web Site usability questionnaire (PHWSUQ) for older adults. Stud Health Technol Inform 2006;122:38-43. [Medline: 17102214]

20. Selter A, Tsangouri C, Ali SB, Freed D, Vatchinsky A, Kizer J, et al. An mHealth app for self-management of chronic lower back pain (Limbr): pilot study. JMIR Mhealth Uhealth 2018 Sep 17;6(9):e179 [FREE Full text] [doi: 10.2196/mhealth.8256] [Medline: 30224333]
21. Farnia T, Jaulent M, Steichen O. Evaluation criteria of noninvasive telemonitoring for patients with heart failure: systematic review. J Med Internet Res 2018 Jan 16;20(1):e16 [FREE Full text] [doi: 10.2196/jmir.7873] [Medline: 29339348]
22. Anderson K, Burford O, Emmerton L. Mobile health apps to facilitate self-care: a qualitative study of user experiences. PLoS One 2016;11(5):e0156164 [FREE Full text] [doi: 10.1371/journal.pone.0156164] [Medline: 27214203]
23. Choi WS, Park J, Choi JY, Yang JS. Stakeholders' resistance to telemedicine with focus on physicians: Utilizing the Delphi technique. J Telemed Telecare 2019 Jul;25(6):378-385. [doi: 10.1177/1357633X18775853] [Medline: 29792080]
24. Lee H, Park J, Choi SW, Yoon YE, Park HE, Lee SE, et al. Impact of a telehealth program with voice recognition technology in patients with chronic heart failure: feasibility study. JMIR Mhealth Uhealth 2017 Oct 2;5(10):e127 [FREE Full text] [doi: 10.2196/mhealth.7058] [Medline: 28970189]
25. Ware P, Dorai M, Ross HJ, Cafazzo JA, Laporte A, Boodoo C, et al. Patient adherence to a mobile phone-based heart failure telemonitoring program: a longitudinal mixed-methods study. JMIR Mhealth Uhealth 2019 Feb 26;7(2):e13259 [FREE Full text] [doi: 10.2196/13259] [Medline: 30806625]
26. Chaudhry SI, Mattera JA, Curtis JP, Spertus JA, Herrin J, Lin Z, et al. Telemonitoring in patients with heart failure. N Engl J Med 2010 Dec 9;363(24):2301-2309 [FREE Full text] [doi: 10.1056/NEJMoa1010029] [Medline: 21080835]
27. Deek H, Chang S, Newton PJ, Noureddine S, Inglis SC, Arab GA, et al. An evaluation of involving family caregivers in the self-care of heart failure patients on hospital readmission: Randomised controlled trial (the FAMILY study). Int J Nurs Stud 2017 Oct;75:101-111. [doi: 10.1016/j.ijnurstu.2017.07.015] [Medline: 28772186]
28. Piette JD, Striplin D, Marinec N, Chen J, Aikens JE. A randomized trial of mobile health support for heart failure patients and their informal caregivers: impacts on caregiver-reported outcomes. Med Care 2015 Aug;53(8):692-699 [FREE Full text] [doi: 10.1097/MLR.0000000000000378] [Medline: 26125415]
29. McMurray JJ, Adamopoulos S, Anker SD, Auricchio A, Böhm M, Dickstein K, Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology, ESC Committee for Practice Guidelines. ESC guidelines for the diagnosis and treatment of acute and chronic heart failure 2012: The Task Force for the Diagnosis and Treatment of Acute and Chronic Heart Failure 2012 of the European Society of Cardiology. Developed in collaboration with the Heart Failure Association (HFA) of the ESC. Eur J Heart Fail 2012 Aug;14(8):803-869 [FREE Full text] [doi: 10.1093/eurjhf/hfs105] [Medline: 22828712]
30. Cowie MR, Bax J, Bruining N, Cleland JG, Koehler F, Malik M, et al. e-Health: a position statement of the European Society of Cardiology. Eur Heart J 2016 Jan 1;37(1):63-66 [FREE Full text] [doi: 10.1093/eurheartj/ehv416] [Medline: 26303835]
31. Rubinstein A, Miranda JJ, Beratarrechea A, Diez-Canseco F, Kanter R, Gutierrez L, GISMAL group. Effectiveness of an mHealth intervention to improve the cardiometabolic profile of people with prehypertension in low-resource urban settings in Latin America: a randomised controlled trial. Lancet Diabetes Endocrinol 2016 Jan;4(1):52-63. [doi: 10.1016/S2213-8587(15)00381-2] [Medline: 26653067]
32. Piette JD, Striplin D, Marinec N, Chen J, Trivedi RB, Aron DC, et al. A mobile health intervention supporting heart failure patients and their informal caregivers: a randomized comparative effectiveness trial. J Med Internet Res 2015 Jun 10;17(6):e142 [FREE Full text] [doi: 10.2196/jmir.4550] [Medline: 26063161]
33. Meeker D, Goldberg J, Kim KK, Peneva D, Campos HD, Maclean R, et al. Patient Commitment to Health (PACT-Health) in the Heart Failure Population: a focus group study of an active communication framework for patient-centered health behavior change. J Med Internet Res 2019 Aug 6;21(8):e12483 [FREE Full text] [doi: 10.2196/12483] [Medline: 31389339]

Abbreviations

BP: blood pressure
CHF: chronic heart failure
ECG: electrocardiogram
HCF: hospital-community-family

Edited by G Eysenbach; submitted 23.12.18; peer-reviewed by JB Park, O Steichen; comments to author 11.04.19; revised version received 06.06.19; accepted 26.09.19; published 13.12.19.

Please cite as:
Guo X, Gu X, Jiang J, Li H, Duan R, Zhang Y, Sun L, Bao Z, Shen J, Chen F
A Hospital-Community-Family–Based Telehealth Program for Patients With Chronic Heart Failure: Single-Arm, Prospective Feasibility Study
JMIR Mhealth Uhealth 2019;7(12):e13229
URL: https://mhealth.jmir.org/2019/12/e13229
doi:10.2196/13229
PMID:31833835

Original Paper

Usefulness of Modern Activity Trackers for Monitoring Exercise Behavior in Chronic Cardiac Patients: Validation Study

Cyrille Herkert[1], MD; Jos Johannes Kraal[1], PhD; Eline Maria Agnes van Loon[1], MD; Martijn van Hooff[2], MSc; Hareld Marijn Clemens Kemps[1], MD, PhD

[1]Máxima Medical Center, Flow, Center for Prevention, Telemedicine and Rehabilitation in Chronic Disease, Eindhoven, Netherlands

[2]Máxima Medical Center, Department of Sports Medicine, Eindhoven, Netherlands

Corresponding Author:
Cyrille Herkert, MD
Máxima Medical Center
Flow, Center for Prevention, Telemedicine and Rehabilitation in Chronic Disease
Dominee Theodor Fliednerstraat 1
Eindhoven, 5631 BM
Netherlands
Phone: 31 408888200
Email: cyrille.herkert@mmc.nl

Abstract

Background: Improving physical activity (PA) is a core component of secondary prevention and cardiac (tele)rehabilitation. Commercially available activity trackers are frequently used to monitor and promote PA in cardiac patients. However, studies on the validity of these devices in cardiac patients are scarce. As cardiac patients are being advised and treated based on PA parameters measured by these devices, it is highly important to evaluate the accuracy of these parameters in this specific population.

Objective: The aim of this study was to determine the accuracy and responsiveness of 2 wrist-worn activity trackers, Fitbit Charge 2 (FC2) and Mio Slice (MS), for the assessment of energy expenditure (EE) in cardiac patients.

Methods: EE assessed by the activity trackers was compared with indirect calorimetry (Oxycon Mobile [OM]) during a laboratory activity protocol. Two groups were assessed: patients with stable coronary artery disease (CAD) with preserved left ventricular ejection fraction (LVEF) and patients with heart failure with reduced ejection fraction (HFrEF).

Results: A total of 38 patients were included: 19 with CAD and 19 with HFrEF (LVEF 31.8%, SD 7.6%). The CAD group showed no significant difference in total EE between FC2 and OM (47.5 kcal, SD 112 kcal; *P*=.09), in contrast to a significant difference between MS and OM (88 kcal, SD 108 kcal; *P*=.003). The HFrEF group showed significant differences in EE between FC2 and OM (38 kcal, SD 57 kcal; *P*=.01), as well as between MS and OM (106 kcal, SD 167 kcal; *P*=.02). Agreement of the activity trackers was low in both groups (CAD: intraclass correlation coefficient [ICC] FC2=0.10, ICC MS=0.12; HFrEF: ICC FC2=0.42, ICC MS=0.11). The responsiveness of FC2 was poor, whereas MS was able to detect changes in cycling loads only.

Conclusions: Both activity trackers demonstrated low accuracy in estimating EE in cardiac patients and poor performance to detect within-patient changes in the low-to-moderate exercise intensity domain. Although the use of activity trackers in cardiac patients is promising and could enhance daily exercise behavior, these findings highlight the need for population-specific devices and algorithms.

(JMIR Mhealth Uhealth 2019;7(12):e15045) doi:10.2196/15045

KEYWORDS

cardiac diseases; activity trackers; energy metabolism; physical activity; validation studies

Introduction

Background

Improving physical fitness and physical activity (PA) levels are core components of cardiac rehabilitation (CR) and secondary prevention in patients with coronary artery disease (CAD) and chronic heart failure (CHF) [1,2]. Exercise-based CR in these patients has documented positive effects on psychological well-being, hospitalization, and mortality by slowing the progression of CAD and CHF, while also combatting risk factors as hypertension, dyslipidemia, and psychological stress [1,3,4].

Besides exercise training, enhancing daily PA and reducing sedentary behavior are also effective for the prevention of repetitive cardiac events and therefore highly recommended by the current guidelines [5]. Yet, despite these recommendations, patients with CAD and CHF are characterized by a less active lifestyle compared with individuals not diagnosed with these conditions. Subsequently, this results in further deterioration of their activity level, social participation, prognosis, and quality of life [6-8]. One of the explanations for low activity levels in these patients is that secondary prevention programs and center-based CR typically do not focus on increasing daily activity behavior and reduction of sedentary time, as shown by high sedentary levels in post-CR patients comparable with patients who have not participated in CR [9]. This may be because of the fact that CR programs mainly focus on short-term improvements of exercise capacity rather than preparation for self-management and improving exercise behavior in daily life, causing a relapse into sedentary behavior after the rehabilitation program [10].

Home-based exercise programs such as telerehabilitation, may be successful methods for improving PA and reducing sedentary behavior on long term [11-15]. Telerehabilitation programs do thereby not only focus on prescribing exercise sessions but also on monitoring PA levels during the day to give appropriate feedback on PA behavior. To monitor and promote PA successfully during cardiac telerehabilitation, reliable and nonobtrusive devices to assess energy expenditure (EE) need to be available. In fact, safety and clinical effectivity of medical devices are high on the agenda of the European Society of Cardiology, mainly focusing on high risk (implantable) devices such as coronary artery stents or implantable cardioverter defibrillators [16]. At this stage, less attention is payed to the clinical effectiveness of wearable sensors and its algorithms in cardiac patients. Since nowadays, exercise prescriptions in cardiac patients are being made based on commercially available activity trackers, more information on the validity of these devices in cardiac patients is urgently needed, as medical advice based on invalid physiological parameters might be potentially harmful.

Important prerequisites for activity trackers include high accuracy and responsiveness. Accuracy is defined as the closeness of agreement between the device measurement and the true value [17]. Responsiveness of a device is the ability to detect within-patient changes of exercise intensity over time and is therefore highly important in cardiac patients to monitor progression, and for daily coaching [18,19].

Kraal et al found that combining heart rate (HR) with accelerometer data provides a higher accuracy for measuring EE than using accelerometer or HR data alone in patients with CAD [20]. However, in this study, a hip-worn accelerometer was combined with a chest belt to record HR, which is not feasible for all-day use. Wrist-worn activity trackers can provide a nonobtrusive solution combining HR and accelerometer data for calculation of EE. However, little information is available about their accuracy and responsiveness. Studies that did validate such devices showed mixed results [21-24]. These studies, however, used early device models and were performed in a population without cardiac conditions. The results of these studies cannot be extrapolated to cardiac patients because of factors such as chronotropic incompetence and use of HR lowering medication such as beta blockers. Moreover, to our knowledge, there has been no previous investigation to the responsiveness of wrist-based devices combining HR with accelerometer data, especially not in cardiac patients.

Objectives

The aim of this study is to investigate the accuracy and responsiveness of 2 commercially available wrist-worn activity trackers to calculate EE in patients with CAD and CHF. This study may provide important information whether 2 modern activity trackers, Fitbit Charge 2 (FC2) and Mio Slice (MS), can be used for measuring EE to monitor PA levels in cardiac patients.

Methods

Study Population

Patients (aged ≥18 years) were included based on their diagnosis to form 2 patient groups: patients with stable CAD with preserved left ventricular ejection fraction (LVEF) and patients with stable heart failure with reduced ejection fraction (HFrEF). These 2 patient categories were selected because HFrEF patients generally have lower activity levels compared with CAD patients with preserved LVEF. Household activities, for example, can be experienced as more intensive by HFrEF patients than by CAD patients with preserved LVEF. Also, HFrEF patients often suffer from chronotropic incompetence, yielding a difference in HR variation during activities. Therefore, both groups were analyzed separately. The participants were recruited via their cardiologist in the outpatient clinic of the Máxima Medical Center, the Netherlands, and were randomly selected from a list of patients who participated in previous studies and gave informed consent to be contacted for participation in future research projects. Patients were excluded if they suffered from hemodynamic significant valvular disease, permanent atrial fibrillation or peripheral vascular, neurological, or orthopedic conditions impairing exercise capacity. All patients provided written informed consent. The study was approved by the local medical ethical committee of the Máxima Medical Center and was conducted in accordance with the declaration of Helsinki.

Protocol

Participants completed a laboratory protocol consisting of 14 low-to-moderate intensity activities, which was modified from a previous study with cardiac patients [20]. The protocol comprised sedentary and household activities, treadmill walking on 3 different speeds, cycling on 3 different loads, and walking up and down the stairs. Walking speeds and cycling loads were adjusted for the HFrEF group. The total duration of the protocol was 39 min (resting time excluded). An overview of the protocol is shown in Table 1. Resting HR was measured using a chest belt (Polar T31, Polar) at the start of the protocol. Between each activity, the patients received recovery time which lasted until the HR reached resting HR. The protocol was performed at the gym of the physical therapy department in the Máxima Medical Center and was supervised by a medical doctor and an assistant.

The room temperature was approximately 20°C. EE calculated by the activity trackers was noted at the start and at the end of an activity. To ensure continuous HR tracking during each activity, the workout mode was turned on in both activity trackers when an activity was started.

Table 1. Activity protocol.

Activity type and activity	Duration in minutes
Sedentary activities	
Sitting	5
Standing	2
Typing	3
Household activities	
Table cleaning	3
(Un)loading the dishwasher	3
Vacuuming	3
Stairs	
Ascending	1
Descending	1
Cycling (ergometer); load	
CAD[a] 0 Watt; HFrEF[b] 0 Watt	3
CAD 40 Watt; HFrEF 25 Watt	3
CAD 70 Watt; HFrEF 50 Watt	3
Walking (treadmill); speed/inclination	
CAD 4 km/h; HFrEF 3 km/h	3
CAD 5.5 km/h; HFrEF 4.5 km/h	3
CAD 4 km/h 5% slope; HFrEF 3 km/h 5% slope	3

[a]CAD: coronary artery disease.

[b]HFrEF: heart failure with reduced ejection fraction.

Criterion Measure

Breath-by-breath oxygen uptake (VO_2) and carbon dioxide production (VCO_2) were measured during the entire length of the protocol using the Oxycon Mobile (OM; CareFusion). The OM is a light-weighted mobile device consisting of a facemask and a gas analyzer unit with battery attached to the patients back via a shoulder belt system. Real time data measured by the device were sent to a computer with corresponding software. Before the start of the protocol, automatic volume and gas calibration was performed and ambient conditions were checked. The OM has been validated before by comparing it with the golden standard, the Douglas Bag, and has been found reliable as a criterion measure [25].

Devices

Fitbit Charge 2

The FC2 (Fitbit Inc) is a wrist-worn activity tracker consisting of a 3-axial accelerometer, an altimeter, and an optical HR tracker. EE calculation is based on a combination of basal metabolic rate (which is calculated by gender, age, height, and weight), activity counts during the activities, and, as claimed by the manufacturer on HR [26,27]. The Fitbit was worn on the dominant wrist following the recommendations of the manufacturer. The activity tracker was connected via Bluetooth to the Fitbit app in which parameters such as date of birth, length, weight, gender, wrist orientation, and handedness were entered for each patient. The app was supplied with the most recent firmware updates.

Mio Slice

The MS (MIO Global) is a wrist-worn activity tracker consisting of a 2-axial accelerometer and an optical HR tracker. The activity tracker was worn on the nondominant wrist. Age, length, weight, gender, and wrist orientation were entered in the corresponding app for every patient. Afterwards, the activity tracker was synchronized with the Mio app via Bluetooth. The app contained the latest version of firmware. Information on algorithms used to calculate EE is not provided by the manufacturer. The manufacturer claims that both activity counts and HR are used for EE calculation when the workout mode is activated [28].

Data Analysis

Raw data from the breathing analysis was exported and imported together with the values from the FC2 and MS in a custom-made MATLAB analysis program (R2018a [9.4.0.813654],

Mathworks). The entire activity bouts were analyzed. First, the EE was calculated from breath-by-breath measurements using the Weir equation, as follows [29]:

$$EE=[(3.941\times VO_2)+(1.11\times VCO_2)]\times 1.1440$$

Then, outliers (eg, coughing) in the EE data were detected using a Hampel filter. Outliers were replaced if the value exceeded 3 standard deviations from the median of itself and 3 neighboring data points of that median value [30]. Thereafter, the data were cubic spline interpolated to 1 second values following a 1-Hz low pass fourth Butterworth filter with a cutoff frequency of 0.04 Hz [31,32].

Statistical Analysis

To achieve 80% power to detect an intraclass correlation coefficient (ICC) of 0.75 (excellent agreement) under the alternative hypothesis that the ICC is 0.35 (poor agreement), a sample size of 19 subjects per study group (ie, CAD and HFrEF) was calculated.

Descriptive statistics were used to describe the population regarding baseline clinical characteristics. Normality of data was assessed by visual inspection of histograms and by interpreting Skewness and Kurtosis [33]. Accuracy of FC2 and MS was assessed by calculating mean EE and mean differences in EE compared with the criterion measure (OM). The values were calculated per activity and over the total protocol (resting time included). To identify if agreement (between the activity trackers and the criterion measure) was between reasonable limits (set at 10% error zone), one-sample T-tests were performed using mean differences (device minus OM) compared with zero. In addition, Bland-Altman plots were created to illustrate the level of agreement between estimated EE and criterion EE with mean bias and 95% upper and lower limits of agreement (LoA). Data falling outside the LoA were inspected; however, no clear reason was found why these data were different. Therefore, these data were included in the analysis. The ICC using 2-way mixed models with absolute agreement was used for assessment of reliability of the devices for each activity and total protocol. An ICC below 0.4 was considered poor, between 0.4 and 0.59 fair, between 0.6 and 0.74 good, and above 0.75 was considered as excellent reliability [34]. The root mean square error (RMSE) of FC2 and MS was calculated for the total protocol in both groups. Responsiveness of FC2 and MS was assessed by using a paired T-test during walking at different speeds and cycling at different loads. All data analyses were performed using SPSS software (Version 22.0, SPSS Inc). Significance level was set at $P<.05$ for all analyses.

Results

A total of 38 patients were included and completed the protocol. The group was equally divided in CAD patients (n=19, age 61.4 years, SD 6.9 years) and patients with HFrEF (n=19, age 65.1 years, SD 6.6 years, LVEF 31.8%, SD 7.6%). In both groups, the majority of patients were using HR lowering medication (14 CAD patients [14/19, 74%] and 17 HFrEF patients [17/19, 89%]). Pulmonary diseases were present in 2 HFrEF patients (1 patient with mild bronchiectasis and 1 patient with COPD treated by the general practitioner). Additional patient characteristics are shown in Table 2. All data recordings of 2 patients (1 CAD and 1 HFrEF) and stair walking activities of 3 additional (2 CAD and 1 HFrEF) were excluded from the analysis, because of a failure in OM measurement.

Table 2. Patient characteristics.

Characteristics	CAD[a] (N=19)	HFrEF[b] (N=19)
Age (years), mean (SD)	61.4 (6.9)	65.1 (6.6)
Gender, n (%)		
Male	14 (74)	17 (89)
Female	5 (26)	2 (11)
Height (cm), mean (SD)	176 (6.8)	177 (5.4)
Weight (kg), mean (SD)	84.3 (12.1)	86.7 (13.7)
BMI[c] (kg/m^2), mean (SD)	27.1 (3.1)	27.7 (4.2)
LVEF[d] (%), mean (SD)	60.5 (4.5)	31.8 (7.6)
NYHA[e] classification I/II/III/Unknown, n (%)	—[f]	6(32)/11(58)/1(5)/1(5)
Heart failure etiology, n (%)		
Ischemic	—	11 (58)
Nonischemic	—	8 (42)
Medication, n (%)		
Beta-blocker	12 (63)	15 (79)
Calcium channel blocker (non-DHP[g])	2 (11)	0 (0)
Amiodarone	0 (0)	4 (21)
Ivabradine	0 (0)	1 (5)

[a]CAD: coronary artery disease.

[b]HFrEF: heart failure with reduced ejection fraction.

[c]BMI: body mass index.

[d]LVEF: left ventricular ejection fraction.

[e]NYHA: New York Heart Association.

[f]Not applicable.

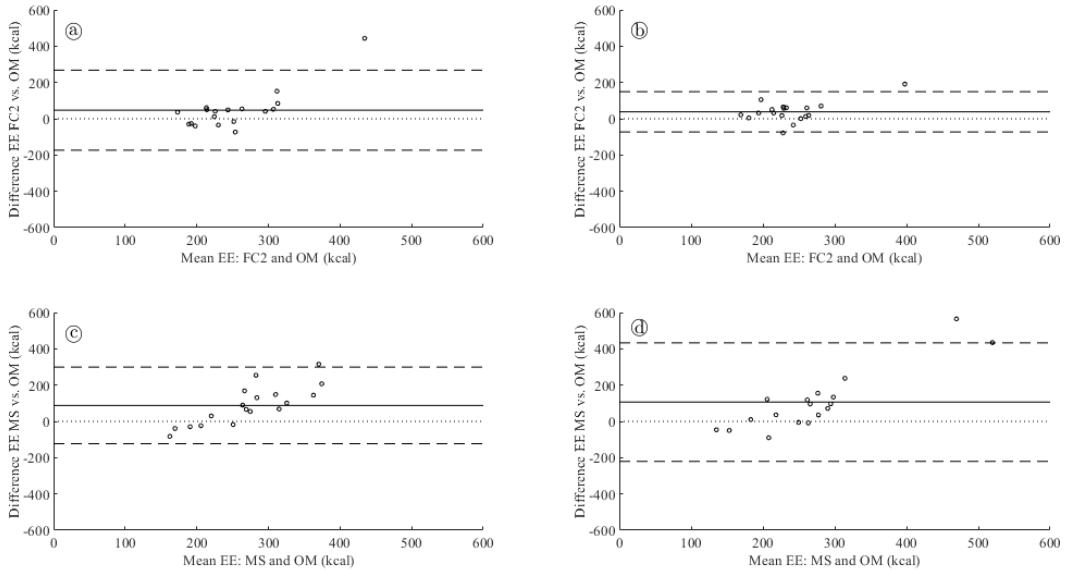
[g]DHP: dihydropyridine.

Accuracy

Coronary Artery Disease Group

The Multimedia Appendix 1 demonstrates the accuracy of EE measurement by FC2 and MS for participants with CAD. Mean (SD) EE in the CAD group over the total protocol was 228.1 (37.0) kcal, 275.6 (113.5) kcal, and 316.2 (113.3) kcal for OM, FC2, and MS, respectively. MS significantly overestimated EE over the total protocol (mean difference 88.1 kcal, *P*=.003). FC2 showed a nonsignificant overestimation in total EE (mean difference 47.5 kcal, *P*=.09). Most sedentary activities were underestimated by both FC2 and MS, and cycling activities were underestimated by FC2 at all loads. Bland-Altman plots based on total EE illustrate the overestimation with wide LoA for both devices (see Figure 1). The MS showed an increasing bias when EE levels are higher. The ICCs for the total protocol were low for both devices (FC2 0.10 and MS 0.12). The RMSE was 119.2 kcal and 136.7 kcal for FC2 and MS, respectively.

Figure 1. Bland-Altman plots for total energy expenditure. The solid horizontal line corresponds to the mean difference, whereas the dashed horizontal lines correspond to limits of agreement. The dotted line is the line of equality. (a) Comparison of Oxycon Mobile with Fitbit Charge 2 for patients with CAD. (b) Comparison of Oxycon Mobile with Fitbit charge 2 for patients with HFrEF. (c) Comparison of Oxycon Mobile with Mio Slice for patients with CAD. (d) Comparison of Oxycon Mobile with Mio Slice for patients with HFrEF. CAD: coronary artery disease; EE: energy expenditure; FC2: Fitbit Charge 2; HFrEF: heart failure with reduced ejection fraction; MS: Mio Slice; OM: Oxycon Mobile.



Heart Failure With Reduced Ejection Fraction Group

The Multimedia Appendix 2 demonstrates the accuracy of EE measurement by FC2 and MS for participants with HFrEF. Mean (SD) EE over the total protocol in the HFrEF group was 218.2 (42.3) kcal, 256.4 (69.3) kcal, and 324.4 (174.6) kcal for OM, FC2, and MS, respectively. Both devices significantly overestimated EE (mean difference FC2 38.2 kcal, MS 106.2 kcal, *P*=.01 and *P*=.02, respectively) with a similar pattern of underestimation in sedentary activities and cycling activities. Bland-Altman plots based on total EE illustrate the overestimation with wide LoA for both devices (see Figure 1). The MS showed wider LoA (lower LoA −220.3 kcal and upper LoA 432.7 kcal) and an increasing bias when EE levels were higher. The ICCs for the total protocol were low for both devices (FC2 0.42 and MS 0.11). The RMSE was 66.9 kcal and 193.6 kcal for FC2 and MS, respectively.

Responsiveness

Table 3 shows the ability of FC2 and MS to detect within patient changes in walking and cycling activities.

Coronary Artery Disease Group

FC2 was able to detect a difference between cycling at 0 versus 40 watts (mean difference 3.3 kcal, *P*=.003) and between walking at 4 km/h with a 5% slope versus 5.5 km/h (mean difference 4.4 kcal, *P*=.01) in patients with CAD. However, no significant differences were observed for the other walking and cycling activities. The MS was able to detect differences at all cycling loads; however, it was not able to detect any differences in walking speeds/inclination. Note that the difference in EE between walking 4 km/h with a 5% slope and 5.5 km/h was nonsignificant as measured by the OM.

Heart Failure With Reduced Ejection Fraction Group

FC2 was not able to detect changes at any walking speeds or cycling loads in the HFrEF group. MS was able to detect within-patient changes at cycling 0 versus 50 watts (mean difference 4.7 kcal, *P*=.02), at cycling 25 versus 50 watts (mean difference 3.6 kcal, *P*=.02) and at walking 3 km/h with a 5% slope versus 4.5 km/h (mean difference 3.0 kcal, *P*=.03). Note that the difference in EE between walking 3 km/h with a 5% slope and 4.5 km/h was nonsignificant as measured by the OM.

Table 3. Responsiveness of Fitbit Charge 2 and Mio Slice.

Group and activity	Oxycon Mobile, Mean difference (kcal[a])	*P* value	Fitbit Charge 2, Mean difference (kcal)	*P* value	Mio Slice, Mean difference (kcal)	*P* value
CAD[b] (N=18)						
Cycling						
0 versus 40 watts	2.7	<.001	3.3	.003	3.2	<.001
0 versus 70 watts	5.4	<.001	2.6	.11	5.1	<.001
40 versus 70 watts	2.7	<.001	0.7	.67	1.9	.03
Walking						
4 km/h versus 4 km/h 5% slope	2.5	<.001	1.8	.15	2.2	.36
4 km/h versus 5.5 km/h	2.5	<.001	2.6	.15	3.8	.31
4 km/h 5% slope versus 5.5 km/h	0.1	.71	4.4	.01	1.6	.42
HFrEF[c] (N=18)						
Cycling						
0 versus 25 watts	1.2	<.001	0.3	.88	1.1	.23
0 versus 50 watts	3.0	<.001	1.0	.46	4.7	.02
25 versus 50 watts	1.9	<.001	1.3	.40	3.6	.02
Walking						
3 km/h versus 3 km/h 5% slope	1.1	.002	1.9	.16	.8	.66
3 km/h versus 4.5 km/h	1.5	.001	.3	.89	2.2	.16
3 km/h 5% slope versus 4.5 km/h	0.4	.27	2.2	.08	3.0	.03

[a]kcal: kilocalories.

[b]CAD: coronary artery disease.

[c]HFrEF: heart failure with reduced ejection fraction.

Discussion

This study is the first to evaluate the accuracy and responsiveness of wrist-worn activity trackers combining HR monitoring and accelerometry for EE calculation in patients with CAD with preserved LVEF and patients with HFrEF. Poor accuracy was observed for both devices in predicting EE, with MS performing worse than FC2. MS provides a higher responsiveness than FC2 with regard to the ability to detect changes in cycling load, but both devices performed poorly with respect to detecting within-patient differences in walking speed.

Accuracy

Both FC2 and MS significantly overestimated EE over the total activity protocol with a tendency of greater bias when EE increased. Other studies using wrist-worn Fitbit models that combine accelerometer data and heart data showed mixed results. The results from our study were in line with previous research from Bai et al which showed a whole-trial overestimation of EE (mean absolute percentage error 32.9%) of FC HR [23]. Regarding exercise intensity, the device showed an underestimation of sedentary activities and overestimation of light physical activities and aerobic activities, similar to our results. Dooley et al also found significant overestimation of EE at baseline, light, and moderate intensity treadmill activities by FC HR [35]. In contrast, other studies showed an underestimation of EE by wrist-worn devices [21,22]. Differences in study outcomes might be due to variation in study design and population. The above-mentioned studies have been performed in noncardiac participants, so results cannot be extrapolated to cardiac patients using beta-blocker medication and chronotropic incompetence. To our knowledge, only 1 study evaluated the accuracy of a wrist-worn Fitbit device (Fitbit Flex, 2013) in a population in which 59% had coronary heart disease [36]. This study showed a 10% overestimation of minutes of moderate to vigorous physical activity (MVPA), which was in line with our study. Although they found a high correlation between Fitbit and the criterion measure for minutes of MVPA (r=0.74 total population, r=0.71 cardiac patients), ICC's and Bland-Altman analysis were not performed to evaluate accuracy.

Possible causes for the limited accuracy of EE estimation by different wrist-worn devices include poor quality of HR and accelerometer assessment and inadequate algorithms to calculate EE (ie, not well tailored to the target population). Yet, the algorithms to calculate EE are usually not provided by the manufacturer. As patient characteristics such as length, weight,

and exercise modality are fixed, EE estimates are most likely determined by the accuracy of the (accelerometer and HR) sensors and the reliability of the algorithms related to HR and activity counts. However, both the reliability of the algorithms and the accuracy of both sensors were not evaluated in this study. Concerning the accuracy of the HR sensor, Wallen et al concluded that Mio Alfa and FC HR slightly underestimate HR, within an expectable range (ICC Fitbit 0.78; ICC Mio 0.91) [21]. Other studies found similar results for Fitbit Surge, FC HR, and Mio Alfa [24,37]. In total, 2 studies found that the error of Mio Alfa was comparable with the reference method; however, considerable variability was observed when dividing protocol activities into different types, and the device did not perform well as speed or load of an activity increased [38,39]. Despite reasonable accuracy of HR assessment by previous Fitbit and Mio models, it remains unclear how HR is incorporated in EE equations. Moreover, it is questionable if the relation between changes in HR and EE is comparable between people who do not have a cardiac condition and patients with CAD or HFrEF. Therefore, EE estimates might be even more inaccurate in our study compared with these studies with individuals without cardiac conditions.

Another factor that may have influenced the accuracy of EE estimation is the location of the accelerometer. Waist placement is generally considered favorable as the sensor is close to the center of body mass and is able to detect whole body movement. A recent review evaluating the influence of body placement to accuracy of EE estimation concluded that wrist placement generally leads to overestimation and torso placement to underestimation of EE, with a greater mean error for devices placed at the wrist [40]. Hand placement on the bars of the treadmill might also contribute to the devices not being able to provide an accurate EE measurement in our study. Wrist-based wearable devices might have difficulties with detecting activities when specific characteristics of an activity (swing of the arm during walking) are not detected. Nonetheless, our findings for indoor conditions are valuable as patients will also use the treadmill during rehabilitation, leisure time sports activities, or even at home.

Nevertheless, our study clearly showed that EE estimates, using algorithms for commercially available wrist-worn devices, should be interpreted with caution in cardiac patients. Therefore, to improve the utility of these devices for this population, extraction of raw HR and accelerometer data are needed to be able to develop adequately tailored algorithms. However, most manufacturers of activity trackers do not provide this opportunity.

Responsiveness

Whereas, MS was shown to be useful for detecting changes in cycling load, changes in walking speed and inclination were not detected. Furthermore, FC2 was not capable in detecting changes in both walking and cycling activities. Although the ability to detect changes in intensity within specific activities is an important feature of an activity tracker, previous research on responsiveness is scarce. Price et al concluded the hip-worn Fitbit One is able to detect gross changes in walking and running speed [41]. Compared with Price et al [41], we tested responsiveness for smaller differences in walking speed, which may explain the difference with our findings. In addition, Gusmer et al showed the hip-worn Fitbit Ultra was able to detect changes in slow and brisk walking, where slow walking was defined as minus 10% of a self-selected comfortable speed and brisk walking as plus 10% of the self-selected speed [42]. Personalized changes in gait speed and device placement might contribute to the detection of differences in walking speed. To our knowledge, there are no studies available which evaluate responsiveness during cycling, which is relevant because cycling is a major component of Dutch CR and during daily life activities [43]. Given the fact that daily exercise behavior consists of a combination of walking, cycling, and other physical activities, our results show that EE measured by both devices should not be used for monitoring purposes and exercise prescription in patients with CHF and CAD. However, wearable devices have demonstrated promising results in motivating and engaging PA and exercise behavior [44].

Strengths and Limitations

This study is the first to evaluate both the accuracy and responsiveness of wrist-worn activity trackers, which combine HR and accelerometer data in a cardiac population. The responsiveness of a device is a very important feature when implemented in practice, such as in cardiac telerehabilitation. The study is limited by not evaluating test-retest reliability. This would have given a more complete overview of the overall device validity. Moreover, patients were tested in a laboratory setting, so it is not sure whether these results can be extrapolated to free-living conditions. However, because we mimicked free-living conditions by creating a protocol consisting of daily life activities, we expect little differences with a free-living validation study. Another limitation is that HR boundaries were not personalized for each patient. As we did not assess the maximum HR for each individual patient, default settings of the activity trackers were used, which could have influenced the calculation of EE.

Conclusions

Both wrist-worn activity trackers demonstrated low accuracy in estimation of EE in patients with CAD and HFrEF. Importantly, both devices also showed poor performance to detect within-patient changes in the low-to-moderate exercise intensity domain. Notwithstanding the fact that the use of activity trackers in cardiac patients might stimulate daily exercise behavior, these findings highlight the need for population-specific devices and algorithms.

Authors' Contributions

CH, JJK, and HMCK contributed to the conception and design of the study. CH and JJK contributed to the acquisition of data. CH, EMAvL, JJK, and MvH contributed to the analysis of data. All authors contributed to interpretation of data. CH drafted the

manuscript. All authors critically revised the manuscript. All gave final approval and agreed to be accountable for all aspects of work ensuring integrity and accuracy.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Accuracy of energy expenditure measurement by Fitbit Charge 2 and Mio Slice, for participants with coronary artery disease.
[PDF File (Adobe PDF File), 202 KB - mhealth_v7i12e15045_app1.pdf]

Multimedia Appendix 2
Accuracy of energy expenditure measurement by Fitbit Charge 2 and Mio Slice, for participants with heart failure with reduced ejection fraction.
[PDF File (Adobe PDF File), 190 KB - mhealth_v7i12e15045_app2.pdf]

References

1. Ades PA. Cardiac rehabilitation and secondary prevention of coronary heart disease. N Engl J Med 2001 Sep 20;345(12):892-902. [doi: 10.1056/NEJMra001529] [Medline: 11565523]
2. Piepoli MF, Corrà U, Adamopoulos S, Benzer W, Bjarnason-Wehrens B, Cupples M, et al. Secondary prevention in the clinical management of patients with cardiovascular diseases. Core components, standards and outcome measures for referral and delivery: a policy statement from the cardiac rehabilitation section of the European Association for Cardiovascular Prevention & Rehabilitation. Endorsed by the Committee for Practice Guidelines of the European Society of Cardiology. Eur J Prev Cardiol 2014 Jun;21(6):664-681. [doi: 10.1177/2047487312449597] [Medline: 22718797]
3. Anderson L, Oldridge N, Thompson DR, Zwisler A, 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]
4. Dalal HM, Taylor RS, Jolly K, Davis RC, Doherty P, Miles J, et al. The effects and costs of home-based rehabilitation for heart failure with reduced ejection fraction: the REACH-HF multicentre randomized controlled trial. Eur J Prev Cardiol 2019 Feb;26(3):262-272 [FREE Full text] [doi: 10.1177/2047487318806358] [Medline: 30304644]
5. ter Hoeve N, Sunamura M, Stam HJ, Boersma E, Geleijnse ML, van Domburg RT, et al. Effects of two behavioral cardiac rehabilitation interventions on physical activity: a randomized controlled trial. Int J Cardiol 2018 Mar 15;255:221-228 [FREE Full text] [doi: 10.1016/j.ijcard.2017.12.015] [Medline: 29425564]
6. van den Berg-Emons RJ, Bussmann JB, Balk AH, Stam HJ. Factors associated with the level of movement-related everyday activity and quality of life in people with chronic heart failure. Phys Ther 2005 Dec;85(12):1340-1348. [Medline: 16305272]
7. van den Berg-Emons HR, Bussmann JH, Balk A, Keijzer-Oster D, Stam H. Level of activities associated with mobility during everyday life in patients with chronic congestive heart failure as measured with an 'Activity monitor'. Phys Ther 2001 Sep 1;81(9):1502-1511. [doi: 10.1093/ptj/81.9.1502] [Medline: 28206626]
8. Walsh JT, Charlesworth A, Andrews R, Hawkins M, Cowley AJ. Relation of daily activity levels in patients with chronic heart failure to long-term prognosis. Am J Cardiol 1997 May 15;79(10):1364-1369. [doi: 10.1016/s0002-9149(97)00141-0] [Medline: 9165159]
9. Prince SA, Blanchard CM, Grace SL, Reid RD. Objectively-measured sedentary time and its association with markers of cardiometabolic health and fitness among cardiac rehabilitation graduates. Eur J Prev Cardiol 2016 May;23(8):818-825. [doi: 10.1177/2047487315617101] [Medline: 26607698]
10. Hu GC, Chuang YC, Chien KL. Clinical efficacy and safety of exercise training among chronic heart failure patients. J Intern Med Taiwan 2010 Jun;21(3):165-171 [FREE Full text]
11. ter Hoeve HN, Huisstede BM, Stam HJ, van Domburg RT, Sunamura M, van den Berg-Emons RJ. Does cardiac rehabilitation after an acute cardiac syndrome lead to changes in physical activity habits? Systematic review. Phys Ther 2015 Feb;95(2):167-179. [doi: 10.2522/ptj.20130509] [Medline: 25278337]
12. Smith KM, Arthur HM, McKelvie RS, Kodis J. Differences in sustainability of exercise and health-related quality of life outcomes following home or hospital-based cardiac rehabilitation. Eur J Cardiovasc Prev Rehabil 2004 Aug;11(4):313-319. [doi: 10.1097/01.hjr.0000136414.40017.10] [Medline: 15292765]
13. Smith KM, McKelvie RS, Thorpe KE, Arthur HM. Six-year follow-up of a randomised controlled trial examining hospital versus home-based exercise training after coronary artery bypass graft surgery. Heart 2011 Jul;97(14):1169-1174. [doi: 10.1136/hrt.2010.202036] [Medline: 21561899]
14. Frederix I, Vanhees L, Dendale P, Goetschalckx K. A review of telerehabilitation for cardiac patients. J Telemed Telecare 2015 Jan;21(1):45-53. [doi: 10.1177/1357633X14562732] [Medline: 25475219]

15. Frederix I, Solmi F, Piepoli MF, Dendale P. Cardiac telerehabilitation: a novel cost-efficient care delivery strategy that can induce long-term health benefits. Eur J Prev Cardiol 2017 Nov;24(16):1708-1717. [doi: 10.1177/2047487317732274] [Medline: 28925749]
16. European Society of Cardiology. 2013. Position on the Recast of the European Union Medical Device Directives. URL: https://www.escardio.org/static_file/Escardio/EU-Affairs/ESC-response-EU-draft-regulation.pdf [accessed 2019-11-08]
17. Menditto A, Patriarca M, Magnusson B. Understanding the meaning of accuracy, trueness and precision. Accred Qual Assur 2006;12(1):45-47. [doi: 10.1007/s00769-006-0191-z]
18. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol 2010 Jul;63(7):737-745. [doi: 10.1016/j.jclinepi.2010.02.006] [Medline: 20494804]
19. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Responsiveness. In: Measurement in Medicine. Cambridge: Cambridge University Press; 2011:202-226.
20. Kraal JJ, Sartor F, Papini G, Stut W, Peek N, Kemps HM, et al. Energy expenditure estimation in beta-blocker-medicated cardiac patients by combining heart rate and body movement data. Eur J Prev Cardiol 2016 Nov;23(16):1734-1742. [doi: 10.1177/2047487316667786] [Medline: 27625154]
21. Wallen MP, Gomersall SR, Keating SE, Wisløff U, Coombes JS. Accuracy of heart rate watches: implications for weight management. PLoS One 2016;11(5):e0154420 [FREE Full text] [doi: 10.1371/journal.pone.0154420] [Medline: 27232714]
22. Chowdhury EA, Western MJ, Nightingale TE, Peacock OJ, Thompson D. Assessment of laboratory and daily energy expenditure estimates from consumer multi-sensor physical activity monitors. PLoS One 2017;12(2):e0171720 [FREE Full text] [doi: 10.1371/journal.pone.0171720] [Medline: 28234979]
23. Bai Y, Hibbing P, Mantis C, Welk GJ. Comparative evaluation of heart rate-based monitors: Apple Watch vs Fitbit Charge HR. J Sports Sci 2018 Aug;36(15):1734-1741. [doi: 10.1080/02640414.2017.1412235] [Medline: 29210326]
24. Shcherbina A, Mattsson CM, Waggott D, Salisbury H, Christle JW, Hastie T, et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J Pers Med 2017 May 24;7(2):pii: E3 [FREE Full text] [doi: 10.3390/jpm7020003] [Medline: 28538708]
25. Rosdahl H, Gullstrand L, Salier-Eriksson J, Johansson P, Schantz P. Evaluation of the Oxycon Mobile metabolic system against the Douglas bag method. Eur J Appl Physiol 2010 May;109(2):159-171. [doi: 10.1007/s00421-009-1326-9] [Medline: 20043228]
26. Fitbit Inc. 2019. How does Fitbit estimate how many calories I've burned? URL: https://help.fitbit.com/articles/en_US/Help_article/1381 [accessed 2019-11-08]
27. Fitbit Inc. 2019. What are active minutes? URL: https://help.fitbit.com/articles/en_US/Help_article/1379/?q=active+minutes&l=en_US&fs=Search&pn=1 [accessed 2019-11-08]
28. Amazon. 2015. Forum: Can anyone explain total calories vs active calories from the Mio Go app? URL: https://www.amazon.com/ask/questions/Tx1IIQOP5QJE14E/1 [accessed 2019-11-08]
29. Weir JB. New methods for calculating metabolic rate with special reference to protein metabolism. J Physiol 1949 Aug;109(1-2):1-9 [FREE Full text] [doi: 10.1113/jphysiol.1949.sp004363] [Medline: 15394301]
30. Pearson RK. Outliers in process modeling and identification. IEEE Trans Contr Syst Technol 2002;10(1):55-63. [doi: 10.1109/87.974338]
31. Robergs RA, Burnet A. Methods used to process data from indirect calirometry and their application to VO2max. J Exerc Physiol 2003;6(2):44-57 [FREE Full text]
32. Robergs RA, Dwyer D, Astorino T. Recommendations for improved data processing from expired gas analysis indirect calorimetry. Sports Med 2010 Feb 1;40(2):95-111. [doi: 10.2165/11319670-000000000-00000] [Medline: 20092364]
33. Kim H. Statistical notes for clinical researchers: assessing normal distribution (2) using skewness and kurtosis. Restor Dent Endod 2013 Feb;38(1):52-54 [FREE Full text] [doi: 10.5395/rde.2013.38.1.52] [Medline: 23495371]
34. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess 1994;6(4):284-290. [doi: 10.1037/1040-3590.6.4.284]
35. Dooley EE, Golaszewski NM, Bartholomew JB. Estimating accuracy at exercise intensities: a comparative study of self-monitoring heart rate and physical activity wearable devices. JMIR Mhealth Uhealth 2017 Mar 16;5(3):e34 [FREE Full text] [doi: 10.2196/mhealth.7043] [Medline: 28302596]
36. Alharbi M, Bauman A, Neubeck L, Gallagher R. Validation of Fitbit-Flex as a measure of free-living physical activity in a community-based phase III cardiac rehabilitation population. Eur J Prev Cardiol 2016 Sep;23(14):1476-1485. [doi: 10.1177/2047487316634883] [Medline: 26907794]
37. Stahl SE, An H, Dinkel DM, Noble JM, Lee J. How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? BMJ Open Sport Exerc Med 2016;2(1):e000106 [FREE Full text] [doi: 10.1136/bmjsem-2015-000106] [Medline: 27900173]
38. Spierer DK, Rosen Z, Litman LL, Fujii K. Validation of photoplethysmography as a method to detect heart rate during rest and exercise. J Med Eng Technol 2015;39(5):264-271. [doi: 10.3109/03091902.2015.1047536] [Medline: 26112379]
39. Parak J, Korhonen I. Evaluation of wearable consumer heart rate monitors based on photopletysmography. Conf Proc IEEE Eng Med Biol Soc 2014;2014:3670-3673. [doi: 10.1109/EMBC.2014.6944419] [Medline: 25570787]

40. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, et al. Accuracy of Fitbit devices: systematic review and narrative syntheses of quantitative data. JMIR Mhealth Uhealth 2018 Aug 9;6(8):e10527 [FREE Full text] [doi: 10.2196/10527] [Medline: 30093371]
41. Price K, Bird SR, Lythgo N, Raj IS, Wong JY, Lynch C. Validation of the Fitbit One, Garmin Vivofit and Jawbone UP activity tracker in estimation of energy expenditure during treadmill walking and running. J Med Eng Technol 2017 Apr;41(3):208-215. [doi: 10.1080/03091902.2016.1253795] [Medline: 27919170]
42. Gusmer RJ, Bosch TA, Watkins A, Ostrem J, Dengel D. Comparison of FitBit Ultra to ActiGraph GT1M for assessment of physical activity in young adults during treadmill walking. Open Sport Med J 2014;8(1):11-15. [doi: 10.2174/1874387001408010011]
43. Vromen T, Spee RF, Kraal JJ, Peek N, van Engen-Verheul MM, Kraaijenhagen RA, et al. Exercise training programs in Dutch cardiac rehabilitation centres. Neth Heart J 2013 Mar;21(3):138-143 [FREE Full text] [doi: 10.1007/s12471-013-0374-2] [Medline: 23338553]
44. Lewis ZH, Lyons EJ, Jarvis JM, Baillargeon J. Using an electronic activity monitor system as an intervention modality: a systematic review. BMC Public Health 2015 Jun 24;15:585 [FREE Full text] [doi: 10.1186/s12889-015-1947-3] [Medline: 26104189]

Abbreviations

CAD: coronary artery disease
CHF: chronic heart failure
CR: cardiac rehabilitation
EE: energy expenditure
FC2: Fitbit Charge 2
HFrEF: heart failure with reduced ejection fraction
HR: heart rate
ICC: intraclass correlation coefficient
LoA: limits of agreement
LVEF: left ventricular ejection fraction
MS: Mio Slice
OM: Oxycon Mobile
PA: physical activity
RMSE: root mean square error

Edited by G Eysenbach; submitted 15.06.19; peer-reviewed by K Ng, V Cornelissen; comments to author 01.08.19; revised version received 22.08.19; accepted 24.09.19; published 19.12.19.

Please cite as:
Herkert C, Kraal JJ, van Loon EMA, van Hooff M, Kemps HMC
Usefulness of Modern Activity Trackers for Monitoring Exercise Behavior in Chronic Cardiac Patients: Validation Study
JMIR Mhealth Uhealth 2019;7(12):e15045
URL: http://mhealth.jmir.org/2019/12/e15045/
doi:10.2196/15045
PMID:31855191

Original Paper

Improving Pacific Adolescents' Physical Activity Toward International Recommendations: Exploratory Study of a Digital Education App Coupled With Activity Trackers

Olivier Galy[1*], PhD; Kalina Yacef[2,3*], PhD; Corinne Caillaud[3,4*], PhD

[1]Interdisciplinary Laboratory for Research in Education, EA 7483, School of Education, The University of New Caledonia, Noumea, New Caledonia
[2]School of Computer Science, The University of Sydney, Sydney, Australia
[3]Charles Perkins Centre, The University of Sydney, Sydney, Australia
[4]Faculty of Health Sciences, The University of Sydney, Sydney, Australia
[*]all authors contributed equally

Corresponding Author:
Olivier Galy, PhD
Interdisciplinary Laboratory for Research in Education, EA 7483
School of Education
The University of New Caledonia
Campus de Nouville
Noumea
New Caledonia
Phone: 687 815 602
Email: olivier.galy@unc.nc

Abstract

Background: The prevalence of overweight and obesity in children and adolescents has dramatically increased in the Pacific Island countries and territories over the last decade. Childhood overweight and obesity not only have short-term consequences but are also likely to lead to noncommunicable diseases in adulthood. A major factor contributing to the rising prevalence is an insufficient amount of daily moderate-to-vigorous physical activity (MVPA). In the Pacific region, less than 50% of children and adolescents meet the international recommendations of 11,000 steps and 60 min of MVPA per day. Although studies have shown the potential of digital technologies to change behaviors, none has been proposed to guide adolescents toward achieving these recommendations.

Objective: The aims of this study were (1) to investigate whether a technology-based educational program that combines education, objective measures of physical activity (PA), and self-assessment of goal achievement would be well received by Pacific adolescents and help change their PA behaviors toward the international PA recommendations and (2) to create more insightful data analysis methods to better understand PA behavior change.

Methods: A total of 24 adolescents, aged 12 to 14 years, participated in a 4-week program comprising 8 1-hour modules designed to develop health literacy and physical skills. This self-paced user-centered program was delivered via an app and provided health-related learning content as well as goal setting and self-assessment tasks. PA performed during the 4-week program was captured by an activity tracker to support learning and help the adolescents self-assess their achievements against personal goals. The data were analyzed using a consistency rate and daily behavior clustering to reveal any PA changes, particularly regarding adherence to international recommendations.

Results: The consistency rate of daily steps revealed that the adolescents reached 11,000 steps per day 48% (approximately 3.4 days per week) of the time in the first week of the program, and this peaked at 59% (approximately 4.1 days per week) toward the end of the program. PA data showed an overall increase during the program, particularly in the less active adolescents, who increased their daily steps by 15% and ultimately reached 11,000 steps more frequently. The consistency of daily behavior clustering showed a 27% increase in adherence to international recommendations in the least active adolescents.

Conclusions: Technology-supported educational programs that include self-monitored PA via activity trackers can be successfully delivered to adolescents in schools in remote Pacific areas. New data mining techniques enable innovative analyses of PA engagement based on the international recommendations.

(JMIR Mhealth Uhealth 2019;7(12):e14854) doi:10.2196/14854

KEYWORDS

exercise; eHealth; adolescents; health education; noncommunicable diseases; iEngage; data mining; movement; food; Melanesia

Introduction

Background

The populations of the Pacific Island countries and territories (PICTs) have undergone a rapid lifestyle transition, with impacts on health. Despite a history of colonization, tribal and rural living has persisted, but today's urbanization, industrialization, and mechanization are causing many transformations [1], notably a decrease in physical activity (PA) and dietary changes because of imported and processed foods [2]. These phenomena have affected the main PICT communities (Melanesian, Polynesian, Asian, and European) [3,4], resulting in an alarming increase in overweight and obesity in children and adolescents [5] that often persists into adulthood [6] and leads to noncommunicable diseases [7]. Early intervention is, therefore, crucial to reverse these trends, and both PA—especially moderate-to-vigorous physical activity (MVPA)—and healthy dietary behaviors in childhood are notably associated with a healthy adult lifestyle [3]. Research indicates that less than 50% of children aged between 13 and 15 years living in the Pacific meet the international recommendations of 11,000 steps and 60 min of MVPA per day [2]. In New Caledonia, 35% of children aged between 11 and 16 years are overweight or obese [3,4], and 45% and 20% of these adolescents living in rural and urban regions, respectively, meet the recommendations [8]. Although these data are based on self-reports, which are known to have a large margin of error when estimating PA levels, they underline the need for action to encourage children and adolescents to become more physically active [9]. This is a challenge for the PICTs and indeed many countries worldwide.

Data are limited for the PICTs, but a recent Australian study found that initiatives to increase adolescent MVPA have been unsuccessful [10]. The low MVPA was attributed in part to perceived lack of support, poor motivation, and low physical competence [11]. Important drivers for children's and adolescents' engagement in PA include health knowledge, personally organized PA, and competence in diverse PA types [12]. Systematic reviews of Western studies have concluded that future initiatives need to focus on school interventions involving families or communities and should embed multiple components, such as the guided use of technology [13,14].

Furthermore, digital interventions aiming to increase PA have shown promising results, particularly when they involve self-monitoring and feedback via activity trackers [15]. However, the optimal choice and appropriate use of the trackers is critical; Kerner and Goodyear [16] indicated that a bracelet and app designed for adults but used for adolescents without educational support conveyed inappropriate messages (eg, *be fit or be fat*), resulting in demotivation and negative feelings.

We explored a technology-supported educational program focused on PA in a Pacific Island rural school environment. Our program (called iEngage) was designed to improve adolescents' PA knowledge and skills and to help them understand their activity tracker data, which provided objective feedback on their PA. Data mining enabled us to develop a comprehensive approach to analyzing PA trends throughout the program. First, clustering took into account all the components of the international recommendations (number of steps, time spent at each intensity, and days over 11,000 steps). Second, a sliding 7-day window captured the frequency and regularity of the PA behaviors.

Objectives

The first aim was to investigate whether a technology-based educational program that combines education, objective measures of PA, and self-assessment of goal achievement would be well received by Pacific adolescents and help change their PA behaviors toward the international PA recommendations. The second aim was to create more insightful data analysis methods to better understand PA behavior change.

Methods

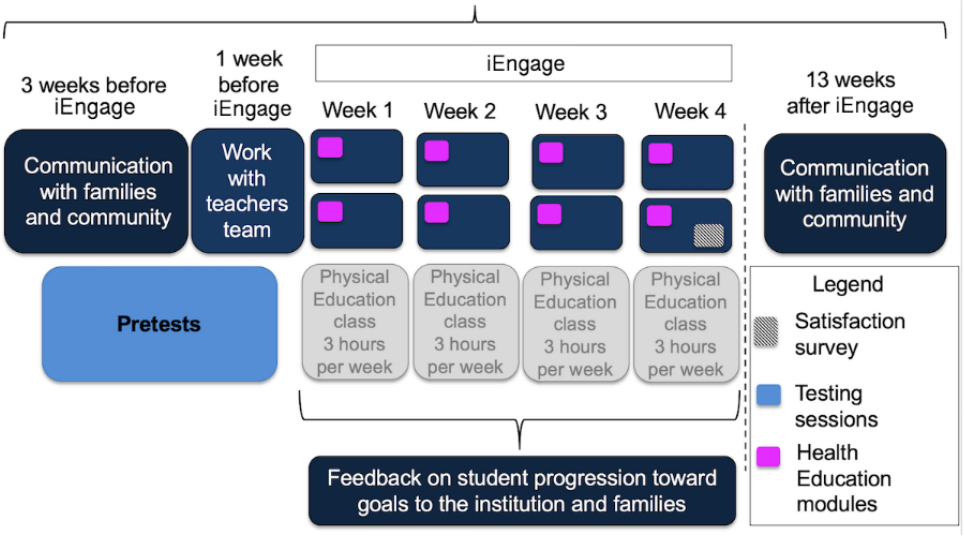
Study Design and Participants

This exploratory study was conducted in a rural school in New Caledonia on Lifou Island. A total of 24 adolescents aged 12 to 14 years participated in this pilot. Parents gave informed written consent before their child's participation in the study, which met all legal requirements and the criteria of the Declaration of Helsinki. The protocol was approved by the ethics committee of the University of New Caledonia and the consultative ethics committee of New Caledonia (CEC-NC03-2016).

Our study design is illustrated in Figure 1. Before the study, we communicated extensively with the school principal, the teaching team, and the wider community (including families) to explain the framework of the iEngage educational program.

Moreover, 3 weeks before the program, anthropometric and physical fitness data including aerobic capacity, speed, and agility were assessed during scheduled physical education classes as described below. Once anthropometry and physical fitness testing were completed, participants were equipped with research-grade activity sensors (GENEActiv) for 5 consecutive school days before the iEngage program to measure baseline PA behavior. They then started the iEngage program, which lasted for 4 weeks, with 2 1-hour modules per week. Throughout the program, they continuously wore a commercial activity tracker (Misfit Shine 2, United States), which served as the educational support tool. The 2 activity trackers served very different purposes and collected PA data in different formats. To measure baseline PA, we needed a device meeting the research-grade criterion, and to support learning during the program, we needed a device that was commercially and easily available for future wide-scale deployment.

Figure 1. Study design. Pretests and communication with families began 3 weeks before the start of the educational program; coordination with the teaching team was finalized 1 week prior, along with baseline measures. The digital program lasted for 4 weeks, with biweekly health education modules held in class (in pink). Feedback to families about the program occurred in the week following the end of the program.



iEngage: A Technology-Enabled Educational Program

The iEngage program targets health literacy, PA-related skills, and sugar-focused nutrition guidelines. Health literacy topics include the definition of PA, classifications of PA intensities (light, moderate, and vigorous), physical fitness parameters and their interpretation, rating one's own effort during exercise, sedentary behaviors and their effects on health, health definition, well-being (physical, mental, and social), sugar as a source of energy, and guidelines related to sugar consumption. iEngage v1.0 was delivered via an app (powered by BePatient) comprising 8 1-hour modules; it was self-paced and had learning activities as well as goal setting and self-assessment tasks. Each module proposed learning activities, quizzes, and brief 2- to 5-min PA sessions that generally focused on a particular series of movements: sprints, jumping, walking, running, squats, sit-and-reach, or push-ups. The total PA per adolescent over the 8 modules was 21 min; thus, these prescribed PA sessions had no significant direct impact on the MVPA levels and could not have contributed to an artificial increase in MVPA. PA was captured by the Misfit Shine 2 activity tracker and conveyed to the child via the Misfit app exclusively during the iEngage modules to support learning and provide help in self-assessing achievement against personal goals. Figure 2 presents screenshots of the app and pictures of children in activity during the program.

At the end of each module, the adolescents were guided to set their own goals for the next module. These were 2-fold: first, they chose an individual "objective" in terms of number of steps and intended intensity (eg, "I plan to do 10,000 steps and 45 min of MVPA every day"), and then, they selected a "mission," which was a specific task usually involving family, friends, or health professionals (eg, "I will discuss what I've learned today with my family" or "I will play soccer with friends"). At the start of the following module, they indicated whether they had achieved their goals. In the last module, they were encouraged to set long-term goals.

A game challenge was designed to foster engagement in learning activities and goal achievement. The adolescents were assigned to 5 groups of 3 or 4, represented by an animal mascot. Groups won points when members achieved individual goals (objectives and missions) and for collective learning achievements during the modules; the results were tallied at the end of each week.

Figure 2. App screenshot and pictures of children in activity during modules.

Procedures

Preprogram Measurements

Anthropometry

Height was measured to the nearest 0.5 cm using a portable stadiometer (Leicester Tanita HR 001, Tanita Corporation, Tokyo, Japan). Body weight was determined using a scale (Tanita HA 503, Tanita Corporation, Tokyo, Japan) to the nearest 0.1 kg, with the adolescents in light clothing. Body mass index (BMI) was calculated by dividing mass in kilograms by height squared in meters. The BMI z score was then calculated. The percentage of fat body mass (FBM) was estimated from the skinfold thickness of the sum of 4 skin areas (biceps, triceps, subscapular, and supra-iliac) measured on the right side of the body with Harpenden skinfold calipers, expressed in millimeters. This enabled us to determine the FBM and lean body mass) in kg. The detailed anthropometric methods are described in the study by Galy et al [17].

Physical Fitness Tests

Physical fitness tests included a time trial of 30-m sprints, with 5 m, 10 m, and 15 m lap times recorded using photocell gates (Brower Timing Systems, Salt Lake City, UT, United States; accuracy of 0.01 sec) placed 1 m above the ground. The *t* test was used to determine the adolescents' agility. Trials were recorded using photocell gates in the same conditions as for the sprints. The maximal aerobic speed (MAS) test individually assessed running speed over gradually accelerating 1-min increments. When a child stopped the test at maximal effort, the last stage reached was recorded and converted to MAS as described in the study by Galy et al [17].

Baseline Physical Activity (GENEActiv)

Baseline PA was obtained via GENEActiv activity trackers (validated in children [18]; Kimbolton, Cambs, PE28 0LF) positioned on the nondominant wrist for 5 days before the program. The datasets from these trackers contained 60-Hz 3-dimensional accelerometer data. Raw data were processed into 1-sec epoch signal vector magnitude datapoints of daytime activity (8 am to 8 pm daily) and were then categorized into PA levels: sedentary, light PA, and MVPA, for each second using Phillips's cut points for children [19]. We focused on bouts of sedentary activity or MVPA occurring with a minimum duration [18], as these behaviors were targeted by the program. A bout was defined as a continuous episode of PA at a specific range of intensity, and the length of a bout was the number of seconds spent at that intensity during that episode. Thresholds for sedentary bouts (60 sec minimum) and MVPA (3 sec minimum) were based on the literature [18] and the analysis of Diaz and Yacef [20].

Physical Activity During the Program (Misfit)

The Misfit activity trackers were worn continuously over the 4-week program, and data were available to the researchers in a daily aggregated form for each child: this included the total daily steps and the PA sessions that were detected with duration, steps, and calories spent during each session. We reverse-engineered the PA intensities from calories spent using the following criteria, which were found empirically to match Misfit's categories: PA sessions that consumed less than 2 calories per minute were considered light, between 2 and 3.5 calories were considered moderate, and over 3.5 calories were considered vigorous. These empirically defined cut points are aligned with those of Colley et al [21] for light and moderate PA and were slightly lower for vigorous activity, suggesting that our calculation was on the conservative side. This had no bearing on our study as we were interested in MVPA.

As noted earlier, these trackers were not used here for scientific validation but for learning purposes. Nevertheless, to assess data accuracy, we compared their data with those of the GENEActiv trackers, by asking children to wear them concurrently on the same wrist during the first week: the daily totals were consistent, although the GENEActiv trackers sometimes captured more finely detailed data. This indicates

that the Misfit trackers may sometimes underestimate very small activity bouts, yet still provided a close lower bound of the children's activity. Their continuous data across the 28-day program data were, therefore, useful to explore.

Feedback From Adolescents

At the end of each module, the adolescents had the opportunity to provide quick feedback on their experience via a survey question: "How much fun was this module?"

At the end of the program, through a multiple-choice quiz, they were also asked to indicate their intention to engage in healthier PA behaviors in the coming months.

Statistical Analysis

Anthropometry and Physical Fitness Analysis

Anthropometry and physical fitness data are presented as mean (SD). Means and SDs for the physical fitness tests were calculated for the whole group; these tests were MAS; speed at 5, 10, 15 m; and an agility test.

Analysis of Baseline Physical Activity (GENEActiv)

The mean and SD of daily times spent being sedentary and engaged in light, moderate, and vigorous PA were calculated on the minimum bout-filtered daily accumulated data. Analyses were conducted using R Core Team 3.1.0. [22].

Data Mining Methods

Analysis of Program Physical Activity (Misfit)

We analyzed the PA data in several ways to gain insight into the PA behavior changes during the program. We first identified the types of daily behavior observed through *daily behavioral clustering*. We then devised a *consistency rate* function to track the regularity of the participants' healthy PA behaviors over a sliding window of 7 days. We assumed that although the adolescents might not meet the PA recommendations every day, a noteworthy sign of progress is the fact that they meet them more regularly at the end of the program than at the beginning. These methods are summarized below.

Daily Behavior Clustering

In addition to the binary achievement of minimum daily steps, we explored different shapes of the PA behaviors regarding the number of steps, the time spent at various intensities, and how these behaviors changed daily. The behavior clustering method thus revealed the types of PA behaviors over a day. The activity tracker data, for each child and each day, were mapped into vectors with 5 features: 1, 2, and 3—the total daily time spent in light, moderate, and vigorous PA, respectively; 4—the total number of steps; and 5—a binary feature that was true if at least 11,000 steps were accumulated on that day. Each of these vectors was, therefore, a representation of the PA behavior of 1 child on a given day. Numerical features were normalized.

These vectors of daily PA behaviors were then clustered (Kmeans, k=3 determined using the elbow method to find the best balance between the number of clusters and the sum of square errors within clusters).

Consistency Rate

The consistency rate captures the frequency and regularity of desirable behaviors in relation to a specific target. It represents the average value of achievement of that target over a certain number of days, here 7, and is computed as shown in Figure 3, where *n* is the number of days (ie, 7), *R(i)* is the degree of target achievement on day *i*, and *P(i)* is a binary function showing whether behavioral data were available on day *i*.

The purpose of *P(i)* was to ensure that the data were only averaged for the number of days on which data were captured and to exclude those days when the adolescents did not wear their device (loss, malfunction, etc). For instance, an adolescent who had not once met the recommendations over the prior 7 continuous days would have a rate of 0%, whereas one who did so consistently would have a rate of 100%. If no data were recorded in the 7-day period, the value was excluded. As this 7-day window moved each day, we were able to follow trends with the consistency rate.

To follow progress, we applied this method to 2 targets: (1) daily number of steps over 11,000 and (2) daily PA behavior within the desirable clusters. For the first target, R(i)=1 if the daily number of steps were achieved on day i, 0 otherwise. For the second one, each cluster was assigned a value within the range (0;1) corresponding to the activity level of that cluster (where 1: most active and meeting all international recommendations), and R(i) was the value of the corresponding behavior cluster on day i.

Figure 3. Consistency equation.

$$Consistency(n) = \frac{\sum_{i=n-6}^{n} R(i)}{\sum_{i=n-6}^{n} P(i)}$$

Results

Anthropometry, Physical Fitness, and Physical Activity Before the iEngage Program

Descriptive data collected before the iEngage program are presented in Table 1. Accelerometry results from GENEActiv show that participants spent an average of 112 min per day in sedentary behaviors lasting at least 60 sec. Daily (8 am-8 pm) baseline PA intensities expressed in minutes (sedentary, light, and MVPA) showed that they spent 122 min in light PA and 36 min in MVPA in bouts lasting at least 3 sec.

Table 1. Descriptive anthropometric, physical fitness, and physical activity data before the iEngage program (N=24).

Individual characteristics	Value, mean (SD)
Anthropometric variables	
Height, m	1.58 (0.06)
Weight, kg	56.45 (13.09)
Age, years	11.88 (0.57)
FBM[a], %	27.99 (7.52)
FBM, kg	39.99 (6.76)
BMI[b], %	22.33 (4.60)
BMI z score	1.20 (1.02)
Waist, cm	81.58 (10.84)
Waist to height ratio	0.51 (0.06)
Physical fitness variables	
Maximal aerobic speed, $km.h^{-1}$	10.90 (1.10)
5 m sprint, sec	1.36 (0.10)
10 m sprint, sec	2.96 (0.33)
30 m sprint, sec	5.38 (0.44)
Agility test, sec	13.02 (3.09)
Mean time spent per day at the corresponding intensity in minutes	
Sedentary activity (60-sec bouts)	112.20 (19.30)
Light physical activity (3-sec bouts)	121.90 (5.80)
Moderate-to-vigorous physical activity (3-sec bouts)	35.30 (3.80)

[a]FBM: fat body mass.

[b]BMI: body mass index.

Daily Behavior Clustering

Behavior clustering identified 3 types of daily PA behaviors, shown in Table 2: the less active cluster (cluster 0), well under the daily recommended number of steps and minimal MVPA; the active cluster (cluster 1), where daily step count was reached but well under 60 min of MVPA; and the very active cluster (cluster 2), with active days and well over the recommended step number and time in MVPA.

Table 2. Centroids of daily behavior clusters (k=3) during the 4-week iEngage program.

Feature	Mean cluster 0: less active cluster (n=180)	Mean cluster 1: active cluster (n=229)	Mean cluster 2: very active cluster (n=37)	Mean overall (n=446)
Light-intensity activity, min	39	92	85	64
Moderate-intensity activity, min	5.30	16	57	14
Vigorous-intensity activity, min	2.30	4	52	7
Number of steps	7850	13,860	18,260	11,140
Over 11,000 steps	No	Yes	Yes	No

Consistency Rate

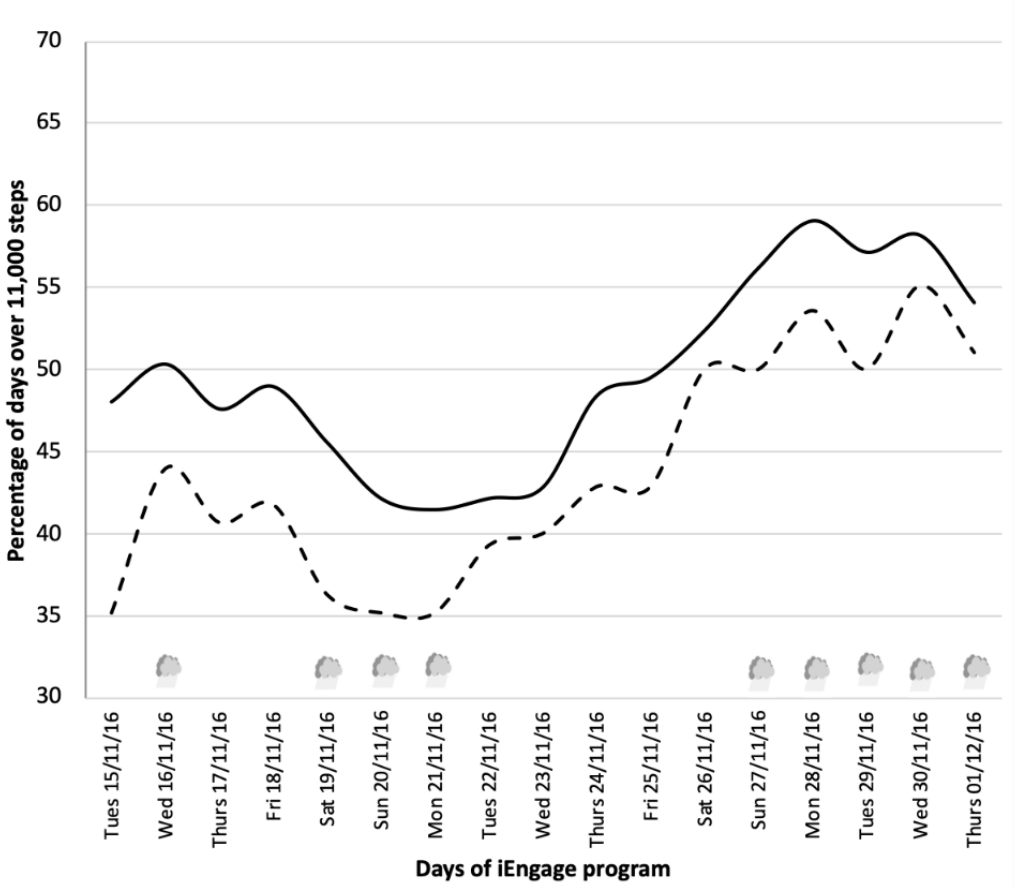
Progress was assessed according to 2 daily targets: (1) reaching 11,000 steps and (2) being in the most active daily PA behavior cluster.

Target 1: Achievement of Daily Steps

Overall analysis showed that, on average, adolescents achieved 11,197 (SD 1376) steps per day during the 4-week program. The consistency rate of achieving 11,000 steps per day improved throughout the program (Figure 4). Consistency rate analysis showed that 11,000 steps were achieved for 48% of the days in week 1 (approximately 3.4 days/week). This increased to 54% (approximately 3.8 days/week) in week 4, with a peak of 59% (approximately 4.1 days/week) shortly before. More interestingly, the adolescents who were the least active in week 1 (ie, achieving less than 50% of daily recommendations in the first week) increased their consistency rate from 35% (approximately 2.4 days/week) at program start to 51%

(approximately 3.5 days per week) at program end, with a peak of 53.6% (approximately 3.7 days/week).

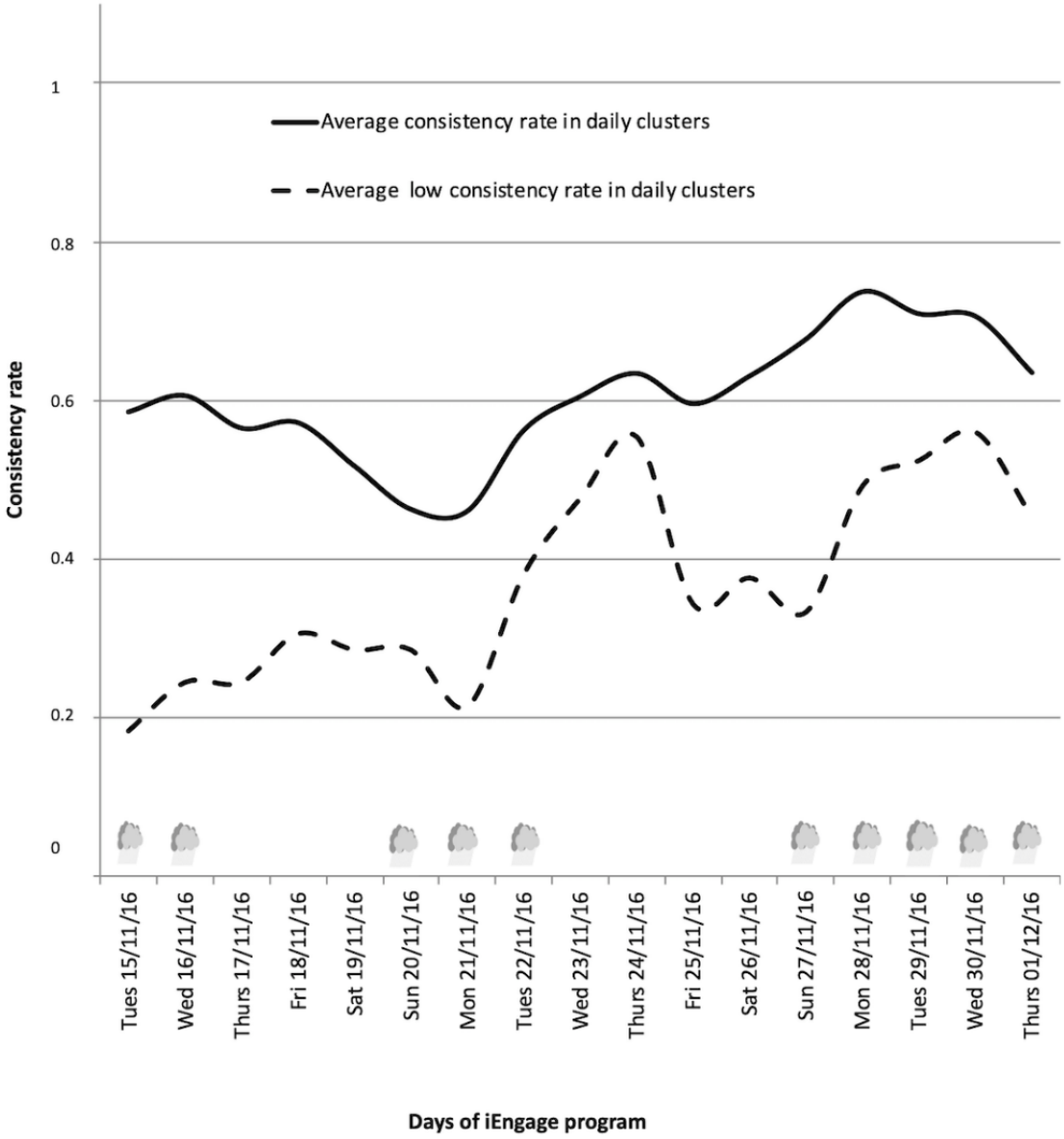
Figure 4. Average consistency rate (in %) of daily steps during the iEngage program. Consistency rate is calculated at the end of a 7-day window sliding over time; hence, the data presented in the graph start on day 7. Rainy days (N=10, between 0.3 and 14.3 mm.3h^{-1}) are represented by a cloud symbol.



Target 2: Improvement in Daily Physical Activity Behavior

The daily behavior clusters provided a more refined analysis of progress, taking into account the time spent at each PA intensity (light, moderate, and vigorous), the number of steps, and the achievement of daily recommended steps. With 3 clusters, R(n) was a 3-value function: 0 for cluster 0 (less active), 0.5 for cluster 1, and 1 for cluster 2 (very active, meeting all recommendations). A consistency value of 0 meant that every day of the past week was spent in the less active cluster, and a value of 1 meant that every day was spent in the very active cluster. A value of 0.5 could mean that the adolescents split their time between cluster 0 and cluster 2 (alternating less active and very active days) or every day in cluster 1 (active but under the recommendations). The analysis of weekly consistency rate using these clusters is presented in Figure 5. On average, the adolescents started at 0.60 and finished at 0.63 (with a peak at 0.71) at the end of the iEngage program. More interestingly, the least active adolescents in week 1 (defined as being in the less active cluster more than 50% of the time) started with a consistency rate of 0.18 and ended with 0.45 (peak at 0.51), indicating a 27% increase over the program for these adolescents.

Figure 5. Average consistency rate (from 0 to 2) of daily behavior clusters during the iEngage program. Rainy days (N=10, between 0.3 and 14.3 mm.3h^{-1}) are represented by a cloud symbol.



Feedback From Adolescents

The percentage of satisfaction with the modules (ie, module rated as "fun") was 95% across the program (for all modules). At the end of the program, adolescents were invited to set some goals for the future. Overall 85% (18/21) of adolescents declared intentions to achieve daily MVPA: 52% (11/21): "I intend to achieve 60 min of MVPA per day" and 33% (7/21) "I intend to achieve 30 min of MVPA per day". Furthermore, 62% (13/21) declared "I will discuss with my family and friends about what I learned," 62% (13/21) "I will seek assistance from close friends to reach my objectives," and 0% "I don't know/I am not motived." To the question "Do you think you will achieve it?," the answer was "Yes" for 38% (8/21) of the adolescents, "I'm not sure" for 57% (12/21), and "I don't think so" for 4% (1/21).

Discussion

This study showed that deploying a technology-supported educational program in a school context targeting PA behaviors in a remote area of the PICTs is feasible. In addition, we provide new data mining techniques for tracking adolescents' adherence to international PA recommendations through daily behavior clustering and consistency rate.

The 95% Melanesian population was from a remote and rural area of New Caledonia. The overweight prevalence and the physical fitness level before the program (Table 1) were consistent with previous observations in the region [7,11,23-25]. PA data recorded before the program, although in a small group, showed for the first time that these adolescents performed only approximately 30 min of MVPA daily, which is half the recommended volume (Table 1). This is similar to other Asia-Pacific countries such as Australia [10] and highlights the need for intervention for New Caledonian adolescents [8].

PA was analyzed using novel integrative approaches. First, it was analyzed per day and per adolescent rather than averaged over the observed period. This aligns with how recommendations are expressed and allowed the detection of changes on a daily basis. Second, a behavior clustering technique was used to combine the PA components to characterize the daily PA of an adolescent: volume (daily steps), time spent at each PA intensity, and achievement of internationally recommended daily steps. These objectively measured data enabled us to identify 3 daily PA behavior clusters (Table 2): days spent in clusters 0 and 1 were below the international recommendations of 60 min of MVPA per day, whereas days in cluster 2 (only 37/446 days) were very active and well over the international recommendations. Third, changes in PA were measured by calculating the consistency of PA over a 7-day sliding window. PA was measured via 2 indicators: a simple one based on the achievement of the recommended daily steps (Figure 4) and a more comprehensive one using the daily behavior clusters (Figure 5). Figure 4 shows that the least active adolescents in the first week (eg, achieving the daily recommended steps less than half the time) progressively increased their steps by 15.6% over the 4-week program. Furthermore, the consistency rate and daily behavior clustering methods that take into account durations, intensities, and recommendations were useful for interpreting the PA patterns. This integrative approach showed that the adolescents improved PA behaviors over the program (+27%; Figure 5), especially those who were the least active in the first week. Indeed, these adolescents more often met the international recommendations for daily steps and PA intensity by week 4. This approach provided better daily global analyses of PA patterns in adolescents based on the recommendations [26].

Overall, PA steadily increased, showing levels approaching the international recommendations, especially for those who were less active at the beginning of the study. We highlight that PA was realized outside the iEngage modules, ie during the adolescents' daily life, as the program only prescribed at most 4 min per module (the remaining time in each session was spent learning via the app or discussing concepts with peers).

The school educational team, including the principal, the teachers, and the school nurse, as well as the families and community (Melanesian custom) were very involved in the program, as we intended it. Indeed, a New Zealander study showed that contextualized learning facilitated knowledge translation in programs promoting scientific and health literacy [27]. The adolescents were more able to decide whether and how to incorporate scientific evidence to enhance their current and future well-being. The program encouraged them to share knowledge and skills with peers, family, and the broader community (including teachers and school nurse) and to involve them in achieving health goals. For example, as the efficacy of adolescent health interventions depends in part on family involvement [14], these adolescents were encouraged to identify a "mission" at the end of each module to engage parents, carers, siblings, or friends in the activity and/or discuss health information and recommendations. We put in place an information strategy and meetings with parents along with important people in the community, eg, religious leaders and elders. Some of them accepted to convey messages of support for the program based on their teaching discipline or role in the community. They did so in French and the local language (*Drehu*) via videos that were incorporated into the digital program. This was particularly important because of the key role of spiritual leaders in Pacific communities [28]. In particular, we invited "custom keepers" and religious leaders to talk in *Drehu* about the importance of PA and nutrition as ways to build healthy lifestyles and be part of healthy communities. We believe that this community involvement contributed to the adolescents' strong appreciation for the program and their intentions to continue their engagement after the program end, and these too were positive outcomes of this exploratory study.

Some study limitations need to be underlined. This exploratory study was conducted in 1 class with 24 children for a period of 4 weeks and no control group was included, which limits any generalization. Although Misfit data were recorded continuously during the program, enabling us to obtain unique information on the children's behavior, we did not measure the pre-post effects of the intervention via accelerometry, and this may have limited the interpretation. However, at the start of the program and for 5 consecutive days, the Misfit data were compared for data accuracy with the GENEActiv data. The daily totals were consistent, indicating that although the Misfit trackers sometimes underestimated small activity bouts, they still provided a close lower bound of the children's activity.

In conclusion, an integrative approach in the sports sciences, combining user-centered and community-based education and digital technologies (wearable trackers) to build health-related skills and change PA behaviors, seems promising in the PICTs. Our novel data mining technique provides an innovative way to assess adherence to international PA recommendations. Although several improvements are needed, this exploratory study showed that an electronic health program can be deployed in a remote Pacific area with a strong traditional culture and suggests the potential for implementing it at a larger scale for different ages, populations, and cultures.

Acknowledgments

The authors would like to thank BePatient [29] for their contribution to this project via their digital platform. The authors would also like to thank the school teaching team of Wé in Lifou Island and the administrative staff for their help and support in the investigation, especially Dr P Zongo, P Monjo, A Mildred-Canet, E Cullum, A Le Nestour, J McBroom, and the Vice-Rectorat of New Caledonia. This research program was funded by a grant from the French Ministry of Foreign Affairs through the Pacific Fund to CC, KY, and OG.

Authors' Contributions

OG, KY, and CC equally contributed to this study. They created the educational component of the iEngage program and designed and coordinated the study. KY developed the model for data analysis with input from CC and OG. OG drafted the manuscript with critical input from KY and CC. All authors revised the manuscript and approved the final version.

Conflicts of Interest

None declared.

References

1. Hughes R, Lawrence M. Globalization, food and health in Pacific Island countries. Asia Pac J Clin Nutr 2005;14(4):298-306 [FREE Full text] [Medline: 16326635]
2. Kessaram T, McKenzie J, Girin N, Merilles OE, Pullar J, Roth A, et al. Overweight, obesity, physical activity and sugar-sweetened beverage consumption in adolescents of Pacific islands: results from the Global School-Based Student Health Survey and the Youth Risk Behavior Surveillance System. BMC Obes 2015;2:34 [FREE Full text] [doi: 10.1186/s40608-015-0062-4] [Medline: 26401344]
3. Colditz GA, Peterson LL. Obesity and cancer: evidence, impact, and future directions. Clin Chem 2018 Jan;64(1):154-162 [FREE Full text] [doi: 10.1373/clinchem.2017.277376] [Medline: 29038151]
4. Frayon S, Cherrier S, Cavaloc Y, Wattelez G, Lerrant Y, Galy O. Relationship of body fat and body mass index in young Pacific Islanders: a cross-sectional study in European, Melanesian and Polynesian groups. Pediatr Obes 2018 Jun;13(6):357-364. [doi: 10.1111/ijpo.12229] [Medline: 28752614]
5. Central Intelligence Agency. The Work Factbook: Country Comparison: Obesity-adult-prevalence rate URL: https://www.cia.gov/library/publications/the-world-factbook/rankorder/2228rank.html [accessed 2019-11-13]
6. Malina RM. Tracking of physical activity and physical fitness across the lifespan. Res Q Exerc Sport 1996 Sep;67(3 Suppl):S48-S57. [doi: 10.1080/02701367.1996.10608853] [Medline: 8902908]
7. NCD Risk Factor Collaboration (NCD-RisC). Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. Lancet 2017 Dec 16;390(10113):2627-2642 [FREE Full text] [doi: 10.1016/S0140-6736(17)32129-3] [Medline: 29029897]
8. Zongo P, Frayon S, Antoine-Jonville S, Wattelez G, Le Roux P, Hue O, et al. Anthropometric characteristics and physical fitness in rural and urban 11- to 16-year-old Melanesian adolescents: a cross-sectional study in new Caledonian schools. Asia Pac J Public Health 2017 Oct;29(7):589-598. [doi: 10.1177/1010539517735414] [Medline: 29094632]
9. Ekelund U, Luan J, Sherar LB, Esliger DW, Griew P, Cooper A, International Children's Accelerometry Database (ICAD) Collaborators. Moderate to vigorous physical activity and sedentary time and cardiometabolic risk factors in children and adolescents. J Am Med Assoc 2012 Feb 15;307(7):704-712 [FREE Full text] [doi: 10.1001/jama.2012.156] [Medline: 22337681]
10. Hardy LL, Okely AD, Dobbins TA, Booth ML. Physical activity among adolescents in New South Wales (Australia): 1997 and 2004. Med Sci Sports Exerc 2008 May;40(5):835-841. [doi: 10.1249/MSS.0b013e318163f286] [Medline: 18408616]
11. Wattelez G, Frayon S, Cavaloc Y, Cherrier S, Lerrant Y, Galy O. Sugar-sweetened beverage consumption and associated factors in school-going adolescents of new Caledonia. Nutrients 2019 Feb 21;11(2):pii: E452 [FREE Full text] [doi: 10.3390/nu11020452] [Medline: 30795633]
12. Martins J, Marques A, Sarmento H, Carreiro da Costa F. Adolescents' perspectives on the barriers and facilitators of physical activity: a systematic review of qualitative studies. Health Educ Res 2015 Oct;30(5):742-755. [doi: 10.1093/her/cyv042] [Medline: 26324394]
13. van Sluijs EM, McMinn AM, Griffin SJ. Effectiveness of interventions to promote physical activity in children and adolescents: systematic review of controlled trials. Br Med J 2007 Oct 6;335(7622):703 [FREE Full text] [doi: 10.1136/bmj.39320.843947.BE] [Medline: 17884863]
14. Jourdan D, Christensen JH, Darlington E, Bonde AH, Bloch P, Jensen BB, et al. The involvement of young people in school- and community-based noncommunicable disease prevention interventions: a scoping review of designs and outcomes. BMC Public Health 2016 Oct 26;16(1):1123 [FREE Full text] [doi: 10.1186/s12889-016-3779-1] [Medline: 27784301]
15. Brannon E, Cushing C. A systematic review: is there an app for that? Translational science of pediatric behavior change for physical activity and dietary interventions. J Pediatr Psychol 2015 May;40(4):373-384. [doi: 10.1093/jpepsy/jsu108] [Medline: 25502745]
16. Kerner C, Goodyear VA. The motivational impact of wearable healthy lifestyle technologies: a self-determination perspective on Fitbits with adolescents. Am J Health Educ 2017;48(5):287-297. [doi: 10.1080/19325037.2017.1343161]
17. Galy O, Zongo P, Chamari K, Chaouachi A, Michalak E, Dellal A, et al. Anthropometric and physiological characteristics of Melanesian futsal players: a first approach to talent identification in Oceania. Biol Sport 2015 Jun;32(2):135-141 [FREE Full text] [doi: 10.5604/20831862.1140428] [Medline: 26060337]

18. Schaefer CA, Nigg CR, Hill JO, Brink LA, Browning RC. Establishing and evaluating wrist cutpoints for the GENEActiv accelerometer in youth. Med Sci Sports Exerc 2014 Apr;46(4):826-833 [FREE Full text] [doi: 10.1249/MSS.0000000000000150] [Medline: 24121241]
19. Phillips LR, Parfitt G, Rowlands AV. Calibration of the GENEA accelerometer for assessment of physical activity intensity in children. J Sci Med Sport 2013 Mar;16(2):124-128. [doi: 10.1016/j.jsams.2012.05.013] [Medline: 22770768]
20. Diaz C, Yacef K. Detecting behaviour changes in accelerometer data. Knowledge Discovery in Healthcare Data (KDH@IJCAI-ECAI Stockholm), CEUR-WS 2018;2128:21-26 [FREE Full text]
21. Colley RC, Tremblay MS. Moderate and vigorous physical activity intensity cut-points for the Actical accelerometer. J Sports Sci 2011 May;29(8):783-789. [doi: 10.1080/02640414.2011.557744] [Medline: 21424979]
22. R: The R Project for Statistical Computing. 2014. URL: http://www.R-project.org/ [accessed 2019-11-13]
23. Dietz WH. Periods of risk in childhood for the development of adult obesity--what do we need to learn? J Nutr 1997 Sep;127(9):1884S-1886S. [doi: 10.1093/jn/127.9.1884S] [Medline: 9278575]
24. Wate JT, Snowdon W, Millar L, Nichols M, Mavoa H, Goundar R, et al. Adolescent dietary patterns in Fiji and their relationships with standardized body mass index. Int J Behav Nutr Phys Act 2013 Apr 9;10:45 [FREE Full text] [doi: 10.1186/1479-5868-10-45] [Medline: 23570554]
25. Thompson-McCormick JJ, Thomas JJ, Bainivualiku A, Khan AN, Becker AE. Breakfast skipping as a risk correlate of overweight and obesity in school-going ethnic Fijian adolescent girls. Asia Pac J Clin Nutr 2010;19(3):372-382 [FREE Full text] [Medline: 20805082]
26. Plowman SA, Meredith MD. The Cooper Institute. 2013. Fitnessgram/Activitygram Reference Guide URL: https://www.cooperinstitute.org/vault/2440/web/files/662.pdf [accessed 2019-11-13]
27. Bay JL, Vickers MH, Mora HA, Sloboda DM, Morton SM. Adolescents as agents of healthful change through scientific literacy development: a school-university partnership program in New Zealand. Int J STEM Educ 2017;4(1):15 [FREE Full text] [doi: 10.1186/s40594-017-0077-0] [Medline: 30631671]
28. Schultz J, Utter J, Mathews L, Cama T, Mavoa H, Swinburn B. The Pacific OPIC Project (Obesity Prevention in Communities): action plans and interventions. Pac Health Dialog 2007 Sep;14(2):147-153. [Medline: 19588622]
29. BEPATIENT. URL: https://www.bepatient.com/en/ [accessed 2019-11-13]

Abbreviations

BMI: body mass index
FBM: fat body mass
MAS: maximal aerobic speed
MVPA: moderate-to-vigorous physical activity
PA: physical activity
PICTs: Pacific Island countries and territories

Edited by G Eysenbach; submitted 29.05.19; peer-reviewed by A Ramachandran, B Fisher, A Videira-Silva, C Young, J Alvarez Pitti; comments to author 18.07.19; revised version received 11.08.19; accepted 24.09.19; published 11.12.19.

Please cite as:
Galy O, Yacef K, Caillaud C
Improving Pacific Adolescents' Physical Activity Toward International Recommendations: Exploratory Study of a Digital Education App Coupled With Activity Trackers
JMIR Mhealth Uhealth 2019;7(12):e14854
URL: http://mhealth.jmir.org/2019/12/e14854/
doi:10.2196/14854
PMID:31825319

Original Paper

A Stress Relief App Intervention for Newly Employed Nursing Staff: Quasi-Experimental Design

I-Chiu Chang[1], PhD; Wei-Chen Cheng[1,2], MS; Wen-Chuan Kung[3], MS, RN

[1]Department of Information Management, National Chung Cheng University, Chia-Yi, Taiwan

[2]Information Technology Office, Tainan Municipal An-Nan Hospital, Tainan, Taiwan

[3]Nursing Department, Hsinchu MacKay Memorial Hospital, Hsinchu City, Taiwan

Corresponding Author:
Wen-Chuan Kung, MS, RN
Nursing Department
Hsinchu MacKay Memorial Hospital
No 690, Sec 2
Guangfu Rd, East Dist
Hsinchu City, 300
Taiwan
Phone: 886 3 688 9595
Email: 6444@mmh.org.tw

Abstract

Background: Most newly employed nurses have limited practical experience, lack problem-solving abilities, and have low resistance to stress, and therefore often opt to resign from the nursing profession.

Objective: This study aimed to assess the effectiveness of a stress relief app (SR_APP) to monitor the stress levels of newly employed nurses.

Methods: We conducted a quasi-experiment to assess changes in stress levels of newly employed nurses at a case hospital, in which the experimental group used the SR_APP and the control group did not. In-depth interviews were conducted to reveal insights regarding their stress. The app usage experiences of experimental group members were assessed via a questionnaire.

Results: All the participants appreciated the experiment and were interested to know more about managing their stress. The experimental group members showed significant differences in heart rate variability scores before and after using the SR_APP, and they reported high levels of intention to use and satisfaction with regard to the SR_APP.

Conclusions: The SR_APP can be effective in helping newly employed nurses to manage their stress.

(JMIR Mhealth Uhealth 2019;7(12):e15785) doi:10.2196/15785

KEYWORDS

nursing staff; occupational stress; mobile app

Introduction

Occupational Stress

Work-related stress is a prevalent condition in the nursing profession. A recent literature review showed that studies of nursing stress covered areas from developed to developing countries, departments from general wards to emergency departments, subjects from newly graduated nurses to nurse managers, and topics from the antecedents to the impact of work-related stress [1-11]. Furthermore, significant positive correlation was found between work stress and the demission rate among nursing staff [12]. However, the methods of stress assessment used by those studies are mainly through stress questionnaires and lack objective and quantitative data as references.

Physiological monitoring markers for objective measurements of stress include heart rate, blood pressure, and heart rate variability (HRV). The quantitative analysis of HRV can portray stress indices and circumvent such shortcomings as the intentional concealment of problems during completion of the questionnaire [13]. HRV analysis is widely used as part of the self-adjustment practice for workers in high-stress conditions, students facing examination anxiety, and those in occupations that focus on professional peak performance, such as athletes and artists [14]. This study included HRV as a more objective measurement of stress in nurses. However, various factors can

interfere with HRV, including heart rate, age, gender, day and night rhythm, and acute and chronic disease [15]. Therefore, in-depth interviews were conducted to reveal insights regarding their stress.

Managing Occupational Stress

Stress can be viewed as a personal weakness, and employees are simply obliged to endure workplace stress. A study by Ke et al concludes that high work stress may contribute to nurses overdosing on sedatives, hypnotics, and antipsychotics, and younger nurses becoming more susceptible to drug abuse [16]. Another viewpoint is that stress is both an organizational and personal issue. Therefore, organizations should adopt active and preventive strategies to manage workplace stress. A study by Dewe [17] proposed a 3-level prevention strategy to manage workplace stress. The goal of primary prevention is to eliminate, decrease, or control the number or intensity of stress sources; to modify systems; and to redesign workflow so that the organization's staff can improve their productivity and increase their motivation. The goal of secondary prevention is to educate and train individuals to identify stress sources and respond to stress effectively. The goal of tertiary prevention is to treat individuals who have already been exposed to stress sources and have experienced physical and mental damage because of it. This study adopted Dewe's strategy to manage workplace stress of newly employed nurses.

Information technology usage can have a positive effect on performance of organizations, from commercial, manufacturing, to financial ones [18-20]. Similarly, hospitals and clinics have introduced information technology (eg, electronic medical records) with rapid dissemination, processing, and storage capabilities to improve upon the conventional manual workflow and reduce the number of flaws and errors [21-23]. This helps improve the service quality of the entire health care system and increase administrative efficiency. This study considers workplace stress to be a common problem that is both organizational and personal and proposes a stress relief app (SR_APP) in conjunction with the traditional consulting mechanisms for stress management.

Methods

Introduction to SR_APP

This study selected a regional teaching hospital in northern Taiwan as its case hospital. This hospital is a religious hospital and places more emphasis on the physical, mental, and spiritual care of their employees. Meanwhile, this hospital is committed to technological advancement with cloud-based medical and nursing information management, which makes this case hospital the leading regional hospital in Taiwan. The SR_APP was created by a project team consisting of 2 representatives from nursing management and administration, 2 representatives from among newly employed nursing staff, 1 clinical psychologist, and 1 external information systems expert. The representatives from nursing management and administration had the job titles of supervisor and head nurse, and each had more than 10 years of experience in counseling and training courses for newly employed staff. The 2 newly employed nurses had been working at the hospital for 1 and 3 months when this study was conducted. The clinical psychologist possesses a professional psychology license and had 9 years of clinical counseling experience at the time of study. The medical information system expert was a professor in information systems from a national university with research interests in the development of nursing systems in hospitals.

The design of the SR_APP started with generating users' requirements from the project team. The information technology staff translated those requirements into system functions and implemented, tested, and revised the app. In total, it took 1.5 months to launch a stable version of the SR_APP. The requirements generated by the project team for the SR_APP were that the system should inform supervisors, staff, and psychologists via email when the staff member's stress monitoring data exceed the threshold value and when relaxation techniques are used, allow reviewing of historical data, be password protected, have different levels of authorization to access data, and be easy to operate. The information technology staff created the SR_APP according to the aforementioned requirements. The app screenshots are shown in Figure 1.

As the case hospital was granted an International Organization for Standardization specification for an information security management system (ISO 27001) certificate, the research data were stored in the hospital database and accessed under authority control. The nursing manager who was in charge of the newly employed nurses' learning program had access to the database. Each participant could only access their own data. If they accidently deleted the app, they could reload the app using the Quick Response code on the hospital's Web page and review the data they had. The data would be kept for 3 years according to the institutional review board (IRB) regulations.

Figure 1. SR-App screenshots.

Research Design

As newly employed nursing staff reported for work at the case hospital, those who met the inclusion criteria (no endocrine or cardiovascular disease, no history of smoking, and not currently on any medication) were invited to participate in the quasi-experiment. Those who agreed to participate were told the aim and methods of the study and were asked to fill out the consent forms and personal background information forms. They were then assigned to the experimental or control group. Ethical approval for the study was obtained from the IRB of MacKay Memorial Hospital—IRB serial number: 14MMHIS045.

Two times a week, the participants arrived at the experiment room after work, rested for 10 min, and conducted stress assessment using ANSWatch Model TS-0411 as the HRV analyzer. When the stress level of a given participant in the experimental group exceeded the threshold value, an email was sent to that participant's mobile phone to introduce relaxation techniques. No such intervention was conducted in the control group. If a participant's stress level was consistently higher than the threshold value, emails were sent to the appropriate supervisor and counselor to arrange a counseling meeting. Participants' stress levels were continuously measured during the entire time the experiment was conducted. The experiment room setting is shown in Figure 2.

Figure 2. The experiment room setting.

The HRV values of both groups were compared to test the following hypothesis:

> *H1: Newly employed nursing staff members who use the SR_APP exhibit higher HRV levels than staff members who do not.*

After the experiment, in-depth interviews were conducted to allow participants to describe their true feelings about the experiment and the effectiveness of the interventions used in this study. Semistructured, open-ended questions were used to collect the feedback. All the participants agreed to be interviewed after work on the scheduled date. Each interview lasted for approximately 20 to 60 min. A second interview was conducted if there were further questions or additional intact and accurate information had to be obtained. Interview outlines for both groups included questions about their motivation to participate in this study, feedback regarding the experimental process and impacts, and suggestions for improving the study (eg, the experimental process and questionnaire or interview). In addition, the experimental group's experience of using the SR_APP and their suggestions to improve the SR_APP were assessed using the widely accepted measurement shown in Table 1.

Table 1. The items for evaluating the SR_APP.

Variable	Items
System quality	
User-friendly	SR_APP is user-friendly; SR_APP is easy to use
Interactivity	The SR_APP provides quick feedback; The SR_APP gives me a variety of choices for relaxing
Information quality	
Accuracy	The SR_APP provides accurate information; I am satisfied with the accuracy of the SR_APP
Immediacy	Information provided by the SR_APP is timely; The SR_APP provides up-to-date information
Integrity	Information from the SR_APP is valuable; Information from the SR_APP is complete; The SR_APP covers my decompression needs
Perceived playfulness	
Pleasure	When interacting the SR_APP, I am not aware of any noise; Using the SR_APP gives me enjoyment; Using the SR_APP keeps me happy
User satisfaction	I am satisfied with the "stress alleviation techniques" function in the SR_APP; I am satisfied with the "stress alleviation methods" function in the SR_APP; I am satisfied with the "emotion diary" function in the SR_APP; I am satisfied with all of the functions in this app

The statistics software of Statistical Package for Social Sciences version 19 was used to analyze the collected data. The basic information of the participants was summarized using descriptive statistics. An independent *t* test and 1-way analysis of variance were used to test for significant differences on work stress among different categories. Finally, the Wilcoxon test

was used to compare differences between the experimental and control groups before and after intervention measures.

Results

Participants Demographic Information

A total of 28 participants were enrolled in this study, of which 21 completed 3 months of stress monitoring (8 in the experimental group and 13 in the control group). The completion rate was 75% (21/28), and 7 participants did not complete the stress monitoring because of the sudden illness of a family member, personal emotional reasons, poor adjustment, and/or resignation from work. The basic information of the participants is shown in Table 2.

Most participants were young females (95%, 20/21) aged between 21 and 25 years (76%, 16/21), who had university or equivalent education, with no work experience (66%, 14/21) and no relevant experience (76%, 16/21). All participants were unmarried, not on medication, nonsmokers, and had no children and no history of acute and chronic disease (eg, hypertension, diabetes, cardiovascular, and endocrine disease).

Table 2. Basic information of the participants.

Measure categories	Experimental group (n=8), n (%)	Control group (n=13), n (%)	Total (N=21), n (%)
Age (years)			
<20	2 (25)	3 (23)	5 (23)
21-25	6 (75)	10 (76)	16 (76)
Gender			
Female	7 (87)	13 (100)	20 (95)
Male	1 (12)	0 (0)	1 (4.8)
Education			
College	4 (50)	5 (38)	9 (42)
Bachelor's	4 (50)	8 (61)	12 (51)
Ward			
Medical Ward	3 (37)	2 (15)	5 (23)
Surgery Ward	3 (37)	4 (30)	7 (33)
Emergency	2 (25)	6 (46)	8 (38)
Outpatient	0 (0)	1 (7)	1 (4)
Relevant experience[a]			
No	5 (62)	11 (84)	16 (76)
Yes	3 (37)	2 (15)	5 (23)
Working experience			
No	8 (100)	6 (46)	14 (66)
Yes	0 (0)	7 (53)	7 (33)
Marriage			
Unmarried	8 (100)	13 (100)	21 (100)
Child			
None	8 (100)	13 (100)	21 (100)
Disease history[b]			
None	8 (100)	13 (100)	21 (100)
Medication			
None	8 (100)	13 (100)	21 (100)
Smoking			
None	8 (100)	13 (100)	21 (100)

[a]The prior experience related to the job in current division.

[b]Acute and chronic disease.

Results of Wilcoxon Test

As the sample size was small, the nonparametric Wilcoxon test was used to compare differences in HRV between the experimental and control groups before and after the intervention measures. The HRV values were ranked, and those ranks were summed for each group. The detailed results shown in Table 3 indicate that before the intervention, members of experimental group had higher stress levels than those of the control group (*P*=.02), and after the intervention, both groups had indifferent HRV scores (*P*=.93).

In other words, the experimental group improved more than the control group. To further explore the differences within the group, the pre- and postaverage scores of HRV for both groups were tested using a paired *t* test. The results are shown in Table 4.

Both groups had significantly higher HRV scores after the intervention, and the experimental group improved about 3 times more than the control group did.

Table 3. Wilcoxon test for experimental and control group comparison.

Group	HRV[a] difference			
	Average rank	Sum of rank	M-H test	*P* value
Pre			127.50	.02
Experimental	16.58	298.50		
Control	25.90	647.50		
Post			221.500	.93
Experimental	22.19	399.50		
Control	21.86	546.50		

[a]HRV: heart rate variability.

Table 4. Paired t test for experimental and control group heart rate variability score comparison.

Group	HRV[a] difference		
	Mean (SD)	*t* test (*df*)	*P* value
Experimental		–3.661 (7)	<.001
Pre	27.78 (6.495)		
Post	73.72 (13.775)		
Control		–3.158 (12)	.002
Pre	32.20 (5.164)		
Post	47.16 (19.893)		

[a]HRV: heart rate variability.

Follow-Up Interviews

As aforementioned, certain factors can interfere with the HRV assessment, including heart rate, age, gender, day and night rhythm, and endocrine disease. This study found other factors such as ad hoc assignments, conditions of patients, and critics from clinical preceptors to interfere in the participants' HRV. Therefore, in-depth interviews were conducted to help understand the insights from the interventions used in this study. The interview data were written into transcripts, and open coding of interview records was conducted based on the subject's original intention. Meaningful information from the raw data was grouped under similar codes. The in-depth interviews results showed that most newly employed nursing staff valued this experiment as a unique experience and were interested in understanding the changes in their stress levels during the employment process. Therefore, they looked forward to receiving their 3-month test results, regardless of whether they were in the experimental or control groups. Some of the feedback is transcribed below:

> *I want to know my stress level results.*
>
> *Although I knew that I would definitely experience stress, I still eagerly want to know test results of my stress level.*
>
> *Now that I have seen my stress level results, I feel that this is quite interesting.*
>
> *Initially, I was curious about the final result and felt that it was fun, which was why I agreed to participate in the study!*
>
> *This is my first job, so I want to know how great is the stress that I will experience.*

During each stress assessment, the facilitator observed that the newly employed staff showed fatigue, which was evident in their faces. The facilitator was worried that retaining the subjects for tests would bother them. It turned out that some nurses took

advantage of that period of time to emotionally recover from a busy day, though the stress test lasted about 10 min only. Some nurses tried to recall handing over tasks that they had forgotten to perform. Others were so relaxed that they fell asleep. Some nurses persistently complained or even broke down and cried. It seems that this period of time provided an opportunity for newly employed staff to change their mood before returning home.

> *Staying here for a stress level test enabled me to think about whether I had missed anything.*
>
> *Apologies, I was too tired and fell asleep during testing!*
>
> *During this period, I can try to think through what my senior stuff taught me today.*
>
> *I appreciate this time period to be alone.*
>
> *I feel like I cannot take this anymore...*
>
> *I really feel that I am not suitable for this job as I cannot do anything well.*
>
> *I feel that I will be a burden to my senior staff if I stay.*

Newly employed staff experienced work stress everywhere in the workplace, and some stressors were found in the prior studies. Stressors included too many staff names in the department to remember, insufficient staff reporting to work, no common topics for chatting with coworkers, and unfamiliarity with techniques and workflow, as well as receiving new patients, transferring patients, emergency situations, and patients with serious conditions. Additional stress came from mutual disturbances with roommates whose schedules clashed, inconvenient meal times, and lack of parental support.

> *The department is large with too many senior staff names to remember. I am afraid of identifying the wrong person and feeling embarrassed.*
>
> *The night shift is fine but there are only two staff members, and I do not know what to chat about and feel awkward throughout the long night.*
>
> *I am most afraid of receiving new patients as we have to be busy for a while even for one patient.*
>
> *Every time a new patient arrives, we become very busy.*
>
> *Patients have stayed for a long period of time, and therefore, we fear having to transfer them.*
>
> *I am frightened when the patients' conditions suddenly change.*
>
> *Although my seniors have taught me some techniques, I am still not familiar with them.*
>
> *When CPR is performed on the patients, I can only stand aside in a daze and not help as I do not recall what to do.*
>
> *I am easily awakened by slight noises or light. My roommate also works in shifts, so we often disturb each other.*
>
> *I cannot eat well because my meal time always changes. I am afraid of disturbing others when I get up and can only lie in a daze.*
>
> *My mother has been pressuring me to quit, and I do not know what to do.*

Most newly employed nursing staff adopted similar methods to relieve stress, such as having a large meal and sleeping off tiredness. Those with stamina left would chat with classmates, friends, and family members regarding work problems and find more courage to continue the nursing profession when they realized that others also encounter similar problems. Only a few nurses would exercise, go shopping, or watch movies to adjust their moods.

> *While I experienced a lot of stress from working, going home to sleep is my most commonly used method to relieve the stress.*
>
> *I feel that sleeping leads to a new starting point.*
>
> *I will have a big meal to get rid of the bad mood from being scolded at work, and feel better.*
>
> *I feel better after having good food.*
>
> *Having a big meal after receiving my salary makes me feel better!*
>
> *When my work is not going well, I will chat with my colleagues who have the same problem and quit thinking that I am not suitable.*
>
> *We newly employed staff usually get together and complain about the seniors. After we find out that each department has similar problems, we then feel that we are not particularly pitiful.*
>
> *When I complain to my family members, they will say that it's the same everywhere.*
>
> *I will go jogging after work. This enables me to slim down and release some stress.*
>
> *I will go shopping and watch movies with my friends.*

The experimental group had more alternatives by which to relieve their stress, as follows:

> *In the beginning, I did not use the app. After I experienced more stress, I used it quite often.*
>
> *For a long while, I would burst out crying in my dorm after work, and using the app helped me fall asleep.*
>
> *If I could have had this app during my internship, I would enjoy my current job more.*
>
> *I would follow the steps in the app to relax and felt quite comfortable, so I used it very often.*

Feedback of SR_APP Usage

More information regarding app usage was gathered via the questionnaire. The reliability and validity of the items are as follows. The Cronbach alpha values obtained were .851 for system quality, .833 for information quality, .898 for perceived entertainment, and .844 for usage satisfaction. The overall Cronbach alpha value was .896, which is above .8, indicating that the questionnaire has good reliability. The questionnaire was structured based on a literature review, the study objectives,

and the investigators' practical experience. An expert panel was convened to assess the content validity. The average scores for system quality, information quality, perceived entertainment, and user satisfaction with the SR_APP were 4.22, 3.95, 4.13, and 4.25, respectively, as measured on a 5-point Likert scale where 5 represents highly satisfied and 1 represents highly dissatisfied. In conclusion, the users evaluated the SR_APP system as satisfactory.

Discussion

Principal Findings

HRV analysis is widely used in self-adjustment methods for nonmedical workers in high-stress conditions [14]. Endukuru and Tripathi [24] indicated that the components of HRV were sensitive to stress in all healthy individuals. Eller et al [25] found a significant correlation between work stress syndrome and a reduction in the HRV. Some outpatient clinics that specialize in psychiatry and the treatment of emotional stress also use HRV analysis to detect patients' emotional stress [26,27]. However, applying HRV analysis to the counseling of newly employed nursing staff is rare.

We found that when stress levels exceeded the threshold value, the HRV value decreased, which confirmed the findings of prior studies. The work stress of both groups was reduced significantly. Furthermore, the experimental group's usage of the app resulted in larger differences in HRV before and after intervention, which indicates that the SR_APP can be used as an additional tool in conjunction with the traditional consulting mechanisms for stress management. An information system is useless if no one uses it. Therefore, the high satisfaction levels and strong usage intention reported by the experimental group after using the SR_APP confirms the value of adding this tool to the conventional stress relaxing methods for newly employed nurses.

A high demission rate not only increases the costs for training new staff but also increases the work burden of existing staff, thereby decreasing the quality of care and possibly affecting patient safety in severe cases [28,29]. The demission rate of newly employed nursing staff of the case hospital during the period of time this study was conducted was 18.84%, as compared with 23.7% in the same period the previous year. By the time this study was conducted, the national average demission rate for hospitals of the same rank was 19.6%. In other words, this shows that the intervention in this study may have also reduced the demission rate of newly employed nursing staff while alleviating employee stress.

Conclusions

Appropriate technology can help newly employed nursing staff alleviate their work stress when their own stress management skills are poor. The results of this study were inspiring. Not only the case hospital but other chain hospitals would like to adopt the SR_APP. However, the currently used device for measuring the HRV was costly and inefficient. Therefore, the project would continue once the medical engineering department invented an inexpensive and efficient solution for measuring the HRV. By then, the RS-App would be revised accordingly.

Although the researchers strove to make the study as rigorous as possible, some limitations could not be overcome because of time and resource constraints. Therefore, several limitations must be considered when making inferences based on the results of this study. First, as the recruitment of the case hospital during the time this study was conducted was not successful, this resulted in a limited number of subjects being invited to participate in the experiment. Second, some subjects were unable to continue participating in the study for personal and family reasons, which led to an uneven number of participants between the experimental and control groups. Third, the subjects were in constant contact with the facilitators during the study period, meeting twice a week for stress measurements, which might have interfered with the test results, especially when some subjects were emotionally unstable because of the excessive stress. To provide timely consolation and care in such cases, the facilitator might find it difficult to maintain a neutral position. Fourth, subjects calling in sick, changing shifts, going on urgent leave, and forgetting to stay after work also made it difficult to monitor their stress with consistent regularity, leading to missing values in the collected data. Finally, this study examined the changes in stress levels of newly employed nursing staff at a case hospital in northern Taiwan. As each hospital has a different culture and management style, caution must be exercised when generalizing the results of this study to other hospitals.

Future studies can perform comparisons and improve the generalizability of our results by extending the sample to different levels or types of hospitals in different areas. There are other factors that may affect the changes in stress levels in newly employed nursing staff and the leadership of clinical preceptors is one of them. It was observed that the clinical preceptors experience great stress when guiding new nurses. Therefore, examining the correlation between stress in preceptors and in newly employed nursing staff may provide a deeper understanding of the work stress for both sides. Meanwhile, some feedback related to the functions of the SR_APP were collected to facilitate future improvements to the system. For example, the interaction function of the current SR_APP will actively show the stress relief techniques only when the user's stress level exceeds the threshold value. As the reasons why the stress level exceeded the threshold value might be recorded in the participant's emotion diary, with proper linkage to the diary, the SR_APP can provide appropriate suggestions based on the actual cause. Also suggested for inclusion are additional functions such as the ability to inform the team when 1 member has poor interpersonal relationships in the department and a means by which to manage stress caused by emergency situations encountered by the patient. Finally, more personalized stress relief methods should be added to the app, which will increase the effectiveness with which SR_APP can provide stress relief for newly employed nursing staff.

Nursing staff shortage is a global problem that needs government support. Several successful adoptions of hospital-related information systems have been initiated and supported by the Taiwanese government [30,31]. Thus, on the basis of the findings of this study, the government can encourage hospitals

to adopt advanced information technology to provide a friendly work environment for retaining newly employed nursing staff.

Acknowledgments

This study was supported by a grant from the Hsinchu Mackay Memorial Hospital, Taiwan.

Conflicts of Interest

None declared.

References

1. Frögéli E, Rudman A, Gustavsson P. The relationship between task mastery, role clarity, social acceptance, and stress: an intensive longitudinal study with a sample of newly registered nurses. Int J Nurs Stud 2019 Mar;91:60-69. [doi: 10.1016/j.ijnurstu.2018.10.007] [Medline: 30677589]
2. Liu Y, Aungsuroch Y. Work stress, perceived social support, self-efficacy and burnout among Chinese registered nurses. J Nurs Manag 2019 Jul 15. [doi: 10.1111/jonm.12828] [Medline: 31306524]
3. Saquib N, Zaghloul MS, Saquib J, Alhomaidan HT, Al-Mohaimeed A, Al-Mazrou A. Association of cumulative job dissatisfaction with depression, anxiety and stress among expatriate nurses in Saudi Arabia. J Nurs Manag 2019 May;27(4):740-748. [doi: 10.1111/jonm.12762] [Medline: 30784143]
4. Pishgooie AH, Atashzadeh-Shoorideh F, Falcó-Pegueroles A, Lotfi Z. Correlation between nursing managers' leadership styles and nurses' job stress and anticipated turnover. J Nurs Manag 2019 Apr;27(3):527-534. [doi: 10.1111/jonm.12707] [Medline: 30136322]
5. Sonoda Y, Onozuka D, Hagihara A. Factors related to teamwork performance and stress of operating room nurses. J Nurs Manag 2018 Jan;26(1):66-73. [doi: 10.1111/jonm.12522] [Medline: 28744975]
6. Adriaenssens J, Hamelink A, Bogaert PV. Predictors of occupational stress and well-being in First-Line Nurse Managers: a cross-sectional survey study. Int J Nurs Stud 2017 Aug;73:85-92. [doi: 10.1016/j.ijnurstu.2017.05.007] [Medline: 28551478]
7. Chen CH, Wang J, Yang CS, Fan JY. Nurse practitioner job content and stress effects on anxiety and depressive symptoms, and self-perceived health status. J Nurs Manag 2016 Jul;24(5):695-704. [doi: 10.1111/jonm.12375] [Medline: 27086775]
8. Johansen ML, Cadmus E. Conflict management style, supportive work environments and the experience of work stress in emergency nurses. J Nurs Manag 2016 Mar;24(2):211-218. [doi: 10.1111/jonm.12302] [Medline: 25846993]
9. Lavoie S, Talbot LR, Mathieu L, Dallaire C, Dubois M, Courcy F. An exploration of factors associated with post-traumatic stress in ER nurses. J Nurs Manag 2016 Mar;24(2):174-183. [doi: 10.1111/jonm.12294] [Medline: 25827974]
10. Admi H, Eilon-Moshe Y. Do hospital shift charge nurses from different cultures experience similar stress? An international cross sectional study. Int J Nurs Stud 2016 Nov;63:48-57. [doi: 10.1016/j.ijnurstu.2016.08.005] [Medline: 27591723]
11. Ríos-Risquez MI, García-Izquierdo M. Patient satisfaction, stress and burnout in nursing personnel in emergency departments: a cross-sectional study. Int J Nurs Stud 2016 Jul;59:60-67. [doi: 10.1016/j.ijnurstu.2016.02.008] [Medline: 27222451]
12. Yang H, Lv J, Zhou X, Liu H, Mi B. Validation of work pressure and associated factors influencing hospital nurse turnover: a cross-sectional investigation in Shaanxi Province, China. BMC Health Serv Res 2017 Feb 3;17(1):112 [FREE Full text] [doi: 10.1186/s12913-017-2056-z] [Medline: 28158979]
13. Jarczok MN, Jarczok M, Mauss D, Koenig J, Li J, Herr RM, et al. Autonomic nervous system activity and workplace stressors--a systematic review. Neurosci Biobehav Rev 2013 Sep;37(8):1810-1823. [doi: 10.1016/j.neubiorev.2013.07.004] [Medline: 23891906]
14. Williamon A, Aufegger L, Wasley D, Looney D, Mandic DP. Complexity of physiological responses decreases in high-stress musical performance. J R Soc Interface 2013 Dec 6;10(89):20130719 [FREE Full text] [doi: 10.1098/rsif.2013.0719] [Medline: 24068177]
15. Sammito S, Böckelmann I. Factors influencing heart rate variability. ICFJ 2016 May 4;6. [doi: 10.17987/icfj.v6i0.242]
16. Ke YT, Feng IJ, Hsu CC, Wang JJ, Su SB, Huang CC, et al. Nurses have a four-fold risk for overdose of sedatives, hypnotics, and antipsychotics than other healthcare providers in Taiwan. PLoS One 2018;13(8):e0202004 [FREE Full text] [doi: 10.1371/journal.pone.0202004] [Medline: 30089164]
17. Dewe P. EAPs and stress management: from theory to practice to comprehensiveness. Pers Rev 1994;23(7):21-32. [doi: 10.1108/00483489410072217]
18. Cui T, Ye HJ, Teo HH, Li J. Information technology and open innovation: a strategic alignment perspective. Inform Manag 2015;52(3):348-358. [doi: 10.1016/j.im.2014.12.005]
19. Liu H, Wei S, Ke W, Wei KK, Hua Z. The configuration between supply chain integration and information technology competency: a resource orchestration perspective. J Oper Manag 2016;44(1):13-29. [doi: 10.1016/j.jom.2016.03.009]
20. Tjahjadi B, Soewarno N. The mediating effect of intellectual capital, management accounting information systems, internal process performance, and customer performance. Int J Productivity & Perf Mgmt 2019 Jun 18. [doi: 10.1108/ijppm-02-2018-0049]

21. Devaraj S, Ow TT, Kohli R. Examining the impact of information technology and patient flow on healthcare performance: a Theory of Swift and Even Flow (TSEF) perspective. J Oper Manag 2013;31(4):181-192. [doi: 10.1016/j.jom.2013.03.001]
22. Zardini A, Rossignoli C, Campedelli B. The Impact of the Implementation of the Electronic Medical Record in an Italian University Hospital. In: Organizational Innovation and Change. Cham: Springer; 2016:63-73.
23. Appiahene P, Missah YM, Najim U. Evaluation of information technology impact on bank's performance: The Ghanaian experience. Int J Eng Bus Manag 2019;11:184797901983533. [doi: 10.1177/1847979019835337]
24. Endukuru CK, Tripathi S. Evaluation of cardiac responses to stress in healthy individuals-a non invasive evaluation by heart rate variability and stroop test. Int J Sci Res 2016;5:286-289.
25. Eller NH, Blønd M, Nielsen M, Kristiansen J, Netterstrøm B. Effort reward imbalance is associated with vagal withdrawal in Danish public sector employees. Int J Psychophysiol 2011 Sep;81(3):218-224. [doi: 10.1016/j.ijpsycho.2011.06.014] [Medline: 21723888]
26. Chaudhary R, Haq MM, Shah PS. Changes in heart rate variability in depressed patient. Natl J Integr Res Med 2015;6(2):61-65 [FREE Full text]
27. Green KT, Dennis PA, Neal LC, Hobkirk AL, Hicks TA, Watkins LL, et al. Exploring the relationship between posttraumatic stress disorder symptoms and momentary heart rate variability. J Psychosom Res 2016 Mar;82:31-34 [FREE Full text] [doi: 10.1016/j.jpsychores.2016.01.003] [Medline: 26944396]
28. Dechawatanapaisal D. Nurses' turnover intention: The impact of leader-member exchange, organizational identification and job embeddedness. J Adv Nurs 2018 Jun;74(6):1380-1391. [doi: 10.1111/jan.13552] [Medline: 29468718]
29. Lee HF, Chiang HY, Kuo HT. Relationship between authentic leadership and nurses' intent to leave: the mediating role of work environment and burnout. J Nurs Manag 2019 Jan;27(1):52-65. [doi: 10.1111/jonm.12648] [Medline: 30238670]
30. Weng SJ, Lai LS, Gotcher D, Wu HH, Xu YY, Yang CW. Cloud image data center for healthcare network in Taiwan. J Med Syst 2016 Apr;40(4):89. [doi: 10.1007/s10916-016-0430-8] [Medline: 26846750]
31. Chung MH, Ho CH, Wen HC. Predicting intentions of nurses to adopt patient personal health records: a structural equation modeling approach. Comput Methods Programs Biomed 2016 Nov;136:45-53. [doi: 10.1016/j.cmpb.2016.08.004] [Medline: 27686702]

Abbreviations

HRV: heart rate variability
IRB: institutional review board

Edited by G Eysenbach; submitted 07.08.19; peer-reviewed by H Wu, FJ Yang; comments to author 18.09.19; revised version received 02.10.19; accepted 02.10.19; published 18.12.19.

Please cite as:
Chang IC, Cheng WC, Kung WC
A Stress Relief App Intervention for Newly Employed Nursing Staff: Quasi-Experimental Design
JMIR Mhealth Uhealth 2019;7(12):e15785
URL: https://mhealth.jmir.org/2019/12/e15785
doi:10.2196/15785
PMID:31850848

Original Paper

Analysis of the Implementation, User Perspectives, and Feedback From a Mobile Health Intervention for Individuals Living With Hypertension (DREAM-GLOBAL): Mixed Methods Study

Jordan Barsky[1], BKin; Rebekah Hunter[1], BSc; Colin McAllister[2], PEng; Karen Yeates[3], MD; Norm Campbell[4], MD; Peter Liu[5], MD; Nancy Perkins[1], RN; Diane Hua-Stewart[1], MPH; Marion A Maar[6], PhD; Sheldon W Tobe[1,6], MScCH, MD

[1]Department of Medicine, Sunnybrook Health Sciences Centre, Sunnybrook Research Institute, University of Toronto, Toronto, ON, Canada

[2]Perspect Management Consulting, Regina, SK, Canada

[3]Department of Medicine, Queens University, Kingston, ON, Canada

[4]Department of Medicine, University of Calgary, Calgary, AB, Canada

[5]University of Ottawa Heart Institute, Ottawa, ON, Canada

[6]Faculty of Medicine, Northern Ontario School of Medicine, Sudbury, ON, Canada

Corresponding Author:
Sheldon W Tobe, MScCH, MD
Department of Medicine
Sunnybrook Health Sciences Centre, Sunnybrook Research Institute
University of Toronto
Room 380, 1929 Bayview Avenue
Toronto, ON
Canada
Phone: 1 416 616 7043
Email: Sheldon.tobe@sunnybrook.ca

Abstract

Background: DREAM-GLOBAL (Diagnosing hypertension—Engaging Action and Management in Getting Lower Blood Pressure in Indigenous and low- and middle-income countries) studied a SMS text messaging–based system for blood pressure measurement and hypertension management in Canadian Aboriginal and Tanzanian communities. The use of SMS text messages is an emerging point of interest in global health care initiatives because of their scalability, customizability, transferability, and cost-effectiveness.

Objective: The study aim was to assess the effect on the difference in blood pressure reduction of active hypertension management messages or passive health behavior messages. The system was designed to be implemented in remote areas with wireless availability. This study described the implementation and evaluation of technical components, including quantitative data from the transmission of blood pressure measurements and qualitative data collected on the operational aspects of the system from participants, health care providers, and community leadership.

Methods: The study was implemented in six remote Indigenous Canadian and two rural Tanzanian communities. Blood pressure readings were taken by a community health worker and transmitted to a mobile phone via Bluetooth, then by wireless to a programmed central server. From the server, the readings were sent to the participant's own phone as well. Participants also received biweekly tailored SMS text messages on their phones. Quantitative data on blood pressure reading transmissions were collected from the study central server. Qualitative data were collected by surveys, focus groups, and key informant interviews of participants, health care providers, and health leadership.

Results: In Canada, between February 2014 and February 2017, 2818 blood pressure readings from 243 patients were transmitted to the central server. In Tanzania, between October 2014 and August 2015, 1165 readings from 130 patients were transmitted to the central server. The use of Bluetooth technology enabled the secure, reliable transmission of information from participants to their health care provider. The timing and frequency were satisfactory to 137 of 187 (73.2%) of participants, supporting the process of sending weekly messages twice on Mondays and Thursdays at 11 am. A total of 97.0% (164/169) of the participants surveyed said they would recommend participation in the DREAM-GLOBAL program to a friend or relative with hypertension.

Conclusions: In remote communities, the DREAM-GLOBAL study helped local health care providers deliver a blood pressure management program that enabled patients and community workers to feel connected. The technical components of the study were implemented as planned, and patients felt supported in their management through the SMS text messaging and mobile health program. Technological issues were solved with troubleshooting. Overall, the technical aspects of this research program enhanced clinical care and study evaluation and were well received by participants, health care workers, and community leadership.

Trial Registration: Clinicaltrials.gov NCT02111226; https://clinicaltrials.gov/ct2/show/NCT02111226.

(JMIR Mhealth Uhealth 2019;7(12):e12639) doi:10.2196/12639

KEYWORDS

blood pressure; hypertension; mHealth; population groups; short message service

Introduction

Background

There are over 1 billion individuals living with hypertension around the world, of which two-third are in developing countries [1]. Achieving the rates of hypertension control in low- and middle-income countries (LMICs), and in remote areas and vulnerable populations equal to those provided in high-income countries, remains a treatment gap that has not yet been bridged. The Global Alliance for Chronic Diseases funded and facilitated global collaborations in hypertension implementation research, and it sponsored this research, which was focused on improving the management of hypertension in LMICs and Canada's Indigenous communities [2].

The prevalence of cardiovascular disease, obesity, diabetes, and other related conditions is significantly higher in Indigenous populations in Canada than the rest of the population. Although the prevalence of hypertension is slightly lower, rates of hypertension awareness, treatment, and control are also lower [3]. This is because of the social determinants of health, including physical inactivity, smoking, and diets that are high in sodium and low in fresh fruits/vegetables, and those specific to Indigenous peoples; this also includes poverty and the impact of colonization and the Residential School system, creating barriers to accessing health care, and an increase in behavioral risk factors for cardiovascular disease [4,5]. The ubiquity of mobile phones and mobile health presents an opportunity to overcome some of these barriers.

The use of SMS text messaging in health care interventions is becoming a point of interest because of its scalability and capacity for adaptation to various populations at a low cost [6]. Studies are finding that they are acceptable to recipients and beneficial in the management of chronic disease, but outcomes remain mixed [5,7].

Objectives

The DREAM-GLOBAL, Diagnosing hypertension—Engaging Action and Management in Getting Lower Blood Pressure in Indigenous and low- and middle-income countries, research study was a randomized controlled trial comparing only the blood pressure lowering effect of passive health behavior messages with both active blood pressure management messages and passive health behavior messages [8]. The active messages were specific to the control or lack of control of the participant's blood pressure. To keep the patient's primary health care provider in the loop, the central server also sent faxes with the blood pressure readings to their office. This paper provides a technical review of the DREAM-GLOBAL intervention, discussing the development, implementation, and feedback of the system on the basis of responses from study participants, health care providers, and community health workers.

Methods

Design and SMS Text Message Development

The DREAM-GLOBAL study randomized participants to either active or passive SMS text messages, and the protocol has been previously published [8]. The study took place in 6 Canadian First Nations communities and 2 rural communities in Tanzania. The SMS text messages used in the study were developed from Canadian hypertension clinical practice guidelines and modified to be culturally safe and appropriate for the target population. This was carried out both in Canadian First Nations communities and in rural communities in Tanzania. All communities engaging in the study were invited to provide input on the messages to ensure that the language used was culturally appropriate, and cultural factors were taken into consideration when designing the intervention [9,10]; these factors included social, political, and historical aspects to better understand cultural differences and how these factors can affect health inequities in these rural areas of interest. An empirical decision on the frequency of the messages was made to send them twice weekly on Mondays and Thursdays at 11 am, which also helped to avoid holidays. A central server was programmed to send the messages to participants after registering with the program. The messages were not altered once the study began. The registration process was conducted by local community health workers in each community and participants recruited through hypertension screening days. The individual blood pressure measurements were taken using an automated blood pressure device with Bluetooth transmission capability. Mobile apps were developed for BlackBerry and Android wireless phones to receive the blood pressure measurements and transmit them to the central server. Upon registration into the system, the central server randomized patients to receive either passive or active SMS text messages. For a diagram of the circle of care model and the technological components, see Multimedia Appendix 1.

Participants

The details of how the communities were identified and assessed for research readiness have been previously described [9]. The population size of the individual First Nations communities

ranged from 500 to 4000. The communities were self-governed; they oversaw their own health care, provided partly by provincial and federal health ministries. Some of the remote First Nations communities had only obtained cell towers as recently as 2 years before the study; therefore, the prevalence of mobile phone use at the time was low in these communities. Participants were community members aged 18 years or older, with uncontrolled hypertension and on or off medications. They had to either have a mobile phone capable of receiving SMS text messages or be willing to carry and learn to use a basic flip phone for the study duration. Participants had to have a current primary health care provider. Blood pressure was measured with the A&D UA-767PBT-C monitor (A&D Medical), paired with a BlackBerry or an Android mobile phone app, to identify listed participants and facilitate the wireless transfer of the blood pressure to the central server.

Ethics and Safety

Cellular safety information was outlined to the participants before the study, including adherence to a ban on handheld devices while driving. Participants were advised not to share their phone with others if they did not want anyone to know about their participation in the study, and they were also advised not to send important personal information (ie, bank account or credit card number).

Ethics approval was obtained from the Sunnybrook Health Sciences Centre Research Ethics Board (Approval Number: 953-2013) and from participating University and community ethics boards. The study was carried out according to the principles of Good Clinical Practice, the Declaration of Helsinki, and the TriCouncil Policy Statement on ethical conduct for research involving humans [11-13]. The study adhered to the principles of Ownership, Control, Access, and Possession [14]. The technology and movement of data adhered to the principles of the Personal Information Protection and Electronic Documents Act.

Training on Use of Technology

Before the training and study launch, usability testing was conducted within the study investigator's clinics to confirm reliability of the technology to record and transmit the blood pressure readings to the central server (see Multimedia Appendix 2 for training slides). A goal of the DREAM-GLOBAL technology was to develop a mobile phone app for a trained community health worker or community health nurse to link the mobile phone to the blood pressure measurement device and allow transmission of the reading to the central server. One blood pressure measurement device could therefore be used for multiple participants. Participants only needed an SMS text messaging–capable mobile phone. One community health worker could therefore manage many participants who did not require training on blood pressure measurement. The community health worker and home and community care nurses, after a 2-day training session, were able to recruit, consent, register participants, and take the blood pressure readings, according to guidelines, with the automated oscillometric blood pressure device [15]. The components of this training program are found in Multimedia Appendix 2. The primary outcome of the study was the difference in systolic and diastolic blood pressure from the baseline period to the last 2 months of measurement between randomized groups.

Training to carry out the study was conducted in each community by study team members at planned education sessions. Training community health workers in the study procedures included information on using the Bluetooth-enabled blood pressure machines, enrolling patients into the program, teaching patients how to use their cellular device, and communicating with their patients. They were also provided with an instructional video on how to use the mobile app and blood pressure device to take the blood pressure readings and transmit them to the central server. Ongoing support for the community health workers was offered in the form of regular follow-up visits by the research team and troubleshooting support. In-service trainings and workshops were also provided to the local primary health care providers about the study.

Technology

The DREAM-GLOBAL system in Canada comprised a Bluetooth-enabled blood pressure monitor (A&D UA-767PBT) and a BlackBerry Bold (BlackBerry Inc) mobile phone, with the DREAM-GLOBAL mobile app installed. In Tanzania, an Android mobile phone was used. Blood pressure readings from the UA-767PBT blood pressure monitor were transmitted via Bluetooth to the community health worker's mobile phone and from the mobile phone to a secure central server in Canada. The central server was programmed to assess the blood pressure readings as normal or high. The server was programmed to wait for up to 3 blood pressure readings, for an individual patient, over 5 min. The average of the 3 readings was then calculated by the central server and then transmitted to the patient's primary health care provider by fax and to the patient's own phone by SMS. If the blood pressure was high, an SMS text message was sent to the participant, with advice to contact their health care provider over the coming week (see Multimedia Appendix 3 for the list of SMS text messages). The messages were based on the Hypertension Canadian Clinical Practice Guidelines, and they were developed by clinicians with extensive knowledge of hypertension management [10].

Evaluation

To evaluate components of the technology and processes of this study, both qualitative and quantitative data were collected. Qualitative data were collected from research notes and from key informant interviews and focus groups, including reflective discussion sessions with research and community teams. This focused on the design, the mechanisms, and context of the intervention, attempting to document local knowledge and expertise within the community stakeholders. Information-gathering tools included developing a community profile, an interview guide to facilitate discussion of topics, and a focus group guide to lead the dialogue on community-specific issues related to the intervention [10]. Quantitative evaluation included what information was delivered, as well as the quality and quantity. During the study, there was an evaluation of the acceptability and accessibility to the SMS text messages, the blood pressure measurement process, including data transmission, the impact of the messages on participants, and the impact of the study process on health care providers. After

completion of the study, the main outcome was an assessment of impact on blood pressure control and change of knowledge from the SMS text messages [16]. We have published on the formative data, with early community engagement and readiness [9], and the SMS development process [17]. The change of knowledge is being assessed, and it will be in a separate publication. This paper focuses on the participant, community, and health care provider response to the study processes and technology and the fidelity and quantity of the transmission of study data.

To evaluate the effectiveness of the system and solicit feedback on the technology and study implementation, survey/questionnaires, including simple short answer questions regarding their experiences with the technology, were provided to participants, including community health workers. The questionnaires were designed separately for participants, health care providers, and community leadership. The Participant Evaluation Form for SMS text messages and consent for this evaluation were administered by either the community health worker or the home and community care nurse for that community or the DREAM-GLOBAL clinical coordinator on visits to the communities. Participants were invited to do this feedback, and they provided additional consent to share their responses. Participants were contacted by SMS text messages, by direct phone, or contact through family members to come to the health center for this poststudy evaluation. In total, 184 of 243 participants returned for this final close-out visit. To reinforce the importance of maintaining health information privacy, information was provided at the start of the study, when the general confidentiality form was signed, reviewing that none of their personal information would be disclosed without written consent to anybody other than an investigator, a potential investigator, the client, and an applicable independent ethics committee review board.

Results

Participant Feedback

In Canada, the blood pressure lowering study was carried out with 6 First Nations communities in 3 geographic regions, including Manitoulin Island (Ontario), James Bay coast (Quebec), and the north shore of New Brunswick. In Tanzania, a preliminary portion of the study was carried out in 2 rural villages in Hai District, Kilimanjaro Region, to demonstrate that blood pressure can be measured and transmitted through the system. In Canada, a total of 2818 patient blood pressure readings from 243 patients were transmitted to the central server between February 2014 and February 2017. In Tanzania, with the pilot blood pressure measurement study, 1165 readings from 130 patients were transmitted to the central server between October 2014 and August 2015. The study server was programmed to wait for up to 3 readings and then send the mean to the participant's own phone. This also provided a record about the participant's blood pressure and allowed the participant to confirm that it was the measurement just taken. There were no incidents of participants finding incorrect readings transmitted from, for example, the participants who had their blood pressure measured immediately before the study.

Participants overwhelmingly agreed that the program should be recommended to a friend or relative. Following the study, after presentation of the study progress, all 6 communities supported continuing the program. Only 20 of these participants did not have any blood pressure readings beyond 2 months of enrolling into the study.

An evaluation form was provided with open-ended questions. The results of the Participant Evaluation Form for SMS text messages are found in Table 1.

On the questionnaire, 168 of 187 (89.8%) participants responded that the messages were clear and felt that they made sense. For 137 of 187 (73.2%) participants, the timing and frequency were satisfactory, supporting the process of sending messages on Mondays and Thursdays at 11 am. In response to a probe about privacy concerns about having their blood pressure readings and other health messages sent to their phones, 92 of 121 respondents (76.0%) felt that this was not a problem. Behavior changes noted by participants attributed to the SMS text messages included increased exercise, diet changes to give up sweetened beverages, more awareness of sodium and reading product labels, smoking cessation, and appreciation of the stress reduction messages. In 1 community, the mental health worker shared that their clients were more accessible and receptive if they were participating in the study. Participants also expressed that they felt supported, particularly if they were taking medications and had a greater understanding. Those not on medication noted that they would be more comfortable starting therapy if necessary. Almost one-third of participants shared the messages with family and friends and looked at the texts at their convenience.

An early complaint from the mobile phone app users was the potential to keep the last participant's name open on the app, when the next participant was ready to have their blood pressure assessed. This would lead to the next participant's blood pressure being assigned to the previous participant, and it became apparent when the next participant did not receive a confirmatory SMS text message with the blood pressure reading. To address this, the instructions on how to use the app when measuring patients one after another were revised, and the users were updated on the new process. The most frequent complication was a failure to transmit readings by Bluetooth transmission from the monitors to the BlackBerry phones. This was found to be because of the BlackBerry operating system software in older devices, which did not update automatically. This required resetting the device. Some noted that the BlackBerry mobile phone's version of the DREAM-GLOBAL app was not user friendly—the BlackBerry screens were small, and some parts of the app were difficult to access. In addition, occasional failure of BlackBerry function because of extreme cold or trouble with battery charge was observed. In the early months of study implementation, it was discovered that some BlackBerry devices had not been configured with the correct access point name (APN) setting for the service provider; APN settings were then reconfigured correctly through a troubleshooting process. For Tanzania, none of these problems occurred with the Android software and devices. Updates to the Android software and larger mobile phone screens made it easier to see the roster of participants. More space was allocated for

participant's names to make it easier to distinguish among participants with similar names. When a participant's phone was replaced, the new number had to be input into the registration software on the Web to ensure the participant continued to receive SMS text messages.

Table 1. Evaluation questions: technology and processes.

Target	Intervention	Technology
Patients	Can the patient use the technology, affordability, accessibility?; Is the server/software successfully sending the correct messages?; Training and support required	Fidelity, dose, reach; Recruitment and retention of patients
Providers: community health workers, doctors	Communication and collaboration between health staff and research team; Patient enrollment; Management of hypertension according to clinical practice guidelines?	What training and support were required?; Was study information making it to the patient chart?
Community	Are the communities engaged in the research?; Awareness of stakeholders with the study and willingness to promote	Do the communities have sufficient technology and bandwidth?

Community Health Workers

The community health workers were asked to provide feedback on their overall experience with the program, from site training to the use of technology. Data are summarized in Table 2.

The community health workers and nurses noted that they were seeing their clients participating in the study more frequently and that they became more comfortable with blood pressure measurement. In addition, participants knew their blood pressure numbers and were better able to address other health care issues, particularly related to mental health issues. They felt that the program led them to have better communication with their physicians and that they had greater confidence about the appropriate therapy for hypertension and knowing when action was needed. They also expressed appreciation for the site visits and ongoing telephone conference communication follow-up. The technological training was well received by the community health workers; in follow-up interviews after the training, almost all either agreed or strongly agreed that the training was helpful and gave them confidence in their ability to perform their role. The community health worker feedback from the site initiation and training is summarized in Table 3 (from 11 community health workers trained across the 6 communities).

After completion of the study, site visits with band leadership and community health leadership and elders were conducted to assess interest and for approvals to continue the program. A total of 5 of the communities remained engaged in the research. A total of 1 community did not allow follow-up with the community leadership during the study, but it did welcome a poststudy review and supported continuing the program after the study completion.

Health care providers shared that they noted that their patients participating in the study appreciated being contacted on their cell phones by the SMS text messages. They felt that their patients were more informed and engaged in their disease management and more likely to follow-up with their physicians. Many were surprised to see that their patients were calling them and making appointments, as per the instructions from their cell phone. It was noted that the white coat effect for blood pressure was identified through the community measurement of blood pressure, and more people, who would otherwise not visit, visited the health center.

Table 2. Results of participant evaluation for SMS text messages.

Questions and answers	Values, n (%)
Were the messages you received clear enough? Did they make sense to you? (N=187)	
Yes	168 (89.8)
No	19 (10.1)
Were the timing and frequency of messages okay for you? (N=187)	
Yes	137 (73.2)
No	42 (22.4)
__[a]	8 (4.2)
Was your experience with the phone positive? (N=179)	
Yes	102 (56.9)
No	36 (20.1)
—	41 (22.9)
Did you have any concerns about receiving health messages on your phone (ie, privacy/confidentiality)? (N=121)	
Yes	7 (5.7)
No	92 (76.0)
—	22 (18.1)
What did you like best about receiving SMS text messages? (N=144)	
Information/content	68 (47.2)
Motivation/reminders	43 (29.8)
Easy to read and understand	19 (13.1)
Useful/came in handy	10 (6.9)
Felt cared for	4 (2.7)
What did you like least about receiving text messages? (N=52)	
Messages too repetitive	20 (38.4)
Disliked content	11 (21.1)
Could not use phone	8 (15.4)
Messages too long	6 (11.5)
Messages too frequent	3 (5.7)
Generic messages	2 (3.8)
Text was too small	2 (3.8)
Would you recommend participating in this program to a friend or relative with hypertension? (N=169)	
Yes	165 (97.6)
No	3 (1.7)

[a]Not applicable or participants did not receive messages.

Table 3. Community health workers' feedback from site training and initiation of study (N=11).

Community health workers' excerpts	Strongly disagree, n (%)	Disagree, n (%)	Neutral, n (%)	Agree, n (%)	Strongly agree, n (%)
Uptake of knowledge					
I now have a good understanding of my role in the DREAM-GLOBAL[a] study	0 (0)	0 (0)	0 (0)	4 (36)	7 (64)
Uptake of skills					
I feel confident that I will be able to explain DREAM-GLOBAL to community members	1 (9)	0 (0)	1 (9)	4 (36)	5 (46)
I feel confident in how to use the Manual of Operations for screening and enrolling patients	0 (0)	0 (0)	0 (0)	7 (64)	4 (36)
I feel confident in how to identify, enroll, and register participants into the study	1 (9)	0 (0)	1 (9)	5 (46)	4 (36)
I feel confident in using Canadian Hypertension Education Program (Hypertension Canada) guidelines to take and assess blood pressure readings	0 (0)	0 (0)	0 (0)	6 (54)	5 (46)
I feel confident that I will be able to take and submit blood pressure readings using the BlackBerry device	0 (0)	0 (0)	1 (9)	6 (55)	4 (36)
I feel that I know how to use the safety protocols and process for submitting case report forms	0 (0)	0 (0)	0 (0)	8 (73)	3 (27)
This training was a good way to exchange information and learn about the DREAM-GLOBAL study	1 (9)	0 (0)	0 (0)	2 (18)	8 (73)

[a]DREAM-GLOBAL: Diagnosing hypertension—Engaging Action and Management in Getting Lower Blood Pressure in Indigenous and low- and middle-income countries.

Community

Overall, despite some initial issues, the technology performed as hoped in delivering the SMS text messages to patients, which garnered a positive reception from the communities involved. It also demonstrated that communities had enough technological resources to manage the study. The fact that the community health worker time to manage the study in each community was granted by the health leadership indicates the commitment and engagement of each community. All communities involved requested to have the program continue after the study's completion.

Discussion

Principal Findings

DREAM-GLOBAL was designed to increase the capacity for affordable, evidence-based, guidelines-driven hypertension management interventions at the patient, provider, and community level. This study demonstrated that information could be obtained and transmitted with fidelity from participants in remote settings. It also confirmed that patients were willing and able to participate in the program [18]. It was also possible for community health care providers, including nonmedical health workers, to train study participants to successfully use the technology. The pairing of Bluetooth and wireless technology enabled the secure, reliable, and widespread transmission of information from participants to their health care provider, including the transmission of guideline-based SMS text messages to study participants.

Insights

Community health workers and professional health care providers found that the technology was able to enhance their communication with participants and between each other. The process assisted with participant engagement in their own health care, which has also been demonstrated [19]. Training and support requirements were pragmatic. With achievement of improved blood pressure control, there is at least indirect evidence that study information was positively impacting on participant health care [16]. A theme emerging from community health workers was that the DREAM-GLOBAL technology could be more user friendly. For example, the Blackberry operating system was outdated and often had to be reset. The Android mobile phones and Android app used in Tanzania had much fewer problems. The BlackBerry mobile phone even failed to work in extreme cold weather.

The DREAM-GLOBAL system used minimal resources, requiring only the server, devices, and bulk SMS transmission. Transmission of blood pressure data has also recently been validated [20]. This system can easily be scaled to additional participants, with minimal additional costs. Assignment of blood pressure measurement to the community health worker addresses the realities of health care worker shortages in rural and remote communities. Shifting measurement from a health care provider to community health workers is a way to strengthen and expand community resources. The technology therefore linked the community health center, community health worker, home and community care nurses, and local physicians, thus supporting members within the circle of care.

Conclusions

The DREAM-GLOBAL study has resulted in an innovative system that has the ability to provide blood pressure screening and connect patients to health care providers in the community who can diagnose and treat hypertension. The technical platform provided a solution for hypertension management in low-resource settings, with the use of task shifting in blood pressure measurement and meaningful SMS text messages that promote health behavior change. The system has the potential to further improve patient care, as it can be easily adapted to regional needs and scaled up to include other aspects of chronic disease management.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Circle of care diagram for Diagnosing hypertension—Engaging Action and Management in Getting Lower Blood Pressure in Indigenous and low- and middle-income countries.
[PNG File , 17 KB - mhealth_v7i12e12639_app1.png]

Multimedia Appendix 2
Combined training documents for Diagnosing hypertension—Engaging Action and Management in Getting Lower Blood Pressure in Indigenous and low- and middle-income countries.
[PDF File (Adobe PDF File), 2880 KB - mhealth_v7i12e12639_app2.pdf]

Multimedia Appendix 3
Diagnosing hypertension—Engaging Action and Management in Getting Lower Blood Pressure in Indigenous and low- and middle-income countries short message service text messages.
[PDF File (Adobe PDF File), 148 KB - mhealth_v7i12e12639_app3.pdf]

Multimedia Appendix 4
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 3255 KB - mhealth_v7i12e12639_app4.pdf]

References

1. Lim SS, Vos T, Flaxman AD, Danaei G, Shibuya K, Adair-Rohani H, et al. A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2224-2260 [FREE Full text] [doi: 10.1016/S0140-6736(12)61766-8] [Medline: 23245609]
2. Foulds HJ, Warburton DE. The blood pressure and hypertension experience among North American indigenous populations. J Hypertens 2014 Apr;32(4):724-734. [doi: 10.1097/HJH.0000000000000084] [Medline: 24609208]
3. Gionet L, Roshanafshar S. Statistics Canada. 2015. Health at a Glance URL: https://www150.statcan.gc.ca/n1/pub/82-624-x/2013001/article/11763-eng.htm [accessed 2019-09-11]
4. Government of Canada. 2014. A Statistical Profile on the Health of First Nations in Canada: Determinants of Health, 2006 to 2010 URL: https://www.canada.ca/en/indigenous-services-canada/services/first-nations-inuit-health/reports-publications/aboriginal-health-research/statistical-profile-health-first-nations-canada-determinants-health-2006-2010-health-canada-2014.html [accessed 2019-09-11]
5. Bobrow K, Brennan T, Springer D, Levitt NS, Rayner B, Namane M, et al. Efficacy of a text messaging (SMS) based intervention for adults with hypertension: protocol for the StAR (SMS Text-message Adherence suppoRt trial) randomised controlled trial. BioMed Central Public Health 2014 Jan 11;14:28 [FREE Full text] [doi: 10.1186/1471-2458-14-28] [Medline: 24410738]
6. Cole-Lewis H, Kershaw T. Text messaging as a tool for behavior change in disease prevention and management. Epidemiol Rev 2010;32:56-69 [FREE Full text] [doi: 10.1093/epirev/mxq004] [Medline: 20354039]
7. Leon N, Surender R, Bobrow K, Muller J, Farmer A. Improving treatment adherence for blood pressure lowering via mobile phone SMS-messages in South Africa: a qualitative evaluation of the SMS-text Adherence SuppoRt (StAR) trial. BioMed Central Fam Pract 2015 Jul 3;16:80 [FREE Full text] [doi: 10.1186/s12875-015-0289-7] [Medline: 26137844]
8. Yeates K, Campbell N, Maar MA, Perkins N, Liu P, Sleeth J, et al. The effectiveness of text messaging for detection and management of hypertension in indigenous people in Canada: protocol for a randomized controlled trial. J Med Internet Res Res Protoc 2017 Dec 19;6(12):e244 [FREE Full text] [doi: 10.2196/resprot.7139] [Medline: 29258978]

9. Maar M, Yeates K, Barron M, Hua D, Liu P, Lum-Kwong MM, et al. I-RREACH: an engagement and assessment tool for improving implementation readiness of researchers, organizations and communities in complex interventions. Implement Sci 2015 May 4;10:64 [FREE Full text] [doi: 10.1186/s13012-015-0257-6] [Medline: 25935849]
10. Maar MA, Yeates K, Perkins N, Boesch L, Hua-Stewart D, Liu P, et al. A framework for the study of complex mHealth interventions in diverse cultural settings. JMIR Mhealth Uhealth 2017 Apr 20;5(4):e47 [FREE Full text] [doi: 10.2196/mhealth.7044] [Medline: 28428165]
11. -. ICH Harmonised Tripartite Guideline: Guideline for Good Clinical Practice. In: International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. 1997 Presented at: ICH'97; July 17, 1996; Japan URL: https://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6/E6_R1_Guideline.pdf
12. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. J Am Med Assoc 2013 Nov 27;310(20):2191-2194. [doi: 10.1001/jama.2013.281053] [Medline: 24141714]
13. Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council. Tri-Council Policy Statement Ethical Conduct for Research Involving Humans. Canada: Government of Canada; 2014.
14. Boffa J, King M, McMullin K, Long R. A process for the inclusion of Aboriginal people in health research: lessons from the determinants of TB transmission project. Soc Sci Med 2011 Mar;72(5):733-738. [doi: 10.1016/j.socscimed.2010.10.033] [Medline: 21316828]
15. Leung AA, Daskalopoulou SS, Dasgupta K, McBrien K, Butalia S, Zarnke KB, Hypertension Canada. Hypertension Canada's 2017 guidelines for diagnosis, risk assessment, prevention, and treatment of hypertension in adults. Can J Cardiol 2017 May;33(5):557-576. [doi: 10.1016/j.cjca.2017.03.005] [Medline: 28449828]
16. Tobe SW, Yeates K, Campbell NR, Maar MA, Perkins N, Liu PP, et al. Diagnosing hypertension in indigenous Canadians (DREAM-GLOBAL): a randomized controlled trial to compare the effectiveness of short message service messaging for management of hypertension: main results. J Clin Hypertens (Greenwich) 2019 Jan;21(1):29-36 [FREE Full text] [doi: 10.1111/jch.13434] [Medline: 30474909]
17. Maar MA, Yeates K, Toth Z, Barron M, Boesch L, Hua-Stewart D, et al. Unpacking the black box: a formative research approach to the development of theory-driven, evidence-based, and culturally safe text messages in mobile health interventions. JMIR Mhealth Uhealth 2016 Jan 22;4(1):e10 [FREE Full text] [doi: 10.2196/mhealth.4994] [Medline: 26800712]
18. Green BB, Anderson ML, Ralston JD, Catz S, Fishman PA, Cook AJ. Patient ability and willingness to participate in a web-based intervention to improve hypertension control. J Med Internet Res 2011 Jan 20;13(1):e1 [FREE Full text] [doi: 10.2196/jmir.1625] [Medline: 21371993]
19. Hacking D, Haricharan HJ, Brittain K, Lau YK, Cassidy T, Heap M. Hypertension health promotion via text messaging at a community health center in South Africa: a mixed methods study. JMIR Mhealth Uhealth 2016 Mar 10;4(1):e22 [FREE Full text] [doi: 10.2196/mhealth.4569] [Medline: 26964505]
20. Wetterholm M, Bonn SE, Alexandrou C, Löf M, Lagerros YT. Validation of two automatic blood pressure monitors with the ability to transfer data via Bluetooth. J Med Internet Res 2019 Apr 17;21(4):e12772 [FREE Full text] [doi: 10.2196/12772] [Medline: 30994459]

Abbreviations

APN: access point name
DREAM-GLOBAL: Diagnosing hypertension—Engaging Action and Management in Getting Lower Blood Pressure in Indigenous and low- and middle-income countries
LMICs: low- and middle-income countries

Edited by G Eysenbach; submitted 31.10.18; peer-reviewed by A Karakosta, K Bobrow, S Lear; comments to author 31.03.19; revised version received 07.08.19; accepted 31.08.19; published 09.12.19.

Please cite as:
Barsky J, Hunter R, McAllister C, Yeates K, Campbell N, Liu P, Perkins N, Hua-Stewart D, Maar MA, Tobe SW
Analysis of the Implementation, User Perspectives, and Feedback From a Mobile Health Intervention for Individuals Living With Hypertension (DREAM-GLOBAL): Mixed Methods Study
JMIR Mhealth Uhealth 2019;7(12):e12639
URL: https://mhealth.jmir.org/2019/12/e12639
doi:10.2196/12639
PMID:31815678

Original Paper

Development and Local Contextualization of Mobile Health Messages for Enhancing Disease Management Among Community-Dwelling Stroke Patients in Rural China: Multimethod Study

Enying Gong[1,2*], MSc; Wanbing Gu[1*], MSc; Erdan Luo[1,3], MSc; Liwei Tan[1,4], MSc; Julian Donovan[5], MD; Cheng Sun[1], PhD; Ying Yang[1], MSc; Longkai Zang[1,6], BSc; Peng Bao[1,7], MBBS; Lijing L Yan[1,8], PhD, MPH

[1]Global Health Research Center, Duke Kunshan University, Kunshan, China

[2]School of Population and Global Health, The University of Melbourne, Melbourne, Australia

[3]Institute of Chinese Medical Sciences, University of Macau, Taipa, Macau, China

[4]School of Public Health, Fudan University, Shanghai, China

[5]Northumbria Healthcare National Health Service Foundation Trust, Wallsend, United Kingdom

[6]Second School of Clinical Medicine, Wuhan University, Wuhan, China

[7]Ningxia Medical University, Yinchuan, China

[8]Duke Global Health Institute, Durham, NC, United States

[*]these authors contributed equally

Corresponding Author:
Lijing L Yan, PhD, MPH
Global Health Research Center, Duke Kunshan University
No 8 Duke Road
Kunshan,
China
Phone: 86 512 3665 7057
Email: lijing.yan@dukekunshan.edu.cn

Abstract

Background: Rural China has experienced an increasing health burden because of stroke. Stroke patients in rural communities have relatively poor awareness of and adherence to evidence-based secondary prevention and self-management of stroke. Mobile technology represents an innovative way to influence patient behaviors and improve their self-management.

Objective: This study is part of the System-Integrated Technology-Enabled Model of Care (the SINEMA trial) to improve the health of stroke patients in resource-poor settings in China. This study aimed to develop and pilot-test a mobile phone message–based package, as a component of the SINEMA intervention.

Methods: The SINEMA trial was conducted in Nanhe County, Hebei Province, China. A total of 4 villages were selected for pretrial contextual research and pilot study. The 5 stages for developing the mobile phone messages were as follows: (1) conducting literature review on existing message banks and analyzing the characteristics of these banks; (2) interviewing stroke patients and caregivers to identify their needs; (3) drafting message contents and designing dispatching algorithms for a 3-month pilot testing; (4) collecting feedback from pilot participants through questionnaire survey and in-depth interviews on facilitators and barriers related to their acceptance and understanding of messages; and (5) finalizing the message-based intervention based on participants' feedback for the SINEMA trial.

Results: On the basis of 5 existing message banks screened out of 120 papers and patients' needs identified from 32 in-depth interviews among stroke patients and caregivers, we developed a message bank containing 224 messages for a pilot study among 54 community-dwelling stroke patients from 4 villages. Of 54 participants, 51 (response rate: 94.4%) completed the feedback survey after receiving daily messages for 3 months. Participants' mean age was 68 years (SD 9.2), and about half had never been to school. We observed a higher proportion of participants who were in favor of voice messages (23/42, 54%) than text messages (14/40, 35%). Among participants who received voice messages (n=43) and text messages (n=40), 41 and 30, respectively, self-reported a full or partial understanding of the contents, and 39 (39/43, 91%) and 32 (32/40, 80%), respectively, rated the messages as helpful. Analyses of the 32 interviews further revealed that voice messages containing simple and single-theme

content, in plain language, with a repeated structure, a slow playback speed, and recorded in local dialect, were preferred by rural stroke patients. In addition, the dispatching algorithm and tools may also influence the acceptance of message-based interventions.

Conclusions: By applying multiple methodologies and conducting a pilot study, we designed and fine-tuned a voice message–based intervention package for promoting secondary prevention among community-dwelling stroke patients in rural China. Design of the content and dispatching algorithm should engage both experts and end users and adequately consider the needs and preferences of recipients.

(JMIR Mhealth Uhealth 2019;7(12):e15758) doi:10.2196/15758

KEYWORDS

phone messages; stroke; secondary prevention; rural population; China

Introduction

Stroke Secondary Prevention in Rural China

Stroke is the leading cause of death and disability in China. According to recent estimates, stroke-related deaths reached around 1.1 million in China, accounting for 23% of total deaths in rural areas and 21% in urban areas [1]. In addition, it is estimated that there are 12.4 million stroke patients aged older than 40 years living in China, which causes an enormous burden to the health care system and society [2]. Compared with urban areas, rural China has experienced a more rapidly increasing burden, with the age-standardized prevalence of stroke increasing 2.5-fold over the past three decades [3].

Stroke patients in rural areas have relatively low awareness of and poor adherence to the secondary prevention and self-management of stroke. Effective secondary prevention of stroke, including lifestyle modification and a combination of medical therapy (eg, antiplatelet, antilipid, and antihypertensive therapy), has been well studied as the *best buy* for stroke patients [4]. However, adherence to these effective preventive strategies is poor among stroke patients, with more than half of patients reported discontinuing their secondary prevention medications within 3 months of hospital discharge [5,6]. In addition, the fragmented primary health care system and limited capacities of health care providers in rural China further restrain the availability and quality of the health care services that stroke patients can gain access to [7]. Therefore, there is an urgent need to develop a low-cost and effective strategy to increase the awareness of and adherence to secondary prevention measures of stroke in rural China.

Message-Based Intervention

With the development of mobile communication technology and widespread adoption of mobile phones, mobile health (mHealth) has the potential to empower patients to improve their self-management of chronic conditions and related risk factors [8,9]. A previous literature review illustrated that compared with high-income countries where more advanced mobile technologies are used, low- and middle-income countries have relied more on phone messages for intervention delivery because of its accessibility, lower cost, and better acceptance [10]. Review studies and meta-analysis have illustrated the modest effect of message-based interventions in promoting medication adherence and lifestyle modifications and improving health outcomes among patients with chronic conditions, although further studies with longer follow-up period and larger scale were suggested [8,9]. The evidence on message-based studies targeted at stroke patients was very limited, with a few studies focused on recently discharged stroke patients [11], but almost none on community-dwelling patients. In addition, there were only a few studies that tried to understand the feasibility and impact of message-based interventions among patients with cardiovascular diseases in China, and most of these studies focused on urban patients [12]. The acceptability and feasibility of applying message-based interventions among rural patients still need to be further investigated.

In addition, some of the previous reviews attempted to examine why some of the message-based interventions work or do not work. Factors related to message contents and dispatching algorithms, such as message frequency, content personalization, and dispatching approach in 2 ways, were examined, but the results were mixed [13,14]. Studies also pointed out that lacking theoretical constructs may lead to the difficulty in understanding the mechanisms behind the message-based interventions in chronic disease management. Watkins et al conducted a realist review by mapping the intervention components with psychological theories and frameworks [15]. The authors found that information provided via messages to increase patients' knowledge could motivate patients to believe the relevance of the issue to their health and the potential risk, which may reduce the threat as suggested by the health belief model [16]. Some other psychological behavior change theories, such as the transtheoretical model or a taxonomy model of behavior change techniques, including goal settings and self-regulation, were also applied in the design of messages [15,17].

Objectives

This study is part of the System-Integrated Technology-Enabled Model of Care (the SINEMA trial) to improve the health of stroke patients in resource-poor settings in China. The protocol of the SINEMA trial has already been published [18]. This study aimed to describe the development and contextualization process of this message-based intervention package from field-based research and pilot studies. The results of the effectiveness of the message-based intervention in the SINEMA main trial is beyond the scope of this paper.

Methods

Study Setting

The SINEMA trial and this pilot study were conducted in Nanhe County, Hebei Province, China, with an intention to be adapted

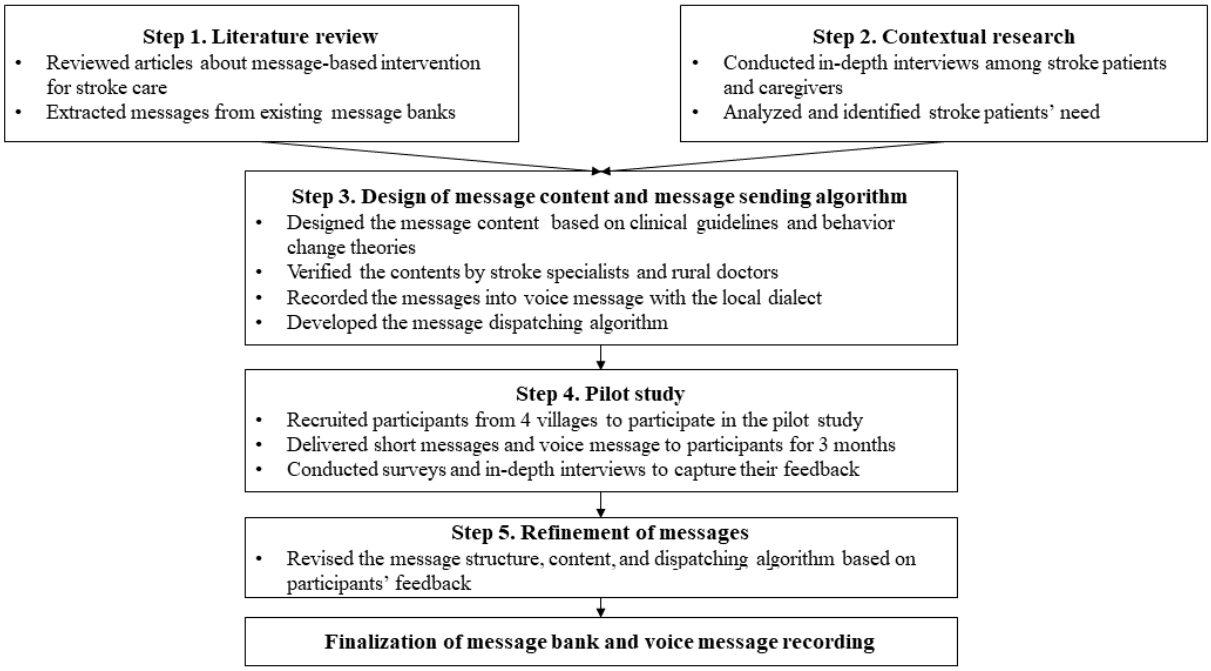
to other resource-limited settings. Nanhe County is an economically poor provincial county with an annual disposable income per capita of 11,030 RMB (less than half of the national average) [19]. Although no specific estimation on stroke burden is available in this county, previous studies have shown a high burden of stroke in this province and rural settings compared with other parts of China [20].

To ensure the study design and blind generalization of the intervention model, we aimed to identify 4 villages out of 218 villages within the geographical area to implement the field-based contextual research and pilot study, and these 4 villages were not eligible to be included in the trial. We received assistance from the local Center for Disease Control and Prevention to identify 4 villages where village doctors were willing to participate. These 4 villages comprised 2 villages from a township that had an insufficient number of villages to be eligible for the main trial and 2 villages from eligible township but with insufficient residents to meet the inclusion criteria of the main trial. Therefore, we were able to identify typical villages but also leave the potential eligible large villages for the main trial.

Message Design Process

We used multiple methodologies, including literature review, expert consultation, qualitative in-depth interviews, and field-based pilot study followed by surveys and interviews, in developing the messages to ensure that the message contents were built based on research evidence and in line with the clinical guidelines and were suitable for the local context. The development process for the mobile phone message banks consisted of the following 5 stages: (1) conducting a literature review on existing message banks targeting people with stroke, (2) interviewing stakeholders to identify the needs of stroke patients, (3) creating and designing the message contents and message sending algorithm, (4) conducting pilot testing of the messages among patients in 4 villages, and (5) refining and finalizing the message bank and sending algorithm based on the lessons learned from the pilot study. Figure 1 outlines the design process.

Figure 1. Overview of key stages of message development.



Stage 1. Literature Review on Existing Message Banks

To assist the design of the message bank, we conducted a literature review based on the PubMed database to search for existing text message banks published between October 1, 2011, and October 1, 2016. Our own design of the messages began in November 2016. Key search terms included stroke and text messages (for search terms and review criteria, see Multimedia Appendix 1). A total of 2 reviewers screened the searched papers, and a snowball review by reviewing titles and abstracts of references of searched papers was also conducted to increase the scope of the review. Existing message banks were extracted and translated into Chinese as a reference for the design.

Stage 2. Contextual Research

To identify the needs of stroke patients, we visited the field 3 times over the period of 5 months from May to September 2016 and conducted in-depth interviews among stakeholders, including stroke patients and family caregivers. Semistructured interview guides were developed, and questions in the interview included patients' health status and needs for managing their conditions. Stroke patients were identified with the assistance

of village doctors with the criteria that the participants were aged 18 years or older, had a history of stroke diagnosed at the county- or higher-level hospitals, were in clinically stable condition, had basic communication abilities, and were willing to participate in the study. Stroke patients' family members who were at home while the research team visited and were willing to participate were also interviewed separately. The number of participants interviewed was determined based on the saturation theory, whereby participants were recruited until no further new information was acquired through the interview process. We expect to interview about 4 patients from each village.

Stage 3. Design of Message Content and Message Sending Algorithm

We designed the contents of messages on the basis of the existing message banks from the literature review, the identified needs of stroke patients from contextual research, and the behavior change theories such as the health belief model and the transtheoretical model [16,21]. First, the structure and focus of the message bank were identified so that messages could tap into key dimensions based on patients' needs. We then grouped the existing messages into 6 categories based on their contents, including management of metabolic risk factors, medication adherence, tobacco and alcohol control, dietary change, exercise and rehabilitation, and psychological support. The content of the messages was modified to be suitable for stroke patients in rural China. The messages were then verified by stroke specialists working in tertiary hospitals and recorded in the local dialect.

We also designed the message dispatching algorithm by setting the sending time and frequency based on the daily habits of stroke patients. This algorithm was linked with a digital health management system that was designed to support the delivery of the digital health components of SINEMA intervention.

Stage 4. Pilot Study

A pilot study was conducted in 4 villages to test the SINEMA intervention model, including the acceptance of the message-based intervention among stroke patients. Participants who were aged older than 18 years, had a history of stroke but in a clinically stable condition, and able to communicate via mobile phone were eligible to participate in the study. In each village, village doctors screened stroke patients in their villages and provided the list to the study team. Participants were invited by village doctors and recruited by the research team. Before the commencement of the pilot study, a structured questionnaire, including questions on participants' demographic characteristics and disease history, designed based on previous studies [22,23], was administered by the research team through face-to-face interviews.

During the 3-month pilot study, participants and their caregivers received text messages at 3 pm every 2 days, and the participants also received voice messages with the same content at 7 pm on alternating days when they did not receive the text messages. At the end of the pilot study, participants completed a short questionnaire administered by their village doctors. Questions in the survey included whether they had read or listened to the messages, their understanding of the contents of the messages, and their perspectives on the helpfulness of these messages. In-depth interviews were also conducted by the research team among selected participants to seek detailed feedback.

Stage 5. Refinement of the Messages

After the pilot study, the research team summarized the feedback from participants and refined the message contents and message dispatching algorithm. To optimize the acceptance and understanding of the messages, the research team revised the language in each message based on participants' feedback and preference. Village doctors and physicians from county hospitals were invited to verify the messages to ensure that local contexts were taken into full consideration, and the terms used in the messages were understandable by the target population.

Data Analysis

To apply the interview results to message development, we analyzed the interview transcripts using thematic analysis approach, which is a method for identifying, analyzing, and reporting patterns within data and is able to both reflect the reality and explain the undersurface meaning [24]. Following the thematic analysis approach, we conducted the following steps: the research team first read the transcripts of the interviews to be familiar with the data, then generated the key concepts for coding based on the interview guide. Data were coded line by line and grouped into categories using NVivo 11 software (QSR International). Each category of coding was then further reviewed and examined. The groups of key themes were refined before finalizing the major themes. Discussions within the research team were organized after each key step to ensure the trustworthiness and define the final themes. Quotations used in this paper were translated from Chinese to English and then back translated into Chinese to increase the accuracy of the interpretation.

A descriptive analysis was conducted for survey data. Continuous variables were reported as mean and SD if normally distributed or as median with 25th and 75th percentiles if not. Counts and proportions were reported for categorical variables. Survey data were analyzed using STATA software (version 15; StataCorp LLC).

Ethical Considerations

All research activities, including both contextual research and pilot study, were approved by the institutional review board of Duke Kunshan University. All participants in the study provided informed consent before taking part in the study.

Results

Summary

In this section, we first describe the main findings from stages 1, 2, and 4 explained above: the literature review, contextual research, and the pilot study. We then describe how we developed and refined the message bank based on these findings (stages 3 and 5). Finally, we briefly introduce the actual -message-based intervention adopted in the trial.

Message Bank Review and Extraction

After the systematic search and snowball review of cited papers, we found 5 existing SMS banks. Using these 5 completed SMS banks, we translated messages into Chinese, resulting in a total of 224 messages as the basis for further design. Multimedia Appendix 1 shows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses diagram of the literature review and the contents and features of these 5 identified message banks.

Results From Contextual Research on Stroke Patients' Needs

During the 3 contextual research visits, we interviewed 22 stroke patients and 10 caregivers. Through analyses of these interviews, we identified the major needs of patients and the required contents of messages.

Content 1: General Knowledge About Stroke Secondary Prevention

Although almost all patients involved in the interview expressed a strong desire to recover from stroke, most of them had limited awareness and knowledge of secondary prevention of stroke, which was insufficient and not comprehensive enough to help them overcome all challenges that they faced in managing their conditions. There were very few sources available to patients regarding evidence-based information related to stroke secondary prevention:

> *No physicians told us what we should pay attention to when I was in the hospital. They [physicians] did not care about it. Only the village doctor told us some [information].* [Patient #8]

Village doctors, considered as the key source of information and the first contact of care, also had the minimal capacity to provide intensive health education to patients because of their existing high workload of serving more than 1000 residents in the village. Therefore, general knowledge related to stroke secondary prevention was required in the message bank.

Content 2: Promoting Adherence to Guideline-Recommended Activities

In the interview, many patients admitted that they were not able to adhere to all secondary prevention medications. The main reasons for nonadherence to medication use included their forgetfulness because of declining memory, impaired mental function, or busy working lives. We found that most of the patients we interviewed lived alone or with their spouse alone, without any other caregivers to remind them to take their medications:

> *I often forgot to take medicine. He [the patient's husband] usually put all the medicines on my table and reminded me to take them...If he also forgot, we would forget about it totally.* [Patient #11]

Considering the lack of reminders and the suboptimal adherence among most stroke patients, we decided that the message should act as a reminder for patients to promote their adherence to treatment.

Content 3: Providing Specific Guidance on Recommendations for Physical Activity

Through the interview, we also found that, even for people who were aware of the importance of stroke secondary prevention, they had little knowledge of how to perform physical activities or rehabilitation. For example, almost no patients knew the proper frequency and intensity of physical activities that they should achieve, nor the strategies to protect themselves from injury during exercise. In addition, walking was the most common way of exercise for most patients involved in the interview, and few patients knew how to train their upper limbs in their physical activities to promote their recovery of daily functions:

> *He [the physician] just told me to walk more.* [Patient #2]

> *I do not have other forms of exercise. I just walk for a while every morning in the village...I usually sit or sleep at home for most of the remaining day.* [Patient #4]

Therefore, we believe that providing more tips and guidance on how to undertake the guideline-recommended physical activity and rehabilitation will match the needs of stroke patients.

Results From the Pilot Study

A total of 54 stroke patients from 4 villages in 2 townships in Nanhe County participated in the pilot study from February to May 2017. Among them, 51 participants (response rate: 94.4%) completed the feedback survey after the pilot study. Table 1 shows the characteristics of the participants. The average age of participants was 68.0 years (SD 9.2). Most of them were males (33/51, 65%), aged older than 65 years (33/51, 53%), married (40/51, 78%), with no education or having never graduated from primary school (27/51, 53%), had suffered ischemic stroke (45/51, 88%), and experienced more than 1 stroke event (31/51, 61%).

Among participants who responded to the feedback survey, 40 (40/51, 78%) and 43 (43/51, 84%) stated that they had successfully received text messages or voice messages during the pilot study, respectively. Table 2 shows the self-reported acceptance and understanding of messages for these participants. We observed a higher proportion of participants who were in favor of voice messages: 54% (23/43) of participants listened to the voice messages all the time, whereas only 35% (14/40) of participants self-reported their frequent text-message reading. Of 43 participants, 41 (95%) who listened to the voice messages self-reported that they could understand the content entirely or partially, whereas 30 of 41 (75%) participants reported that they were able to fully or partially understand the contents in the text messages. In addition, 39 of 43 (91%) participants who had ever received voice messages and 32 of 40 (80%) participants who had ever received text messages rated that the messages were helpful. In addition, 27 participants reported that their family caregivers had received text messages, and 11 of them reported that their family members had read the messages to them.

Table 1. Demographic characteristics and disease history of participants who responded to the feedback survey (N=51).

Characteristics	n (%)
Demographic characteristics	
Gender	
Male	33 (65)
Female	18 (35)
Age group (years)	
≥65	33 (65)
<65	18 (35)
Marital status	
Married	40 (78)
Divorced, widowed, or unmarried	11 (22)
Education	
Never been to school	22 (43)
Less than primary school	5 (10)
Primary school	10 (20)
Primary high school	10 (20)
High school and above	4 (8)
Occupation	
Unemployed	7 (14)
Self-employed including doing farm work	39 (77)
Employed	4 (8)
Others	1 (2)
Main caregiver	
Spouse	35 (69)
Children	7 (14)
Others	9 (18)
Disease history	
Stroke type	
Ischemic stroke	45 (88)
Hemorrhage stroke	5 (10)
Cannot remember	1 (2)
Suffered more than 1 stroke event	
Yes	31 (61)
No	20 (39)

Table 2. Participants' acceptance and perceptions of voice messages and text messages (of 54 participants, 51 completed the feedback survey, 40 responded that they had received text messages, and 43 responded that they had received voice messages).

Acceptance and perception	Voice messages (N=43), n (%)	Text messages (N=40), n (%)
The frequency of reading/listening to the entire message		
All the time	23 (54)	14 (35)
Sometime	20 (47)	14 (35)
Never	0 (0)	12 (30)
Level of understanding toward message contents		
Entire message	18 (42)	12 (30)
A part of the message	23 (54)	18 (45)
Cannot understand at all	2 (5)	10 (25)
Whether think the messages were helpful		
Yes	39 (91)	32 (80)
No	3 (7)	4 (10)
Do not know	1 (2)	1 (3)
Missing	—[a]	3 (8)

[a]3 participants who received text messages did not respond to this question. No missing values among participants who received voice messages on this question.

Qualitative Results From the Pilot Study

After the pilot study, we interviewed 8 participants. Their feedback covered a variety of perspectives, including the content of the message, the form, and algorithms of message dispatching, and other barriers related to receiving and understanding the messages. This feedback guided our decision making in optimizing the message contents and sending the algorithm to fit the local context.

Message Contents

In line with the quantitative results, we also found that not all participants were able to understand the messages fully. Barriers to acceptance included the broad scope and complex content of the information covered in the messages, the relevance of messages toward their health conditions, the complexity of the terms and languages, and their own capacity to read and memorize information.

Many patients told us that they usually had poor memory and would easily forget things. They could usually understand the messages, but they would forget most of the contents after a short while if too much information was covered in a single message:

> *I could understand the message when I listened to it, but I would forget most of them quickly after I put down my phone.* [Patient #6]

In addition to the preference for simple information within a message, participants also reported that some terms in the messages were not in plain language, which made the messages more difficult to understand.

In terms of the impact of messages on helping them in managing the risk factors, some of the participants reported that not all the information was relevant to them. For example, messages related to smoking or drinking were only applicable to a small proportion of participants we interviewed:

> *He [the patient] drank every day before the stroke occurrence, but he quit after it. He doesn't dare to drink anymore.* [Spouse of Patient #2]

Message Dispatch Form and Algorithms

Almost all patients who participated in the interview showed greater interest in and spoke highly of the voice messages, whereas very few participants reported that they had read text messages. This was mainly because of participants' illiteracy in reading text or their technical capability in checking text messages on the phone:

> *My child bought a feature phone for me, and I know how to dial or pick up a call...I never read text messages. I am illiterate.* [Patient #2]

Participants also identified the potential factors that may facilitate their acceptance and understanding toward voice messages. Almost all participants involved in the interviews agreed that they preferred messages using the local dialect rather than standard Mandarin. Some of the participants also mentioned that a slower playback speed of the voice messages might improve their understanding of the message contents. In addition, participants recommended repeating the messages or extending the duration of messages so that they could better absorb the information when listening:

> *Sometimes I was doing my housework when I picked up the phone. She [the voice message] spoke too fast, and the message ended before I could concentrate my mind to listen to it.* [Patient #4]

In terms of the frequency and timing of message dispatch, most participants said that they were willing to receive the messages daily. In addition, participants suggested that receiving the

message in the morning would be more helpful, especially for reminding them of taking medicines:

> *I often forget the medicines I should take in the morning because I usually have housework to do. But I usually can remember to take the medication at night before going to bed.* [Patient #1]

> *I would like to take the medicines, but I inevitably forget to do it sometimes... It would be, of course, good if someone can remind me every day.* [Patient #2]

Phone Scams

From our interview, we also found some other factors that may influence the implementation of the message-based intervention. Patients reported that they would reject phone calls from unknown numbers because they were afraid that it could be a phone scam. The numbers used to send the voice messages were different every day at the beginning of the pilot study, so some of the participants were cautious about the messages received and refused to listen to the voice messages. Participants mentioned that it would be easier for them to get the voice messages if the same number was used.

Major Refinement of Messages

On the basis of the results from the pilot study, we made a series of decisions regarding the contents and dispatching algorithms of the messages to fit the local context and our target population better. Considering the prevalence of illiteracy and poor technical capacity in reading text messages, we decided to only use voice message delivery for the main study. Table 3 summarizes critical points where we optimized the message contents and algorithm.

To increase stroke patients' understanding of the messages, we simplified the themes, structure, and language of each message. We highlighted the focus on medication adherence and physical activities that were appropriate for most stroke patients. Only a small proportion of the total messages were supplemented by general information related to stroke and management of other risk factors. In addition, we further refined each message by retaining only one key piece of information and used the same structure for each message (the program name, followed by a reminder sentence alternating between taking medicine and exercise regularly, and then specific health education information). The application of the health belief model [16] and the transtheoretical model on the stage of change [21] was carefully considered in refining each health education information and the order of messages. Furthermore, we invited physicians in the township hospitals to review all messages and amend any clinical or academic terms into plain language that would be commonly used and understood by the target population. Multimedia Appendix 2 shows some examples of the revised messages.

We also improved the message recording and dispatch algorithm to better suit the preferences of the target population. We continued to record voice messages using the local dialect and selected a slower playback speed, with the message being repeated twice per message delivery. In addition, we decided not to dispatch any text messages but rather send out voice messages daily. We changed the message dispatch time from 7 pm to 7 am so that the messages can better remind patients of medication taking. Furthermore, we optimized our message dispatch system so that a single consistent sender phone number would be displayed on participants' phones.

Table 3. Comparison of key aspects of message optimization.

Aspects of message optimization	Pilot study	Main study
Message content		
Themes	A broad scope of information, including management of metabolic risk factors, medication adherence, tobacco and alcohol control, dietary change, physical activity, and psychological recovery	Focused on medication adherence and physical activity and supplemented with other information on metabolic risk factor management and stroke in general
Structure	Random structure for each message	Same structure for all messages: *program name + reminder sentence + health education information*
Key information	Multiple key information within 1 message	Single key information for 1 message
Languages	With some professional terms	Simple plain language
Verification	By stroke specialists in first-tier hospitals	By village doctors and township physicians
Health behavior change theory	Health belief model and the transtheoretical model (stage of change)	Health belief model and the transtheoretical model (stage of change)
Message recording		
Speed	Normal speaking speed	Slower than the normal speaking speed
Repeating	No repetition	Repeated once
Dialect	Local dialect	Local dialect
Message dispatch algorithms		
Text message	3 pm every 2 days	No text message
Voice message	7 pm every 2 days	7 am every day
Receiver	Patients and caregivers	Patients only
Senders	Random phone number	Single consistent phone number

Messages Delivered Through System-Integrated Technology-Enabled Model of Care Trial

A message bank containing 92 voice messages was developed to support the SINEMA main trial. These messages were repeated for 4 times to support the year-long intervention. The voice messages had been delivered to more than 600 stroke patients who participated in the SINEMA trial in rural China since July 2017. During the year-long intervention, a total of more than 100,000 messages were sent to participants and some caregivers in the intervention arm if they had a phone and were willing to receive the messages.

Discussion

Principal Findings

In this study, we applied a multiple methodological study approach to develop mHealth messages to promote secondary prevention of stroke among community-dwelling stroke patients in rural China. We created the initial messages based on existing message banks from previous studies; involved patients in the initial study design in capturing their needs; evaluated patients' acceptance through a 3-month pilot study, including quantitative surveys and in-depth interviews; and finally, refined and optimized the message contents and dispatching algorithm based on the feedback provided by participants.

This study is one of the few studies that have developed message-based intervention among people with cardiovascular diseases [25] and the first ever for stroke patients in rural China. Our study provides important information to researchers and policy makers about the feasibility of delivering voice message intervention for community-dwelling stroke patients in resource-limited settings. Our study further revealed that voice messages were favored over text messages among stroke patients in rural China, as voice messages do not require reading skills but only basic ability to receive phone calls. This finding is consistent with previous studies that demonstrated the preference for voice messages among the rural population in Bangladesh and Cambodia, who also had low rates of reading literacy [26,27]. Thus, we believe that voice messages have great potential to deliver interventions, especially among populations with low socioeconomic status.

Message-based interventions have been increasingly applied to promote chronic disease management. However, less attention had been paid to content design than its effectiveness. A number of studies have tested the effectiveness of phone-based interventions implemented among various populations worldwide [25,28,29], but only a few studies have described in detail the development process of message-based interventions and their lessons learned. Studies have found insufficient use of behavior change theories in mHealth messages and lack of involvement of the target population in the design process [30]. The modest effect and inconsistent results of message-based interventions suggest that we cannot overlook the design process. In our study, the health belief model [16] and the transtheoretical model (stage of change) [21] were embedded in the message contents by emphasizing the perceived threat of poor adherence to medication use, improving their perceived

benefits in behavior change in line with the secondary prevention recommendation, providing them cues to action and encouraging their maintenance throughout different stages. In addition, considering the education level and health literacy of stroke patients in rural China, we did not involve our target population in the initial drafting of messages as some other studies did [31], but instead, we conducted interviews to understand their needs, tested their acceptance through a pilot study, and refined the messages based on their feedback.

Our study also revealed that the design of messages should adequately consider the local context and characteristics of the target audience. We found that message structure, language, complexity and relevance, and repetition are factors that influence patients' acceptance and understanding. This finding is in line with some of the previous studies. For example, Muench and Baumel [32] analyzed the components of digital triggers in messages and differentiated triggers into 5 elements, including the sender, delivery approach, time of delivery, frequency, and trigger contents. Our study results also showed that a reliable sender, an accepted form of the message dispatch, optimal timing based on patients' daily routine, and simple but relevant key messages were essential for the acceptance of message-based interventions. In addition, we also discovered certain factors that are specific to voice messages. For example, the speed of audio playback, the pattern of repetition, and the use of dialect may also improve audiences' understanding of message contents [26].

Limitations and Strengths

Our study has some limitations. First, in this pilot study, the acceptance of messages was measured using a self-designed short survey, rather than actual statistics from the message dispatch system in the number of messages received and reviewed. Thus, self-report bias could be introduced, as participants may have overrated their acceptance and understanding of messages when interviewed by village doctors because of recall or social desirability bias. In addition, although the questionnaire was designed based on previous studies in China, we did not conduct reliability and validity testing for the questionnaire; thus, we were not able to rule out all potential measurement errors from the study. Second, the participants involved in our contextual research and pilot study were conveniently sampled with a relatively small sample size based on set inclusion criteria. For example, more than 60% of our participants were male, and such proportion is higher than the proportion in general stroke patients, although the nationwide epidemiological study found a higher age-specific prevalence of stroke among men than that among women for the age groups of above 40 years [1]. Although participants may not be fully representative of all stroke patients in rural areas of China, we believe that the views they shared were common among stroke patients in rural China.

Our study also has several strengths. First, we have applied multiple methodologies, including literature review, expert consultation, qualitative in-depth interviews, and a pilot study, including a quantitative survey and personal interviews. By doing so, we were able to ensure a deeper involvement of the target population in the study design and gain a more comprehensive understanding of their needs, acceptance, and preferences. Second, the messages were developed by researchers and verified by health care professionals and village doctors. Such collaboration ensured that the messages were designed based on the clinical evidence and behavior change theories with conscientious consideration of local context. By applying these strategies, the contents of messages could be made reliable, feasible, and understandable. Third, the results of the initial design and pilot study were incorporated into the refinement of both the final content and delivery algorithm, potentially ensuring better acceptance and effectiveness of the message-based intervention.

Conclusions

In summary, our study demonstrates the potential of using voice message interventions to improve chronic disease prevention and management among people with low education level in resource-limited settings. The development process indicates the importance of contextualization, which includes involving the target audience early and considering their preference and characteristics in the design process. The findings on the preference of participants for voice messages and the potential factors related to the acceptance add new evidence to the literature. These findings have general implications beyond stroke care in rural China and will be informative to other researchers who plan to develop message-based interventions for other disease conditions or in other research settings.

We recommend future studies to describe the message development procedures so that researchers could have better insight on the underlining mechanisms of the message design and the contextual environment that enabled the intervention implementation, which will facilitate the construction of the evidence based on the impact of the message-based interventions. In addition, more studies are needed to understand further the generalizability of the key aspects of message optimization that we found through our study among other populations. With such knowledge, we could better design the content and dispatching algorithms of the message-based interventions with higher adoption and maintenance.

Acknowledgments

The authors would like to thank Yun Zhou, Zixiao Li, Yilong Wang, Ninghua Wang, Bin Xie, Jianmin Yao, Mobai Hou, Li Wang, Lihui Hu, Bichuan Lu, Pengfei Dai, and Guodong Fan, who were involved in the verification of the messages. The authors also appreciate all stroke patients, caregivers, village doctors, township physicians, and county managers, who have provided advice and feedback about the mHealth message system to the research team throughout the study. The authors would also like to thank the leaders and staff from the Nanhe County Center for Disease Prevention and Control, Nanhe County People's Hospital, and the Nanhe County Health and Family Planning Commission, who supported the implementation of the study. The study was

jointly funded by the Medical Research Council, Wellcome Trust, the Economic and Social Research Council, and the Department for International Development of the United Kingdom (grant number: MR/N015967/1). These sponsors have no role in either the preparation or approval of the manuscript.

Authors' Contributions

LLY is the principal investigator of the study and guided the design, refinement, and manuscript writing process. EG and WG drafted the manuscript, and other coauthors contributed to the manuscript revision and approved the final version for publication. EL and LT conducted the literature review of existing message banks; EG and WG conducted data analysis; LT, JD, CS, and YY drafted the initial messages for the pilot study; and EL, LZ, PB, EG, WG, and LLY were involved in the message refinements. All authors critically reviewed and approved the content of the paper.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Key searching terms, procedure, and findings from the literature review on the existing message banks.
[DOCX File , 92 KB - mhealth_v7i12e15758_app1.docx]

Multimedia Appendix 2
Examples of messages.
[DOCX File , 14 KB - mhealth_v7i12e15758_app2.docx]

References

1. Wang W, Jiang B, Sun H, Ru X, Sun D, Wang L, NESS-China Investigators. Prevalence, incidence, and mortality of stroke in China: results from a nationwide population-based survey of 480 687 adults. Circulation 2017 Feb 21;135(8):759-771. [doi: 10.1161/CIRCULATIONAHA.116.025250] [Medline: 28052979]
2. Wang L, Liu J, Yang Y, Peng B, Yilong W. The prevention and treatment of stroke still face huge challenges-brief report on stroke prevention and treatment in China, 2018. Chin Circ J 2019;34(2):105-119 [FREE Full text]
3. Liu L, Liu J, Wang Y, Wang D, Wang Y. Substantial improvement of stroke care in China. Stroke 2018 Dec;49(12):3085-3091. [doi: 10.1161/STROKEAHA.118.022618] [Medline: 30571434]
4. World Health Organization. Prevention of Recurrent Heart Attacks and Strokes in Low and Middle Income Populations: Evidence-based Recommendations for Policy-makers and Health Professionals. Geneva: World Health Organization; 2003.
5. Chen Y, Li L, Zhang Q, Clarke R, Chen J, Guo Y, China Kadoorie Biobank Study. Use of drug treatment for secondary prevention of cardiovascular disease in urban and rural communities of China: China Kadoorie Biobank study of 0.5 million people. Int J Cardiol 2014 Mar 1;172(1):88-95 [FREE Full text] [doi: 10.1016/j.ijcard.2013.12.065] [Medline: 24461961]
6. Jiang Y, Yang X, Li Z, Pan Y, Wang Y, Wang Y, et al. Persistence of secondary prevention medication and related factors for acute ischemic stroke and transient ischemic attack in China. Neurol Res 2017 Jun;39(6):492-497. [doi: 10.1080/01616412.2017.1312792] [Medline: 28420316]
7. Shi Y, Yi H, Zhou H, Zhou C, Xue H, Rozelle S, et al. The quality of primary care and correlates among grassroots providers in rural China: a cross-sectional standardised patient study. The Lancet 2017 Dec;390:S16. [doi: 10.1016/S0140-6736(17)33154-9]
8. Orr JA, King RJ. Mobile phone SMS messages can enhance healthy behaviour: a meta-analysis of randomised controlled trials. Health Psychol Rev 2015;9(4):397-416. [doi: 10.1080/17437199.2015.1022847] [Medline: 25739668]
9. Yasmin F, Banu B, Zakir SM, Sauerborn R, Ali L, Souares A. Positive influence of short message service and voice call interventions on adherence and health outcomes in case of chronic disease care: a systematic review. BMC Med Inform Decis Mak 2016 Apr 22;16:46 [FREE Full text] [doi: 10.1186/s12911-016-0286-3] [Medline: 27106263]
10. Beratarrechea A, Lee A, Willner J, Jahangir E, Ciapponi A, Rubinstein A. The impact of mobile health interventions on chronic disease outcomes in developing countries: a systematic review. Telemed J E Health 2014 Jan;20(1):75-82 [FREE Full text] [doi: 10.1089/tmj.2012.0328] [Medline: 24205809]
11. Kamal AK, Shaikh Q, Pasha O, Azam I, Islam M, Memon AA, et al. A randomized controlled behavioral intervention trial to improve medication adherence in adult stroke patients with prescription tailored Short Messaging Service (SMS)-SMS4Stroke study. BMC Neurol 2015 Oct 21;15:212 [FREE Full text] [doi: 10.1186/s12883-015-0471-5] [Medline: 26486857]
12. Yang X, Kovarik C. A systematic review of mobile health interventions in China: identifying gaps in care. J Telemed Telecare 2019 Jul 18:1357633X19856746. [doi: 10.1177/1357633X19856746] [Medline: 31319759]

13. Thakkar J, Kurup R, Laba T, Santo K, Thiagalingam A, Rodgers A, et al. Mobile telephone text messaging for medication adherence in chronic disease: a meta-analysis. JAMA Intern Med 2016 Mar;176(3):340-349. [doi: 10.1001/jamainternmed.2015.7667] [Medline: 26831740]
14. Sahin C, Courtney KL, Naylor PJ, E Rhodes R. Tailored mobile text messaging interventions targeting type 2 diabetes self-management: a systematic review and a meta-analysis. Digit Health 2019;5:2055207619845279 [FREE Full text] [doi: 10.1177/2055207619845279] [Medline: 31041110]
15. Anstey Watkins J, Goudge J, Gómez-Olivé FX, Huxley C, Dodd K, Griffiths F. mHealth text and voice communication for monitoring people with chronic diseases in low-resource settings: a realist review. BMJ Glob Health 2018;3(2):e000543 [FREE Full text] [doi: 10.1136/bmjgh-2017-000543] [Medline: 29527356]
16. Champion V, Skinner C. The health belief model. In: Glanz K, Rimer BK, Viswanath K, editors. Health Behavior and Health Education: Theory, Research, Practice. Fourth Edition. San Francisco: Josseybass; 2008:45-65.
17. Diez-Canseco F, Zavala-Loayza JA, Beratarrechea A, Kanter R, Ramirez-Zea M, Rubinstein A, et al. Design and multi-country validation of text messages for an mHealth intervention for primary prevention of progression to hypertension in Latin America. JMIR Mhealth Uhealth 2015 Feb 18;3(1):e19 [FREE Full text] [doi: 10.2196/mhealth.3874] [Medline: 25693595]
18. Gong E, Gu W, Sun C, Turner EL, Zhou Y, Li Z, et al. System-integrated technology-enabled model of care to improve the health of stroke patients in rural China: protocol for SINEMA-a cluster-randomized controlled trial. Am Heart J 2019 Jan;207:27-39. [doi: 10.1016/j.ahj.2018.08.015] [Medline: 30408621]
19. China Statistic Press. Hebei Economic Yearbook 2016. 2016. URL: http://www.hetj.gov.cn/res/nj2016/indexch.htm [accessed 2019-01-30]
20. Xu G, Ma M, Liu X, Hankey GJ. Is there a stroke belt in China and why? Stroke 2013 Jul;44(7):1775-1783. [doi: 10.1161/STROKEAHA.113.001238] [Medline: 23674531]
21. Prochaska J, Redding C, Evers K. The transtheoretical model and stages of change. In: Glanz K, Rimer BK, Viswanath K, editors. Health Education and Health Behavior: Theory, Research, Practice. San Francisco: JosseyBass; 2015:125-148.
22. Yan LL, Fang W, Delong E, Neal B, Peterson E, Huang Y, et al. Population impact of a high cardiovascular risk management program delivered by village doctors in rural China: design and rationale of a large, cluster-randomized controlled trial. BMC Public Health 2014 Apr 11;14:345 [FREE Full text] [doi: 10.1186/1471-2458-14-345] [Medline: 24721435]
23. Zhou B, Zhang J, Zhao Y, Li X, Anderson CS, Xie B, et al. Caregiver-delivered stroke rehabilitation in rural China. Stroke 2019 Jul;50(7):1825-1830. [doi: 10.1161/STROKEAHA.118.021558] [Medline: 31177978]
24. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
25. Adler AJ, Martin N, Mariani J, Tajer CD, Owolabi OO, Free C, et al. Mobile phone text messaging to improve medication adherence in secondary prevention of cardiovascular disease. Cochrane Database Syst Rev 2017 Apr 29;4:CD011851 [FREE Full text] [doi: 10.1002/14651858.CD011851.pub2] [Medline: 28455948]
26. Eckersberger E, Pearson E, Andersen K, Hossain A, Footman K, Biswas K, et al. Developing mHealth messages to promote postmenstrual regulation contraceptive use in Bangladesh: participatory interview study. JMIR Mhealth Uhealth 2017 Dec 14;5(12):e174 [FREE Full text] [doi: 10.2196/mhealth.6969] [Medline: 29242175]
27. Bullen PA. Operational challenges in the Cambodian mHealth revolution. J Mob Technol Med 2013;2(2):20-23. [doi: 10.7309/jmtm.2.2.5]
28. Kannisto K, Koivunen M, Välimäki MA. Use of mobile phone text message reminders in health care services: a narrative literature review. J Med Internet Res 2014 Oct 17;16(10):e222 [FREE Full text] [doi: 10.2196/jmir.3442] [Medline: 25326646]
29. Marcolino MS, Oliveira JA, D'Agostino M, Ribeiro AL, Alkmim MB, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
30. Cho Y, Lee S, Islam S, Kim SY. Theories applied to m-Health interventions for behavior change in low- and middle-income countries: a systematic review. Telemed J E Health 2018 Oct;24(10):727-741 [FREE Full text] [doi: 10.1089/tmj.2017.0249] [Medline: 29437546]
31. Toefy Y, Skinner D, Thomsen S. 'Please Don't Send Us Spam!' A participative, theory-based methodology for developing an mHealth intervention. JMIR Mhealth Uhealth 2016 Aug 17;4(3):e100 [FREE Full text] [doi: 10.2196/mhealth.6041] [Medline: 27535589]
32. Muench F, Baumel A. More than a text message: dismantling digital triggers to curate behavior change in patient-centered health interventions. J Med Internet Res 2017 May 26;19(5):e147 [FREE Full text] [doi: 10.2196/jmir.7463] [Medline: 28550001]

Abbreviations

mHealth: mobile health
SINEMA: System-Integrated Technology-Enabled Model of Care

Edited by G Eysenbach; submitted 03.08.19; peer-reviewed by T Alessa, L Shen; comments to author 18.09.19; revised version received 29.09.19; accepted 30.09.19; published 17.12.19.

Please cite as:
Gong E, Gu W, Luo E, Tan L, Donovan J, Sun C, Yang Y, Zang L, Bao P, Yan LL
Development and Local Contextualization of Mobile Health Messages for Enhancing Disease Management Among Community-Dwelling Stroke Patients in Rural China: Multimethod Study
JMIR Mhealth Uhealth 2019;7(12):e15758
URL: https://mhealth.jmir.org/2019/12/e15758
doi:10.2196/15758
PMID:31845901

Tutorial

Strategies for the Successful Implementation of a Novel iPhone Loaner System (iShare) in mHealth Interventions: Prospective Study

William E Yang[1], MD; Erin M Spaulding[2], RN, BSN; David Lumelsky[3], BS; George Hung[1], MD; Pauline Phuong Huynh[1], BA; Kellen Knowles[1], MD; Francoise A Marvel[1,4], MD; Valerie Vilarino[3], BA; Jane Wang[1], MD; Lochan M Shah[1], BA; Helen Xun[1], BS; Rongzi Shan[4], BS; Shannon Wongvibulsin[1,5], PhD; Seth S Martin[1,4], MD, MHS

[1]School of Medicine, Johns Hopkins University, Baltimore, MD, United States

[2]School of Nursing, Johns Hopkins University, Baltimore, MD, United States

[3]Krieger School of Arts and Sciences, Johns Hopkins University, Baltimore, MD, United States

[4]Ciccarone Center for the Prevention of Cardiovascular Disease, Division of Cardiology, School of Medicine, Johns Hopkins University, Baltimore, MD, United States

[5]Department of Biomedical Engineering, School of Medicine, Johns Hopkins University, Baltimore, MD, United States

Corresponding Author:
William E Yang, MD
School of Medicine
Johns Hopkins University
4940 Eastern Ave
Baltimore, MD, 21224
United States
Phone: 1 410 550 0100
Email: wyang@jhmi.edu

Related Article:

This is a corrected version. See correction statement: https://mhealth.jmir.org/2021/9/e31472

Abstract

Background: As smartphone ownership continues to rise, health care systems and technology companies are driven to develop mobile health (mHealth) interventions as both diagnostic and therapeutic tools. An important consideration during mHealth intervention development is how to achieve health equity despite demographic differences in smartphone ownership. One solution is through the recirculation of loaner smartphones; however, best practices for implementing such programs to optimize security, privacy, scalability, and convenience for participants are not well defined.

Objective: In this tutorial, we describe how we implemented our novel Corrie iShare program, a 30-day loaner iPhone and smartwatch recirculation program, as part of a multi-center mHealth intervention to improve recovery and access to guideline-directed therapy following acute myocardial infarction.

Methods: We conducted a prospective study utilizing a smartphone app and leveraged iOS enterprise features as well as cellular data service to automate recirculation.

Results: Our configuration protocol was shortened from 1 hour to 10 minutes. Of 200 participants, 92 (46.0%) did not own an iPhone and would have been excluded from the study without iShare. Among iShare participants, 72% (66/92) returned their loaned smartphones.

Conclusions: The Corrie iShare program demonstrates the potential for a sustainable and scalable mHealth loaner program, enabling broader population reach while optimizing user experience. Implementation may face institutional constraints and software limitations. Consideration should be given to optimizing loaner returns.

(JMIR Mhealth Uhealth 2019;7(12):e16391) doi:10.2196/16391

KEYWORDS

mHealth; digital health; innovation; myocardial infarction; health care disparities; smartphone; mobile phone; smart technology; loaner device; telemedicine

Introduction

Mobile health (mHealth) interventions are promising therapeutic tools [1]. The high rate of smartphone ownership among Americans, up to 77% in 2018 [2], and the popularity of wearable technologies have driven health care systems and technology companies to develop mHealth interventions and create innovative health care delivery methods.

The delivery platform critically influences mHealth functionality. Google's Android and Apple's iOS dominate market share, accounting for 99.8% of smartphones [3]. mHealth app developers need to choose between developing a cross-platform app or native apps. Web-based (ie, *cross-platform*) apps require little additional work to reach both major smartphone platforms. Although *cross-platform* apps have fast time to market and require less maintenance, they are unable to take advantage of all the features that a *native* platform-specific app could. On the other hand, developing multiple native apps, in addition to requiring more time and resources, also becomes hindered by the lack of interoperability between platforms (eg, the Apple Watch cannot be paired with an Android phone).

Platform selection also has implications for health equity due to demographic differences in smartphone ownership. In 2013, those who identified as *black, non-Hispanic*, had less than a college education, or whose annual household income was below US $75,000 were more likely to own an Android phone than an iPhone [4]. As socioeconomic status contributes to health disparities, being inclusive of socioeconomically disadvantaged groups when choosing an mHealth intervention platform offers an opportunity to reduce health disparities.

Our team was cognizant of these factors as we developed the *Corrie* app for the Myocardial infarction COmbined-device Recovery Enhancement (MiCORE) prospective study (Johns Hopkins University: IRB00099938, NCT03760796). We considered multiple approaches to balance inclusiveness and optimal user experience. Although we initially focused on cross-platform development, Corrie was ultimately developed as a native iOS app following the release of the Apple CareKit health app framework in 2016 [5]. iOS secures and encrypts protected health information (PHI), while native app development enables smartwatch integration, interactive notifications, calendar and contacts integration, and Bluetooth.

Prior mHealth studies utilizing loaner smartphones have not described how to implement a loaner model [6-8], details of which are key to ensuring security, privacy, scalability, and sustainability. To address these details and maintain broad population reach, we created the *iShare* loaner program. Through iShare, we enrolled patients who did not own an iPhone into MiCORE by providing each with a reusable loaner iPhone and Apple Watch preloaded with Corrie. In this paper, we describe the development of our novel loaner program, initial findings, and considerations for its generalized application.

Methods

Toward a Sustainable and Scalable Protocol

Overview

Without experience or guidance regarding the best approach to operationalize a smartphone loaner program for a clinical research study, we developed the program iteratively. With each iteration came an increased financial investment, often resulting in a greater-than-anticipated improvement in our processes. Table 1 provides a summary of our protocol's evolution. While our final protocol (see Multimedia Appendix 1) was optimized for the MiCORE study, earlier iterations may be more appropriate for other mHealth studies. Determining optimal configuration settings also would have been challenging without the experience gained during earlier iterations.

At the advent of iShare, rather than purchasing new iPhones, we chose the economical approach of purchasing refurbished iPhones in small batches from third-party resellers. Apple donated Apple Watches for the MiCORE study.

Table 1. Use cases and characteristics of each protocol version.

Characteristic	Version 1.0	Version 2.0	Version 3.0
Use case	Exploratory testing of platform capabilities	Small-scale study	Study requiring scalability
Interactive time	30-60 minutes Lowest level of automation	20-30 minutes	5-10 minutes Highest level of automation
Requirements	Minimum financial investment Study team Apple ID for each phone Participants must have home Wi-Fi USB connection required to reset PIN-locked phone	In addition to previous requirements: Apple Mac computer running Apple Configurator software Low-cost SIMs without cellular service	Apple Mac computer running Apple Configurator software Apple Business Manager (free) Mobile Device Management (subscription fee) Cellular plan for each loaner phone
Features	Unrestricted Apple App Store Activation Lock to prevent circumventing security measures by factory reset Find My iPhone GPS tracking to reduce the risk of theft iOS restrictions to disable iCloud and prevent inadvertent sharing of PHI[a]	Increased consistency in configuration due to automation	Apple Device Enrollment Program to prevent circumventing security measures by factory reset Lost Mode GPS tracking to reduce the risk of theft Mobile Device Management policies to disable iCloud and prevent inadvertent sharing of PHI
Limitations	Manual (error-prone) configuration of all features and app installation Insertion and removal of SIM card required to activate phones after each reset Find My iPhone not helpful outside of Wi-Fi range	Manual configuration of Apple ID and app installation Find My iPhone not helpful outside of Wi-Fi range Apple Configurator unable to share configuration profiles with other computers—single team member performs all setup	Apple Configurator required for first-time configuration of each loaner phone

[a]PHI: protected health information.

Enterprise Tools for Automated Configuration

Apple's Device Enrollment Program (DEP) and Volume Purchasing Program (VPP)—unified under Apple Business Manager [9]—together with Mobile Device Management (MDM) software, allow organizations to manage Apple devices that may be redistributed to multiple people over time. Table 2 summarizes the features provided by each enterprise tool.

During initial phone setup, Apple Configurator was used to enroll iPhones into DEP. DEP then contacted the MDM server, and MDM configured the phone remotely, allowing easy automated return processing by any team member, facilitating scalability. The enterprise tools also eliminated the need for Apple IDs on each device, translating into significant time savings. As a result, new iPhones required only 10 minutes and returned iPhones required only 5 minutes to set up. Although we paid a subscription fee for MDM, the efficiency gained greatly outweighed the cost.

Table 2. Summary of enterprise tools adopted in iShare 3.0 and key benefits provided by each tool.

Enterprise tool	Key features
Apple Device Enrollment Program	Protect ownership of study team phones and deter theft without using Apple IDs Automatic and mandatory configuration of phones with Mobile Device Management server No cost to use
Apple Volume Purchase Program	Enable bulk automated app download and purchase Download apps without Apple IDs No cost to use
Mobile Device Management	Automate phone configuration (ie, security policies and app installation) Update configuration on previously deployed phones over the Internet Use of *Lost Mode*—enterprise equivalent to *Find My iPhone*—to disable and GPS-track missing phones Monitor availability of iShare inventory Easily reset returned phones with a few clicks Variable subscription fee based on vendor and features desired

Cellular Service

We purchased prepaid plans for each phone, which enabled patients without home Wi-Fi to participate in the study; facilitated MDM-initiated resetting of locked returned phones; and allowed activation of Lost Mode on phones not connected to Wi-Fi.

Study Outcomes

We describe in detail the iShare 3.0 protocol. In addition to the feasibility of the loaner smartphone process, we evaluated process outcomes including the following: number of additional participants enrolled, device return rate, and the estimated per-participant investment cost of a loaner phone. In the final MiCORE analyses, which will be presented in future publications, we will compare iShare participants' and iPhone owners' demographics, readmission rates, survey data, and app usage data.

Results

Corrie iShare 3.0 Protocol

Device Procurement and Preparation

Equipment was inventoried and labeled. A SIM card with an active prepaid cellular plan was installed in each phone, and each phone was placed in a protective case. Using Apple Configurator on a Mac computer, each iPhone was updated with the latest version of iOS and provisionally enrolled in Apple DEP and our team's cloud-based MDM server. DEP and MDM automated the installation and updating of security policies and the Corrie app on each iPhone, while also enabling our team to remotely disable and track lost or stolen phones. Figure 1 depicts a schematic of this process.

Configuration profiles disabled Apple iCloud, disabled Apple ID configuration to prevent participants from enabling Activation Lock on loaner Apple Watches, and blocked bandwidth-heavy streaming apps to limit cellular charges. We did not use MDM to enforce PIN security because the Corrie app already requires it. Apple Watches were paired with each iPhone prior to participant enrollment to ensure they were updated, an unattended process that can take up to several hours.

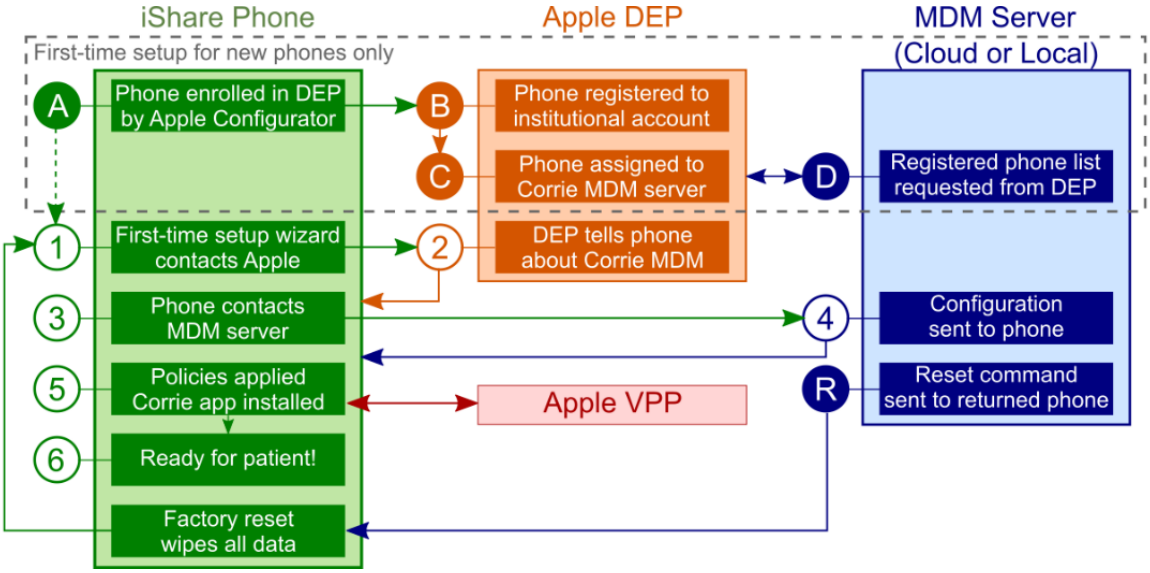
Study Enrollment

Patients were enrolled early in the MiCORE study during their hospitalization for myocardial infarction in accordance with the protocol described previously [10]. Participants with a study-compatible iPhone were assisted with downloading Corrie from the App Store. iPhone owners who did not own an Apple Watch were loaned one, which was paired with their personal device. Participants who did not own a study-compatible iPhone were loaned a team-owned iPhone and prepaired Apple Watch via the Corrie iShare program. Participants began using Corrie while hospitalized and continued to use it at home after hospital discharge. All participants borrowing either an iPhone or Apple Watch were provided with return instructions and a prepaid return package with which to mail equipment back to our team 30 days after hospital discharge. Non-smartphone owners were excluded to remove confounding from the steep learning curve of smartphone adoption.

Study Completion and Device Recycling

A total of 30 days after hospital discharge, participants returned borrowed equipment, which was inventoried and disinfected following a hospital-grade protocol. Using MDM, iPhones were factory reset to ensure no PHI remained. DEP and MDM automated the phone configuration (see Figure 1). We then used MDM to update iOS. The Apple Watch was then paired and updated. Equipment was turned off and placed into storage until needed.

Figure 1. Apple Device Enrollment Program (DEP), Volume Purchasing Program (VPP), and Mobile Device Management (MDM) facilitate automated setup of iShare phones through a series of handoffs. Steps A-D are completed manually the first time an iPhone is adopted into the Corrie iShare program. Steps 1-6 are automatic and require only minimal interaction with the phone. Returned phones are reset with one click through the MDM server (step R), which triggers a factory reset and wipes all data. Phones then automatically proceed through steps 1-6. Physical cleaning and handling steps are not shown in this figure.



Evaluation of Process Outcomes

We initially purchased 24 iPhone 5S phones on a rolling basis. After protocol refinement and site expansion across three states, we upgraded to 28 iPhone 6S Plus phones. The MiCORE study completed enrollment with a target sample size of 200 patients in April 2019. A total of 92 participants out of 200 (46.0%) were enrolled in the iShare program and 108 (54.0%) participants were enrolled with their personal iPhones. The participants who did not own iPhones would have been excluded from the study without the iShare program. Compared to iPhone owners, iShare participants were slightly younger (age range 30-81 years, mean 57.4 [SD 11] vs age range 32-89 years, mean 60.8 [SD 11]). They were also more likely to be women (33/92, 35.9% vs 25/108, 23.2%), of black race (23/92, 25.0% vs 15/108, 13.9%), and insured by Medicaid (18/92, 19.6% vs 4/108, 3.7%).

As expected, some participants were lost to follow-up, and some devices were lost or damaged. Of participants enrolled with a loaner phone, 72% (66/92) returned the phone and 70% (64/92) returned the watch; however, 2 participants (2%) returned damaged equipment (one phone and one watch), while one package (phone and watch) was damaged by the postal service. Estimating the cost of a refurbished phone at US $250, the loss due to phone nonreturn (n=26) or damage (n=2) was US $7000 or US $76 per iShare participant. The incremental cost of our MDM subscription was US $1950 total for 2 years or US $21 per iShare participant. Thus, we spent US $97 for each iShare participant, resulting in a cost savings of 61% compared to purchasing an iPhone for each participant who did not already own one. This estimated cost savings does not take into account the cost of the smartwatch, which Apple donated to our study team.

Final outcomes of MiCORE and iShare will be presented in a future publication.

Discussion

In this tutorial, we shared how our team implemented the Corrie iShare program to maximize patient enrollment in the MiCORE study through the use of loaner smartphones and smartwatches.

Security and Privacy Implications

Data security is paramount for protecting participant privacy. Apple iOS ensures security by encrypting data with the PIN unlock code. We chose to enforce PIN security at the app level rather than by MDM policy; participants could not use Corrie until a PIN had been set on the phone. The advantage of app-based enforcement is that it ensures data security on both iShare phones and personally owned iPhones used in the MiCORE study. While some would advocate for *double enforcement*, we decided to forego an MDM-enforced PIN code, which would have required that our team either not complete the iPhone setup wizard in advance, or set a temporary PIN that would need to be changed at enrollment time. Both situations would prolong study enrollment time without benefit.

To protect participant privacy and ensure that all data were deleted, phones were factory reset after each return. An alternative would have been to clear only user data, which would have reduced turnover time. However, this would have been less complete than a factory reset. While factory reset was initially a laborious and time-consuming process, the automation provided by DEP and MDM mitigated its inconveniences. We also used MDM and VPP to remove the need to preconfigure Apple IDs, eliminating the risk of one participant's data inadvertently being exposed to a future participant through iCloud data synchronization.

Institutional Information Technology Challenges

Overview

As with many novel protocols in large institutions, our initial iShare protocol encountered a number of challenges. While our health system had a process for mobile devices that would be used by patients sequentially while remaining hospitalized, it did not offer guidance for implementing our desired process where a patient uses a loaner phone while hospitalized and continues to use it at home for a time before returning it. We share the lessons we learned to help other mHealth researchers implement a loaner device program.

Apple Enterprise Tools

Apple DEP and VPP are offered to institutions and corporations rather than independent teams within those entities. As a result, it was necessary to identify and collaborate with the information technology (IT) staff member responsible for overseeing DEP and VPP. However, these Apple services are also available to any business at no cost and are not limited to large institutions, making this a valuable resource to independent teams, which could register as small businesses.

Mobile Device Management and Wi-Fi

We launched iShare using a cloud-based MDM solution provided by Jamf Software, LLC. Subsequently, our IT department offered us free access to use the institutional enterprise MDM—VMware AirWatch—which would have saved us the cost of subscribing to Jamf and enabled access to the hospital-authenticated Wi-Fi network. However, unlike Jamf, our institutional MDM required access to our institution's network to manage the phones and thus would not scale to a multi-site study. The interface was also more difficult for nontechnical users to learn. Therefore, we continued to use our team's independent MDM without authenticated Wi-Fi. Since iOS in standby does not stay connected to guest Wi-Fi networks, participants initially struggled to use Corrie with hospital Wi-Fi. This issue was mitigated after we added cellular data service to iShare, but increased cellular data expenses in an environment that otherwise had Wi-Fi.

Loaner Retrieval

We initially asked that participants return loaner equipment 30 days after discharge. Initially, only basic instructions were given to patients on how to return equipment. Our loaner return system evolved to include written instructions, prepaid mailers, and follow-up reminder calls.

We discovered that reminder phone calls made from personal phones were less likely to be answered or returned. Therefore, we switched to using hospital-based phones, which patients were more receptive to answering. Although we did not have a

dedicated phone number or voicemail, either could be useful in the future to facilitate receiving callbacks from patients.

The final step in retrieving unreturned devices was using *Lost Mode* GPS tracking. The vast majority of unreturned phones could not be GPS-tracked because the phone batteries had died. MDM data confirmed these phones had not connected to the Internet for a period of time and were presumed lost. DEP mitigated the risk of theft as participants could not bypass our security restrictions even if phones were reset. Our return rate was higher than one might anticipate in the absence of a return incentive.

Potential methods of increasing the return rate include actively monitoring MDM data and contacting patients if phones are unused for 7 days, rather than waiting until the end of the 30-day period to contact participants who had not returned their phones. Automated email and text reminders to patients or a monetary return incentive would also likely increase returns.

Apple Watch Limitations

Apple Watch setup, although simple, can take several hours to install updates and cannot be automated using Apple Configurator or MDM because of current software limitations, which limits scalability. In addition, a participant could accidentally *take ownership* of an Apple Watch by configuring their Apple ID on the phone, which enables Activation Lock on the watch. While DEP can bypass Activation Lock on iPhones, no bypass is available for Apple Watches; thus, we used MDM to disable Apple ID configuration. However, this reduced the functionality of loaner phones by making some apps and features unavailable.

Conclusions

The Corrie iShare program demonstrates the potential for sustainable and scalable smart technology reuse. A loaner smartphone program enables teams to focus on app development while maximizing study enrollment. Smartphone enterprise features readily adapt to a loaner program that ensures security and privacy. Although the loaner process described here is limited to iOS devices, the overall framework may be generalizable to other platforms. Implementation on other platforms would be an area for future research. In addition, while our work is promising for expanding demographic reach, we excluded patients who did not own a smartphone; thus, further research is necessary to study the feasibility of using a loaner program to reach the 39% of the global population who do not own smartphones [11] and are not typically represented in mHealth research, yet may stand to benefit from these technologies.

Acknowledgments

This study received material support from Apple and iHealth as well as funding from the Maryland Innovation Initiative, Wallace H Coulter Translational Research Partnership, Louis B Thalheimer Fund, Johns Hopkins Individualized Health Initiative, and the American Heart Association. EMS received the following financial support for the research, authorship, and publication of this article: National Institutes of Health (NIH), National Institute of Nursing Research (NINR) Ruth L Kirschstein National Research Service Award (NRSA; grant number: F31 NR017328) and NIH, NINR Predoctoral Fellowship in Interdisciplinary Cardiovascular Health Research (fellowship number: T32 NR012704). SW was supported by the Johns Hopkins School of Medicine Medical Scientist Training Program (T32 NIH Ruth L Kirschstein Institutional NRSA Predoctoral Award) and the NIH Ruth L Kirschstein Individual Predoctoral NRSA for MD/PhD Fellowship (F30).

Authors' Contributions

WEY, EMS, DL, GH, PPH, KK, FAM, VV, JW, LMS, HX, RS, SW, and SSM contributed substantially to the conception and design of this work, data collection, data analysis and interpretation, article drafts, and critical revisions of the article. Final approval of the version to be published was solicited from each author and each agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work were appropriately investigated and resolved. Additionally, WEY took primary responsibility for communication with the journal during the manuscript submission, peer review, and publication process.

Conflicts of Interest

The Corrie app, as described in this work, was developed by FAM, SSM, and Matthias Lee, PhD. They are also founders of, and hold equity in, Corrie Health, which intends to further develop the digital platform. This arrangement has been reviewed and approved by Johns Hopkins University in accordance with its conflict of interest policies. This study received material support from Apple and iHealth as well as funding from the Maryland Innovation Initiative, Wallace H Coulter Translational Research Partnership, Louis B Thalheimer Fund, Johns Hopkins Individualized Health Initiative, and the American Heart Association. SSM also reports research support from the Aetna Foundation, the American Heart Association, the David and June Trone Family Foundation, Google, NIH, Nokia, and the PJ Schafer Memorial Fund. SSM reports personal fees for serving on scientific advisory boards for Akcea Therapeutics, Amgen, Esperion, Novo Nordisk, Quest Diagnostics, Regeneron, and Sanofi. SSM is a coinventor on a pending patent filed by Johns Hopkins University for a system of low-density lipoprotein cholesterol estimation.

Multimedia Appendix 1
Final protocol in step-by-step format.

[DOCX File , 19 KB - mhealth_v7i12e16391_app1.docx]

References

1. Burke LE, Ma J, Azar KM, Bennett GG, Peterson ED, Zheng Y, American Heart Association Publications Committee of the Council on Epidemiology and Prevention, Behavior Change Committee of the Council on Cardiometabolic Health, Council on Cardiovascular and Stroke Nursing, Council on Functional Genomics and Translational Biology, Council on Quality of Care and Outcomes Research, and Stroke Council. Current science on consumer use of mobile health for cardiovascular disease prevention: A scientific statement from the American Heart Association. Circulation 2015 Sep 22;132(12):1157-1213. [doi: 10.1161/CIR.0000000000000232] [Medline: 26271892]
2. Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2019 Jun 12. Mobile fact sheet: Demographics of mobile device ownership and adoption in the United States URL: http://www.pewinternet.org/fact-sheet/mobile/ [accessed 2018-10-24]
3. Gartner. 2017 May 23. Gartner says worldwide sales of smartphones grew 9 percent in first quarter of 2017 URL: https://www.gartner.com/newsroom/id/3725117 [accessed 2018-07-05]
4. Smith A. Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2013 Jun 05. Smartphone ownership 2013 URL: http://www.pewinternet.org/2013/06/05/smartphone-ownership-2013/ [accessed 2019-01-15]
5. Apple Developer. CareKit URL: https://developer.apple.com/carekit/ [accessed 2018-10-24]
6. Domhardt M, Tiefengrabner M, Dinic R, Fötschl U, Oostingh GJ, Stütz T, et al. Training of carbohydrate estimation for people with diabetes using mobile augmented reality. J Diabetes Sci Technol 2015 May;9(3):516-524 [FREE Full text] [doi: 10.1177/1932296815578880] [Medline: 25883165]
7. Oh B, Yi G, Han MK, Kim JS, Lee CH, Cho B, et al. Importance of active participation in obesity management through mobile health care programs: Substudy of a randomized controlled trial. JMIR Mhealth Uhealth 2018 Jan 03;6(1):e2 [FREE Full text] [doi: 10.2196/mhealth.8719] [Medline: 29298749]
8. Athilingam P, Jenkins B, Johansson M, Labrador M. A mobile health intervention to improve self-care in patients with heart failure: Pilot randomized control trial. JMIR Cardio 2017 Aug 11;1(2):e3. [doi: 10.2196/cardio.7848] [Medline: 31758759]
9. Apple. Apple at work URL: https://www.apple.com/business/it/ [accessed 2019-01-23]
10. Spaulding EM, Marvel FA, Lee MA, Yang WE, Demo R, Wang J, et al. Corrie health digital platform for self-management in secondary prevention after acute myocardial infarction. Circ Cardiovasc Qual Outcomes 2019 May;12(5):e005509. [doi: 10.1161/CIRCOUTCOMES.119.005509] [Medline: 31043065]
11. Poushter J, Bishop C, Chwe H. Pew Research Center. Washington, DC: Pew Internet & American Life Project; 2018 Jun 19. Social media use continues to rise in developing countries but plateaus across developed ones: Smartphone ownership on the rise in emerging economies URL: http://www.pewglobal.org/2018/06/19/2-smartphone-ownership-on-the-rise-in-emerging-economies/ [accessed 2019-01-15]

Abbreviations

DEP: Device Enrollment Program
IT: information technology
MDM: Mobile Device Management
MiCORE: Myocardial infarction COmbined-device Recovery Enhancement
NIH: National Institutes of Health
NINR: National Institute of Nursing Research
NRSA: National Research Service Award
PHI: protected health information
VPP: Volume Purchasing Program

Edited by G Eysenbach; submitted 24.09.19; peer-reviewed by E Caiani, E Da Silva; comments to author 04.11.19; revised version received 11.11.19; accepted 13.11.19; published 16.12.19.

Please cite as:
Yang WE, Spaulding EM, Lumelsky D, Hung G, Huynh PP, Knowles K, Marvel FA, Vilarino V, Wang J, Shah LM, Xun H, Shan R, Wongvibulsin S, Martin SS
Strategies for the Successful Implementation of a Novel iPhone Loaner System (iShare) in mHealth Interventions: Prospective Study
JMIR Mhealth Uhealth 2019;7(12):e16391
URL: https://mhealth.jmir.org/2019/12/e16391
doi:10.2196/16391
PMID:31841115

Original Paper

Use of Mobile Health Apps and Wearable Technology to Assess Changes and Predict Pain During Treatment of Acute Pain in Sickle Cell Disease: Feasibility Study

Amanda Johnson[1*], MD, BA; Fan Yang[2*]; Siddharth Gollarahalli[3]; Tanvi Banerjee[2], PhD; Daniel Abrams[4], PhD; Jude Jonassaint[5], RN; Charles Jonassaint[5], PhD, MHS; Nirmish Shah[6], MD

[1]Department of Pediatrics, Duke University, Durham, NC, United States

[2]Department of Computer Science & Engineering, Wright State University, Dayton, OH, United States

[3]North Carolina State University, Raleigh, NC, United States

[4]Engineering Sciences and Applied Mathematics, Northwestern University, Chicago, IL, United States

[5]Social Work and Clinical and Translational Science, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, United States

[6]Division of Hematology, Department of Medicine, Duke University, Durham, NC, United States

[*]these authors contributed equally

Corresponding Author:
Amanda Johnson, MD, BA
Department of Pediatrics
Duke University
2301 Erwin Road
Durham, NC, 27710
United States
Phone: 1 651 207 3255
Email: amanda@ohsu.edu

Abstract

Background: Sickle cell disease (SCD) is an inherited red blood cell disorder affecting millions worldwide, and it results in many potential medical complications throughout the life course. The hallmark of SCD is pain. Many patients experience daily chronic pain as well as intermittent, unpredictable acute vaso-occlusive painful episodes called pain crises. These pain crises often require acute medical care through the day hospital or emergency department. Following presentation, a number of these patients are subsequently admitted with continued efforts of treatment focused on palliative pain control and hydration for management. Mitigating pain crises is challenging for both the patients and their providers, given the perceived unpredictability and subjective nature of pain.

Objective: The objective of this study was to show the feasibility of using objective, physiologic measurements obtained from a wearable device during an acute pain crisis to predict patient-reported pain scores (in an app and to nursing staff) using machine learning techniques.

Methods: For this feasibility study, we enrolled 27 adult patients presenting to the day hospital with acute pain. At the beginning of pain treatment, each participant was given a wearable device (Microsoft Band 2) that collected physiologic measurements. Pain scores from our mobile app, *Technology Resources to Understand Pain Assessment in Patients with Pain*, and those obtained by nursing staff were both used with wearable signals to complete time stamp matching and feature extraction and selection. Following this, we constructed regression and classification machine learning algorithms to build between-subject pain prediction models.

Results: Patients were monitored for an average of 3.79 (SD 2.23) hours, with an average of 5826 (SD 2667) objective data values per patient. As expected, we found that pain scores and heart rate decreased for most patients during the course of their stay. Using the wearable sensor data and pain scores, we were able to create a regression model to predict subjective pain scores with a root mean square error of 1.430 and correlation between observations and predictions of 0.706. Furthermore, we verified the hypothesis that the regression model outperformed the classification model by comparing the performances of the support vector machines (SVM) and the SVM for regression.

Conclusions: The Microsoft Band 2 allowed easy collection of objective, physiologic markers during an acute pain crisis in adults with SCD. Features can be extracted from these data signals and matched with pain scores. Machine learning models can then use these features to feasibly predict patient pain scores.

(JMIR Mhealth Uhealth 2019;7(12):e13671) doi:10.2196/13671

KEYWORDS

pain; sickle cell disease; SCD; machine learning

Introduction

Background

Sickle cell disease (SCD) is a hematologic disorder that can cause a multitude of complications throughout a patient's life, with pain being the most common and a significant cause of morbidity. The pain experienced by SCD patients is often chronic with acute vaso-occlusive crises that are unpredictable and lead to frequent visits to the emergency department (ED) and day hospital for management [1]. Of these patients, 1 in 4 will be admitted and can result in unplanned hospitalizations with missed days from work and school, significantly impairing a patient's quality of life [2]. Acute pain management is palliative, with hydration and pain control via narcotic and nonsteroidal anti-inflammatory drugs (NSAIDs). With pain being inherently subjective, both medical providers and patients express difficulty in determining ideal treatment and management strategies for pain.

In the last several years, there has been an increasing focus on developing and implementing individualized pain plans [3]. However, in addition to the slow adoption of these individualized plans, difficulty also lies in understanding the patient's degree of pain and response to pain management. With at least 1 in 4 patients with SCD seen in the ED being admitted to the hospital, it is critical to determine accurately which patients require additional pain management and which patients can be discharged.

More recently, technology has been leveraged to use mobile apps for recording symptoms in real time and wearable devices to provide more frequent physiologic measurements. The field of mobile health (mHealth) has continued to grow and has been used in a variety of different clinical settings. Many studies have attempted to help patients and providers connect using mobile technology to better understand and treat a multitude of symptoms, including pain [4-6]. Many of the initial mHealth systems and apps are smartphone-based and allow patients to self-report symptoms and activity in addition to recording objective data [7-9].

We previously reported the usefulness and validity of our mHealth app for patients with SCD [7-9]. The app has undergone multiple upgrades in the user interface based on feedback, as we continue to foster patient engagement. We have included additional health and mood questions, and the app was recently expanded to specific patient populations including bone marrow transplant patients [10]. In this study, we used Technology Resources to Understand Pain (TRU-Pain) app, which allows patients to record pain and other symptoms throughout their treatment, as described above [7]. In addition, TRU-Pain now allows the integration of wearable devices such as the Microsoft Band 2 to passively obtain physiologic data such as heart rate (HR), accelerometer activity, and galvanic skin response (GSR) using the AppleCare Kit platform.

In the face of the continued opioid crisis, the search for more objective measures of pain continues to rapidly evolve in medicine, and studies looking at a variety of objective measures to predict pain have been published in recent years. Among these studies, the types of objective data utilized to predict pain vary in invasiveness (vital signs vs neuroimaging) but show promise for utilizing such data to predict pain. Bendall et al [11] examined prehospital vital signs to predict pain severity using ordinal logistic regression and found that elevated respiratory rate, HR, and systolic blood pressure (specifically in older adults) were associated with more severe pain. A more invasive study by Lee et al used multimodal neuroimaging and HR variability with machine learning techniques to predict clinical pain in patients with chronic low back pain [12].

Owing to the growing volume of clinical data and the requirement of high accuracy predictive models, machine learning techniques have been increasingly utilized in medicine. They have been applied to multiple health care domains, from analgesic response prediction to postoperative pain estimation [13-15]. Machine learning techniques have also previously been utilized effectively in SCD studies [16,17]. Our previous study has also shown promising results in pain assessment [18]. Using nurse-obtained vital signs for patients with SCD admitted for pain crisis, our best model predicted pain with an accuracy of 0.429 on an 11-point rating scale (0 to 10) and 0.681 on a 4-point rating scale (none, mild, moderate, and severe) [18]. In these studies, machine learning can be described as a computational method to build efficient and accurate prediction models using known past information [19].

Objectives

We now aim to use physiologic data obtained from a wearable device matched with mobile app and nurse-obtained pain scores to predict pain scores at between-subject level using machine learning techniques. The combination of mobile apps and wearable sensors has been used in several studies to provide novel solutions to different health problems [20-22]. To date, there has been a paucity of research in SCD focused on pain prediction, despite the critical need. The ability to objectively and accurately predict pain severity and onset could result in more prompt and effective treatment of pain crises, leading to improved outcomes, as well as encouraging more diligent use of medications [23,24]. Using our past experience, our hypothesis for this study was as follows: For SCD patients presenting in acute pain, can we feasibly obtain objective data

from a wearable device and then utilize machine learning techniques to accurately predict pain scores?

Methods

Recruitment and Data Collection

Following Duke Institutional Review Board approval, patients presenting for acute pain crisis to the day hospital were approached and asked to participate in the study. A convenience sample of eligible patients who were willing to participate was consented. A small number of patients approached declined to participate, but this specific number was not recorded, and no patients withdrew from the study after consent. Of the 27 patients consented, 20 were included in this study because of insufficient data from the wearable device in 7 patients. Patients were consented Monday through Friday based on the availability of study team members. Study duration was variable based on patient's length of stay in the day hospital. The study included a one-time visit only. Patients might have had other chronic medical conditions but were not excluded based on these conditions, and subgroup analysis was not undertaken.

Following consent, a Microsoft Band 2 wearable was placed on the patient's wrist. The Microsoft Band 2 is a commercially available smart band that is compatible with many smartphones; it has multiple objective sensors including HR monitor, a 3-axis accelerometer and gyrometer, a GSR sensor, and a skin temperature sensor. The physiologic and activity measures utilized in the study are shown in Textbox 1. Overall, we adopted 8 wearable sensor signals to estimate pain scores (HR, R-R interval [RR; time between peak of QRS complex of electrocardiogram to subsequent QRS electrocardiogram peak], GSR, skin temperature, accelerometer [Z axis], angular velocity [Y axis], angular velocity [Z axis], and steps). These 8 signals were chosen partially based on signals readily available on the Microsoft Band 2 as well as previously postulated physiologic correlations with pain. Patients in more pain typically experience higher HR and will move less frequently in the setting of pain [25,26]. Furthermore, greater RR variability has been correlated with better pain treatment outcomes [27]. However, these objective measures have not been well established on their own to correlate with pain. Previous study by our group has supported the use of temperature, systolic blood pressure, diastolic blood pressure, oxygen saturation, and respiratory rate as statistically significant predictors in pain for SCD patients [18].

Patients were also provided with an iPad with the TRU-Pain app to record pain scores and other symptoms in conjunction with nurse-reported pain scores using a visual analog scale from 0 (none) to 10 (worst). Each patient was instructed on the use of the TRU-Pain app. The TRU-Pain app allowed patients to use a slider bar to rate their pain on the visual analog scale from 0 to 10. The app also allowed patients to note other symptoms and rate general health and mood (scale of 0 to 10). The TRU-Pain app implemented these general health and mood measures and a platform upgrade to AppleCare Kit, replacing our previous app, Sickle Cell Disease Mobile Applications to Record Symptoms via Technology. Nursing pain scores were assumed to be entered at the time they were obtained.

Both objective data from the Microsoft Band 2 and the TRU-Pain app were uploaded to a Health Insurance Portability and Accountability Act–compliant Citrix ShareFile cloud-based server. Patients were continuously monitored while in the day hospital, and at the time of discharge, the devices were returned. If patients were admitted, data before transfer were included even if the devices traveled with the patient during admission. Patients were not provided specific questions regarding acceptability and feasibility of participation, but the feasibility of the study was determined by the accuracy of machine learning algorithms in predicting pain scores.

Textbox 1. Physiologic and activity measures from Microsoft Band 2 (values for acceleration in X and Y directions equal that of Z direction—only acceleration in Z direction is included in this study; angular velocity in X direction was not correctly captured and was excluded from the dataset).

- Heart rate
- R-R interval
- Galvanic skin response
- Skin temperature
- Acceleration in X direction
- Acceleration in Y direction
- Acceleration in Z direction
- Angular velocity in X direction
- Angular velocity in Y direction
- Angular velocity in Z direction
- Steps

Data Preprocessing

To apply machine learning analysis on the collected wearable sensor data (physiologic and activity signals from the Microsoft Band 2 in Textbox 1), 3 data preprocessing steps need to be performed: time stamp matching, feature extraction, and feature selection. In time stamp matching, pain scores had to be matched with the wearable sensor data using the time stamp as close to the exact time of data collection as possible. However, the wearable sensor data samples were collected typically per

second, and the pain scores were collected at varying times throughout the stay, with time stamp formatted in hours and minutes only. To complete this best possible match, each pain score was matched with the 1-min long wearable sensor data segment that was tracked at the same hour and minute. By assuming that pain scores usually do not change rapidly within a short period, we also matched the app pain scores without exact time matching to the wearable sensor data when the time stamp difference was less than 10 min.

We obtained 40 matched records containing a 1-min long wearable sensor data segment and a pain score from the mobile app that logged at the same (or approximately the same) period. However, a sample size of 40 was not sufficient for the intended data analysis. To further increase the sample size, we included nurse-documented pain scores in our dataset. Our group made the assumption that nurse-documented scores were similar to patient-reported scores in the app. Nurse-documented pain scores were matched with wearable sensor data using the within 10-min time stamps methodology as described above. By including nurse-documented pain scores, our final dataset contained 107 data samples (40 mobile app and 67 nursing notes).

After time stamp matching, each pain score was mapped to a 1-min long wearable sensor data segment that included 8 signals as mentioned in Textbox 1 above. As the sensor signal was recorded typically every second, a 1-min long segment having 8 signals contained 480 (8×60) data points. It is difficult to process raw sensor signals directly in any analytical task. Therefore, we transformed raw sensor signals to a more suitable data representation format by feature extraction. First, a moving average filter was applied to raw sensor signals to remove noise. The moving average filter is the most common filter in digital signal processing to reduce random noise [28]. Then, 8 statistical features (as described in Table 1) were extracted for each of the 8 signals. These extracted features represented the properties of the original raw signals while reducing the volume of data.

The feature extraction yielded up to total 64 (8×8) features. Given the relatively small sample size (107), a feature selection method was applied (wrapper method) to remove irrelevant or redundant features and to further reduce the number of features [29]. The wrapper method has been reported to be able to improve the predictor performance when compared with variable ranking methods [29]. The basic idea of the wrapper method is selecting the subset of features that yields the best possible performance of a given learning algorithm. A total of 2 types of search strategies are widely adopted in the wrapper method: forward selection and backward elimination. In forward selection, one starts with an empty set and features are progressively added into the subset, whereas in backward elimination, one starts with the full feature set and progressively eliminates the feature with worst performance [30].

Table 2 shows the reduced feature set using the wrapper method with forward selection. A total of 10 features were selected from 5 signals. The table also illustrates the reduced feature set with backward elimination, which contains total 14 features from 7 signals. In both feature selection approaches, no features of acceleration in Z direction (AccZ) were selected, which might be because the information contained in AccZ was already covered by Steps.

Table 1. List of features extracted from wearable signals.

Feature	Description
Mean	Average value of the signal
Standard deviation	Amount of variation of the signal
Mean of derivative	Average rate of change of the signal
RMS[a]	Square root of the mean of the squares of a set of values
Peak to peak	Difference between the maximum and minimum peak
Peak to RMS	The ratio of the largest absolute value to the RMS value
Number of peaks	Number of local maximums (peaks)
Power	Sum of the absolute squares of time-domain samples divided by the length

[a]RMS: root mean square.

Table 2. Signals and reduced feature sets.

Signal	Feature	
	Forward selection	Backward elimination
Heart rate	• Mean of derivative • Number of peaks	• Power
R-R interval	• Number of peaks	• Standard deviation • Peak to RMS[a]
Galvanic skin response	• Mean • Peak to RMS	• Mean • Peak to peak
Steps	• Mean • RMS • Peak to peak	• Number of peaks • Power
Skin temperature	• Peak to RMS	• Power • Mean of derivative • Number of peaks
Angular velocity in Y direction	—[b]	• RMS • Number of peaks
Angular velocity in Z direction	—	• Peak to RMS • Number of peaks

[a]RMS: root mean square.

[b]Not applicable

Machine Learning Techniques

The prediction of numeric pain score, the main study outcome, can be treated as either a regression problem or a classification problem. As the pain scores from app data are float numbers, it is more reasonable to build a regression model to provide continuous estimation of the target variable. More importantly, there is only 1 target variable (pain score) in the regression model. In contrast, there will be 11 classes if pain is treated as a classification problem, as there are 11 distinct possible pain scores (0 to 10). The number of classes can be reduced by employing a sparse rating scale. Using a widely adopted more sparse 4-point rating scale, the 11-point pain scores can be categorized as none (0), mild (1-3), moderate (4-6), and severe (7-10) [31]. However, because of our small sample size, we hold the hypothesis that the regression model is more appropriate than the classification model in this study. We adopted 4 widely used regression algorithms in our analyses: Ridge regression (Ridge), Lasso regression (Lasso), Gaussian process for regression (GPR), and support vector machines for regression (SVR). In addition, we applied support vector machines (SVM) to predict the pain scores using the 4-point rating scale and compared the results with SVR.

For linear models, we utilized Ridge and Lasso [32,33]. Linear models are easy to fit and interpret, but they cannot model the nonlinear relationships between explanatory variables and the outcome variable. The other 2 algorithms are nonlinear models. A Gaussian process (GP) is a collection of random variables such that any finite subset of them has a joint multivariate Gaussian distribution. A GP can be fully specified by a mean function and a positive definite covariance function (or kernel). GPR is one of the Bayesian learning methods in which a previous distribution over the mapping function between inputs and outputs is conditioned on observations (training process). Then, the posterior distribution can be used to make predictions [34]. GPR provides a powerful way to quantify the uncertainty in model estimations to make more robust predictions on new test data. Finally, SVM are usually applied to classification problems. In classification, the SVM model maps the input samples into the feature space, then creates a decision surface among classes with the largest distance to any data point. However, it can also be applied to regression problems where we seek to find a continuous function that maps input variables to output variables, called SVR. For SVR, the goal is to find a function that deviates from the training output by a value no greater than a certain distance for each training point, and at the same time, is as flat as possible [35]. The nonlinearity of the algorithm can be obtained by utilizing kernel modulations.

Results

Overview

A total of 20 adult patients (of 27 consented) had complete data. Median age was 28 years, with a range of 20 years to 66 years (Table 3). A total of 11 (11/20, 55%) patients were female, whereas 9 (9/20, 45%) were male. Moreover, 10 patients (10/20, 50%) had type SS SCD, 8 (8/20, 40%) had type SC, and 2 (2/20, 10%) had S beta thalassemia. The average length of stay in the day hospital was 3.79 (SD 2.23) hours. In addition, 2 patients were subsequently admitted to the hospital. Nursing pain scores decreased in 16 out of 20 patients (80%). Patients had an average decrease in visual analog pain score of 2.75 (SD 2.34). A total

of 11 patients had multiple pain scores through the TRU-Pain app, and 91% (10/11) of the patients had a decrease in pain score, with an average decrease in pain score of 2.69 (SD 2.53).

Patients presenting to the day hospital often receive intravenous fluids, antiemetics, NSAIDs, and opioids. The opioid doses received during their day hospital stay are shown in Table 3. The last 3 columns are the number of visits each patient had to the ED and day hospital as well as admissions over the past calendar year.

Table 3. Patient demographics.

Patient	Age (years)	Sex	Sickle cell disease type	Insurance	Medications	Emergency department visits in prior year	Day hospital visits in prior year	Inpatient stays in prior year
1	21	F[a]	SC[b]	Public[c]	Dilaudid 6 mg; Oxycodone 5 mg	11	1	1
2	25	F	SS[d]	Public	Dilaudid 8 mg	3	8	3
3	24	F	SC	Private	Dilaudid 8 mg	1	4	3
4	40	M[e]	SS	Public	Dilaudid 16 mg; Oxycodone 5 mg	0	4	0
5	48	M	SB+[f]	Public	Dilaudid 9 mg	3	2	2
6	39	M	SS alpha[g]	Public	Dilaudid 12 mg	1	3	0
7	37	F	SC	Public	Dilaudid 9 mg	1	3	1
8	38	F	SC	Public	Dilaudid 8 mg	1	10	2
9[g]	21	M	SS	Public	Dilaudid 4 mg; Dilaudid PCA[h]	14	19	14
10	28	F	SS	Public	Dilaudid 16 mg; Oxycodone 20 mg	5	8	16
11	36	M	SS	Public	Dilaudid 6 mg	23	1	17
12	66	M	SS	Public	Dilaudid 8 mg; Morphine 4 mg	0	0	0
13	44	M	SC	Public	Dilaudid 11 mg	10	12	6
14	28	F	SB0[i]	Public	Dilaudid 8 mg	19	7	12
15	20	F	SC	Public	Dilaudid 9 mg	18	6	10
16	26	F	SS	Public	Dilaudid 13 mg	12	30	4
17	38	F	SS	Public	Dilaudid 16 mg	0	22	2
18	22	M	SC	Private	Dilaudid 8 mg	51	8	3
19	28	M	SC	Public	Dilaudid 8 mg; Oxycodone 10 mg	7	4	8
20	21	F	SS	Public	Dilaudid 5 mg; Oxycodone 10 mg	0	10	7

[a]F: female.

[b]SC: type SC (hemoglobin S and hemoglobin C).

[c]Public: at least some portion of insurance is Medicare or Medicaid.

[d]SS: type SS (hemoglobin S and hemoglobin S).

[e]M: male.

[f]SB+: type S beta thalassemia plus (hemoglobin S and beta thalassemia plus).

[g]SS alpha: type SS with alpha thalassemia (hemoglobin S and hemoglobin S with alpha thalassemia).

[h]PCA: patient-controlled analgesia.

[i]SB0: type S beta thalassemia zero (hemoglobin S and beta thalassemia zero).

Regression Results

A total of 4 regression algorithms were implemented on 2 reduced feature sets. Results were validated using 10-fold cross-validation. Moreover, 2 evaluation metrics were applied to evaluate the performance of algorithms—the root mean square error (RMSE) and Pearson correlation [34]. RMSE is the square root of the average of squared differences between predictions

and actual observations. It is measured on the same scale and has the same units as the pain score. Another metric is the Pearson correlation between predicted values and the actual values, which has a value between +1 and −1, where 0 means no linear correlation and +1 or −1 means total linear correlation. The higher the correlation value, the better the performance of the regression model. Table 4 summarizes the performance of the 4 algorithms on the 2 reduced feature sets.

For the dataset in our study, the standard deviation of 107 pain scores is 2.013, which can be interpreted as the RMSE of using the mean value as the predicted pain values. All the regression models obtained RMSE lower than the mean-only model. With 10 features in the forward selection feature set, the SVR had the best performance as the RMSE of 1.721 and the correlation of 0.522, followed by the GPR obtaining the RMSE of 1.764 and the correlation of 0.475. These results demonstrate the feasibility of using objective wearable sensor measurements to estimate subjective pain scores. With 14 features in the backward elimination feature set, the performance of GPR and SVR is further improved. The SVR model is slightly superior to the GPR model, with an RMSE of 1.430 and correlation of 0.706, respectively, which are also the best performance results obtained using regression methods. These data show that there was a strong association between the subjective pain scores (via app or nurse-obtained) and the predicted pain scores derived from wearable sensor signals.

The result of the SVR model with the best performance can be visualized in Figure 1. It is a scatter plot of the actual pain scores and predicted pain scores using the SVR model with the least squares regression line. The slope value of the least squares regression line is the same as the correlation of 0.706 in Table 4 and demonstrates a strong correlation of values between the actual pain scores and the predicted pain scores.

To better analyze the results of these regression methods, the residual plots of 4 regression models using the backward elimination feature set are illustrated in Figures 2-5. The dashed lines show the positive and negative standard deviation (2.013) of pain scores. The performances of Ridge and Lasso are nearly the same, which can be seen from Figures 2 and 3.

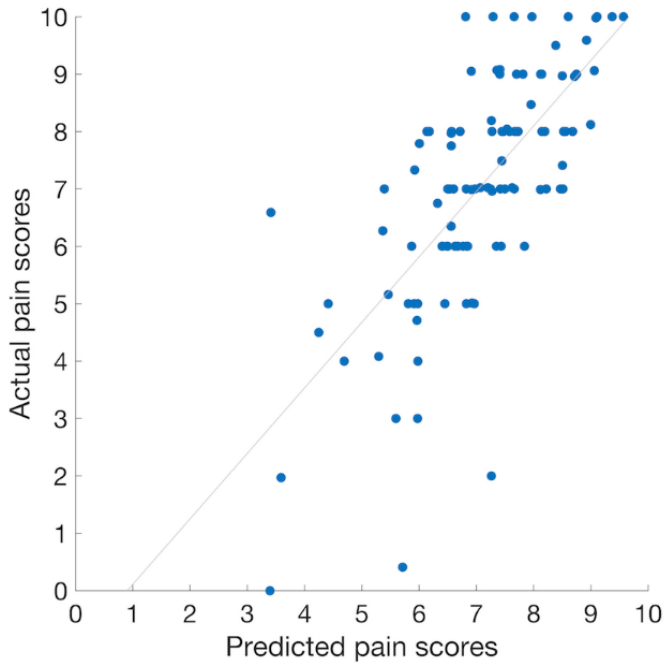
Table 4. Algorithm performances on 2 reduced feature sets using 4 regression methods.

Regression algorithm	Forward selection feature set		Backward elimination feature set	
	RMSE[a]	Correlation	RMSE	Correlation
Ridge	1.853	0.381	1.844	0.370
Lasso	1.871	0.358	1.891	0.370
Gaussian process for regression	1.764	0.475	1.473	0.683
Support vector machines for regression	1.721	0.522	1.430[b]	0.706[b]

[a]RMSE: root mean square error.

[b]Best performed model as described in the text.

Figure 1. Scatter plot of the predicted and actual pain scores using the support vector machines for regression model.



In either Figure 2 or 3, there is a roughly inverted U pattern, suggesting a nonlinear relationship between predictor variables and pain scores. Thus, performances of linear models were notably lower than the other 2 nonlinear models. The

distributions of residuals in Figures 4 and 5 are similar, which explains the comparable performance of the GPR model and the SVR model. The SVR model slightly surpassed the GPR model by having lower extreme residuals. Specifically, there are 2 outliers in both Figures 4 and 5, marked as points 1 and 2 (with actual pain scores of 0.41 and 2, respectively). The reason for the poor performance of these 2 points is the lack of training samples with lower pain values. It suggests that we can further improve our model performances by training the model with more samples having mild and moderate pain scores or having a larger dataset. Although a larger dataset is possible to obtain in future studies, an uneven distribution of pain scores will likely persist when acute pain crises are analyzed, as SCD patients will typically not present to medical care with lower pain scores and will manage minor crises at home [36].

Figure 2. Plot of the residuals versus predicted pain scores using the backward elimination feature set.

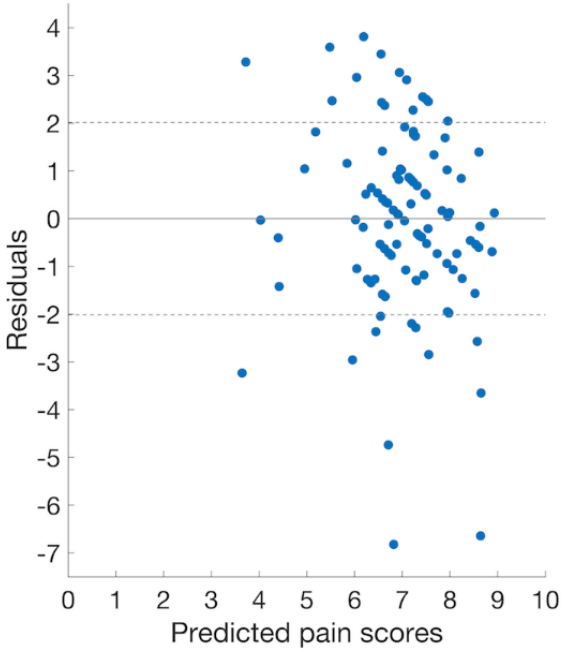


Figure 3. Plot of the residuals versus predicted pain scores using the backward elimination feature set (lasso).

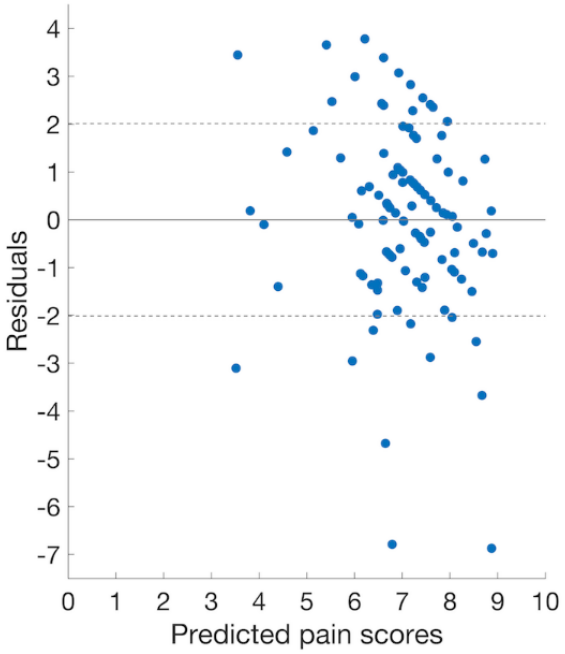


Figure 4. Plot of the residuals versus predicted pain scores using the backward elimination feature set (gaussian process for regression).

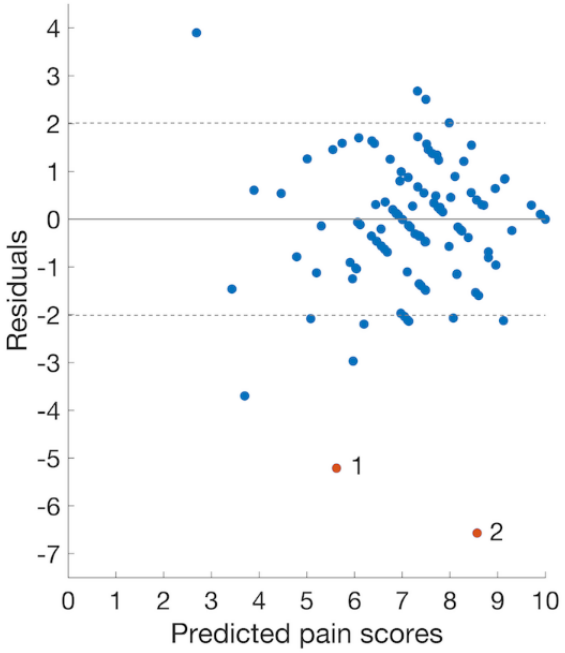
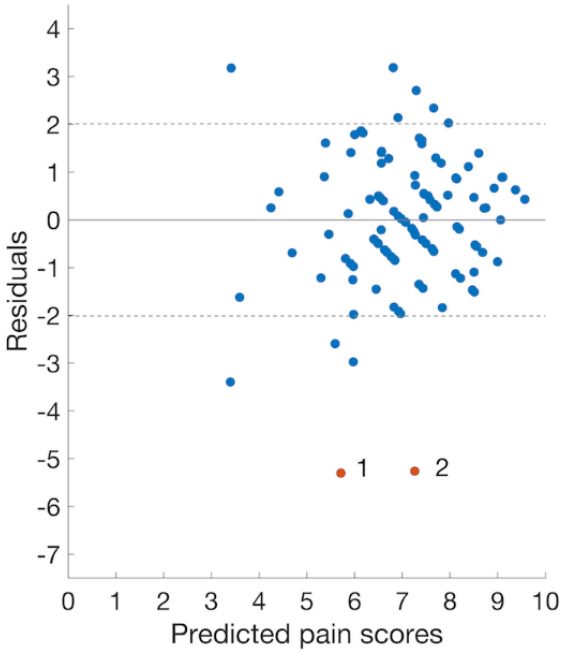


Figure 5. Plot of the residuals versus predicted pain scores using the backward elimination feature set (support vector machines for regression).



Classification Results

To apply classification to the original dataset, pain scores ranging from 0 to 10 were categorized into 4 classes as mentioned above: none (0), mild (1-3), moderate (4-6), and severe (7-10). The number of samples for the 4 pain levels are 2, 4, 34, and 67, respectively and indicates a high-class imbalance among the 4 classes. As patients visit the hospital because of pain management issues, the skewing to higher level pain scores makes clinical sense.

The SVM classifiers were applied on the categorized input dataset and evaluated for accuracy. F1 scores as well as a weighted F1 score were also evaluated. Accuracy is the ratio of correctly predicted pain scores over total number of pain scores. F1 score is the harmonic mean of precision and recall for each pain score, where precision is the ratio of the number of correctly identified entities with this pain score over the total number of this particular pain score predicted by the model. Recall is the ratio of the number of correctly identified entities with this pain score over the total number entities with this pain score in the dataset [35]. The weighted average F1 scores is the average of F1 score among all pain scores weighted by the number of instances of each pain score, and it is a better choice for evaluating datasets with multiple classes [37].

The classification result of the SVM model was compared with that of the best performance model, which was the SVR model

applied on the backward elimination feature set as described above. In the experiment, both SVM and SVR were implemented on the backward elimination feature set. For a fair comparison, the same kernel was used in SVM and SVR. In addition, the continuous predicted pain scores of the SVR model were categorized into 4 classes. In this way, the accuracy, F1 scores, and weighted F1 score were calculated for the SVR model. Table 5 shows the performance comparison between the SVR model and the SVM model. Overall, the SVR model outperformed the SVM model in each evaluation metric.

From Table 5, we can see that the performance of both the SVM and SVR models were affected by the class-imbalance problem, as the F1 scores for no pain and mild pain were much lower than that for the higher pain scores. However, the SVR model can better overcome this issue by treating the outcome as a single continuous variable, as opposed to treating the prediction as a multiclass classifier. The SVR model obtained an F1 score of 0.286 for mild pain even when there were only 4 training samples with mild pain scores in the dataset. In addition, by assuming pain as a continuous variable, there are ordinal relationships between pain levels in SVR. For example, a pain score 5 is greater than pain score 4 in this model. On the contrary, the ordinal relationship is not considered in the SVM model. Treating pain as an ordinal variable is a more reasonable assumption, and it may be another reason why the regression models outperform the classification model. In summary, our results verified the hypothesis that the regression model (SVR) would obtain a higher performance than the classification model (SVM) with a small sample size and when there was a class-imbalance problem in the dataset.

Table 5. Prediction performances on the 4-level pain scale using support vector machines for regression and support vector machines.

Algorithm	Accuracy	F1 score of no pain	F1 score of mild pain	F1 score of moderate pain	F1 score of severe pain	Weighted F1 score
Support vector machines	0.682	0	0	0.537	0.786	0.663
Support vector machines for regression	0.729[a]	0	0.286	0.675	0.803	0.728[a]

[a]Best performed model as described in the text.

Discussion

Principal Findings

This study demonstrates the feasibility of using physiologic data collected on a wearable device and applying these data using machine learning techniques to accurately predict subjective pain scores. The best accuracy was found using the machine learning technique SVR, with an accuracy of 0.729 prediction of pain on a 4-point scale. In addition, for patients treated in the day hospital for pain, we found expected improvement in pain and physiologic measures such as HR from the beginning to the end of their stay.

Our predictive results are encouraging and provide insight into potential techniques to predict pain and the understanding of individual physiologic response to pain and treatment. A few investigators have recently begun to evaluate the potential use of physiologic data to develop digital phenotypes for pain and, subsequently, an individualized pain prediction model. As discussed previously, objective and physiologic data of varying invasiveness have been utilized in medicine to better understand disease processes and symptoms, including SCD. Coates et al have extensively published on objective data in SCD, including spin-tagged magnetic resonance imaging to assess cerebral oxygen extraction and metabolic rate, biventricular dimensions and function to assess cardiac iron load, and the use of a graphical Lasso model to evaluate functional brain connectivity in SCD [38-40]. This group has also published analysis of laboratory measurements of carbon monoxide and heme oxygenase for acute pain crisis prediction [41]. Other groups have studied red blood cell mechanical sensitivity and biomarker signatures of SCD severity [42,43]. The use of machine learning in a variety of areas of medicine including outcome prediction for chemoradiotherapy, breast cancer survival prediction, and early prediction of asthma exacerbations have recently been published [44-46]. However, to date, the combined use of objective and physiologic data with machine learning techniques for pain in SCD is lacking.

Strengths and Limitations

A more objective pain prediction model could significantly help medical providers manage pain crises. As described, data collected from wearable devices can be utilized to improve pain management via advanced machine learning methods. In this analysis, we aimed to build predictive models for pain based on objective, physiologic wearable sensor data. This study is of great value given that the data utilized were obtained from a wearable device and provided minimal to no risk to patients. Furthermore, wearable sensor data were acquired frequently and obtained passively from patients as compared with nurse-documented vitals, which were obtained approximately every 2 hours.

Importantly, wearables and mobile apps (to track symptoms and pain scores over time) paired together to form an mHealth pain prediction system, as in this study, could fairly easily be applied to the inpatient and outpatient settings. mHealth systems are attractive for providers as pain can be tracked on a more frequent basis and can provide more personalized care for patients and potentially prevent ED visits, day hospital visits, and hospital admissions. Further work is needed in this field to continue to develop models with increasing accuracy in predicting pain to help guide management and patient care [47].

There are limitations to our study, including obtaining a convenience sample from our day hospital only and the small number of patients. Patients with SS and SC can be treated the same clinically, but the study included patients with thalassemia who may have a more or less severe phenotype depending on the type of thalassemia. Specific analysis on these patients was not performed for this feasibility study. The study is also limited

given that patients might have had underlying medical conditions that could affect HR, and this was not controlled for in our study. In addition, each patient had pain control achieved through individualized pain protocols, which varied among patients and were administered at various intervals. Therefore, it was impossible to control for these pain medications during this initial study. Medications administered, both opioid and nonopioid, may affect vital sign parameters independently (namely, opioids decreasing HR). The administration of pain medication, however, provides an important future opportunity to also evaluate pre- and postadministration objective datasets for pain prediction. Although all patients were in the day hospital either in a chair or bed, their environment was not completely controlled, and HR changes might have occurred with movement in and out of the bed or chair as well as to use the restroom, and these movements were not accounted for. HR can also vary outside of pain when a patient is at rest based on a multitude of different factors including stress, excitement, and breathing.

In addition, our group had to make the assumption that nurse-documented pain scores and patient-reported pain scores in the app were not dissimilar, but this is also an area for further study. One hypothesis would be that the patients could report a lower pain score to the nurse to look tough, but an alternative hypothesis may be a patient elevating their pain score to be given additional medication. There is also the assumption that the physiologic measures from the wearable device are accurate. We attempted to take data averaged over 1 min (with recordings typically every second) to minimize variability. We chose the Microsoft Band 2 because of the ability to acquire the raw data directly from the wearable and because of previous studies showing its relative accuracy. Stahl et al [48] and Shcherbina et al [49] have reported that wrist-based monitors, including the Microsoft Band, provided an accurate measurement of HR in most activity settings. Xie et al [50] further demonstrated that wearable devices had a high accuracy with respect to HR, number of steps, distance, and sleep duration.

Utilizing mobile devices and technology have great promise as we have discussed, but HR data and other physiologic parameters should be interpreted in the clinical context of the patient's history and exam. For example, a tachycardic patient should be thoroughly evaluated to rule out life-threatening conditions before attributing tachycardia to pain. Although our group has shown that wearable sensor data are feasible to obtain and can be used to create models for predicting pain scores, models and objective vital signs need to be paired with clinical experience and judgment for ideal patient management.

Conclusions

Future directions include refining the predictive model with a larger dataset. We are continuing to troubleshoot our data extraction procedure to minimize lost data. Furthermore, we could attempt to expand our models by examining patient's disease severity (related to number of ED visits, day hospital visits, and hospitalizations per year), length of stay in the day hospital, etc, to obtain a more ideal model for pain score prediction. Given that we combined app pain scores with nursing pain scores, further study is needed to determine if these can be treated as similar scores. Related to medication administration, we could examine HR changes before and after medication, time since last dose, total net dose of medications, etc, and attempt to project pain score and the need for medication before the patient requests medication. This would be an essential part of a real-time pain forecasting system and allow a trial that evaluates the timing of administration of additional doses of opioids based on physiologic and objective data alone. Our initial results indicate promise in pursuing each of these efforts, and our study is a valuable addition to ongoing studies investigating how physiologic and objective data can be used to help providers better understand and treat pain.

Acknowledgments

The authors would like to thank the SCD day hospital staff and patients for their assistance and participation.

Conflicts of Interest

NS is a speaker and consultant at Novartis and a speaker at Alexion. JJ is an Officer of Sicklesoft.

References

1. Brousseau DC, Owens PL, Mosso AL, Panepinto JA, Steiner CA. Acute care utilization and rehospitalizations for sickle cell disease. J Am Med Assoc 2010 Apr 7;303(13):1288-1294. [doi: 10.1001/jama.2010.378] [Medline: 20371788]
2. Cline DM, Silva S, Freiermuth CE, Thornton V, Tanabe P. Emergency Department (ED), ED observation, day hospital, and hospital admissions for adults with sickle cell disease. West J Emerg Med 2018 Mar;19(2):311-318 [FREE Full text] [doi: 10.5811/westjem.2017.9.35422] [Medline: 29560059]
3. Krishnamurti L, Smith-Packard B, Gupta A, Campbell M, Gunawardena S, Saladino R. Impact of individualized pain plan on the emergency management of children with sickle cell disease. Pediatr Blood Cancer 2014 Oct;61(10):1747-1753. [doi: 10.1002/pbc.25024] [Medline: 24962217]
4. Thurnheer SE, Gravestock I, Pichierri G, Steurer J, Burgstaller JM. Benefits of mobile apps in pain management: systematic review. JMIR Mhealth Uhealth 2018 Oct 22;6(10):e11231 [FREE Full text] [doi: 10.2196/11231] [Medline: 30348633]
5. Sundararaman LV, Edwards RR, Ross EL, Jamison RN. Integration of mobile health technology in the treatment of chronic pain: a critical review. Reg Anesth Pain Med 2017;42(4):488-498. [doi: 10.1097/AAP.0000000000000621] [Medline: 28570436]

6. Palermo TM, de la Vega R, Dudeney J, Murray C, Law E. Mobile health intervention for self-management of adolescent chronic pain (WebMAP mobile): protocol for a hybrid effectiveness-implementation cluster randomized controlled trial. Contemp Clin Trials 2018 Nov;74:55-60 [FREE Full text] [doi: 10.1016/j.cct.2018.10.003] [Medline: 30290276]
7. Jonassaint CR, Shah N, Jonassaint J, de Castro L. Usability and feasibility of an mHealth intervention for monitoring and managing pain symptoms in sickle cell disease: The Sickle Cell Disease Mobile Application to Record Symptoms via Technology (SMART). Hemoglobin 2015;39(3):162-168. [doi: 10.3109/03630269.2015.1025141] [Medline: 25831427]
8. Shah N, Jonassaint J, de Castro L. Patients welcome the Sickle Cell Disease Mobile Application to Record Symptoms via Technology (SMART). Hemoglobin 2014;38(2):99-103. [doi: 10.3109/03630269.2014.880716] [Medline: 24512633]
9. Clinical Trials. Patient-Empowered Mobile Technology in Hospitalized Patients (TRU-PAIN) URL: https://clinicaltrials.gov/ct2/show/NCT02895841 [accessed 2019-10-02]
10. Vaughn J, Jonassaint J, Summers-Goeckerman E, Shaw RJ, Shah N. Customization of the TRU-PBMT App (Technology recordings to better understand pediatric blood and marrow transplant). J Pediatr Nurs 2018;42:86-91. [doi: 10.1016/j.pedn.2018.07.007] [Medline: 30219304]
11. Bendall JC, Simpson PM, Middleton PM. Prehospital vital signs can predict pain severity: analysis using ordinal logistic regression. Eur J Emerg Med 2011 Dec;18(6):334-339. [doi: 10.1097/MEJ.0b013e328344fdf2] [Medline: 21407079]
12. Lee J, Mawla I, Kim J, Loggia ML, Ortiz A, Jung C, et al. Machine learning-based prediction of clinical pain using multimodal neuroimaging and autonomic metrics. Pain 2019 Mar;160(3):550-560. [doi: 10.1097/j.pain.0000000000001417] [Medline: 30540621]
13. Nickerson P, Tighe P, Shickel B, Rashidi P. Deep neural network architectures for forecasting analgesic response. Conf Proc IEEE Eng Med Biol Soc 2016 Aug;2016:2966-2969 [FREE Full text] [doi: 10.1109/EMBC.2016.7591352] [Medline: 28268935]
14. Tighe PJ, Harle CA, Hurley RW, Aytug H, Boezaart AP, Fillingim RB. Teaching a machine to feel postoperative pain: combining high-dimensional clinical data with machine learning algorithms to forecast acute postoperative pain. Pain Med 2015 Jul;16(7):1386-1401 [FREE Full text] [doi: 10.1111/pme.12713] [Medline: 26031220]
15. Lötsch J, Ultsch A. Machine learning in pain research. Pain 2018 Apr;159(4):623-630 [FREE Full text] [doi: 10.1097/j.pain.0000000000001118] [Medline: 29194126]
16. Milton JN, Gordeuk VR, Taylor JG, Gladwin MT, Steinberg MH, Sebastiani P. Prediction of fetal hemoglobin in sickle cell anemia using an ensemble of genetic risk prediction models. Circ Cardiovasc Genet 2014 Apr;7(2):110-115 [FREE Full text] [doi: 10.1161/CIRCGENETICS.113.000387] [Medline: 24585758]
17. Desai AA, Zhou T, Ahmad H, Zhang W, Mu W, Trevino S, et al. A novel molecular signature for elevated tricuspid regurgitation velocity in sickle cell disease. Am J Respir Crit Care Med 2012 Aug 15;186(4):359-368 [FREE Full text] [doi: 10.1164/rccm.201201-0057OC] [Medline: 22679008]
18. Yang F, Banerjee T, Narine K, Shah N. Improving pain management in patients with sickle cell disease from physiological measures using machine learning techniques. Smart Health (Amst) 2018 Jun;7-8:48-59 [FREE Full text] [doi: 10.1016/j.smhl.2018.01.002] [Medline: 30906841]
19. Mohri M, Rostamizadeh A, Talwalkar A. Foundations of Machine Learning. Second Edition. Cambridge, Massachusetts: The Mit Press; 2018.
20. Banos O, Villalonga C, Damas M, Gloesekoetter P, Pomares H, Rojas I. PhysioDroid: combining wearable health sensors and mobile devices for a ubiquitous, continuous, and personal monitoring. ScientificWorldJournal 2014;2014:490824 [FREE Full text] [doi: 10.1155/2014/490824] [Medline: 25295301]
21. Miao F, Cheng Y, He Y, He Q, Li Y. A wearable context-aware ECG monitoring system integrated with built-in kinematic sensors of the smartphone. Sensors (Basel) 2015 May 19;15(5):11465-11484 [FREE Full text] [doi: 10.3390/s150511465] [Medline: 25996508]
22. Lieber B, Taylor BE, Appelboom G, McKhann G, Connolly ES. Motion sensors to assess and monitor medical and surgical management of Parkinson disease. World Neurosurg 2015 Aug;84(2):561-566. [doi: 10.1016/j.wneu.2015.03.024] [Medline: 25827041]
23. Payne J, Aban I, Hilliard LM, Madison J, Bemrich-Stolz C, Howard TH, et al. Impact of early analgesia on hospitalization outcomes for sickle cell pain crisis. Pediatr Blood Cancer 2018 Dec;65(12):e27420. [doi: 10.1002/pbc.27420] [Medline: 30151977]
24. Molokie RE, Montminy C, Dionisio C, Farooqui MA, Gowhari M, Yao Y, et al. Opioid doses and acute care utilization outcomes for adults with sickle cell disease: ED versus acute care unit. Am J Emerg Med 2018 Jan;36(1):88-92. [doi: 10.1016/j.ajem.2017.07.037] [Medline: 28802541]
25. Tousignant-Laflamme Y, Rainville P, Marchand S. Establishing a link between heart rate and pain in healthy subjects: a gender effect. J Pain 2005 Jun;6(6):341-347. [doi: 10.1016/j.jpain.2005.01.351] [Medline: 15943955]
26. Hodges PW. Pain and motor control: From the laboratory to rehabilitation. J Electromyogr Kinesiol 2011 Apr;21(2):220-228. [doi: 10.1016/j.jelekin.2011.01.002] [Medline: 21306915]
27. Faye PM, de Jonckheere J, Logier R, Kuissi E, Jeanne M, Rakza T, et al. Newborn infant pain assessment using heart rate variability analysis. Clin J Pain 2010;26(9):777-782. [doi: 10.1097/ajp.0b013e3181ed1058] [Medline: 20973153]
28. Mitra S, Kuo Y. Digital Signal Processing: A Computer-based Approach. New York: McGraw-Hill; 2006.

29. Kohavi R, John GH. Wrappers for feature subset selection. Artif Intell 1997 Dec;97(1-2):273-324. [doi: 10.1016/s0004-3702(97)00043-x]
30. Guyon I, Elisseeff A. An introduction to variable and feature selection. J Mach Learn Res 2003:1157-1182 [FREE Full text]
31. Krebs EE, Carey TS, Weinberger M. Accuracy of the pain numeric rating scale as a screening test in primary care. J Gen Intern Med 2007 Oct;22(10):1453-1458 [FREE Full text] [doi: 10.1007/s11606-007-0321-2] [Medline: 17668269]
32. Rasmussen C. Gaussian processes for machine learning. In: Bousquet O, von Luburg U, Rätsch G, editors. Lecture Notes in Computer Science, Vol 3176. Berlin: Springer; 2003.
33. Smola AJ, Schölkopf B. A tutorial on support vector regression. Stat Comput 2004 Aug;14(3):199-222. [doi: 10.1023/b:stco.0000035301.49549.88]
34. Kuhn M, Johnson K. Applied Predictive Modeling. New York: Springer; 2013.
35. Mitchell T. Machine Learning. New York: McGraw-Hill; 1997.
36. Marco CA, Kanitz W, Jolly M. Pain scores among emergency department (ED) patients: comparison by ED diagnosis. J Emerg Med 2013 Jan;44(1):46-52. [doi: 10.1016/j.jemermed.2012.05.002] [Medline: 22795472]
37. Larsen B, Aone C. Fast and Effective Text Mining Using Linear-Time Document Clustering. In: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 1999 Presented at: KDD'99; August 15-18, 1999; San Diego, California, USA p. 16-22.
38. Bush AM, Coates TD, Wood JC. Diminished cerebral oxygen extraction and metabolic rate in sickle cell disease using T2 relaxation under spin tagging MRI. Magn Reson Med 2018 Jul;80(1):294-303 [FREE Full text] [doi: 10.1002/mrm.27015] [Medline: 29194727]
39. Meloni A, Detterich J, Berdoukas V, Pepe A, Lombardi M, Coates TD, et al. Comparison of biventricular dimensions and function between pediatric sickle-cell disease and thalassemia major patients without cardiac iron. Am J Hematol 2013 Mar;88(3):213-218 [FREE Full text] [doi: 10.1002/ajh.23376] [Medline: 23386313]
40. Coloigner J, Phlypo R, Coates TD, Lepore N, Wood JC. Graph Lasso-based test for evaluating functional brain connectivity in sickle cell disease. Brain Connect 2017 Sep;7(7):443-453 [FREE Full text] [doi: 10.1089/brain.2016.0474] [Medline: 28747064]
41. Gomperts E, Belcher JD, Otterbein LE, Coates TD, Wood J, Skolnick BE, et al. The role of carbon monoxide and heme oxygenase in the prevention of sickle cell disease vaso-occlusive crises. Am J Hematol 2017 Jun;92(6):569-582 [FREE Full text] [doi: 10.1002/ajh.24750] [Medline: 28378932]
42. Simmonds MJ, Suriany S, Ponce D, Detterich JA. Red blood cell mechanical sensitivity improves in patients with sickle cell disease undergoing chronic transfusion after prolonged, subhemolytic shear exposure. Transfusion 2018 Dec;58(12):2788-2796. [doi: 10.1111/trf.14901] [Medline: 30325033]
43. Du M, van Ness S, Gordeuk V, Nouraie SM, Nekhai S, Gladwin M, et al. Biomarker signatures of sickle cell disease severity. Blood Cells Mol Dis 2018 Sep;72:1-9 [FREE Full text] [doi: 10.1016/j.bcmd.2018.05.001] [Medline: 29778312]
44. Deist TM, Dankers FJ, Valdes G, Wijsman R, Hsu I, Oberije C, et al. Machine learning algorithms for outcome prediction in (chemo)radiotherapy: an empirical comparison of classifiers. Med Phys 2018 Jul;45(7):3449-3459 [FREE Full text] [doi: 10.1002/mp.12967] [Medline: 29763967]
45. Montazeri M, Montazeri M, Montazeri M, Beigzadeh A. Machine learning models in breast cancer survival prediction. Technol Health Care 2016;24(1):31-42. [doi: 10.3233/THC-151071] [Medline: 26409558]
46. Finkelstein J, Jeong IC. Machine learning approaches to personalize early prediction of asthma exacerbations. Ann N Y Acad Sci 2017 Jan;1387(1):153-165 [FREE Full text] [doi: 10.1111/nyas.13218] [Medline: 27627195]
47. Cowen R, Stasiowska MK, Laycock H, Bantel C. Assessing pain objectively: the use of physiological markers. Anaesthesia 2015 Jul;70(7):828-847 [FREE Full text] [doi: 10.1111/anae.13018] [Medline: 25772783]
48. Stahl SE, An H, Dinkel DM, Noble JM, Lee J. How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? BMJ Open Sport Exerc Med 2016;2(1):e000106 [FREE Full text] [doi: 10.1136/bmjsem-2015-000106] [Medline: 27900173]
49. Shcherbina A, Mattsson CM, Waggott D, Salisbury H, Christle JW, Hastie T, et al. Accuracy in wrist-worn, sensor-based measurements of heart rate and energy expenditure in a diverse cohort. J Pers Med 2017 May 24;7(2):pii: E3 [FREE Full text] [doi: 10.3390/jpm7020003] [Medline: 28538708]
50. Xie J, Wen D, Liang L, Jia Y, Gao L, Lei J. Evaluating the validity of current mainstream wearable devices in fitness tracking under various physical activities: comparative study. JMIR Mhealth Uhealth 2018 Apr 12;6(4):e94 [FREE Full text] [doi: 10.2196/mhealth.9754] [Medline: 29650506]

Abbreviations

AccZ: acceleration in Z direction
ED: emergency department
GP: Gaussian process
GPR: Gaussian process for regression

GSR: galvanic skin response
HR: heart rate
mHealth: mobile health
NSAID: nonsteroidal anti-inflammatory drug
RMSE: root mean square error
RR: R-R interval
SCD: sickle cell disease
SVM: support vector machines
SVR: support vector machines for regression
TRU-Pain: Technology Resources to Understand Pain

Edited by G Eysenbach; submitted 09.02.19; peer-reviewed by S Creary, A Davoudi, L Crosby, A Majmundar; comments to author 27.04.19; revised version received 22.06.19; accepted 19.07.19; published 02.12.19.

Please cite as:
Johnson A, Yang F, Gollarahalli S, Banerjee T, Abrams D, Jonassaint J, Jonassaint C, Shah N
Use of Mobile Health Apps and Wearable Technology to Assess Changes and Predict Pain During Treatment of Acute Pain in Sickle Cell Disease: Feasibility Study
JMIR Mhealth Uhealth 2019;7(12):e13671
URL: https://mhealth.jmir.org/2019/12/e13671
doi:10.2196/13671
PMID:31789599

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org